

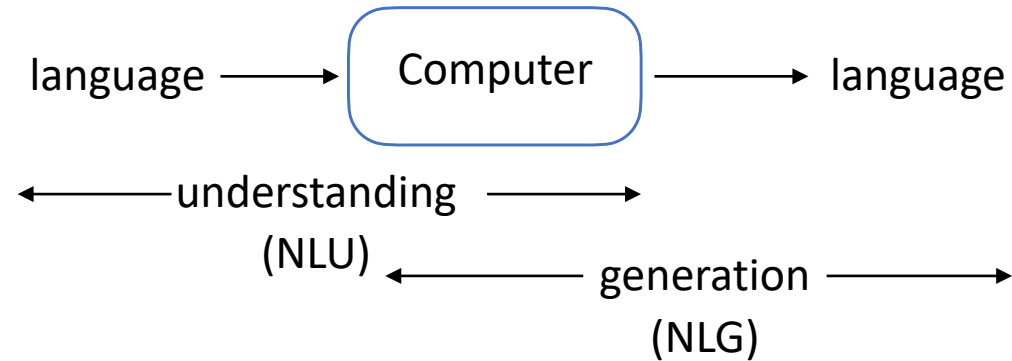
Introduction to NLP

Vardaan Pahuja

Slide credit: Dan Klein

What is NLP?

- Natural Language Processing (NLP) is a field in Artificial Intelligence (AI) devoted to creating computers that use natural language as input and/or output.



Example: Siri

- Siri contains
 - Speech recognition
 - Language Analysis
 - Dialog processing
 - Text to speech

Text Data is Superficial

An iceberg is a large piece of freshwater ice that has broken off from a snow-formed glacier or ice shelf and is floating in open water.



... But Language is Complex



Problem: Ambiguities

Find at least 5 meanings of this sentence:

"I made her duck"

1. I cooked waterfowl for her benefit (to eat)
2. I cooked waterfowl belonging to her
3. I created the (plaster?) duck she owns
4. I caused her to quickly lower her head or body
5. I waved my magic wand and turned her into undifferentiated waterfowl

Problem: Ambiguities

- Headlines:

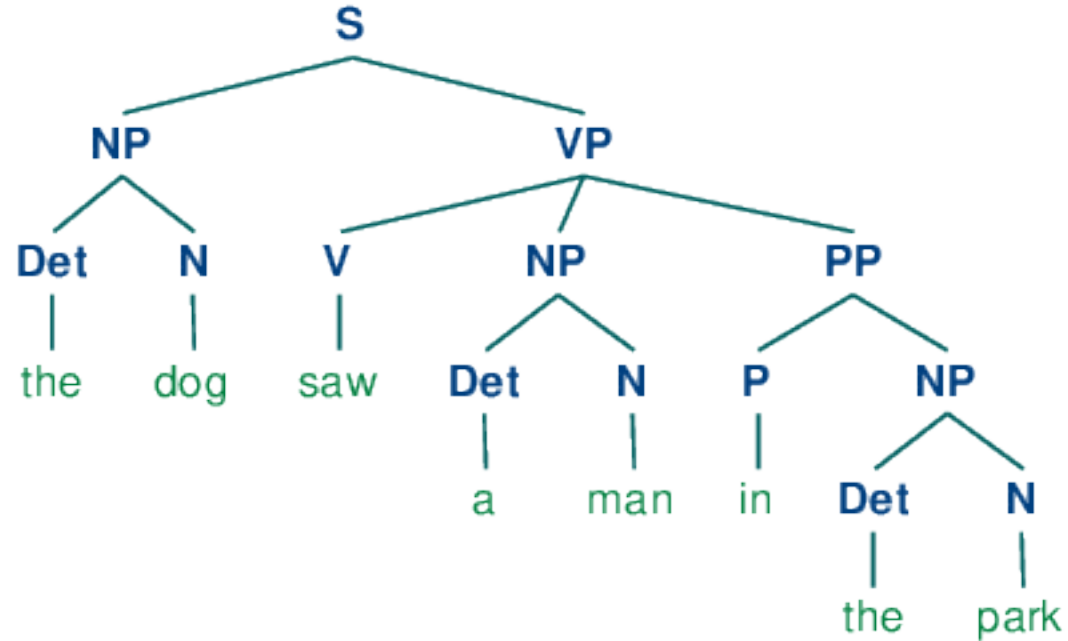
- Juvenile Court to Try Shooting Defendant
- Enraged Cow Injures Farmer with Ax
- Ban on Nude Dancing on Governor's Desk
- Teacher Strikes Idle Kids
- Hospitals Are Sued by 7 Foot Doctors
- Iraqi Head Seeks Arms
- Stolen Painting Found by Tree
- Kids Make Nutritious Snacks
- Local HS Dropouts Cut in Half

- Why are these funny?

Different Levels of Linguistic Analysis

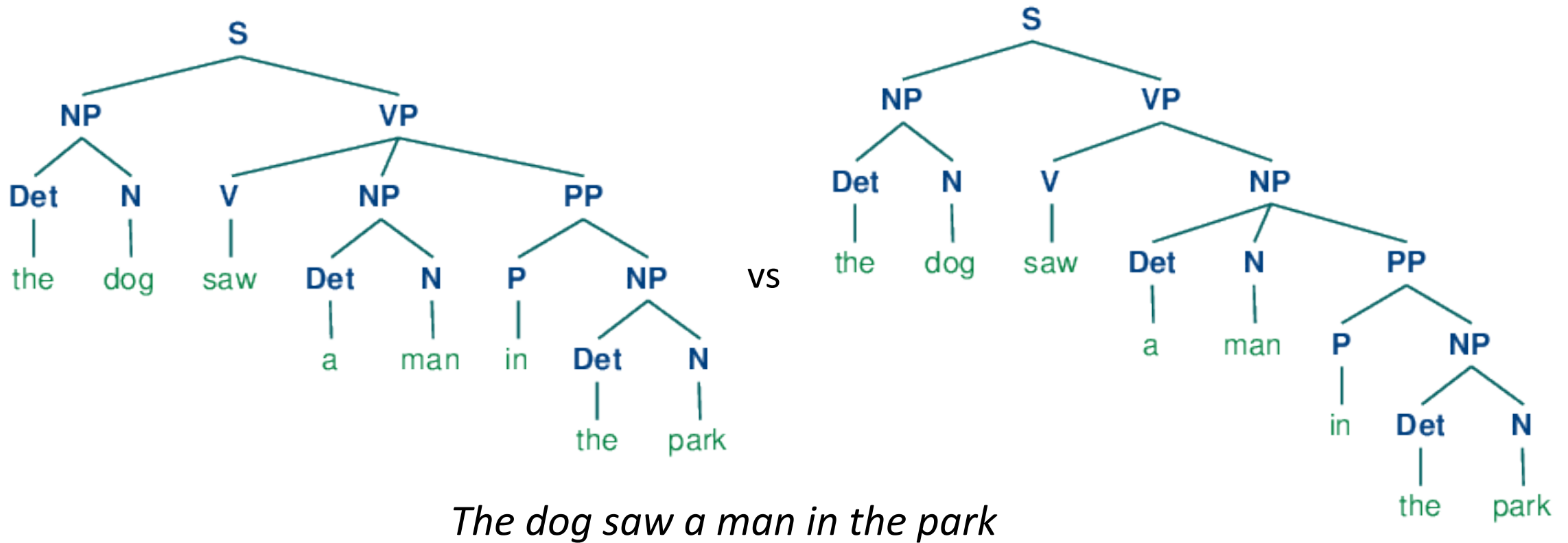
- **Phonology**
 - Speech audio signal to phonemes
- **Morphology**
 - Inflection (e.g. “I”, “my”, “me”; “eat”, “eats”, “ate”, “eaten”)
 - Derivation (e.g. “teach”, “teacher”, “nominate”, “nominee”)
- **Syntax**
 - Part-of-speech (noun, verb, adjective, preposition, etc.)
 - Phrase structure (e.g. noun phrase, verb phrase)
- **Semantics**
 - Meaning of a word (e.g. “book” as a bound volume or an accounting ledger) or a sentence
- **Discourse**
 - Meaning and inter-relation between sentences

Syntactic Analysis



The dog saw a man in the park

Challenges: Ambiguity



Challenges: Ambiguity

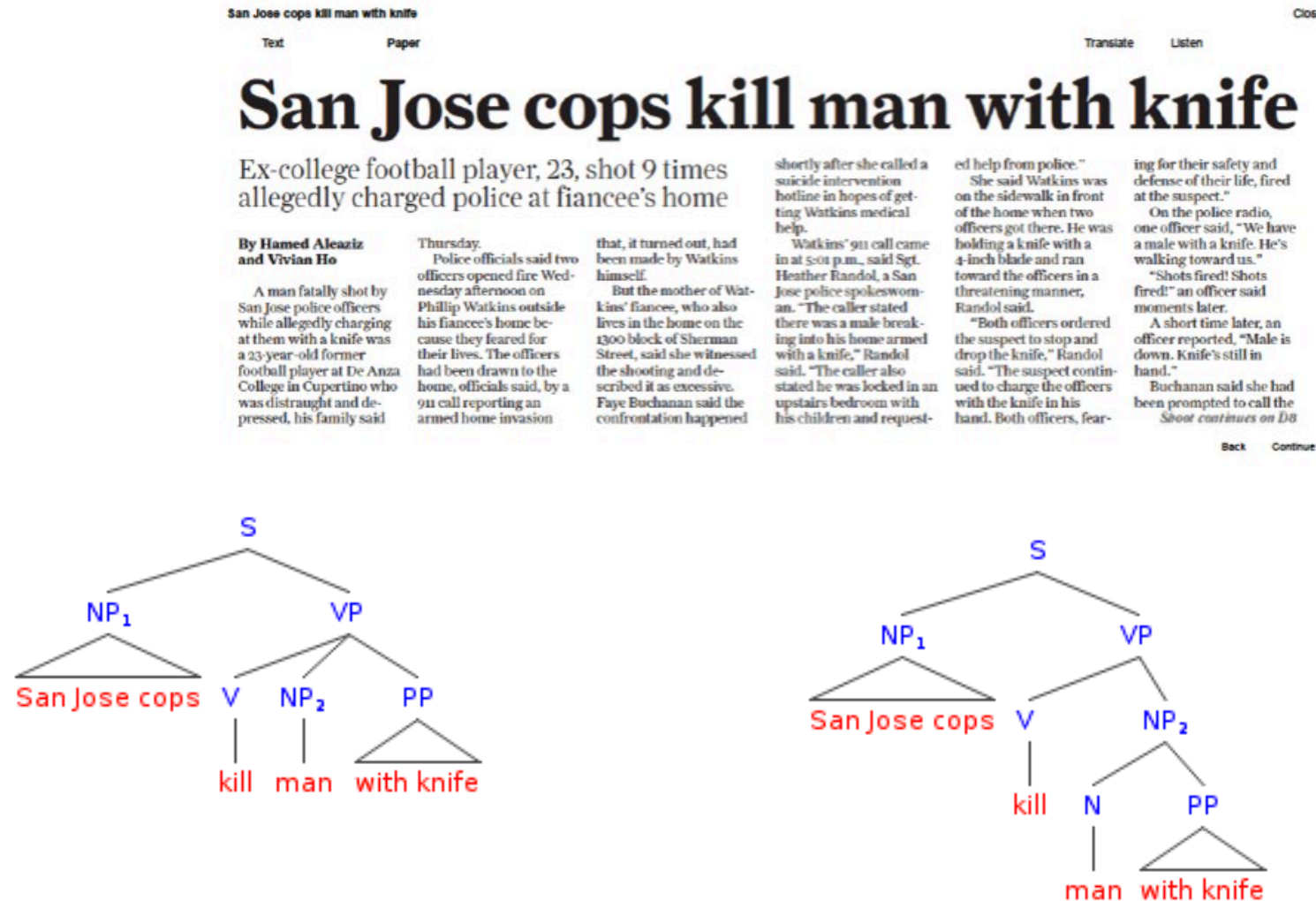


Figure credit: Mark Liberman

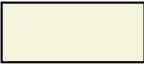
Constituency Parsing Demo

- <https://corenlp.run>

Corpora

- A corpus is a collection of text
 - Often annotated in some way
 - Sometimes just lots of text
 - Balanced corpora
- Examples
 - Newswire collections: 500M+ words
 - Brown corpus: 1M words of tagged “balanced” text
 - Penn Treebank: 1M words of parsed WSJ
 - Canadian Hansards: 10M+ words of aligned French / English sentences
 - The Web: billions of words of who knows what

Data and Knowledge

- Classic knowledge representation worry: How will a machine ever know that...
 - Ice is frozen water?
 - Beige looks like this: 
 - Chairs are solid?
- Answers:
 - 1980: write it all down
 - 2000: get by without it
 - 2020: learn it from data

Word Embeddings

- Representing words by their context
- Distributional semantics
 - “You shall know a word by the company it keeps” (Firth, 1957)

Word Embeddings

expect =

$\begin{pmatrix} 0.286 \\ 0.792 \\ -0.177 \\ -0.107 \\ 0.109 \\ -0.542 \\ 0.349 \\ 0.271 \\ 0.487 \end{pmatrix}$



Word2vec Overview

Idea:

- We have a large corpus of text
- Every word in a fixed vocabulary is represented by a **vector**
- Go through each position t in the text, which has a center word c and context (“outside”) words o
- Use the **similarity of the word vectors** for c and o to calculate the **probability** of o given c (or vice versa)
- **Keep adjusting the word vectors** to maximize this probability

Gender bias in Word Embeddings

- Analogies of embedding vectors

$$\overrightarrow{\text{man}} - \overrightarrow{\text{woman}} \approx \overrightarrow{\text{king}} - \overrightarrow{\text{queen}}$$

- Sexist analogies

$$\overrightarrow{\text{man}} - \overrightarrow{\text{woman}} \approx \overrightarrow{\text{computer programmer}} - \overrightarrow{\text{homemaker}}$$

Gender stereotype *she-he* analogies.

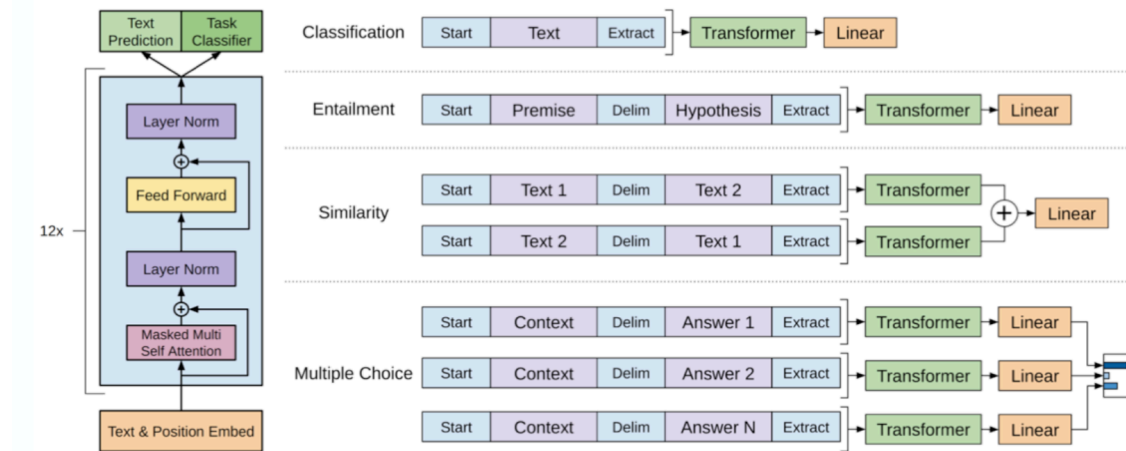
sewing-carpentry	register-nurse-physician	housewife-shopkeeper
nurse-surgeon	interior designer-architect	softball-baseball
blond-burly	feminism-conservatism	cosmetics-pharmaceuticals
giggle-chuckle	vocalist-guitarist	petite-lanky
sassy-snappy	diva-superstar	charming-affable
volleyball-football	cupcakes-pizzas	hairdresser-barber

Gender appropriate *she-he* analogies.

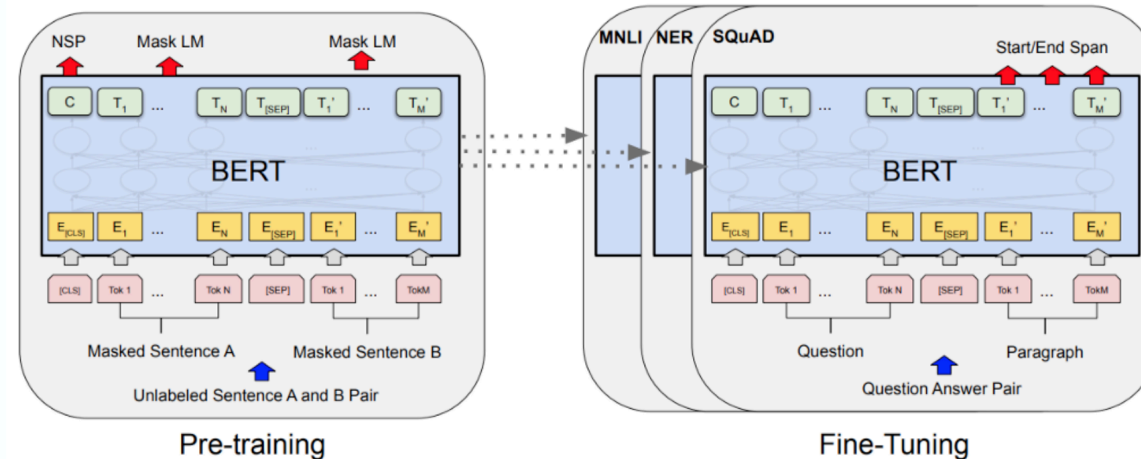
queen-king	sister-brother	mother-father
waitress-waiter	ovarian cancer-prostate cancer	convent-monastery

Contextualized Word Embeddings

OpenAI GPT [Radford et al. 2018]

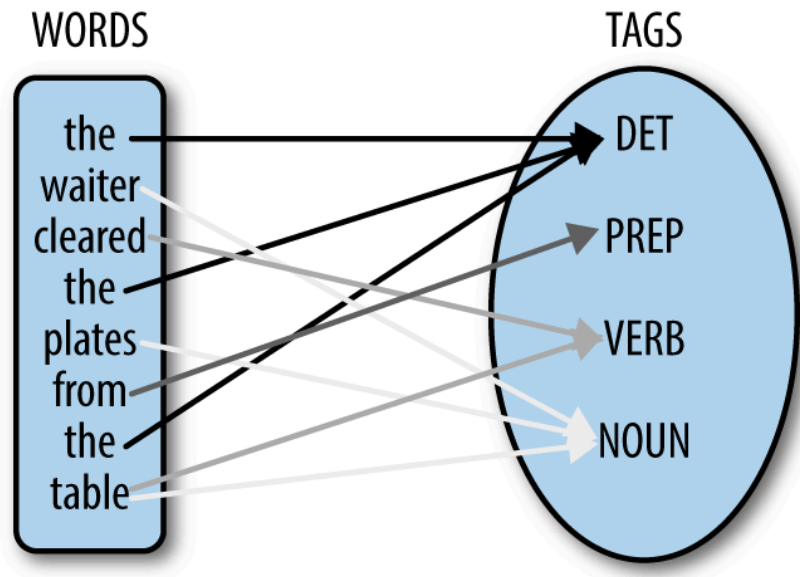


BERT [Devlin et al. 2019]



POS tagging

- Assign Part-of-Speech (POS) tag to each word in a sentence.



Tag	Meaning	English Examples
ADJ	adjective	new, good, high, special, big, local
ADP	adposition	on, of, at, with, by, into, under
ADV	adverb	really, already, still, early, now
CONJ	conjunction	and, or, but, if, while, although
DET	determiner, article	the, a, some, most, every, no, which
NOUN	noun	year, home, costs, time, Africa
NUM	numeral	twenty-four, fourth, 1991, 14:24
PRT	particle	at, on, out, over per, that, up, with
PRON	pronoun	he, their, her, its, my, I, us
VERB	verb	is, say, told, given, playing, would
.	punctuation marks	. , ; !
X	other	ersatz, esprit, dunno, gr8, univeristy

Question Answering

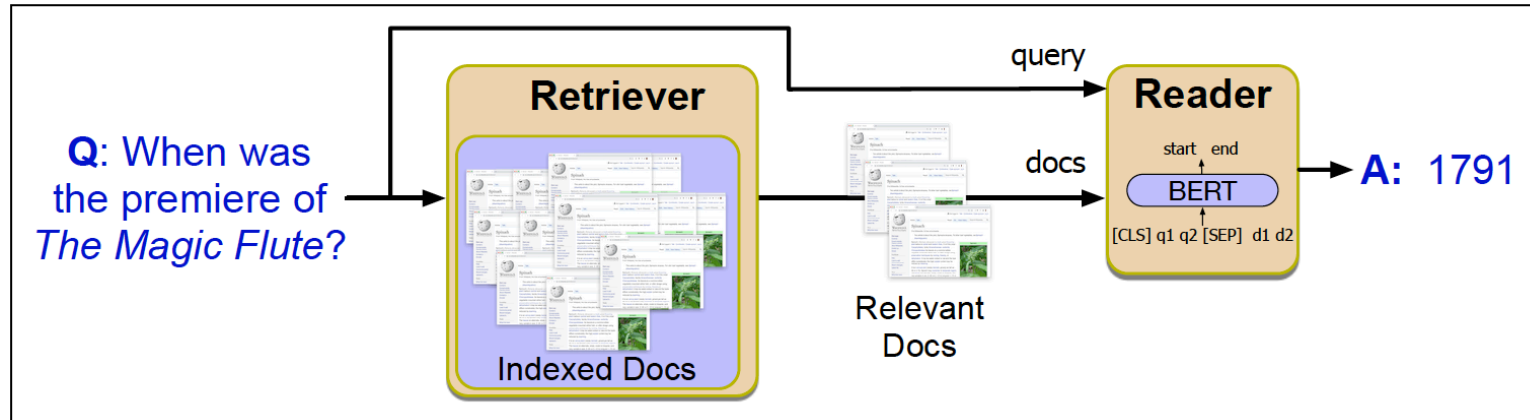
- Question Answering:
 - More than search
 - Ask general comprehension questions of a document collection
 - Can be really easy: “What’s the capital of Wyoming?”
 - Can be harder: “How many US states’ capitals are also their largest cities?”
 - Can be open ended: “What are the main issues in the global warming debate?”
- SOTA: Can do factoids, even when text isn’t a perfect match

Question Answering

- Factoid Question Answering:
 - Questions can be answered with simple facts
 - For example: What is the capital of United States?
- Open-domain QA
- Knowledge-based QA

Open-domain QA

- Question answering based on vast amounts of text (Web or large corpora)



- **Retriever**: retrieve relevant passages using a search engine
- **Reader**: Find the span corresponding to the answer.
- Datasets: TriviaQA, SearchQA etc.
- Demo: <http://qa.cs.washington.edu:2020>

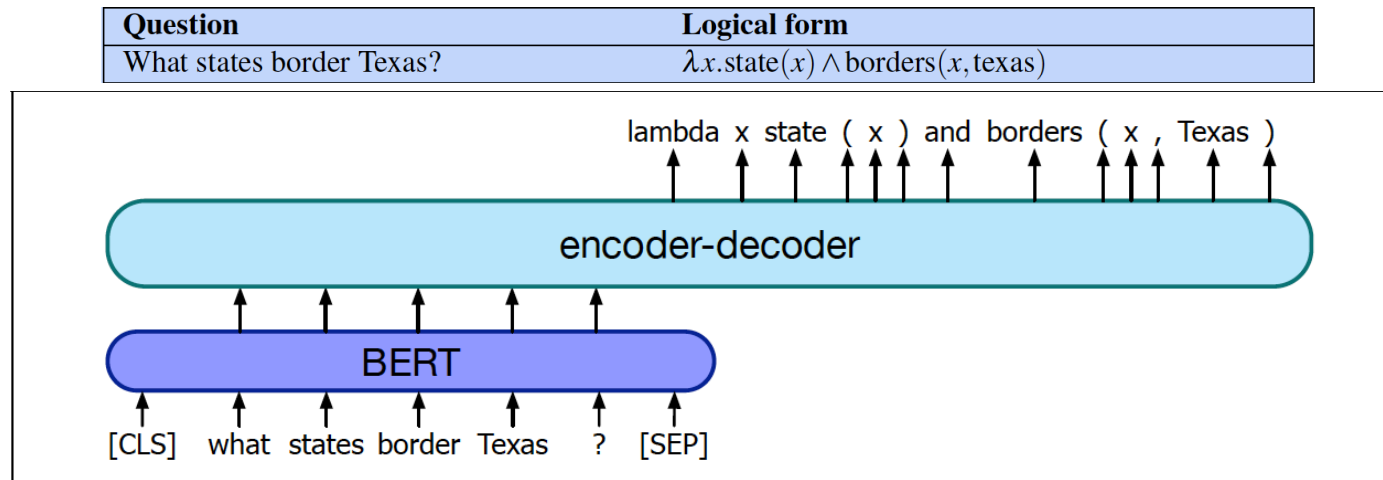
Knowledge-based QA

- Answer a question by converting it into a query over a structured database.
- Use a **semantic parser** to map the question to a structured program to produce an answer

Question	Logical form
What states border Texas?	$\lambda x. \text{state}(x) \wedge \text{borders}(x, \text{texas})$
What is the largest state?	$\text{argmax}(\lambda x. \text{state}(x), \lambda x. \text{size}(x))$
I'd like to book a flight from San Diego to Toronto	<pre>SELECT DISTINCT f1.flight_id FROM flight f1, airport_service a1, city c1, airport_service a2, city c2 WHERE f1.from_airport=a1.airport_code AND a1.city_code=c1.city_code AND c1.city_name= 'san diego' AND f1.to_airport=a2.airport_code AND a2.city_code=c2.city_code AND c2.city_name= 'toronto'</pre>
How many people survived the sinking of the Titanic?	<pre>(count (!fb:event.disaster.survivors fb:en.sinking_of.the.titanic))</pre>

Knowledge-based QA

- Encoder-decoder based Semantic Parser



QA system (Watson)

- IBM's Watson question-answering system won the TV game-show Jeopardy!, surpassing human competitors
- Example question: "Victory in 1805's Battle of Derna on the coast of North Africa inspired a lyric in this song made official in 1929"



Reading Comprehension

- Read a passage of text and answer comprehension questions

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called “showers”.

What causes precipitation to fall?

gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?

graupel

Where do water droplets collide with ice crystals to form precipitation?

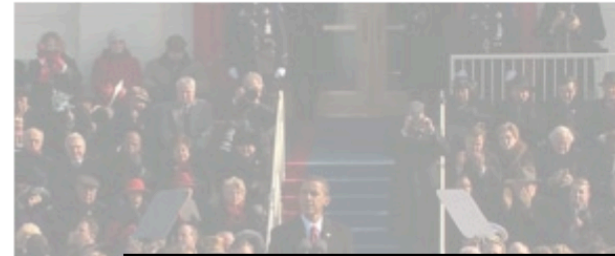
within a cloud

- Demo: <https://www.youtube.com/watch?v=nbsWpcJASMU>

Summarization

- Condensing documents
- 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.

"Stating 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.

Extractive Summaries

Lindsay Lohan pleaded not guilty Wednesday to felony grand theft of a \$2,500 necklace, a case that could return the troubled starlet to jail rather than the big screen. Saying it appeared that Lohan had violated her probation in a 2007 drunken driving case, the judge set bail at \$40,000 and warned that if Lohan was accused of breaking the law while free he would have her held without bail. The Mean Girls star is due back in court on Feb. 23, an important hearing in which Lohan could opt to end the case early.

3 NEWS
310WS.CO.UZ

MADEIRA SAM IS AZEEM LINDA GEORGIE JO

Pickering murder accused warned he would kill - guards

Home National World Sport Entertainment Business Politics Technology Entertainment/So

You are here: » Entertainment » Story 3 News Close | 2 News in a minute | Facebook

Full Story | View Video

Lindsay Lohan pleads not guilty to theft

0 comments | Post Comment

THU 10 FEB 2011 12:10 PM

Lindsay Lohan pleaded not guilty Wednesday to felony grand theft of a US\$2,500 necklace - a case that could return the troubled starlet to jail rather than the big screen.

She entered the plea through her attorney Shawn Chapman Holley after a judge scolded her about following the law.

"Look around this room," District Court Judge Keith Schwartz told Lohan. "Everybody else has to follow the law. You're no different than anyone else. So please, don't push your luck."

Schwartz set bail at US\$40,000. He said it appeared Lohan had violated her probation in a 2007 drunken driving case and warned that if she violated the law while free on bail he would have her arrested and held.

Lohan spoke little during the hearing except to acknowledge her name and that she understood the charge and the possible consequences.

The Mean Girls star was met by a swarm of photographers outside the courtroom as she walked to the judge's chambers.

Address Lindsay Lohan appears in court (Reuters)

VIEW VIDEO | Facebook | Twitter | YouTube

Sign Up to see what your friends like.

Get a quote now.
Save it, then buy online later.

WA today
LIFE & STYLE

Choose: Quizzes & enjoy: Genuine Aussie Hospitality. Complimentary food & wine. More than 100 on-site. The largest domestic.

Beauty | Wellbeing | Fashion | People | Home & Style | Life Matters | Weddings | Shopping | Culture

You are here: Home » Life & Style » People » Article

Lindsay Lohan charged with theft

February 10, 2011

Join the conversation
You're the only person reading this now. Tell your friends.

Top Life & Style articles

1. Tackling the final weight loss hurdle
2. Breaking the feedback loop
3. Time management
4. Western push for natives
5. Under the skin: What's her story?

More Life & Style articles

Story Tools

- Share on Facebook
- Email this story
- Print this story

WA Today Jobs
Accountancy | \$45,000 - \$50,000
East Perth, WA 6004
Adelaide Services

Judge warns Lohan: "Don't push your luck"

Lindsay Lohan pleads not guilty to felony grand theft of a \$2,500 necklace. The US judge said it appeared Lohan violated her probation and gave her a stern warning. Bail was set at \$40,000.

US actress Lindsay Lohan has pleaded not guilty to felony grand theft of a \$2,500 necklace - the most serious charge yet filed against the troubled starlet who has wrestled with drug and alcohol abuse for years.

"Saying it appeared that Lohan had violated her probation in a 2007 drunken driving case, the judge set bail at \$40,000 and warned that, if Lohan was accused of breaking the law while free, he would have her held without bail."

THE DAILY CALLER

HOME | POLITICS | US | ENTERTAINMENT | CRIME | WORLD | SPORTS | BUSINESS | TECH | LIFE | VISA

Log In | Register

Judge tells Lohan she's no star in his courtroom

By Andrew J. Pineda | Published: 5:11 PM 02/10/2011 | Updated: 10:08 AM 02/10/2011

LOS ANGELES (AP) — Lindsay Lohan walked into a courtroom to face a felony grand theft charge looking like a million dollars, only to be told by a judge she was no different than anyone else.

Lohan's arraignment on a charge that she stole a \$2,500 necklace from an upscale jeweler wasn't the first time a judge threatened to throw the troubled starlet in jail. But it was the first time a judge wielded enough power to keep her locked up for a long time.

"You're in a different situation now that a felony has been filed," Los Angeles Superior Court Judge Keith Schwartz said after the actress pleaded not guilty Wednesday.

"Everybody else has to follow the law," Schwartz said, noting that he was giving the actress a tamer version of a lecture he'd delivered to her attorney behind closed doors and away from the dozens of assembled reporters.

"You're no different than anyone else. So please, don't push your luck."

Testing the limits — in the courtroom rather than the big screen — has been Lohan's calling card in recent months.

The Mean Girls star is due back in court on Feb. 23, an important hearing in which Lohan could opt to end the case early. Her attorney, Shawn Chapman Holley, indicated Wednesday that the actress was interested in an early disposition program if the terms are right.

Machine Translation (French)

"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 : Le China 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

Ref: Lina, Reporter, Video



- Translate text from one language to another
- Recombines fragments of example translations

Machine Translation (Japanese)

Google Translate

Japanese - detected → English

Translation

Short-term decisive battle, by-election of the House of Councilors changed the situation

22:00



Early voting for the House of Representatives election, maximum 16.62 million people 10 days from the day after the announcement

22:11



項目	数値
総投票率	65.1%
投票者数	10.49
投票率	8.84
投票率	9.77
投票率	14.71
投票率	15.74

287 new infections confirmed nationwide The number of severely ill cases decreased by 11 to 134

22:13



項目	数値
新規感染者数	172,322人
重症患者数	170,521人
亡者数	1,827人

Tawaman disguised as seismic isolation rubber, the company of the party buys it for 3.7 billion yen

19:16



Kei Komuro failed the NY Bar Exam "I'll do my best" for the February exam

21:35



Unspoken Olympics "Honmani Enka" Hotel President of Uganda Team Accommodation

18:00



"Senior citizens" "Die alone", the era of malicious words Ryo Isobe

17:00



53 minutes ago

1 hour ago

1 hour ago

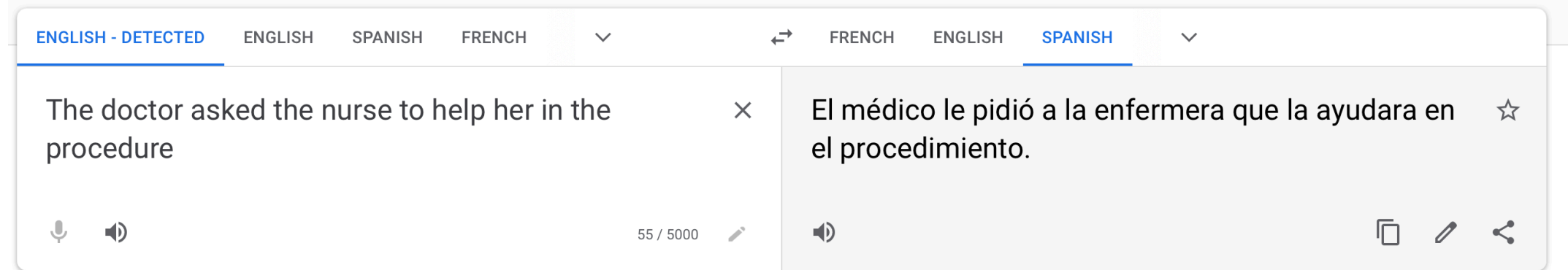
Workshop in Nagasaki to think about support for global hibakusha

Suspected bird flu in Nara is negative Judgment by genetic test of NARO

"I was able to sing with the

Attention information PR

Challenges: Gender bias



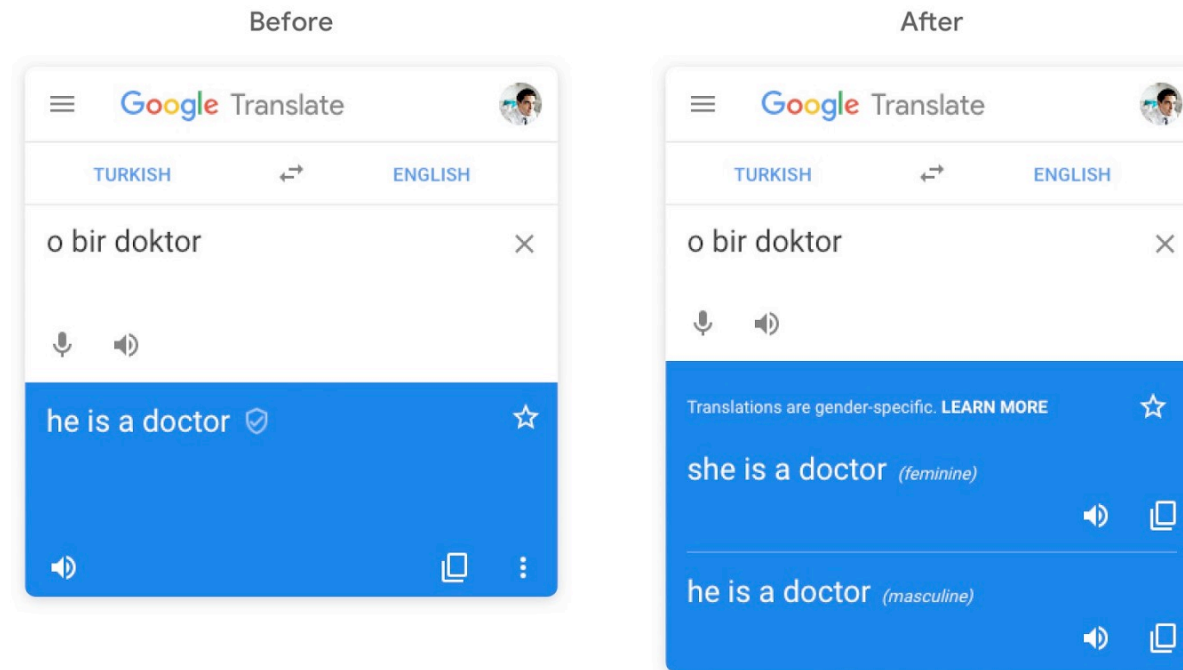
El **médico** le pidió a **la enfermera** que la ayudara en el procedimiento.

El **médico**: (male) doctor

la enfermera: (female) nurse

Challenges: Gender bias

- Recent attempts to mitigate gender biases



Coreference Resolution

- Cluster mentions in text which refer to the same real-world entities

Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.

Information Extraction

- Transform the unstructured information present in texts into structured form.

Citing high fuel prices, United Airlines said Friday it has increased fares by \$6 per round trip on flights to some cities also served by lower-cost carriers. American Airlines, a unit of AMR Corp., immediately matched the move, spokesman Tim Wagner said. United, a unit of UAL Corp., said the increase took effect Thursday and applies to most routes where it competes against discount carriers, such as Chicago to Dallas and Denver to San Francisco.



FARE-RAISE ATTEMPT:	LEAD AIRLINE:	UNITED AIRLINES
	AMOUNT:	\$6
	EFFECTIVE DATE:	2006-10-26
	FOLLOWER:	AMERICAN AIRLINES

IE: Relation Extraction

- Predict the relation corresponding to the relation statement given the entity mentions.

Input: Barack Obama was born in Honolulu, Hawaii as he always said.

Output: born_in

Entity Mentions:

Barack Obama, Honolulu

Input: Eiffel Tower is located in Paris, France.

Output: located_in

Entity Mentions:

Eiffel Tower, Paris

IE: Named Entity Recognition

- Identify the named entities in a sentence.
- For example: U.N. official Ekeus heads for Baghdad.
- [ORG U.N.] official [PER Ekeus] heads for [LOC Baghdad].
- Demo: <https://corenlp.run>

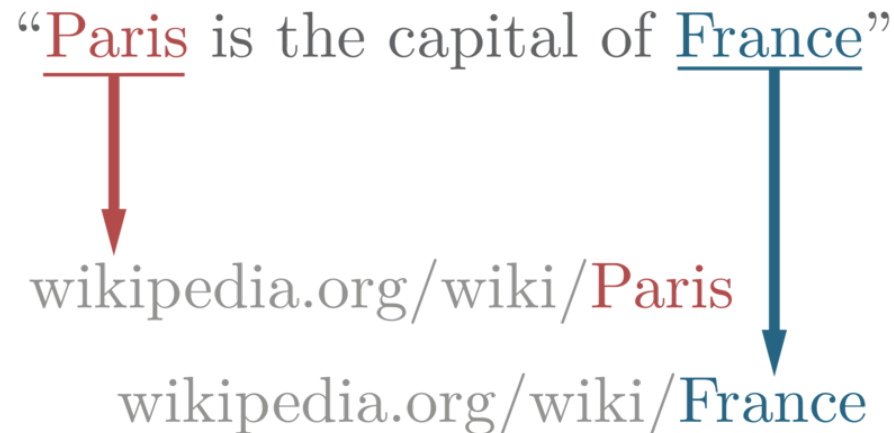
IE: Event Extraction

- Identify mentions of events in texts

[EVENT Citing] high fuel prices, United Airlines [EVENT said] Friday it has [EVENT increased] fares by \$6 per round trip on flights to some cities also served by lower-cost carriers. American Airlines, a unit of AMR Corp., immediately [EVENT matched] [EVENT the move], spokesman Tim Wagner [EVENT said]. United, a unit of UAL Corp., [EVENT said] [EVENT the increase] took effect Thursday and [EVENT applies] to most routes where it [EVENT competes] against discount carriers, such as Chicago to Dallas and Denver to San Francisco.

Entity Linking

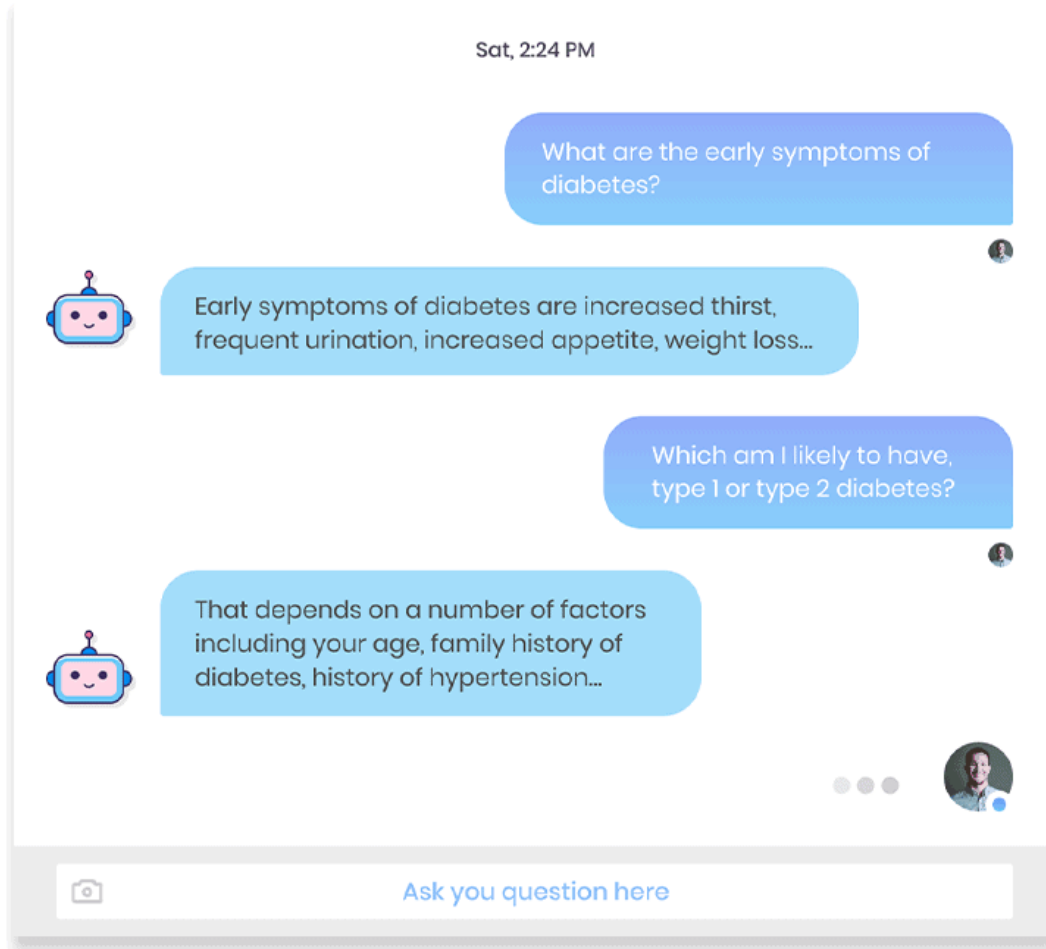
- **Task:** To identify the mentions of entities in a document and link them to entities in a Knowledge Graph.
- It is composed of two sub-tasks:
 - **Candidate Generation:** In this sub-task, the candidates of referent entities are generated for each mention.
 - **Mention Disambiguation:** Given the set of candidate entities, the task is to disambiguate the mention by selecting the most relevant entity.



Dialog Agents

- Digital Assistants
 - Siri
 - Alexa
 - Google Home
 - Cortona
- Task-oriented dialogue agents
 - Use conversation with humans to complete tasks e.g. order food from a restaurant.

Dialog Agents



Machine: *hi*

Human: *hello*

Machine: *could you please let me know what are the operating systems you are using ?*

Human: *i am using linux*

Machine: *ok it was in the ssl vpn on which you stated it was broken*

Machine: *please check if you are logged in to corp network*

Machine: *what is the error that you are running please*

Human: *i am seeing an error related to vpn*

Machine: *what is the error message that you are getting when connecting to vpn using network connect ?*

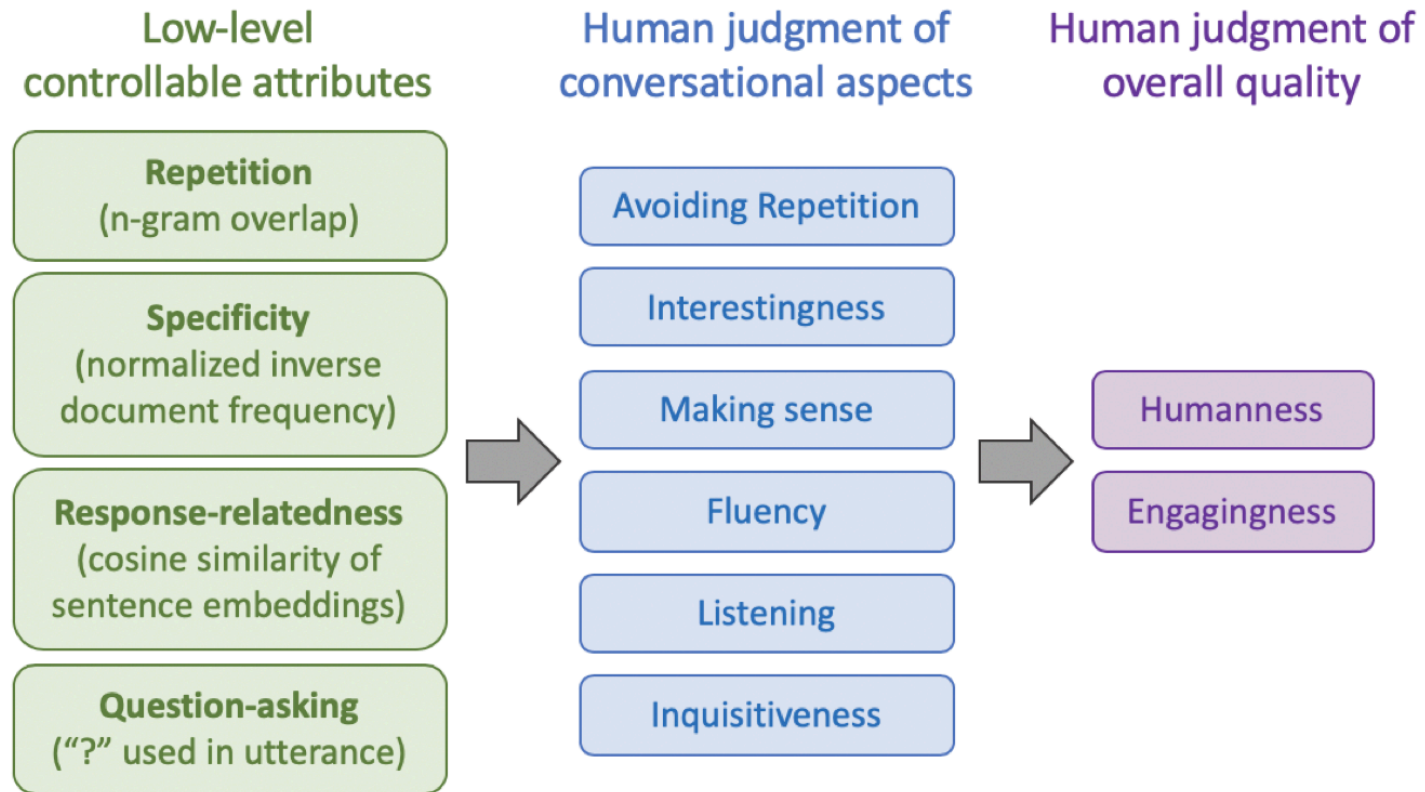
Human: *connection refused or something like that*

Machine: *may i know the version of network connect you connect ?*

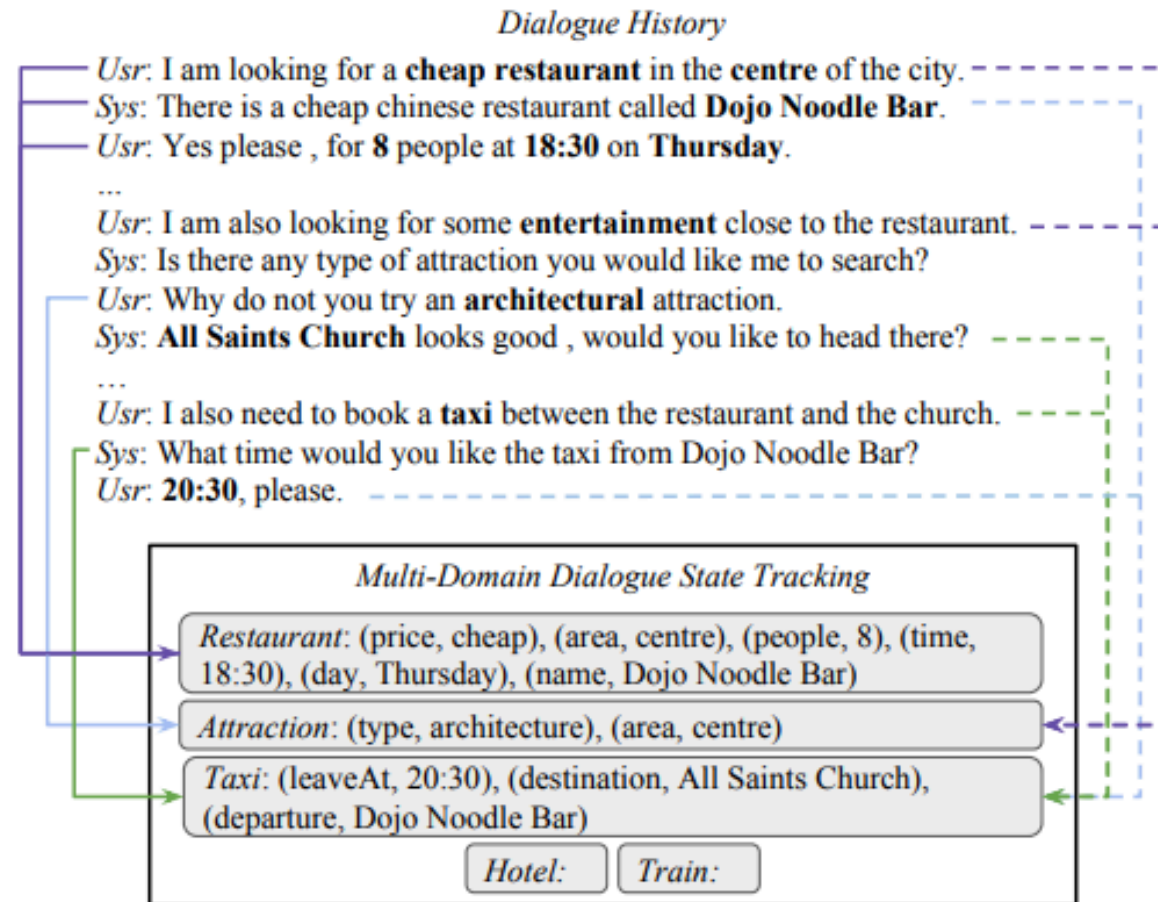
Human: *i am not sure i know that*

Machine: *is the network connect prompting for any user certificate*

Dialog Agents



Task-oriented Dialog



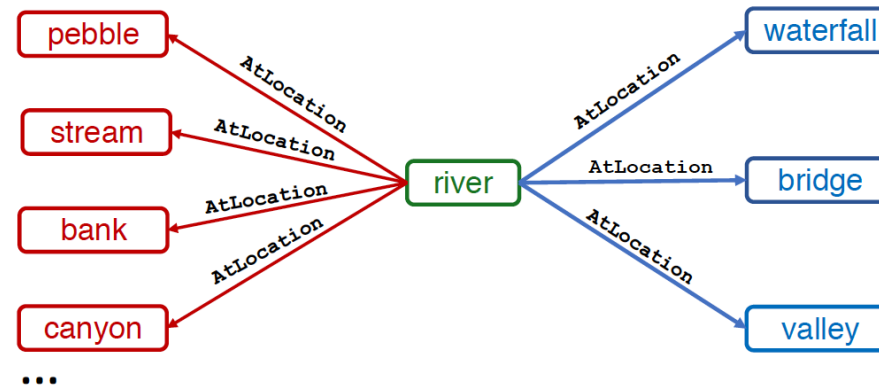
Sentiment Analysis

- Predict the sentiment expressed in an utterance (positive or negative)
- Demo: <http://text-processing.com/demo/sentiment/>
- Tweet sentiment visualization: https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/

Commonsense Reasoning

- Use common-sense to make inference

a) Sample ConceptNet for specific subgraphs



b) Crowd source corresponding natural language questions and two additional distractors

*Where on a **river** can you hold a cup upright to catch water on a sunny day?*

✓ **waterfall**, X **bridge**, X **valley**, X **pebble**, X **mountain**

*Where can I stand on a **river** to see water falling without getting wet?*

X **waterfall**, ✓ **bridge**, X **valley**, X **stream**, X **bottom**

*I'm crossing the **river**, my feet are wet but my body is dry, where am I?*

X **waterfall**, X **bridge**, ✓ **valley**, X **bank**, X **island**

Thank You!