

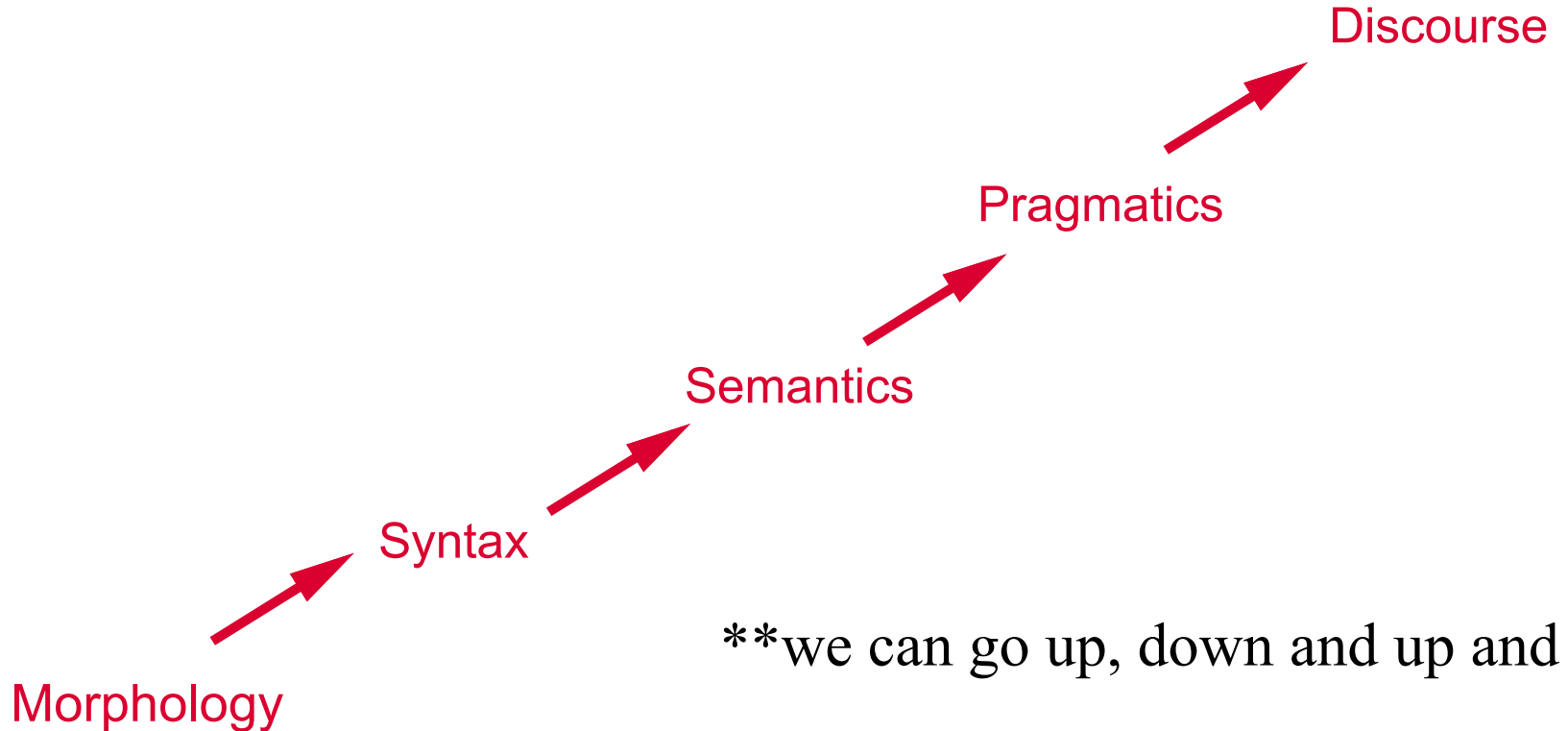
# NLP Tasks

Instructor: Mausam

(Slides adapted from Heng Ji, Dan Klein,  
Jurafsky & Martin, Noah Smith, Luke  
Zettlemoyer)

# NLP Tasks

---

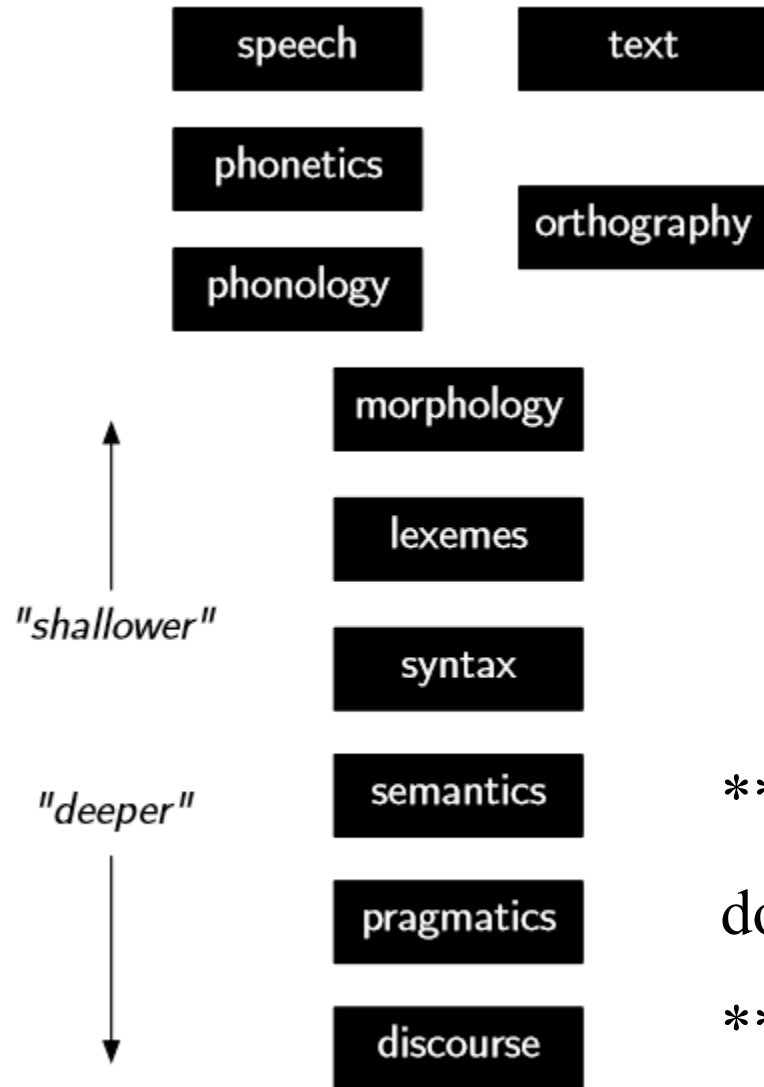


\*\*we can go up, down and up and down and combine steps too!!

\*\*every step is equally complex

# Levels of Linguistic Knowledge

---



\*\*we can go up, down and up and down and combine steps too!!

\*\*every step is equally complex

# Orthography

---

ลูกศิษย์วัดกระทิงยังยึดปิดถนนทางขึ้นไปนมัสการพระบาทเขาคิชฌกูฏ หวัดปะทะ  
กับเจ้าถิ่นที่ออกมาเผชิญหน้าเพราะเดือดร้อนสัญจรไม่ได้ ผวจ.เร่งทุกฝ่ายเจรจา  
ก่อนที่ชื่อเสียงของจังหวัดจะเสียหายไปมากกว่านี้ พร้อมเสนอหยุดจัดงาน 15 วัน....

# Morphology

---

uygarlaştıramadıklarımızdanmışsınızcasına

“(behaving) as if you are among those whom we could not civilize”

TIFGOSH ET HA-LELED BA-GAN

“you will meet the boy in the park”

unfriend, Obamacare, Manfuckinghattan

Suffix, prefix, infix (tmesis), circumfix

arundhati “leftist” roy, fan-frickin'-tastic,

छुड़ाया chuRAyA → छुड़वाया chhurwAyA

# The Challenge of Words

---

- Segmenting text into words
  - Thai example
- Morphological variation
  - Turkish and Hebrew examples
- Words with multiple meanings:
  - bank, mean
- Domain-specific meanings:
  - latex, sport, etc.
- Multiword expressions:
  - make a decision, take out, make up

# POS Tagging

---

ikr smh he asked fir yo last name

so he can add u on fb lololol

# Morphology+Syntax

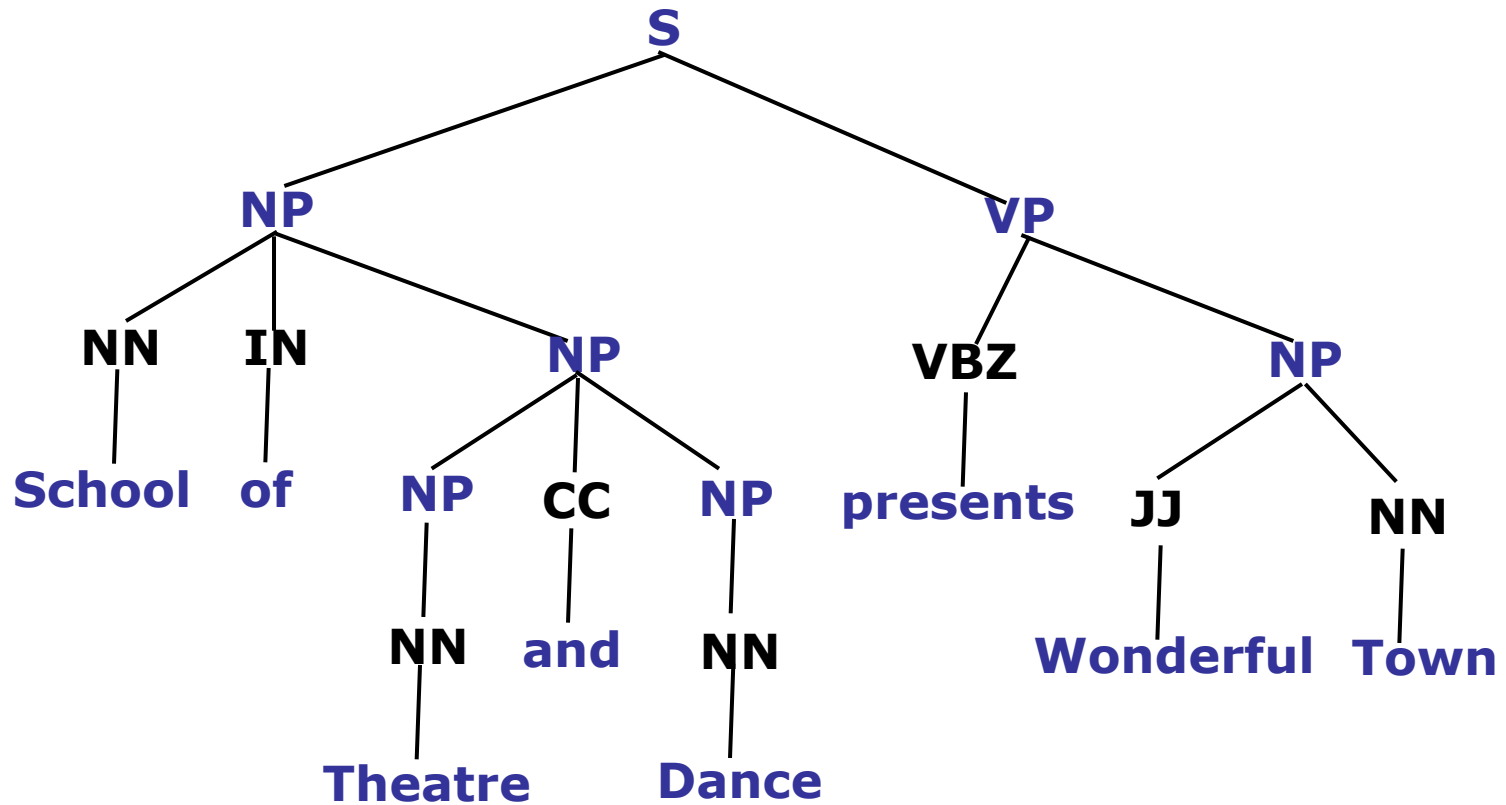
---

A ship-shipping ship, shipping shipping-ships.



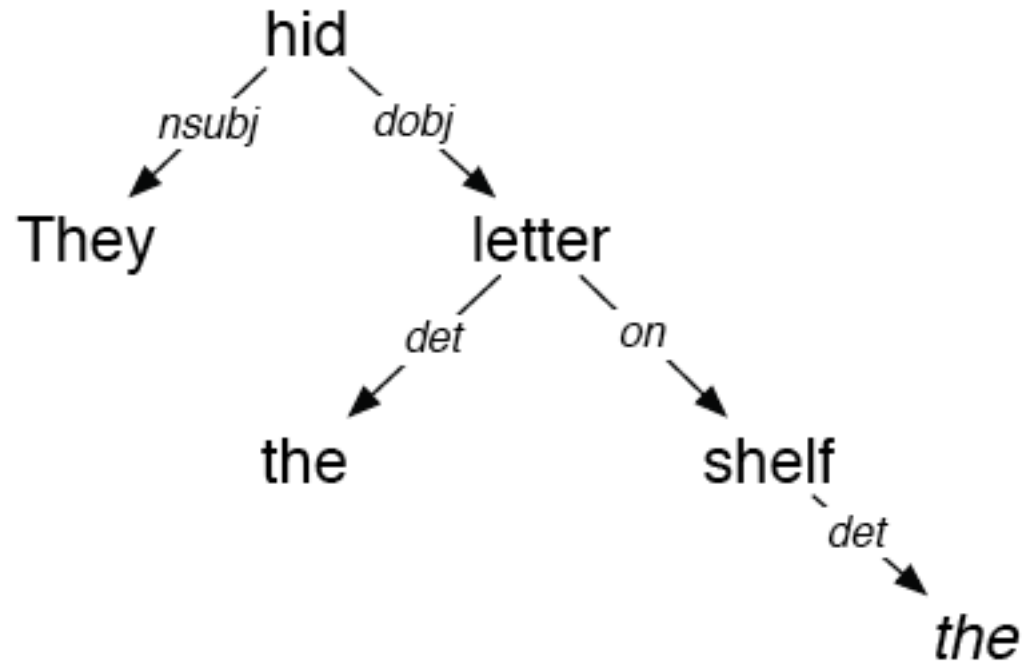
# Constituency Parsing

---



# Dependency Parsing

---



*They hid the letter on the shelf*

# Dependency Relations

---

<b>Argument Dependencies</b>	<b>Description</b>
<b>nsubj</b>	nominal subject
<b>csubj</b>	clausal subject
<b>dobj</b>	direct object
<b>iobj</b>	indirect object
<b>pobj</b>	object of preposition
<b>Modifier Dependencies</b>	<b>Description</b>
<b>tmod</b>	temporal modifier
<b>appos</b>	appositional modifier
<b>det</b>	determiner
<b>prep</b>	prepositional modifier

# Semantics

---

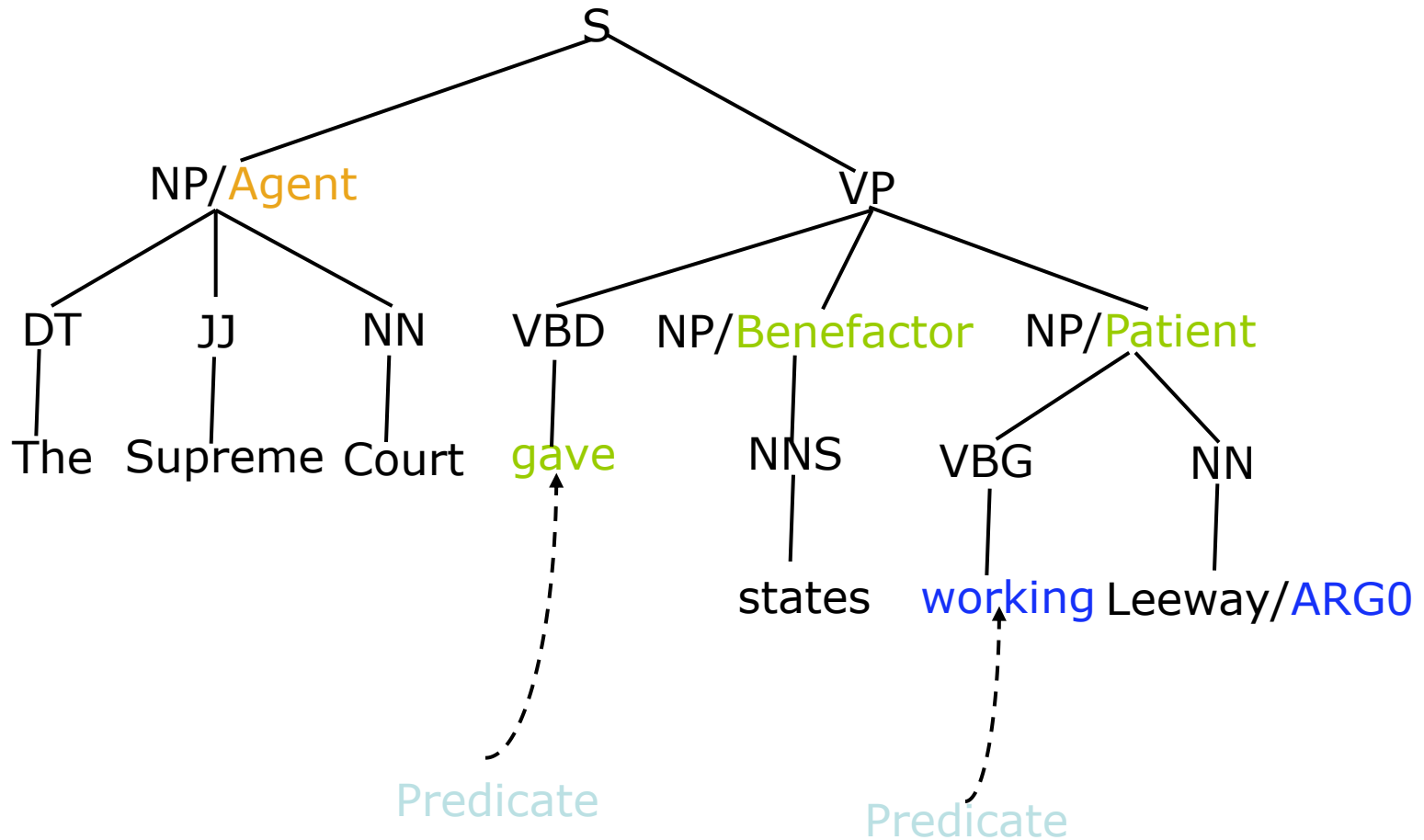
- Lexical Semantics
  - Meanings of words
  - Distributional similarity
- Frame semantics
  - (shallow semantics)
- Sentential semantics
  - (full semantics)

# Lexical Semantics

---

- **Synonym**
  - start-begin, bodoh-dungu, next to-near
- **Antonym**
  - clever x stupid, high x low, big x small, etc.
- **Homonym**
  - identical in spelling and pronunciation
  - bear, bank, ...
- **Homophones**
  - sounds identical but are written differently
  - site-sight, piece-peace.
- **Homograph**
  - written identically but sound differently
  - lead, wind, minute
- **Polysemy**
  - a word which has two (or more) related meanings

# Semantic Role Labeling: Adding Semantics into Trees



# Core Arguments

---

- Arg0 = agent
- Arg1 = direct object / theme / patient
- Arg2 = indirect object / benefactive / instrument / attribute / end state
- Arg3 = start point / benefactive / instrument / attribute
- Arg4 = end point

# Full Sentential Semantics

---

- Formal/Logical Semantics
  - All man are mortal
  - India borders a country that borders Afghanistan.

# Information Extraction

---

- Unstructured text to database entries

New York Times Co. named **Russell T. Lewis**, 45, **president and general manager** of its flagship **New York Times newspaper**, responsible for all business-side activities. He was **executive vice president and deputy general manager**. He succeeds **Lance R. Primis**, who in September was named **president and chief operating officer** of **the parent**.

Person	Company	Post	State
<b>Russell T. Lewis</b>	New York Times newspaper	president and general manager	start
<b>Russell T. Lewis</b>	New York Times newspaper	executive vice president	end
<b>Lance R. Primis</b>	New York Times Co.	president and CEO	start

- **SOTA**: perhaps 80% accuracy for pre-defined tables, 90%+ for single easy fields
- **But remember**: information is redundant!

# KB Slots

Person		Organization
per:alternate_names	per:title	org:alternate_names
per:date_of_birth	per:member_of	org:political/religious_affiliation
per:age	per:employee_of	org:top_members/employees
per:country_of_birth	per:religion	org:number_of_employees/members
per:stateorprovince_of_birth	per:spouse	org:members
per:city_of_birth	per:children	org:member_of
per:origin	per:parents	org:subsidiaries
per:date_of_death	per:siblings	org:parents
per:country_of_death	per:other_family	org:founded_by
per:stateorprovince_of_death	per:charges	org:founded
per:city_of_death		org:dissolved
per:cause_of_death		org:country_of_headquarters
per:countries_of_residence		org:stateorprovince_of_headquarters
per:stateorprovinces_of_residence		org:city_of_headquarters
per:cities_of_residence		org:shareholders
per:schools_attended		org:website

# Knowledge Base Linking (Wikification)

## Shocking Jim Parsons truths revealed after Emmy win

August 29, 2010 | 7:22 pm



**Jim Parsons**

... details after he won Sunday for lead actor in the comedy series. Interesting, or just over the top, decide.

For example: "I'm a big reader of almanacs, or I was, and I like lists -- oh, I sound a bit OCD, don't I?" "The nerd will dance out," he says, when he mixes his character's "genius" thing with the writers' words.

And when he was very young, he played the kolo kolo bird in Rudyard Kipling's "The Elephant's Child," donning a pair of yellow tights



Article Discussion

### James A. Parsons

From Wikipedia, the free encyclopedia

*For other people named James Parsons, see James Parsons (disambiguator)*

**James A. Parsons** (b. ca. 1868 Steuben County, New York - March 4, 1945 Albar

#### Life

He was admitted to the bar of [Nebraska](#) in 1890, and moved back to [New York](#) in 1911, he was appointed Fourth Deputy Attorney General by [Thomas Carmody](#), a



**WIKIPEDIA**  
The Free Encyclopedia

Main page

Contents

Featured content

Current events

Random article

Interaction

About Wikipedia

Community portal

Recent changes

Contact Wikipedia

*Cook County, Illinois from 1960 to 1961.*

Parsons was a federal judge on the [United States District Court for the Northern District of Illinois](#) vacated by [Philip L. Sullivan](#). He was confirmed by the [United States Senate](#) on August 30, 1981. Parsons's service was terminated by the [United States District Court for the Northern District of Illinois](#) called The Benchwarmers that was very critical about Parsons. Goulden claim claimed that Parsons had sat on the bench while drunk and an overwhelming r

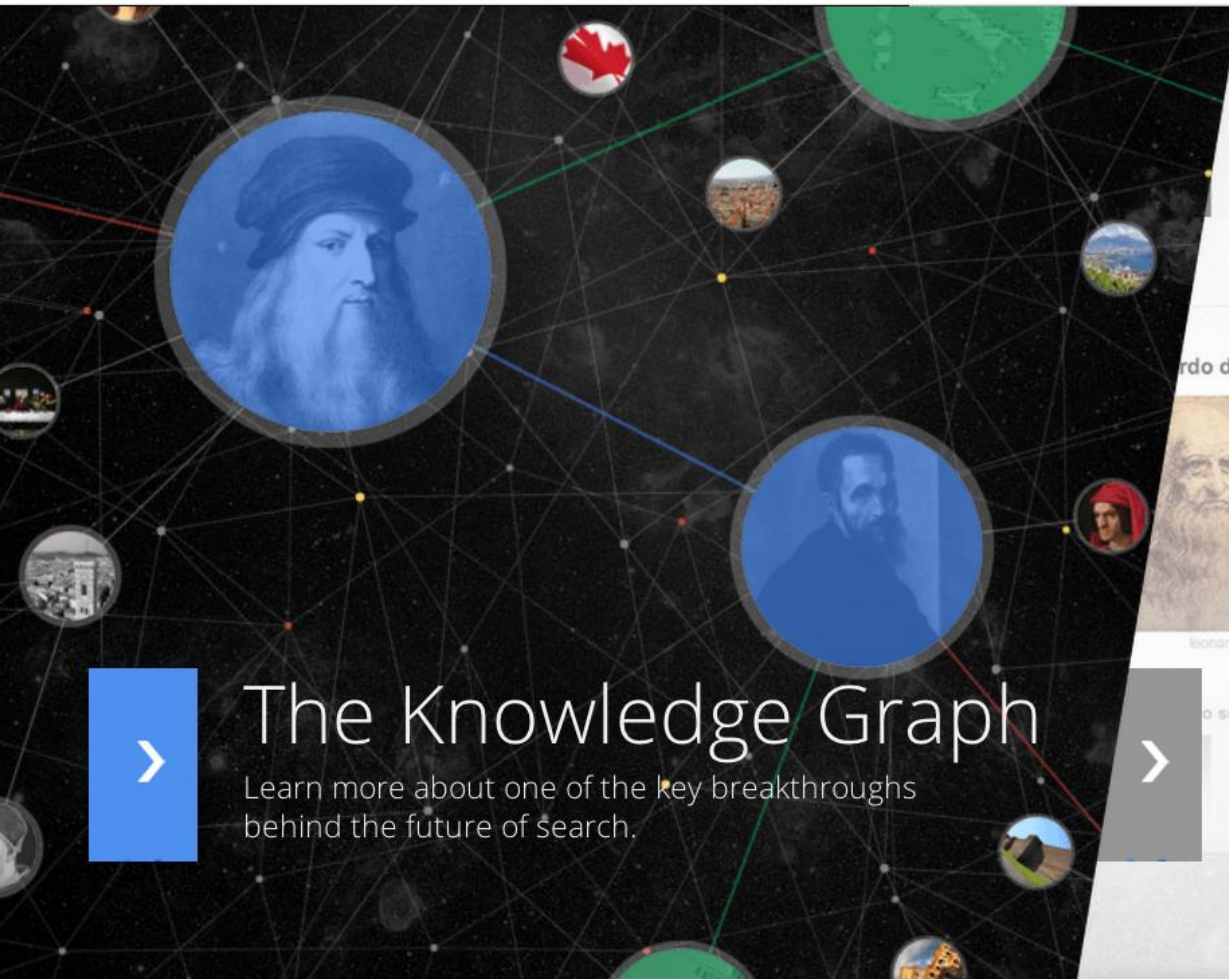
#### Jim Parsons



Query = "James Parsons"

# Google Knowledge Graph (2012)

Home Tips & Tricks **Features** Search Stories Playground Blog Help



## The Knowledge Graph

Learn more about one of the key breakthroughs behind the future of search.



### Leonardo da Vinci



Leonardo di ser Piero da Vinci was an Italian Renaissance polymath: painter, sculptor, architect, musician, scientist, mathematician, engineer, inventor, anatomist, geologist, cartographer, botanist, and writer. [Wikipedia](#)

**Born:** April 15, 1452, [Anchiano](#)

**Died:** May 2, 1519, [Clos Lucé](#)

**Buried:** [Château d'Amboise](#)

**Parents:** [Caterina da Vinci](#), [Piero da Vinci](#)

**Structures:** [Vebjam Sand Da Vinci Project](#)



## See it in action

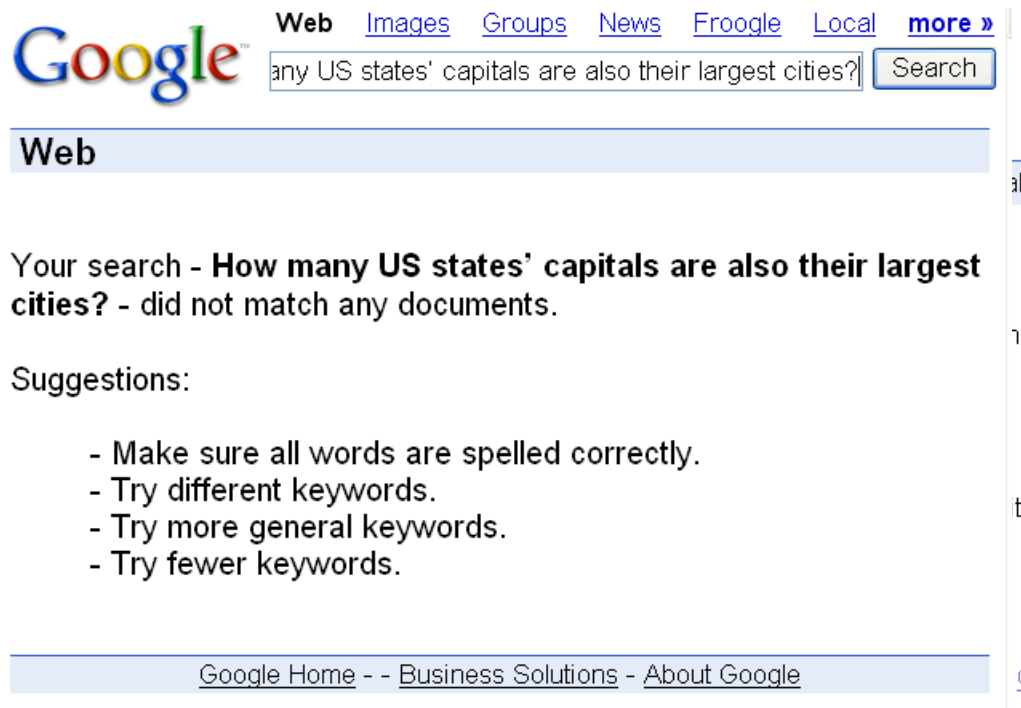
Discover answers to questions you never thought to ask, and explore collections and lists.



# QA / NL Interaction

- Question Answering:
  - More than search
  - Can be really easy: “What’s the capital of Wyoming?”
  - Can be harder: “How many US states’ capitals are also their largest cities?”
  - Can be open ended: “What are the main issues in the global warming debate?”

- Natural Language Interaction:
  - Understand requests and act on them
  - “Make me a reservation for two at Quinn’s tonight”



Google Web Images Groups News Froogle Local more »

any US states' capitals are also their largest cities? Search

Web

Your search - **How many US states' capitals are also their largest cities?** - did not match any documents.

Suggestions:

- Make sure all words are spelled correctly.
- Try different keywords.
- Try more general keywords.
- Try fewer keywords.

Google Home - - Business Solutions - About Google

## [capital of Wyoming: Information From Answers.com](#)

Note: click on a word meaning below to see its connections and related words.

The noun **capital** of **Wyoming** has one meaning: Meaning #1 : the **capital**.

[www.answers.com/topic/capital-of-wyoming](#) - 21k - [Cached](#) - [Similar pages](#)

## [Cheyenne: Weather and Much More From Answers.com](#)


Chey·enne ( shī-ăn ' , -ěň ' ) The **capital** of **Wyoming**, in the southeast part of the state near the Nebraska and Colorado borders.

[www.answers.com/topic/cheyenne-wyoming](#) - 74k - [Cached](#) - [Similar pages](#)

AI Overview



There are **17 U.S. states** where the state capital is also the state's largest city by population, including **Phoenix (Arizona), Atlanta (Georgia), Honolulu (Hawaii), Indianapolis (Indiana), Boston (Massachusetts), Nashville (Tennessee), and Salt Lake City (Utah)**, among others, with Columbus, Ohio, recently joining this group after surpassing Cleveland and Cincinnati. [↗](#)

 Reddit · r/MapPorn  
60+ comments · 7 years ago

The 17 US states (purple) whose capital is also that ...

The **17 US states** (purple) whose capital is also that state's largest city by population, as of 2016 estimates [2000x1237]. [Read more](#)



how many US states' capitals are also their largest cities by area

AI Mode All Images Shopping News Videos Web More Tools

### AI Overview



En



Listen



About 17 U.S. states have capitals that are also their largest cities by population, including Phoenix (AZ), Denver (CO), Boston (MA), and Columbus (OH), while only a few, like Juneau (AK), Oklahoma City (OK), and Nashville (TN), are largest by area due to consolidated city-county governments, with most states having separate, larger economic centers.

# Jeopardy! World Champion



US Cities: Its largest airport is named for a World War II hero; its second largest, for a World War II battle.

<http://www.youtube.com/watch?v=qpKolfTukrA>



# Coreference

---

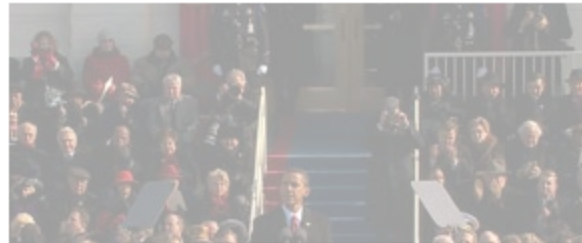


- But **the little prince** could not restrain admiration:
- "Oh! How beautiful you are!"
- "Am I not?" the flower responded, sweetly. "And I was born at the same moment as the sun . . ."
- **The little prince** could guess easily enough that she was not any too modest--but how moving--and exciting--she was!
- "I think it is time for breakfast," she added an instant later. "If **you** would have the kindness to think of my needs--"
- And **the little prince**, completely abashed, went to look for a sprinkling-can of fresh water. So, **he** tended the flower.

# Summarization

- Condensing documents
  - Single or multiple docs
  - Extractive or synthetic
  - Aggregative or representative
- Very context-dependent!
- An example of analysis with generation

WASHINGTON (CNN) -- President Obama's inaugural address was cooler, more measured and reassuring than that of other presidents making it, perhaps, the right speech for the times.



Some inaugural addresses are known for their soaring, inspirational language. Like John F. Kennedy's in 1961: "Ask not what your country can do for you. Ask what you can do for your country."

Obama's address was less stirring, perhaps, but it was also more candid and down-to-earth.

"Starting today," the new president said, "we must begin

#### STORY HIGHLIGHTS

- Obama's address less stirring than others but more candid, analyst says
- Schneider: At a time of crisis, president must be reassuring
- Country has chosen "hope over fear, unity of purpose over ... discord," Obama said
- Obama's speech was a cool speech, not a hot one, Schneider says

CNN

President Obama renewed his call for a massive plan to stimulate economic growth.

[more photos »](#)

aid in his first inaugural in 1933, "The only thing we have to fear is fear itself." Or Bill Clinton, who took office during the economic crisis of the early 1990s. "There is nothing wrong with America that cannot be fixed by what is right with America," Clinton declared at his first inaugural.

[Obama](#), too, offered reassurance.

"We gather because we have chosen hope over fear, unity of purpose over conflict and discord," Obama said.

Obama's call to unity after decades of political division echoed Abraham Lincoln's first inaugural address in 1861. Even though he delivered it at the onset of a terrible civil war, Lincoln's speech was not a call to battle. It was a call to look beyond the war, toward reconciliation based on what he called "the better angels of our nature."

Some presidents used their [inaugural address](#) to set out a bold agenda.

# Machine Translation

## "Il est impossible aux journalistes de rentrer dans les régions tibétaines"

Bruno Philip, correspondant du "Monde" en Chine, estime que les journalistes de l'AFP qui ont été expulsés de la province tibétaine du Qinghai "n'étaient pas dans l'illégalité".

**Les faits** Le dalaï-lama dénonce l'"enfer" imposé au Tibet depuis sa fuite, en 1959

**Vidéo** Anniversaire de la rébellion tibétaine: la Chine sur ses gardes



## "It is impossible for journalists to enter Tibetan areas"

Philip Bruno, correspondent for "World" in China, said that journalists of the AFP who have been deported from the Tibetan province of Qinghai "were not illegal."

**Facts** The Dalai Lama denounces the "hell" imposed since he fled Tibet in 1959

**Video** Anniversary of the Tibetan rebellion: China on guard



- Translate text from one language to another
- Recombines fragments of example translations
- Challenges:
  - What fragments? [learning to translate]
  - How to make efficient? [fast translation search]
  - Fluency vs fidelity

# English -- Russian

---

- *The spirit is willing but the flesh is weak. (English)*
- *The vodka is good but the meat is rotten. (Russian)*

# Language Comprehension/Common-Sense

---

"The rock was still wet. The animal was glistening, like it was still swimming," recalls Hou Xiangang. Hou discovered the unusual fossil while surveying rocks as a paleontology graduate student in 1984, near the Chinese town of Chengjiang. "My teachers always talked about the Burgess Shale animals. It looked like one of them. My hands began to shake." Hou had indeed found a *Naraoia* like those from Canada. However, Hou's animal was 15 million years older than its Canadian relatives.

It can be inferred that Hou Xiangang's "hands began to shake", because he was:

- (A) afraid that he might lose the fossil
- (B) worried about the implications of his finding
- (C) concerned that he might not get credit for his work
- (D) uncertain about the authenticity of the fossil
- (E) excited about the magnitude of his discovery

# ConvAI/ChatBots/Dialog Systems

---



Write a tweet about the importance of learning AI and ML learning skills and embracing the new possibilities of this sector.



"Embracing #AI and #MachineLearning is more than a trend; it's a pathway to the future. 🚀  
Acquiring these skills opens doors to innovation, creativity, and endless possibilities in tech.  
Don't miss out on shaping tomorrow. Invest in learning today! 📖 🌐 #FutureTech #LearnAI"



Write a tweet about the importance of learning AI and ML learning skills and embracing the new possibilities of this sector.

My goal is to entice readers to subscriber to my newsletter and download my guides for using automation to become more productive.



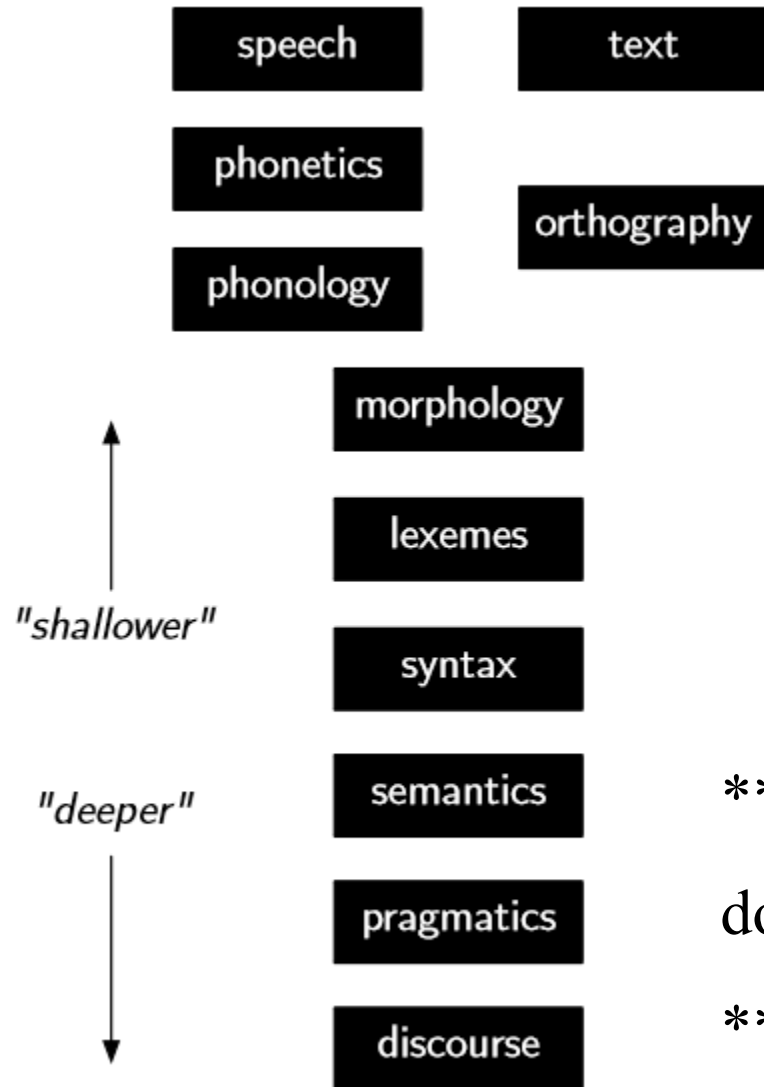
"Unlock the power of #AI and #MachineLearning! 🤖 ✨ Enhance productivity and discover new horizons with our expert guides. Subscribe to our newsletter and download FREE tools to automate your life. Don't just keep up with the future – lead it! 🔗 [Your Link Here] #AutomationMagic"



🔄 Regenerate

# Levels of Linguistic Knowledge

---



\*\*we can go up, down and up and down and combine steps too!!

\*\*every step is equally complex

# Why is NLP hard?

---

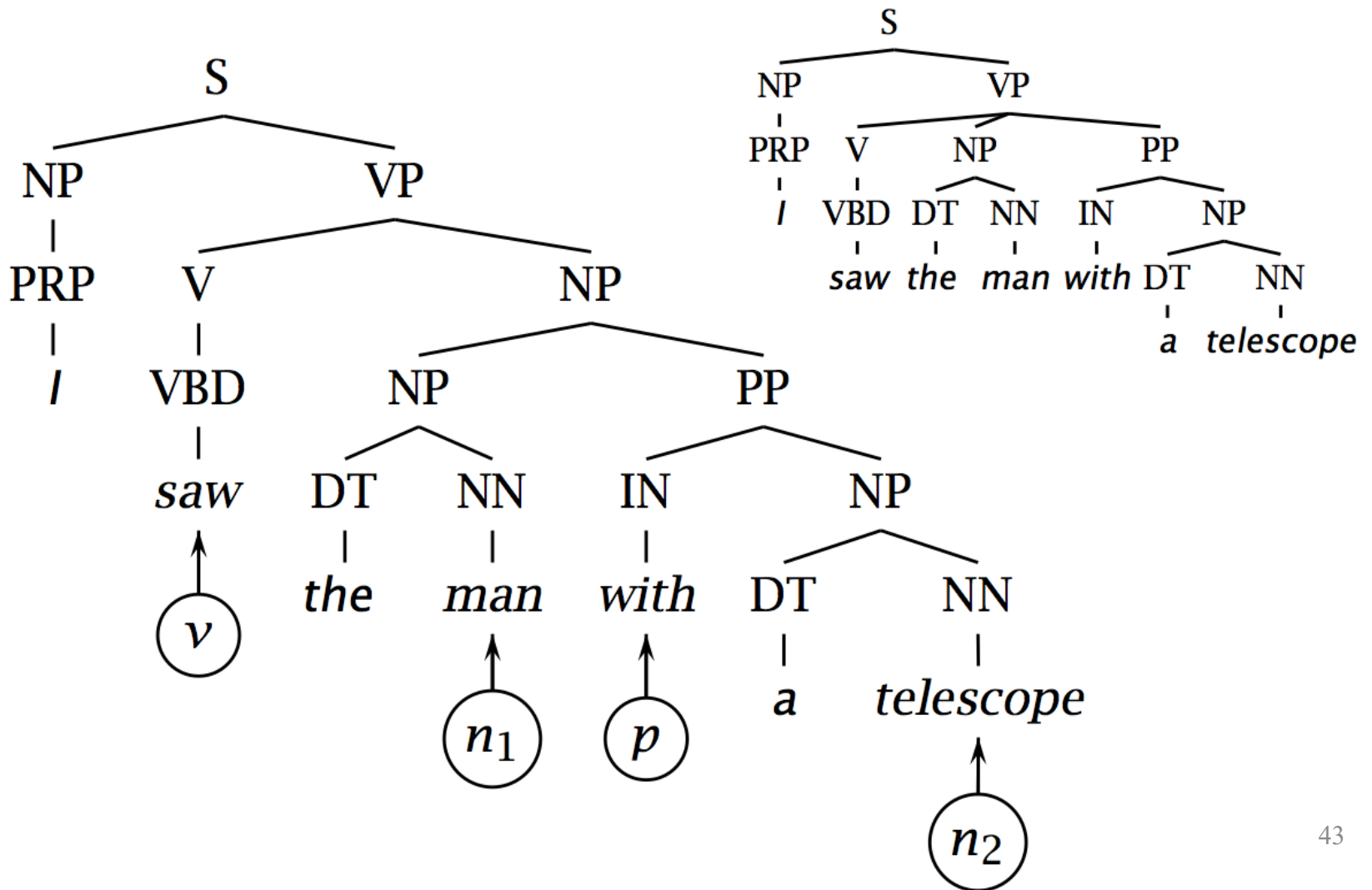
- Ambiguity
- Ambiguity
- Ambiguity
- Sparsity

# Ambiguities

---

- Headlines: Why are these funny?
  - Ban on Nude Dancing on Governor's Desk
  - Iraqi Head Seeks Arms
  - Juvenile Court to Try Shooting Defendant
  - Teacher Strikes Idle Kids
  - Stolen Painting Found by Tree
  - Local High School Dropouts Cut in Half
  - Red Tape Holds Up New Bridges
  - Clinton Wins on Budget, but More Lies Ahead
  - Hospitals Are Sued by 7 Foot Doctors
  - Kids Make Nutritious Snacks

# PP Attachment Ambiguity



# Coordination Ambiguity

---

Modifier	% of Total	Example from Surveys (modifiers underlined)
Noun	46.4	( It ) targeted the project and election <u>managers</u>
Adjective	23.2	... define <u>architectural</u> components and connectors
Prep	15.9	Facilitate the scheduling and performing <u>of works</u>
Verb	5.8	capacity and network resources <u>required</u>
Adverb	4.4	( It ) might be <u>automatically</u> rejected or flagged
Rel. Clause	2.2	Assumptions and dependencies <u>that are of importance</u>
Number	0.7	<u>zero</u> mean values and standard deviation
Other	1.4	increased by the <u>lack of</u> funding and local resources

# Semantic Ambiguity

---

*At last, a computer that understands you like your mother.*

- **Direct Meanings:**
  - It understands you like your mother (does) [presumably well]
  - It understands (that) you like your mother
  - It understands you like (it understands) your mother
- **But there are other possibilities, e.g. mother could mean:**
  - a woman who has given birth to a child
  - a stringy slimy substance consisting of yeast cells and bacteria; is added to cider or wine to produce vinegar
- **Context matters, e.g. what if previous sentence was:**
  - Wow, Amazon predicted that you would need to order a big batch of new vinegar brewing ingredients. 😊

[Example from L. Lee]

# Ambiguities (contd)

---

- Get the cat with the gloves.



# Ambiguity

---

- Find at least 5 meanings of this sentence:
  - I made her duck

# Ambiguity

---

- Find at least 5 meanings of this sentence:
  - I made her duck
- I cooked waterfowl for her benefit (to eat)
- I cooked waterfowl belonging to her
- I created the (plaster?) duck she owns
- I caused her to quickly lower her head or body
- I waved my magic wand and turned her into undifferentiated waterfowl

# Ambiguity is Pervasive

---

- I caused her to quickly lower her head or body
  - **Lexical category:** “duck” can be a N or V
- I cooked waterfowl belonging to her.
  - **Lexical category:** “her” can be a possessive (“of her”) or dative (“for her”) pronoun
- I made the (plaster) duck statue she owns
  - **Lexical Semantics:** “make” can mean “create” or “cook”

# Ambiguity is Pervasive

---

- **Grammar: Make can be:**
  - **Transitive: (verb has a noun direct object)**
    - I cooked [waterfowl belonging to her]
  - **Ditransitive: (verb has 2 noun objects)**
    - I made [her] (into) [undifferentiated waterfowl]
  - **Action-transitive (verb has a direct object and another verb)**
    - I caused [her] [to move her body]

# Ambiguity is Pervasive

---

- **Phonetics!**
  - I mate or duck
  - I'm eight or duck
  - Eye maid; her duck
  - Aye mate, her duck
  - I maid her duck
  - I'm aid her duck
  - I mate her duck
  - I'm ate her duck
  - I'm ate or duck
  - I mate or duck

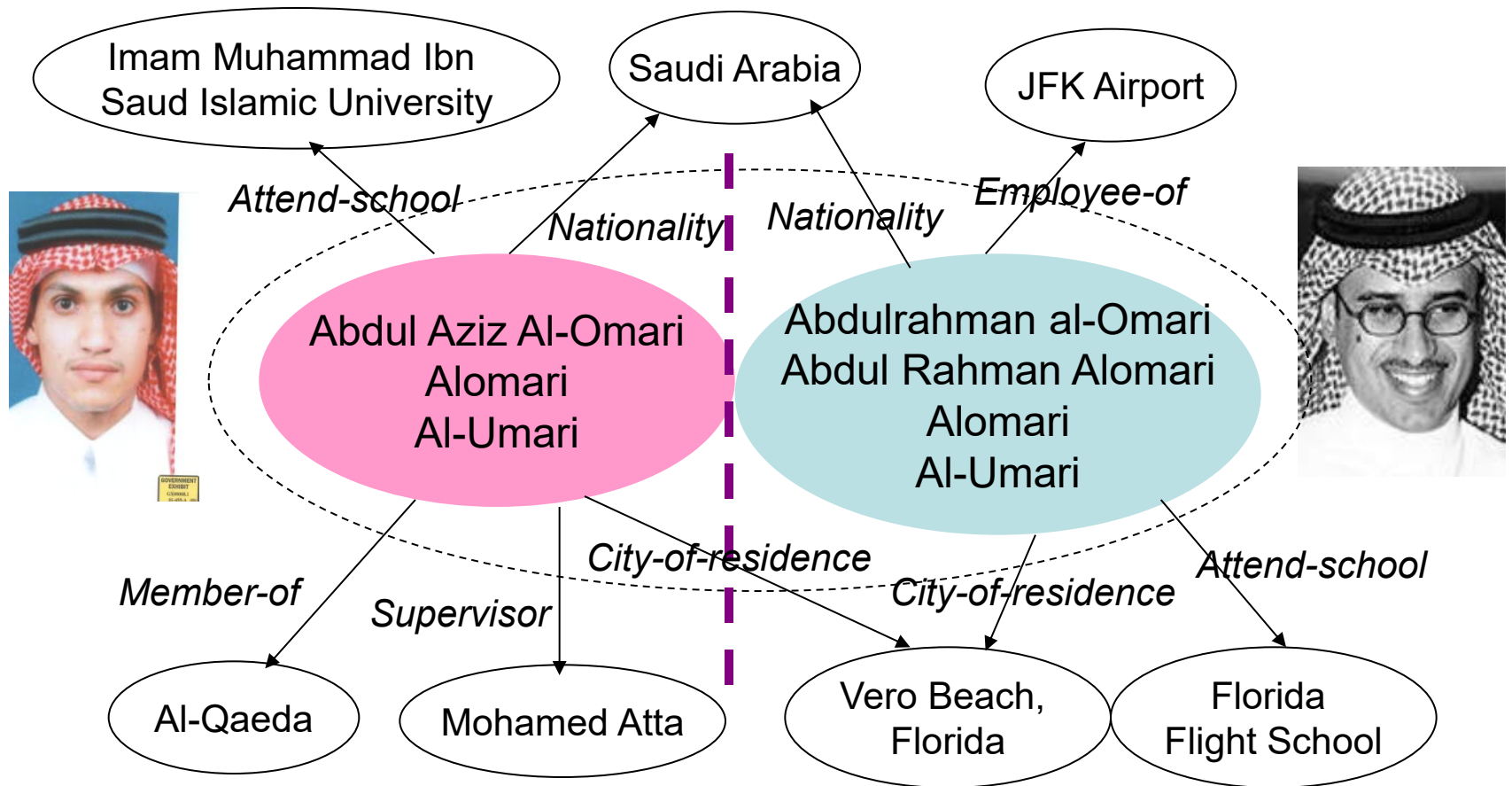
# Abbreviations out of context

- Medical Domain: 33% of abbreviations are ambiguous (Liu et al., 2001), major source of errors in medical NLP (Friedman et al., 2001)

RA	“rheumatoid arthritis”, “renal artery”, “right atrium”, “right atrial”, “refractory anemia”, “radioactive”, “right arm”, “rheumatic arthritis”, ...
PN	“Penicillin”; “Pneumonia”; “Polyarteritis”; “Nodosa”; “Peripheral neuropathy”; “Peripheral nerve”; “Polyneuropathy”; “Pyelonephritis”; “Polyneuritis”; “Parenteral nutrition”; “Positional Nystagmus”; “Periarteritis nodosa”, ...

- Military Domain
  - “*GA ADT 1, USDA, USAID, Turkish PRT, and the DAIL staff met to create the Wardak Agricultural Steering Committee.*”
  - “*DST*” = “*District Stability Team*” or “*District Sanitation Technician*”?

# Uncertainty: Ambiguity Example



# Problem: Sparsity

- New unigram (word), bigram (word pair)
- Solution: embedding (deep learning) solutions

