# Statistical Learning Models for Text and Graph Data
# Lecture 1: Introduction

Yangqiu Song

Hong Kong University of Science and Technology

*yqsong@cse.ust.hk*

September 4, 2019

∗Contents are based on materials created by Chris Manning, Percy Liang, Hongning Wang, Heng Ji, Dan Roth, and Haixun Wang

# Reference Content

- Chris Manning. CS 224N/Ling 237. Natural Language Processing. `https://web.stanford.edu/class/cs224n/`
- Percy Liang. ICML tutorial on Natural Language Understanding: Foundations and State-of-the-Art `https://icml.cc/2015/tutorials/icml2015-nlu-tutorial.pdf`

# Overview

1. Logistics

2. Introduction to NLP
   - Why is NLP Important?
   - Machine Learning for NLP: Algorithms, Tasks, and Challenges
   - The Need of Knowledge Graphs

# Logistics

- Instructor: Yangqiu Song
- Email: `yqsong@cse.ust.hk`
- Office: RM3518 (Lift25/26)
- Canvas (`https://canvas.ust.hk`)

- For CSE students, even if you enroll MATH5471, this course may not satisfy the requirement: "The 3 credits may be satisfied by courses from other Schools"
- Difference between COMP5222 and MATH5471: you can do a survey instead of the course project for MATH students
- NOTE: This is not an entry level course; students should have some machine learning background

# Course Information

- Four reading notes (20%): one paper per week, related to lectures
  - Find long papers in top venues such as ACL, EMNLP, NAACL, ICML, TACL, CL, JAIR, JMLR
  - Write a review about the strength and weakness of the paper
- Mid-term project proposal: title and abstract (10%):
  - Could be a discussion/survey paper for Math students
  - Free research, or
  - A given project: how to combine knowledge graphs to NLP tasks
- Project report (30%)
  - 8 pages not including references in ACL format
  - Consider to submit to ACL this year (deadline: December 9, 2019)
- Final project presentation(10%)
- Final exam (30%)
  - Examples put to Canvas
  - Reduced difficulty level than lecture notes

# Topic Covered

- Sequence modeling: language models, distributed representations
- Document classification: supervised learning, semi-supervised learning
- Topic modeling: SVD, probabilistic models
- Sequence labeling: sequence models, constrained models, posterior regularization
- Graph modeling: graph embedding, random walks, knowledge graphs
- Deep learning
    - Text and graph

# Overview

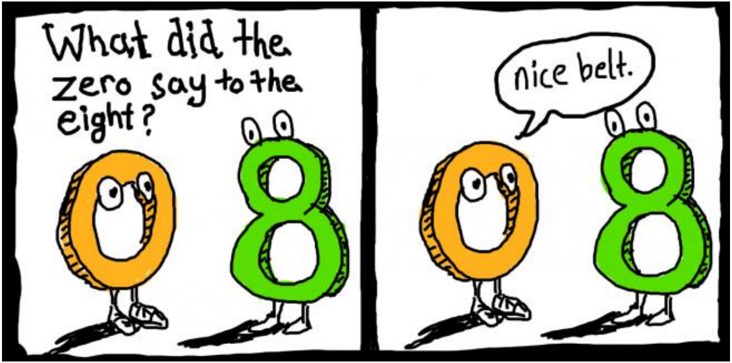
1 Logistics

2 Introduction to NLP

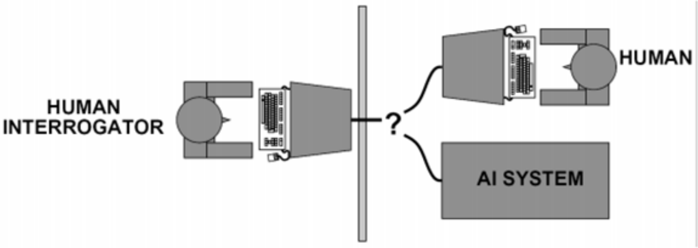
- Why is NLP Important?
- Machine Learning for NLP: Algorithms, Tasks, and Challenges
- The Need of Knowledge Graphs

# Natural Language

- Understanding language is a very complex thing
- But something that humans are amazingly good at

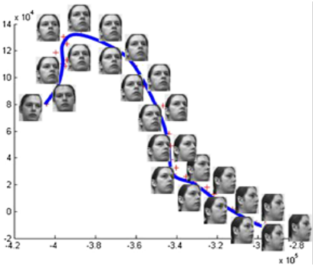


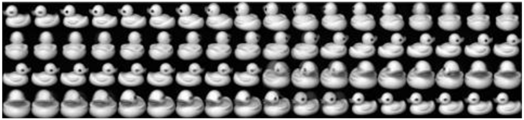
# Artificial Intelligence: Turing Test (1950)



- Replacement of “Can machines think?”
  - “Can machines behave intelligently?”
  - Operational test for intelligent behavior: the Imitation Game (later dubbed the Turing test)
  - Suggested major components of AI: knowledge, reasoning, language understanding, learning

# What's Special about Human Language?

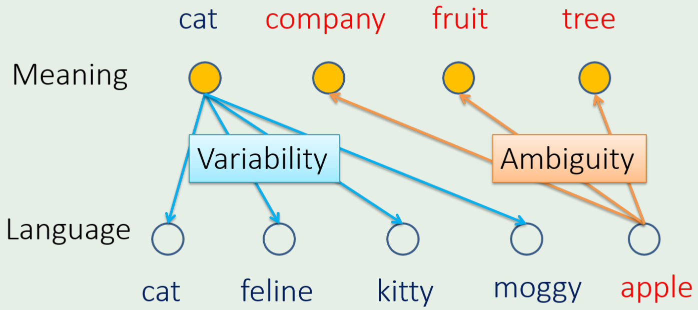




- A human language is a discrete/symbolic/categorical signaling system
- With very minor exceptions for expressive signaling ("I loooove it." "Whoomppaaa")
- Large vocabulary, symbolic encoding of words creates a problem for machine learning – sparsity!

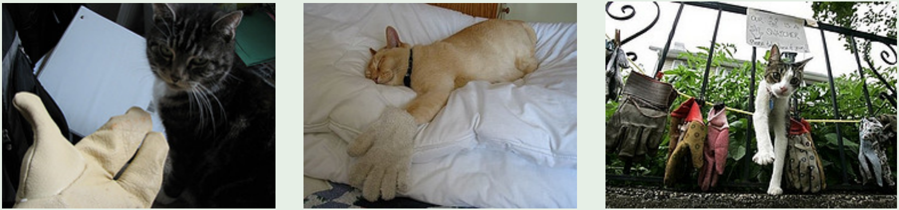
# Why is NLP Difficult?

## Example (variability and ambiguity everywhere)



# Why is NLP Difficult?

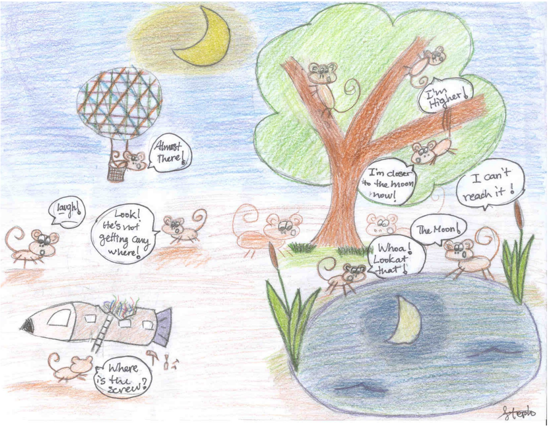
## Example (“Get the cat with the gloves.”)



# The AI Winter

- AI winters: 1974-80 and 1987-93
  - 1966: the failure of machine translation,
  - 1970: the abandonment of connectionism,
  - 1971-75: DARPA's frustration with the Speech Understanding Research program at Carnegie Mellon University,
  - 1973: the large decrease in AI research in the United Kingdom in response to the Lighthill report,
  - 1973-74: DARPA's cutbacks to academic AI research in general,
  - 1987: the collapse of the Lisp machine market,
  - 1988: the cancellation of new spending on AI by the Strategic Computing Initiative,
  - 1993: expert systems slowly reaching the bottom, and
  - 1990s: the quiet disappearance of the fifth-generation computer project's original goals.

# “All models are wrong; but some are useful.” – George E. P. Box



http://www.stat.ucla.edu/~sczhu/research_blog.html

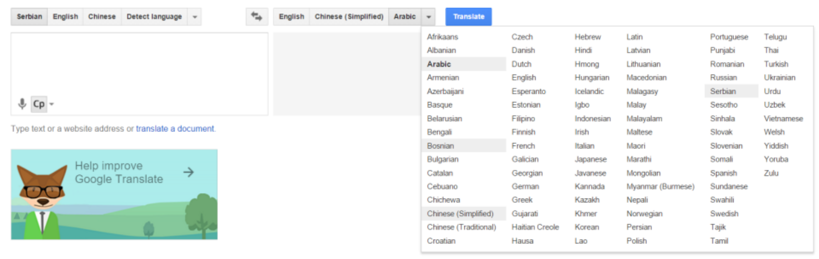
# Texts in the Era of Big Data

- Huge in size
  - Google processes 5.13B queries/day (2013)
  - Twitter receives 340M tweets/day (2012)
  - Facebook has 2.5 PB of user data + 15 TB/day (4/2009) (1PB=$10^{15}$bytes=1000terabytes)
  - eBay has 6.5 PB of user data + 50 TB/day (5/2009)
- 80% data is unstructured (IBM, 2010)
  - Traditional NLP techniques (e.g., parsing) are too slow to handle them
  - Traditional NLP models are based on labeled data in specific domains (WSJ data)

# NLP Enabled by Big Data

## Example (Google Translate)

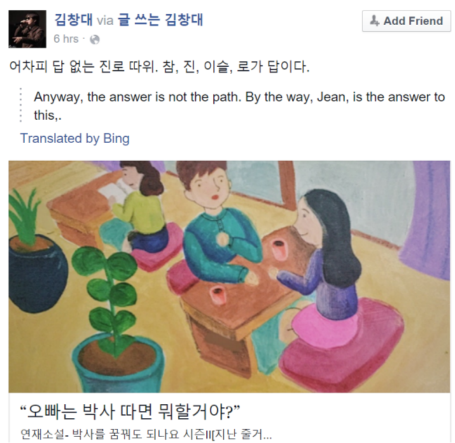
- 1966: the failure of machine translation
- Now: Google Translate can work with more than 100 languages



Type text or a website address or translate a document.

# NLP Enabled by Big Data

## Example (Facebook Translation)



어차피 답 없는 진로 따위. 참, 진, 이슬, 로가 답이다.

Anyway, the answer is not the path. By the way, Jean, is the answer to this,.

Translated by Bing

"오빠는 박사 따면 뭐할거야?"

연재소설- 박사를 꿈꿔도 되나요 시즌II[지난 줄거...

# NLP Enabled by Big Data

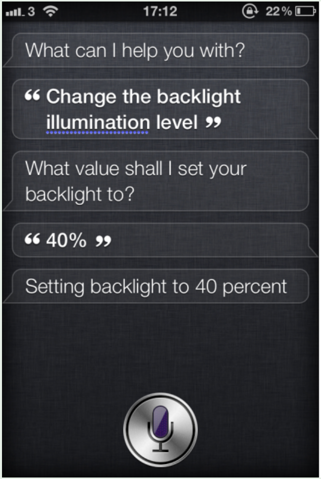
## Example (IBM's Watson)

- 1971–75:DARPA's frustration with the Speech Understanding
- Now: "Watson is aquestion answering (QA) computing system that IBM built to apply advanced
  - natural language processing,
  - information retrieval,
  - knowledge representation,
  - automated reasoning, and
  - machine learning technologies
- to the field of open domain question answering."

In 2011, Watson competed on Jeopardy! against former winners Brad Rutter and Ken Jennings. Watson received the first place prize of $1 million.

# NLP Enabled by Big Data

## Example (Apple's Siri)



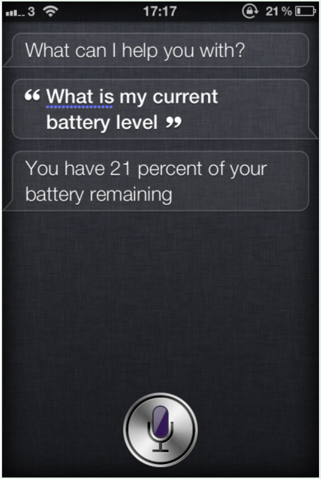
What can I help you with?

“ Change the backlight illumination level ”

What value shall I set your backlight to?

“ 40% ”

Setting backlight to 40 percent



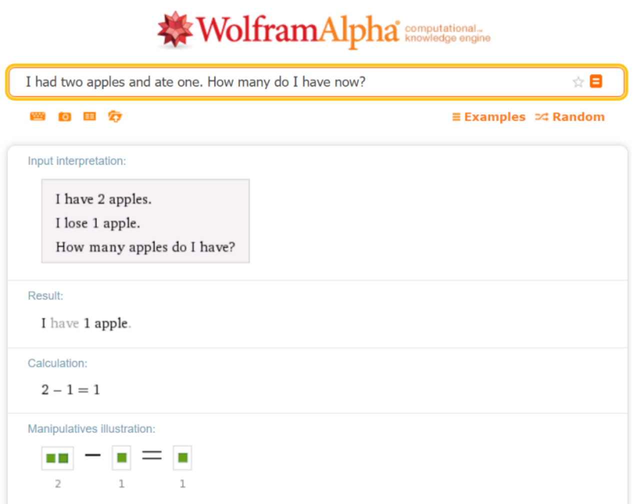
What can I help you with?

“ What is my current battery level ”

You have 21 percent of your battery remaining

# NLP Enabled by Big Data

## Example (WolframAlpha Knowledge Powered QA)



WolframAlpha computational knowledge engine

I had two apples and ate one. How many do I have now?

Input interpretation:

I have 2 apples.
I lose 1 apple.
How many apples do I have?

Result:

I have 1 apple.

Calculation:

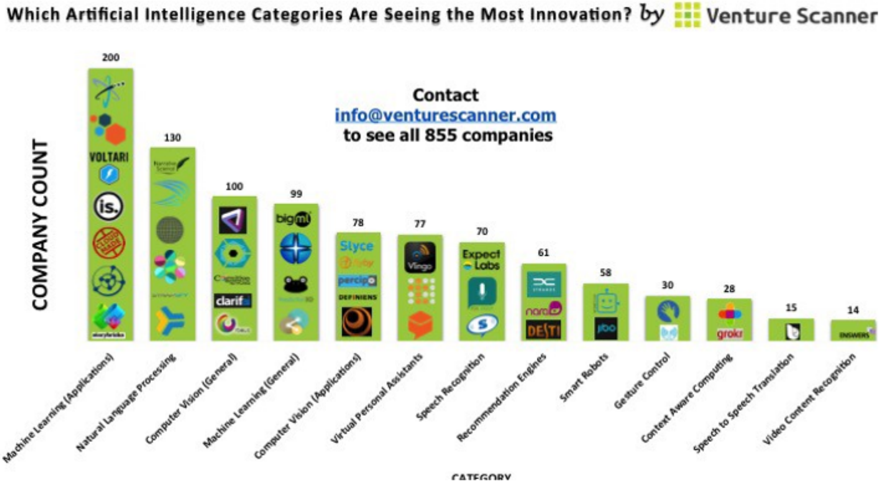
$2 - 1 = 1$

Manipulatives illustration:

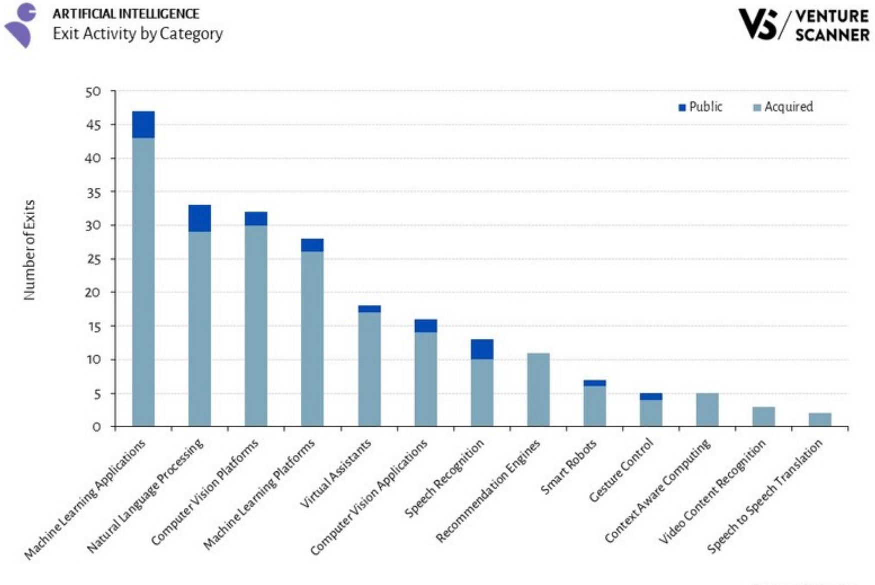
🟩🟩 − 🟩 = 🟩

2 1 1

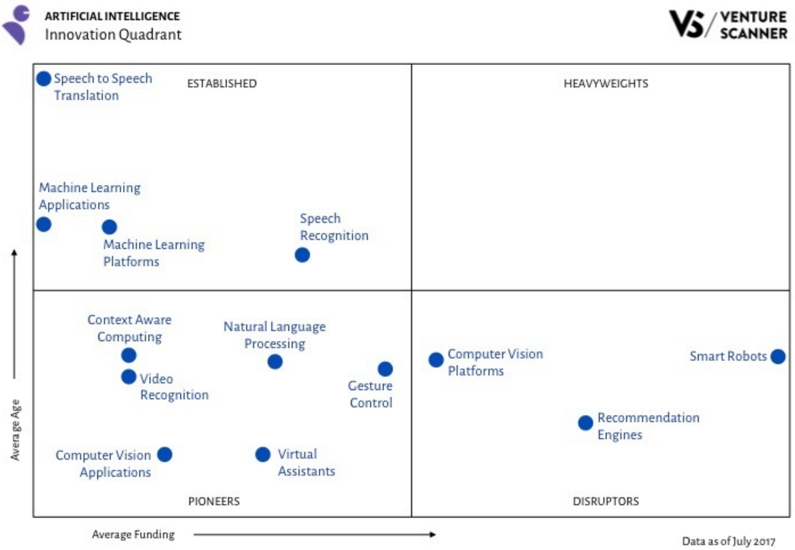
# Startup Companies (2015)



# Number of Exits (Acquisitions and IPOs, 2017)



# Funding Size vs. Company Age (2017)



# Overview

1 Logistics

2 Introduction to NLP

- Why is NLP Important?
- Machine Learning for NLP: Algorithms, Tasks, and Challenges
- The Need of Knowledge Graphs

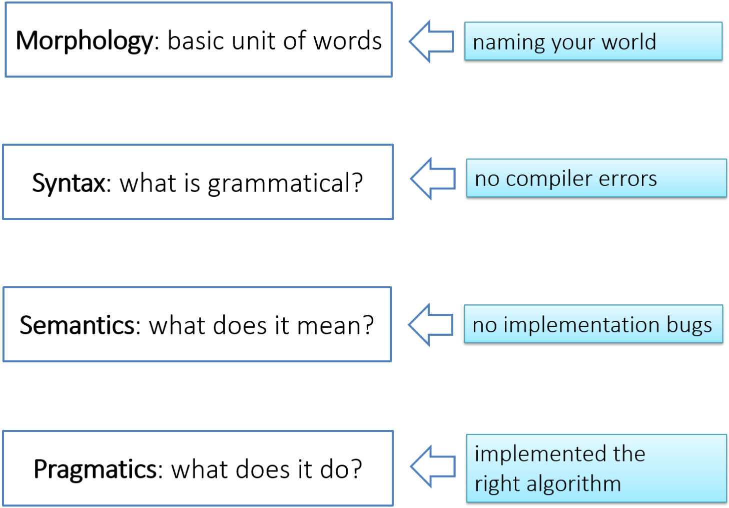
# Statistical Machine Learning

- Natural Language Processing
  - Natural Language Understanding (NLU)
  - Natural Language Generation (NLG)
- Machine learning has been widely used in both NLU and NLG
  - given that we have a lot of data now

# Popular Statistical Machine Learning Algorithms for NLP

- Mid-1970s: Hidden Markov Models (HMMs) for speech recognition → probabilistic models
- Early 2000s: Conditional Random Fields (CRFs) for part-of-speech tagging → structured prediction
- Early 2000s: Latent Dirichlet Allocation (LDA) for modeling text documents → topic modeling
- Mid 2010s: sequence-to-sequence models for machine translation → Deep Learning neural networks with memory/state
- Now: ??? for natural language understanding/generation
    - Reinforcement learning?
    - Turing machines?
    - Knowledge graph reasoning models?

# NLP Tasks: Levels of Linguistic Analysis



**Morphology**: basic unit of words ⇐ naming your world

**Syntax**: what is grammatical? ⇐ no compiler errors

**Semantics**: what does it mean? ⇐ no implementation bugs

**Pragmatics**: what does it do? ⇐ implemented the right algorithm

# Analogy with Programming Languages

- Syntax: no compiler errors
- Semantics: no implementation bugs
- Pragmatics: implemented the right algorithm

- Different syntax, same semantics (5):

$$2 + 3 \Leftrightarrow 3 + 2$$

- Same syntax, different semantics (1 and 1.5):

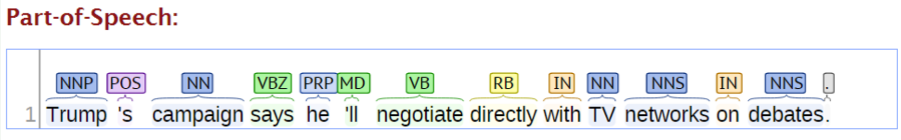
$$3 \;/\; 2 \text{ (Python 2.7)} \not\Leftrightarrow 3 \;/\; 2 \text{ (Python 3)}$$

- Good semantics, bad pragmatics:

correct implementation of deep neural network
for estimating coin flip prob.

# Syntax (1): Part of Speech

## Example (Part of speech)



**Part-of-Speech:**

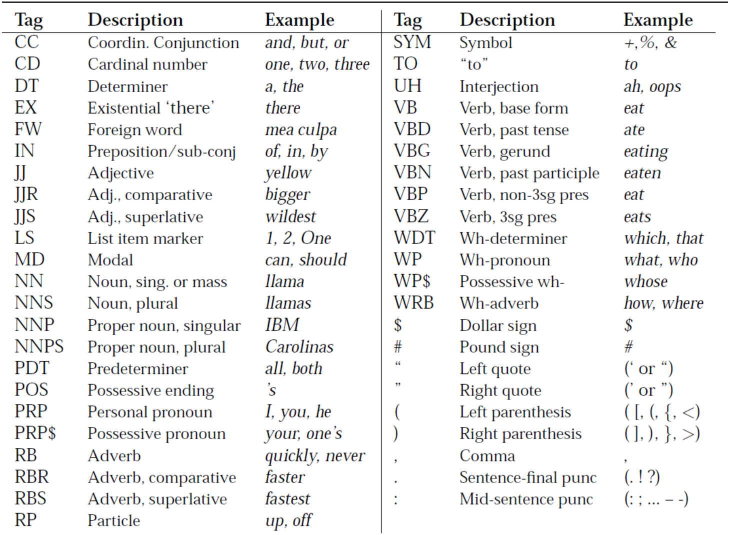
| NNP | POS | NN | VBZ | PRP | MD | VB | RB | IN | NN | NNS | IN | NNS | . |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trump | 's | campaign | says | he | 'll | negotiate | directly | with | TV | networks | on | debates | . |

Tags:

- NN: common noun
- NNP: proper noun
- VB: verb, base form
- VBZ: verb, 3rd person singular
- ...

# Syntax (1): Part of Speech

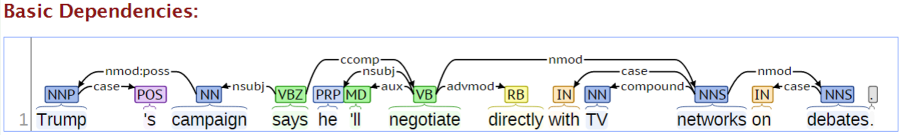
Penn Treebank part-of-speech tags (including punctuation).



| Tag | Description | Example | Tag | Description | Example |
|---|---|---|---|---|---|
| CC | Coordin. Conjunction | *and, but, or* | SYM | Symbol | +,%, & |
| CD | Cardinal number | *one, two, three* | TO | “to” | *to* |
| DT | Determiner | *a, the* | UH | Interjection | *ah, oops* |
| EX | Existential ‘there’ | *there* | VB | Verb, base form | *eat* |
| FW | Foreign word | *mea culpa* | VBD | Verb, past tense | *ate* |
| IN | Preposition/sub-conj | *of, in, by* | VBG | Verb, gerund | *eating* |
| JJ | Adjective | *yellow* | VBN | Verb, past participle | *eaten* |
| JJR | Adj., comparative | *bigger* | VBP | Verb, non-3sg pres | *eat* |
| JJS | Adj., superlative | *wildest* | VBZ | Verb, 3sg pres | *eats* |
| LS | List item marker | *1, 2, One* | WDT | Wh-determiner | *which, that* |
| MD | Modal | *can, should* | WP | Wh-pronoun | *what, who* |
| NN | Noun, sing. or mass | *llama* | WP$ | Possessive wh- | *whose* |
| NNS | Noun, plural | *llamas* | WRB | Wh-adverb | *how, where* |
| NNP | Proper noun, singular | *IBM* | $ | Dollar sign | $ |
| NNPS | Proper noun, plural | *Carolinas* | # | Pound sign | # |
| PDT | Predeterminer | *all, both* | “ | Left quote | (‘ or “) |
| POS | Possessive ending | *’s* | ” | Right quote | (’ or ”) |
| PRP | Personal pronoun | *I, you, he* | ( | Left parenthesis | ( [, (, {, < ) |
| PRP$ | Possessive pronoun | *your, one’s* | ) | Right parenthesis | ( ], ), }, > ) |
| RB | Adverb | *quickly, never* | , | Comma | , |
| RBR | Adverb, comparative | *faster* | . | Sentence-final punc | (. ! ?) |
| RBS | Adverb, superlative | *fastest* | : | Mid-sentence punc | (: ; ... – -) |
| RP | Particle | *up, off* | | | |

# Syntax (2): Dependency Parse Tree

## Example (Dependency parse)

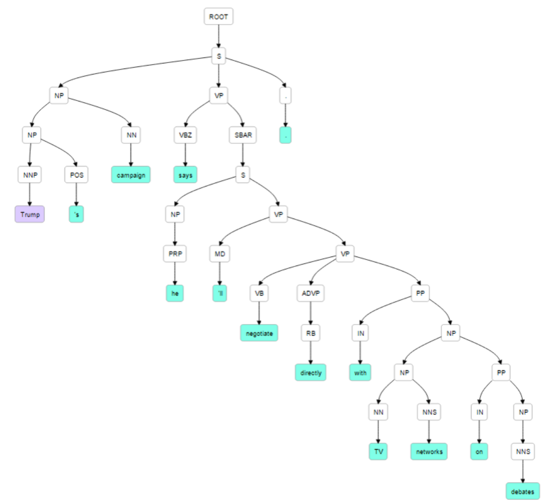


Dependency relations:

- nsubj: subject (nominal)
- advmod: adverbial modifier
- ...

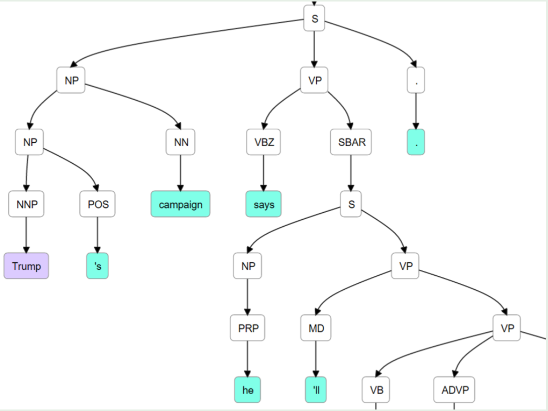
# Syntax (3): Constituency Parse Tree

## Example (Constituency parsing)



# Syntax (3): Constituency Parse Tree

## Example (Constituency parsing)



- POS: possessive ending
- PRP: personal pronoun
- MD: modal; can, should

# Semantics

- Syntax: no compiler errors
- Semantics: no implementation bugs
- Pragmatics: implemented the right algorithm

- Semantics: meanings
  - Lexical semantics: what words mean
  - Compositional semantics: how meaning gets combined

# Semantics (1): What's a Word?

## Example

Words

light

Multi-word expressions: meaning unit beyond a word

light bulb

Morphology: meaning unit within a word

light lighten lightening relight

Polysemy: one word has multiple meanings (word senses)

- The light was filtered through a soft glass window.
- He stepped into the light.
- This lamp lights up the room.
- The load is not light.

# Semantics (1): Synonymy

## Example (Synonymy)

Words:

confusing unclear perplexing mystifying

Sentences (paraphrases):

- I have fond memories of my childhood.
- I reflect on my childhood with a certain fondness.
- I enjoy thinking back to when I was a kid.

Beware: no true equivalence due to subtle differences in meaning; think distance metric

But there's more to meaning than similarity...

# Other Lexical Relations

Hyponymy (is-a):

a cat is a mammal

Meronomy (has-a):

a cat has a tail

Useful for entailment:

Alice is 170cm high and Bob is 180cm high.

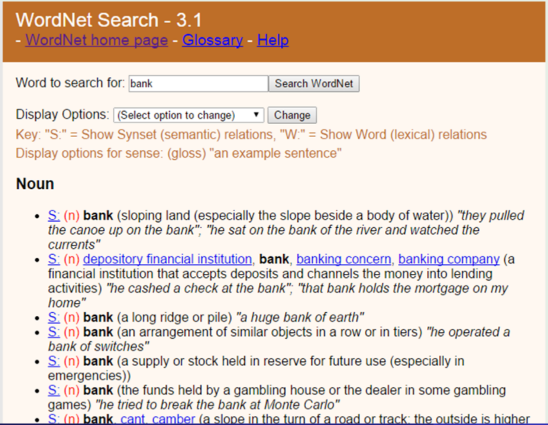
$$\Rightarrow$$

Bob is taller than Alice.

# WordNet (Starting from 1985)

- A machine readable lexical database of English:
- Word senses grouped into synonym sets (“synsets”) linked into a conceptual-semantic hierarchy

## Example (Bank)



WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: bank | Search WordNet

Display Options: (Select option to change) ▾ | Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

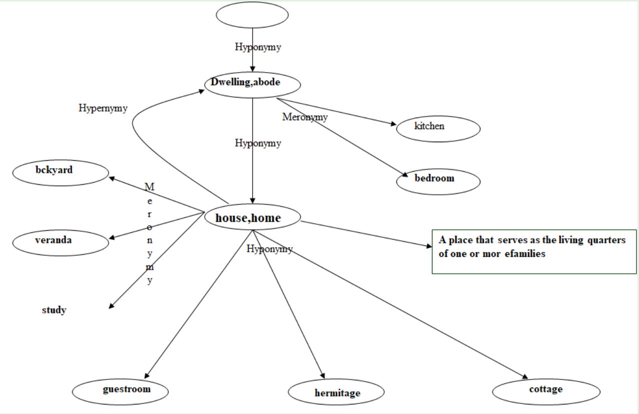
**Noun**

- S: (n) **bank** (sloping land (especially the slope beside a body of water)) *"they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents"*
- S: (n) depository financial institution, **bank**, banking concern, banking company (a financial institution that accepts deposits and channels the money into lending activities) *"he cashed a check at the bank"; "that bank holds the mortgage on my home"*
- S: (n) **bank** (a long ridge or pile) *"a huge bank of earth"*
- S: (n) **bank** (an arrangement of similar objects in a row or in tiers) *"he operated a bank of switches"*
- S: (n) **bank** (a supply or stock held in reserve for future use (especially in emergencies))
- S: (n) **bank** (the funds held by a gambling house or the dealer in some gambling games) *"he tried to break the bank at Monte Carlo"*
- S: (n) **bank**, cant, camber (a slope in the turn of a road or track; the outside is higher

# WordNet (Starting from 1985)

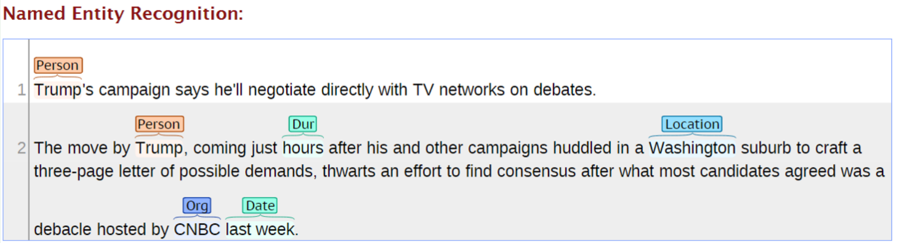
- A machine readable lexical database of English:
- Word senses grouped into synonym sets ("synsets") linked into a conceptual-semantic hierarchy

## Example (Overview)



# Semantics (2): Named Entities (Recognition, Typing, Linking)

## Example (Named Entity Recognition)



**Named Entity Recognition:**

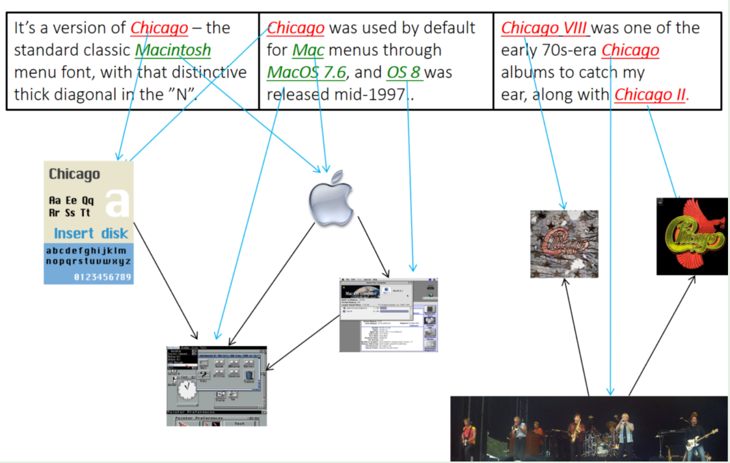
1 Trump's campaign says he'll negotiate directly with TV networks on debates.

2 The move by Trump, coming just hours after his and other campaigns huddled in a Washington suburb to craft a three-page letter of possible demands, thwarts an effort to find consensus after what most candidates agreed was a debacle hosted by CNBC last week.

- Pers: Person
- Location
- Org: Organization
- Date/time

# Semantics (2): Named Entities (Recognition, Typing, Linking)

## Example (Entity Linking, Information Network Analysis)



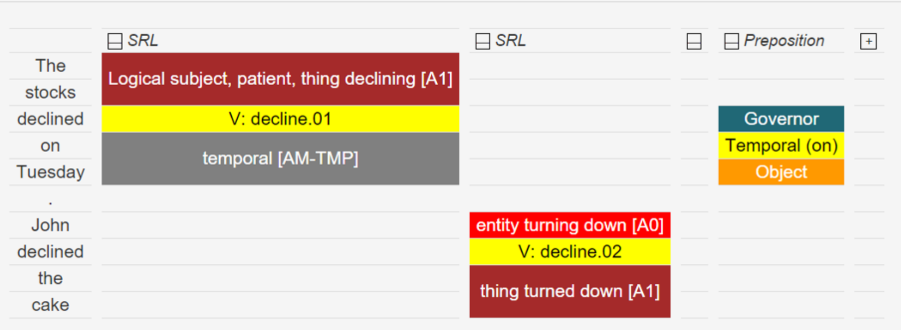
It’s a version of Chicago – the standard classic Macintosh menu font, with that distinctive thick diagonal in the ”N”.

Chicago was used by default for Mac menus through MacOS 7.6, and OS 8 was released mid-1997..

Chicago VIII was one of the early 70s-era Chicago albums to catch my ear, along with Chicago II.

# Semantics (3): Frame based Semantics

## Example (Semantic Role Labeling)

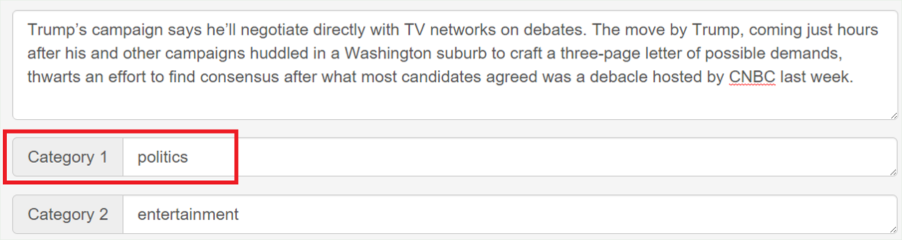


| | ⊟ SRL | ⊟ SRL | ⊟ | ⊟ Preposition | ⊞ |
|---|---|---|---|---|---|
| The | Logical subject, patient, thing declining [A1] | | | | |
| stocks | | | | | |
| declined | V: decline.01 | | | Governor | |
| on | temporal [AM-TMP] | | | Temporal (on) | |
| Tuesday | | | | Object | |
| . | | | | | |
| John | | entity turning down [A0] | | | |
| declined | | V: decline.02 | | | |
| the | | thing turned down [A1] | | | |
| cake | | | | | |

- Predicates
- Arguments
- Senses

# Semantics (4): Topics

## Example (Topics)



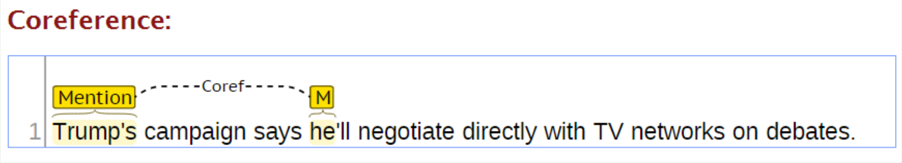
Trump's campaign says he'll negotiate directly with TV networks on debates. The move by Trump, coming just hours after his and other campaigns huddled in a Washington suburb to craft a three-page letter of possible demands, thwarts an effort to find consensus after what most candidates agreed was a debacle hosted by CNBC last week.

| | |
|---|---|
| Category 1 | politics |
| Category 2 | entertainment |

- Classification
- Clustering
- Topic modeling

# Discourse

## Example (Coreference Resolution (Pronoun Resolution))



Coreference:

1 Trump's campaign says he'll negotiate directly with TV networks on debates.

"The Winograd Schema Challenge" (Levesque, 2011)

- The dog chased the cat, which ran up a tree. It waited at the top.
- The dog chased the cat, which ran up a tree. It waited at the bottom.
- Paul tried to call George on the phone, but he wasn't successful.
- Paul tried to call George on the phone, but he wasn't available.

Easy for humans, can't use surface-level patterns

# Discourse

## Example (Shallow Discourse Parser for Document-level Analysis)

- S1: Kemper is the first firm to make a major statement with program trading.
- S2: He added that "having just one firm do this isn't going to mean a hill of beans."

We can add a connective "but" between to above two sentences to indicate "Contrast relationship"

- S1: Senator calls this "the first gift of democracy."
- S2: The Poles might do better to view this as a Trojan Horse.

# Pragmatics

Conversational implicature: new material suggested (not logically implied) by sentence

## Example (Conversational implicature)

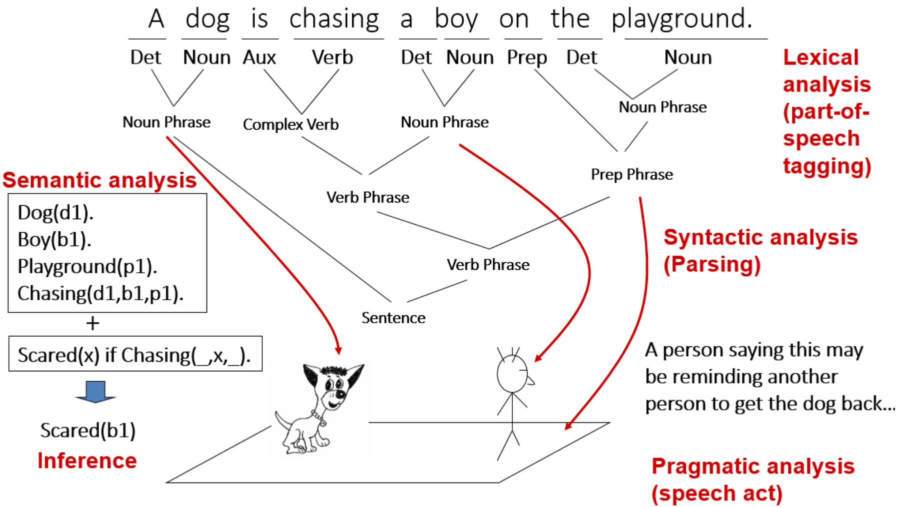
- A: What on earth has happened to the roast beef?
- B: The dog is looking very happy.
- Implicature: The dog at the roast beef.

Presupposition: background assumption independent of truth of sentence

## Example (Presupposition)

- I have stopped eating meat.
- Presupposition: I once was eating meat.

# States of NLP



A dog is chasing a boy on the playground.

Det Noun Aux Verb Det Noun Prep Det Noun

Noun Phrase

Complex Verb

Noun Phrase

Noun Phrase

Prep Phrase

Verb Phrase

Verb Phrase

Sentence

**Lexical analysis (part-of-speech tagging)**

**Syntactic analysis (Parsing)**

**Semantic analysis**

Dog(d1).
Boy(b1).
Playground(p1).
Chasing(d1,b1,p1).

+

Scared(x) if Chasing(_,x,_).

Scared(b1)

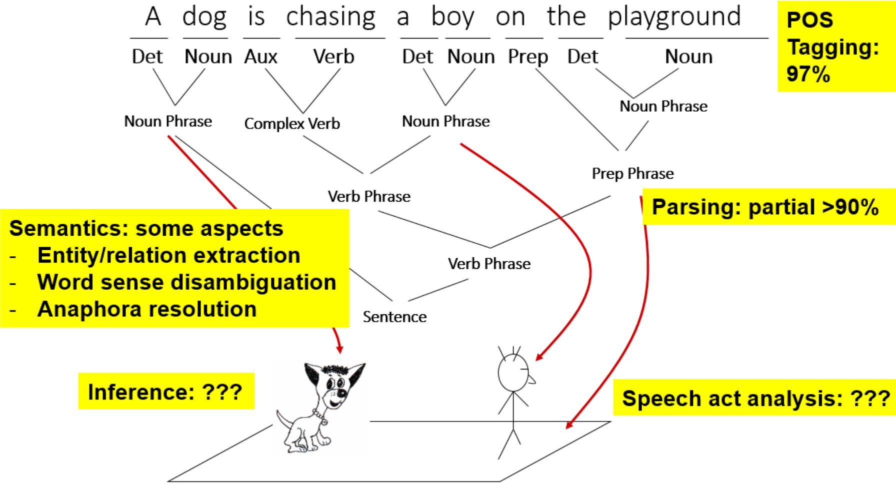
**Inference**

A person saying this may be reminding another person to get the dog back…

**Pragmatic analysis (speech act)**

Hongning Wang@UVa

# States of NLP



A dog is chasing a boy on the playground

Det Noun Aux Verb Det Noun Prep Det Noun

Noun Phrase

Complex Verb

Noun Phrase

Noun Phrase

Prep Phrase

Verb Phrase

Verb Phrase

Sentence

**POS Tagging: 97%**

**Parsing: partial >90%**

**Semantics: some aspects**

- **Entity/relation extraction**
- **Word sense disambiguation**
- **Anaphora resolution**

**Inference: ???**

**Speech act analysis: ???**

CS@UVa

23

# Overview

1. Logistics

2. Introduction to NLP
   - Why is NLP Important?
   - Machine Learning for NLP: Algorithms, Tasks, and Challenges
   - The Need of Knowledge Graphs

# Pragmatics

Semantics: what does it mean literally?
Pragmatics: what is the speaker really conveying?

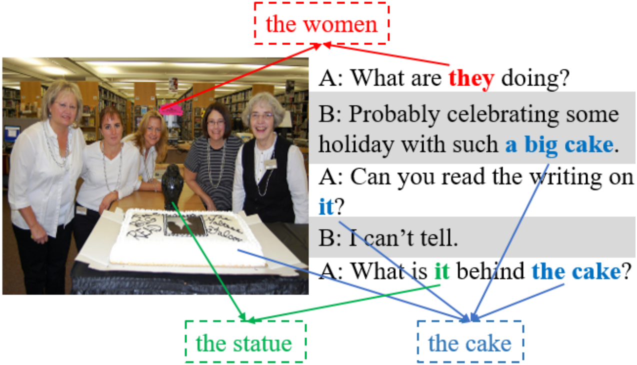
- Underlying principle (Grice, 1975): language is cooperative game between speaker and listener
- Implicature and presupposition depend on people and context (multi-modality and knowledge graph opportunities here) and involve soft inference (machine learning opportunities here)

We need a lot of background knowledge and commonsense knowledge

- We need to combine symbolic reasoning and machine learning!

# Pragmatics

Sometimes we need to ground natural language texts to the world or contexts to make inference


the women
A: What are they doing?
B: Probably celebrating some holiday with such a big cake.
A: Can you read the writing on it?
B: I can't tell.
A: What is it behind the cake?
the statue
the cake

# More about "Commonsense Knowledge"

When we communicate,

- we omit a lot of "common sense" knowledge, which we assume the hearer/reader possesses
- we keep a lot of ambiguities, which we assume the hearer/reader knows how to resolve

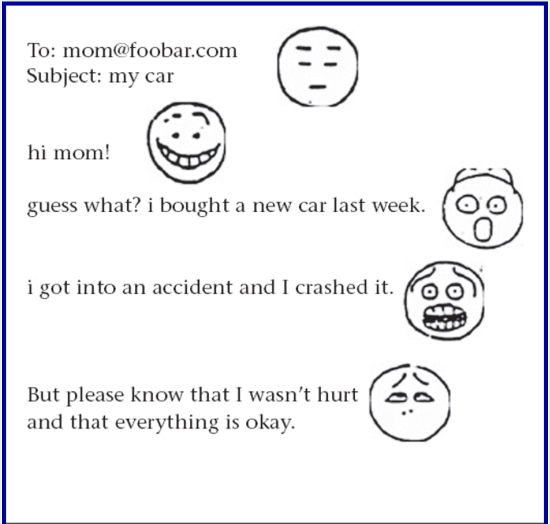
Knowledge about the everyday world that is possessed by all people

## Example (Commonsense Knowledge)

- A lemon is sour.
- To open a door, you must usually first turn the doorknob.
- If you forget someones birthday, they may be unhappy with you.
- A coat is used for keeping warm.
- People want to be respected.
- The last thing you do when you cook dinner is wash your dishes.
- People want good coffee.

# Commonsense Knowledge in Sentiment Analysis

## Example (Sentiment Analysis)



To: mom@foobar.com
Subject: my car

hi mom!

guess what? i bought a new car last week.

i got into an accident and I crashed it.

But please know that I wasn't hurt and that everything is okay.

*Figure 2. Empathy Buddy Reacts to an E-mail.*

# More about Ambiguity

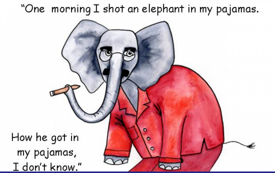
Ambiguity: more than one possible (precise) interpretations

One morning I shot an elephant **in** my pajamas.

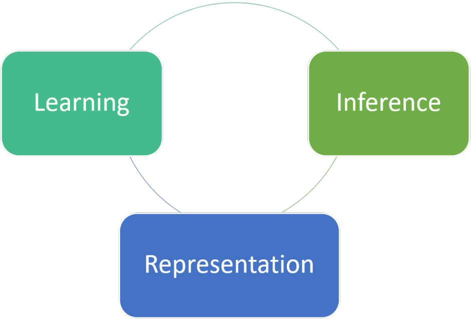
How he got in my pajamas, I don't know. — Groucho Marx

- The joke is based on misdirection, where the listener thinks one thing, and the teller says another
- "One morning I was wearing my pajamas, and I shot an elephant." or
- "One morning, an elephant was wearing my pajamas, and I shot that elephant."

How he got in my pajamas, I don't know. — Groucho Marx


"One morning I shot an elephant in my pajamas.
How he got in my pajamas, I don't know."

# Course Organization



- Representation: language models, word embeddings, topic models, knowledge graphs
- Learning: supervised learning, semi-supervised learning, distant supervision, indirect supervision, sequence models, deep learning, optimization techniques
- Inference: constraint modeling, joint inference, search algorithms

Applications: tasks introduced above

# Summary

1. Logistics
2. Introduction to NLP
   - Why is NLP Important?
   - Machine Learning for NLP: Algorithms, Tasks, and Challenges
   - The Need of Knowledge Graphs

In this course, we will

- Understand the intuition and motivation of how to model text and graph data
- Know popular and state-of-the-art statistical models for NLP
- Build relationships of different algorithms