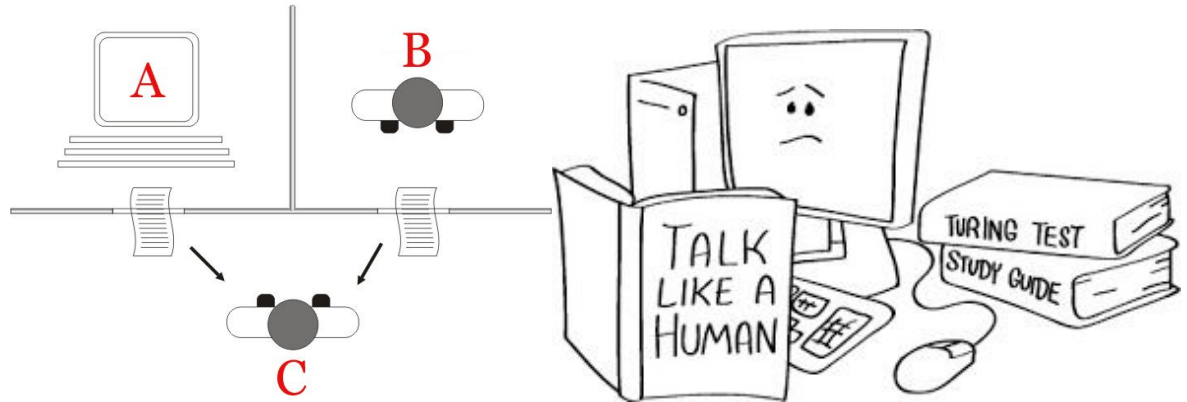
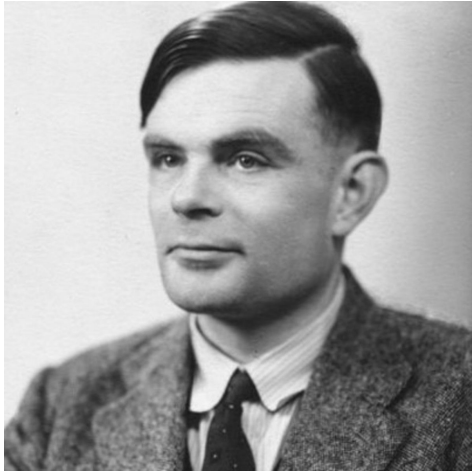


# Introduction

Lecture 1

Language is long considered a core problem of AI

# Turing test (Turing 1950)



Q: Please write me a sonnet on the subject of the Forth Bridge.

A: Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764.

A: (Pause about 30 seconds and then give as answer) 105621.

**Key problem of AI**

With many applications

# Translation

English Spanish Chinese Detect language ▾



English Spanish Chinese (Simplified) ▾

Translate

目前，中文维基百科拥有989,840  
条条目，在所有语言的维基百科中  
列第13位。



39/5000

Currently, there are 989,840 entries  
in Chinese Wikipedia and thirteenth  
in Wikipedia in all languages.



Mùqián, zhōngwén wéijī bǎikē yǒngyǒu 989,840 tiáo tiáomù, zài  
suǒyǒu yǔyán de wéijī bǎikē zhōng liè dì 13 wèi.

# Machine reading

Tesla gained experience in telephony and electrical engineering before emigrating to the United States in 1884 to work for Thomas Edison in New York City. He soon struck out on his own with financial backers, setting up laboratories and companies to develop a range of electrical devices. His patented AC induction motor and transformer were licensed by George Westinghouse, who also hired Tesla for a short time as a consultant. His work in the formative years of electric power development was involved in a corporate alternating current/direct current "War of Currents" as well as various patent battles.

**In what area of the United States did Tesla move to?**

Ground Truth Answers: New York City New York City New York City

**What "war" was Tesla involved in?**

Ground Truth Answers: War of Currents War of Currents War of Currents

**When did Tesla come to the US?**

Ground Truth Answers: 1884 1884 1884

**What other inventor did he work with?**

Ground Truth Answers: Thomas Edison Thomas Edison Thomas Edison

**Where did Tesla work with Edison?**

Ground Truth Answers: New York City New York City New York City

**Who licensed Tesla's induction motor?**

Ground Truth Answers: George Westinghouse George Westinghouse George Westinghouse

<https://rajpurkar.github.io/SQuAD-explorer/explore/1.1/dev/>

## Question

Why are humans enslaved in The Matrix?

## Passage Context

The Matrix is a 1999 science fiction action film written and directed by The Wachowskis, starring Keanu Reeves, Laurence Fishburne, Carrie-Anne Moss, Hugo Weaving, and Joe Pantoliano. It depicts a dystopian future in which reality as perceived by most humans is actually a simulated reality called "the Matrix", created by sentient machines to subdue the human population, while **their bodies' heat and electrical activity are used as an energy source**. Computer programmer "Neo" learns this truth and is drawn into a rebellion against the machines, which involves other people who have been freed from the "dream world."

## Question

Which parts of the shuttle are reused

A reusable launch system (RLS, or reusable launch vehicle, RLV) is a launch system which is capable of launching a payload into space more than once. This contrasts with expendable launch systems, where each launch vehicle is launched once and then discarded. No completely reusable orbital launch system has ever been created. Two partially reusable launch systems were developed, the Space Shuttle and Falcon 9. The Space Shuttle was partially reusable: the orbiter (which included the **Space Shuttle main engines and the Orbital Maneuvering System engines**), and the **two solid rocket boosters** were reused after several months of refitting work for each launch. The external tank was discarded after each flight.



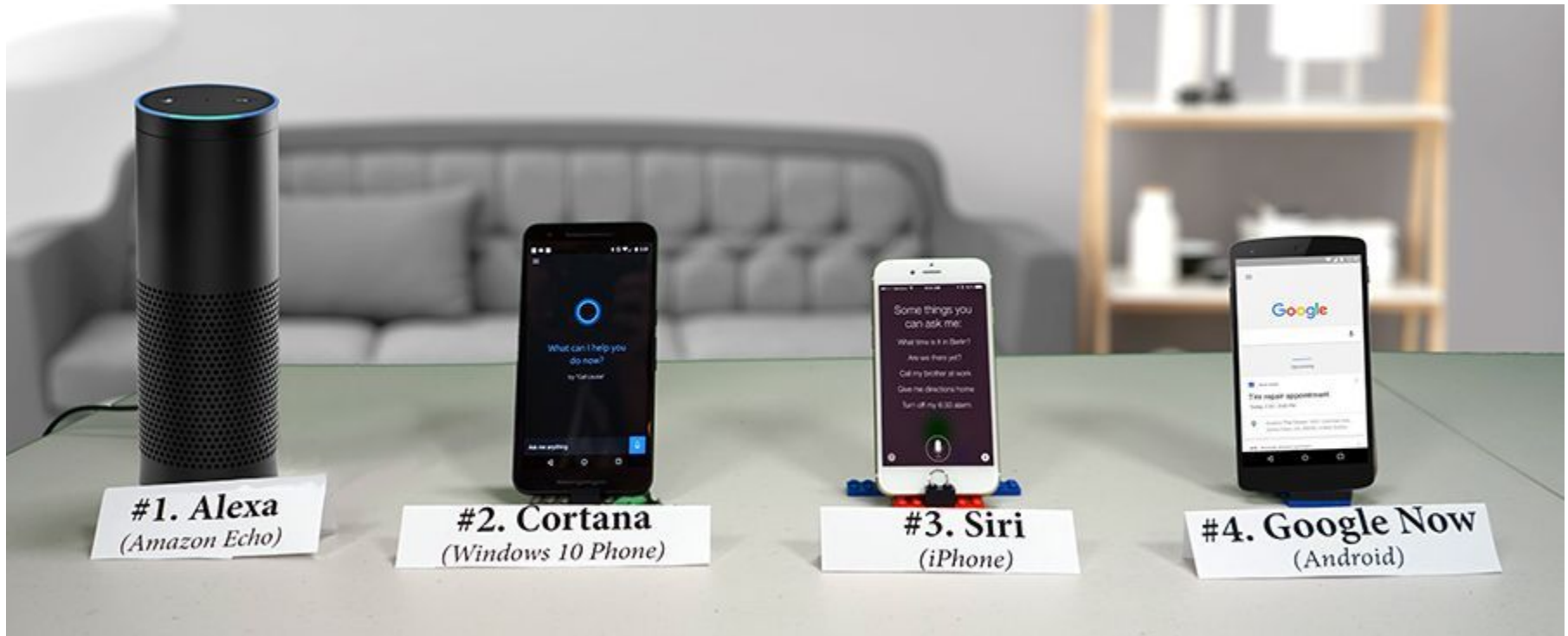
## Question

which part of the shuttle are not reused?

## Passage Context

A reusable launch system (RLS, or reusable launch vehicle, RLV) is a launch system which is capable of launching a payload into space more than once. This contrasts with expendable launch systems, where each launch vehicle is launched once and then discarded. No completely reusable orbital launch system has ever been created. Two partially reusable launch systems were developed, the **Space Shuttle** and Falcon 9. The Space Shuttle was partially reusable: the orbiter (which included the Space Shuttle main engines and the Orbital Maneuvering System engines), and the two solid rocket boosters were reused after several months of refitting work for each launch. The external tank was discarded after each flight.

# Conversational interfaces



# Conversational interfaces

[https://youtu.be/yiQX-\\_Y0gms](https://youtu.be/yiQX-_Y0gms)

The screenshot displays a Google Assistant chat interface. At the top right is a blue microphone icon. The assistant's response is shown in a white bubble with a blue, red, and yellow dot icon: "Hi, how can I help?". The user's input is in a grey bubble: "when is Thanksgiving". The assistant's response is in a white bubble: "Here to help". Below this is a white card with the text "Thursday, November 22" and "Thanksgiving in United States 2018". The user's next input is in a grey bubble: "I meant the Canadian one". The assistant's response is in a white bubble: "Here you go". Below this is another white card with the text "Monday, October 8" and "Thanksgiving 2018 in Canada". At the bottom, there are three buttons: "G Search", "Add to calendar", and "When is Than". At the very bottom, there is a keyboard icon on the left and a microphone icon on the right.

A growing proportion of queries require semantic interpretation.  
Conventional keyword-based retrieval does not suffice!

*how to bike to my office*

```
(TravelQuery  
  (Destination /m/0d6lp)  
  (Mode BIKE))
```

*angelina jolie net worth*

```
(FactoidQuery  
  (Entity /m/0f4vbz)  
  (Attribute /person/net_worth))
```

*weather friday austin tx*

```
(WeatherQuery  
  (Location /m/0vzm)  
  (Date 2013-12-13))
```

*text my wife on my way*

```
(SendMessage  
  (Recipient 0x31cbf492)  
  (MessageType SMS)  
  (Subject "on my way"))
```

*play sunny by boney m*

```
(PlayMedia  
  (MediaType MUSIC)  
  (SongTitle "sunny")  
  (MusicArtist /m/017mh))
```

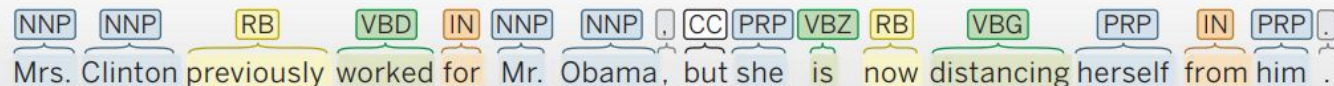
*is REI open on sunday*

```
(LocalQuery  
  (QueryType OPENING_HOURS)  
  (Location /m/02nx4d)  
  (Date 2013-12-15))
```

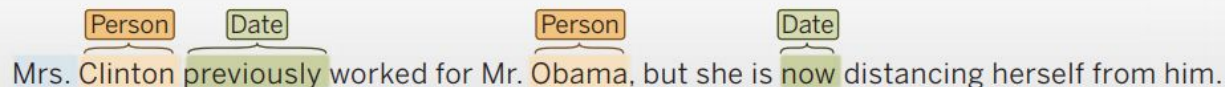
[MacCartney and Potts cs224u]

# What is NLP?

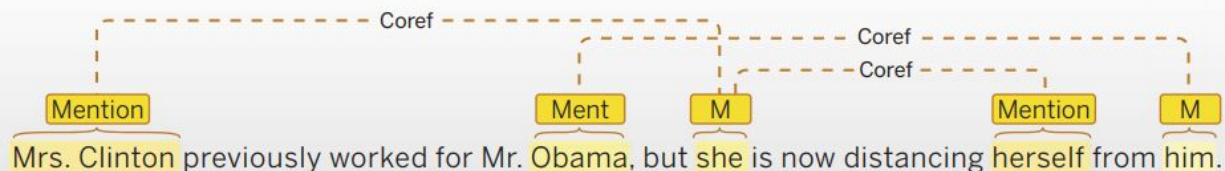
## Part of speech:



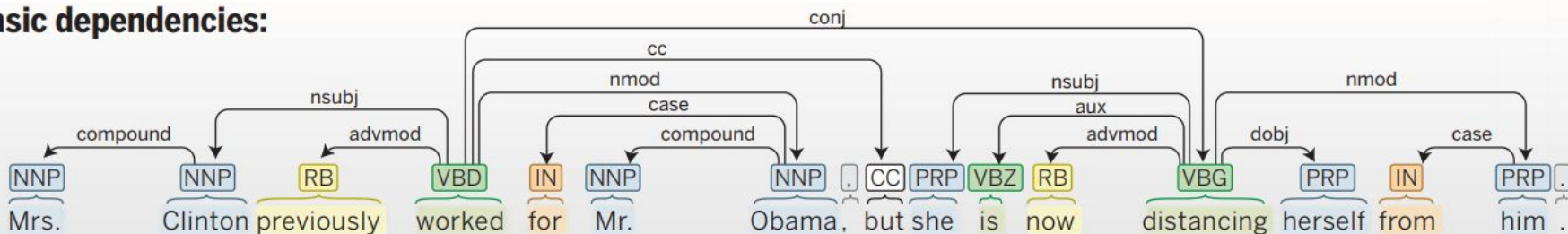
## Named entity recognition:



## Co-reference:



## Basic dependencies:



**Fig. 1. Many language technology tools start by doing linguistic structure analysis.** Here we show output from Stanford CoreNLP. As shown from top to bottom, this tool determines the parts of speech of each word, tags various words or phrases as semantic named entities of various sorts, determines which entity mentions co-refer to the same person or organization, and then works out the syntactic structure of each sentence, using a dependency grammar analysis.

Deep learning progress

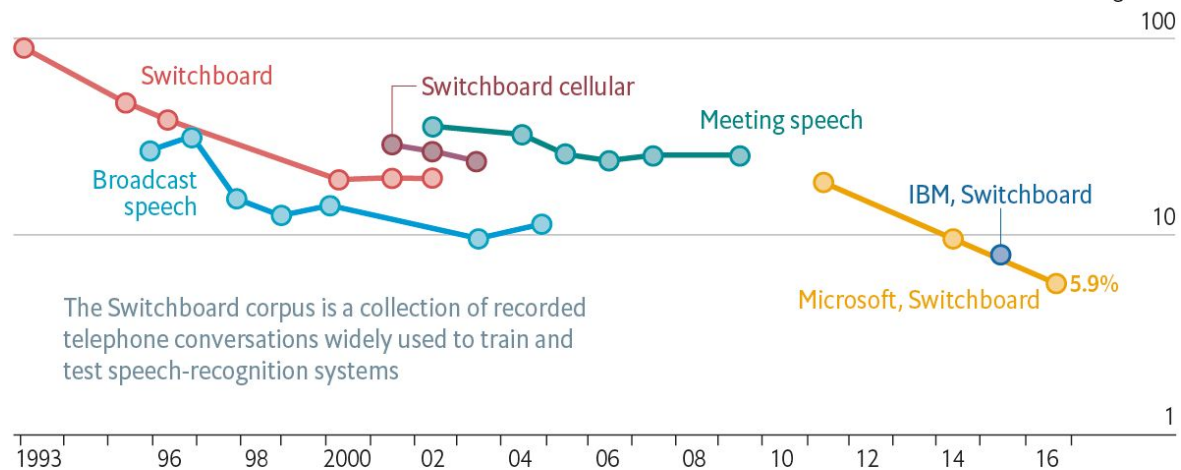
# Speech recognition

Significant process. Increasingly deployed technology.

## Loud and clear

Speech-recognition word-error rate, selected benchmarks, %

Log scale



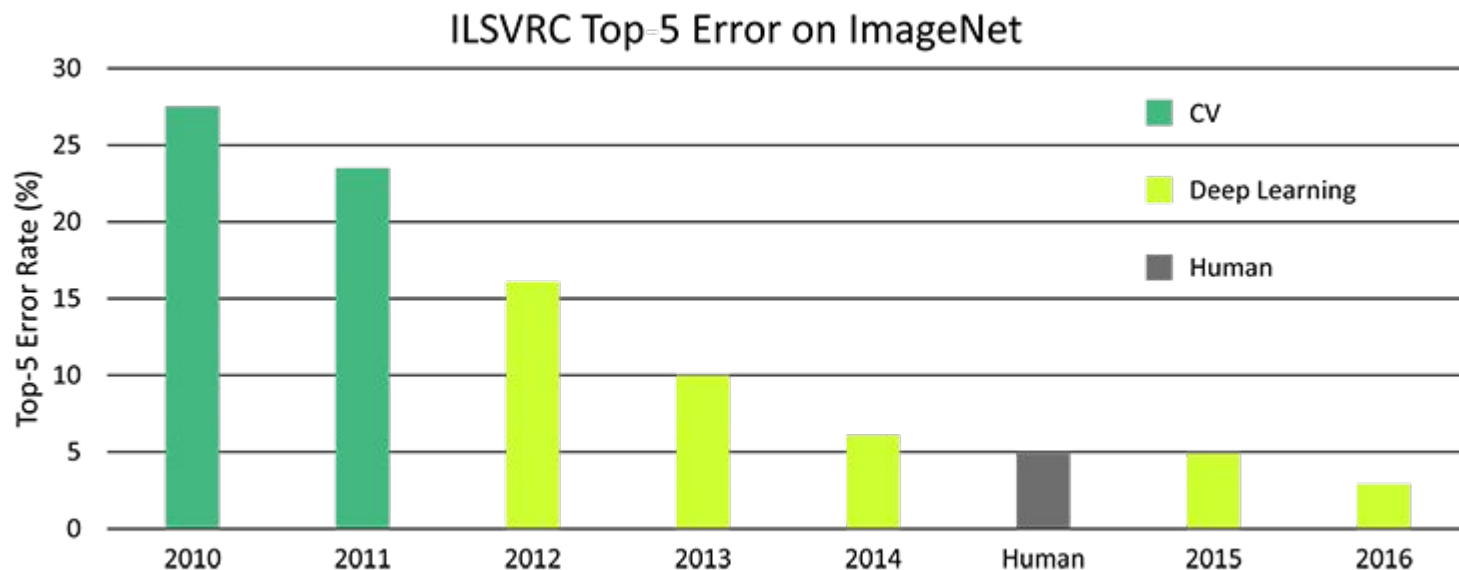
The Switchboard corpus is a collection of recorded telephone conversations widely used to train and test speech-recognition systems

Sources: Microsoft; research papers

# Computer vision

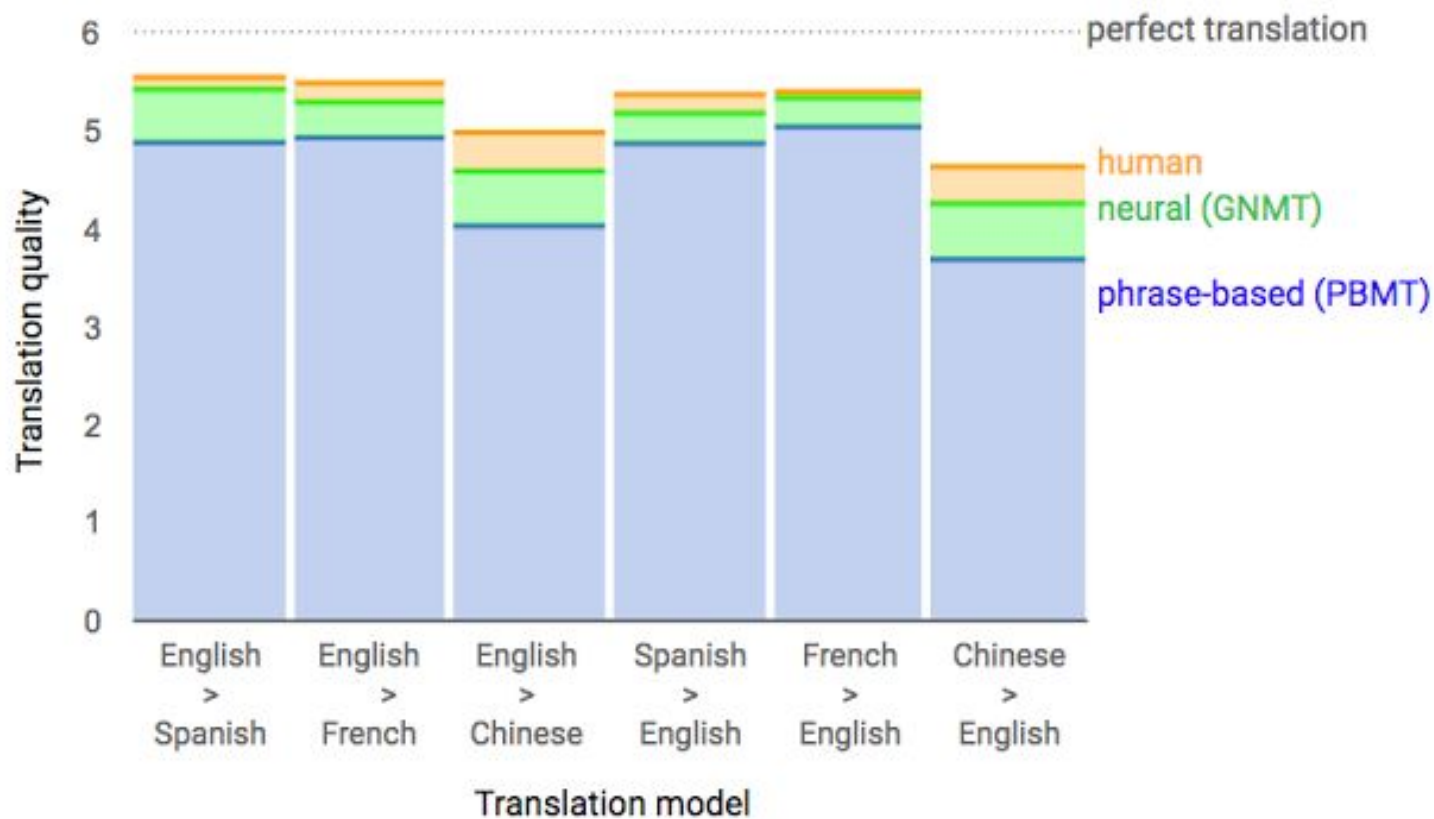
ImageNet error rate from dsiac.org

Hopes high for self-driving cars and other innovations





# Progress on various language pairs



# Michael I. Jordan (UC Berkeley)

**AMA:** “If you got a billion dollars to spend on a huge research project that you get to lead, what would you like to do?”

**[michaelijordan](#)**: I'd use the billion dollars to build a NASA-size program focusing on natural language processing (NLP), in all of its glory (semantics, pragmatics, etc).

Intellectually I think that NLP is fascinating, allowing us to focus on highly-structured inference problems, on issues that go to the core of "what is thought" but remain eminently practical, and on a technology that surely would make the world a better place.

Although current deep learning research tends to claim to encompass NLP, I'm (1) much less convinced about the strength of the results, compared to the results in, say, vision; (2) much less convinced in the case of NLP than, say, vision, the way to go is to couple huge amounts of data with black-box learning architectures.

[http://www.reddit.com/r/MachineLearning/comments/2fxi6v/ama\\_michael\\_i\\_jordan](http://www.reddit.com/r/MachineLearning/comments/2fxi6v/ama_michael_i_jordan)

# Yann LeCun (NYU and Facebook)

Across the industry, it's already reinventing image and speech recognition. But like Google, LeCun and FAIR are pushing for more. The next big frontier, he says, is natural language processing, which seeks to give machines the power to understand not just individual words but entire sentences and paragraphs.

<https://www.wired.com/2014/12/fb/>

# Geoff Hinton (U Toronto and Google)

I think that the most exciting areas over the next five years will be really understanding text and videos. I will be disappointed if in five years' time we do not have something that can watch a YouTube video and tell a story about what happened. In a few years time we will put [Deep Learning] on a chip that fits into someone's ear and have an English-decoding chip that's just like a real Babel fish.

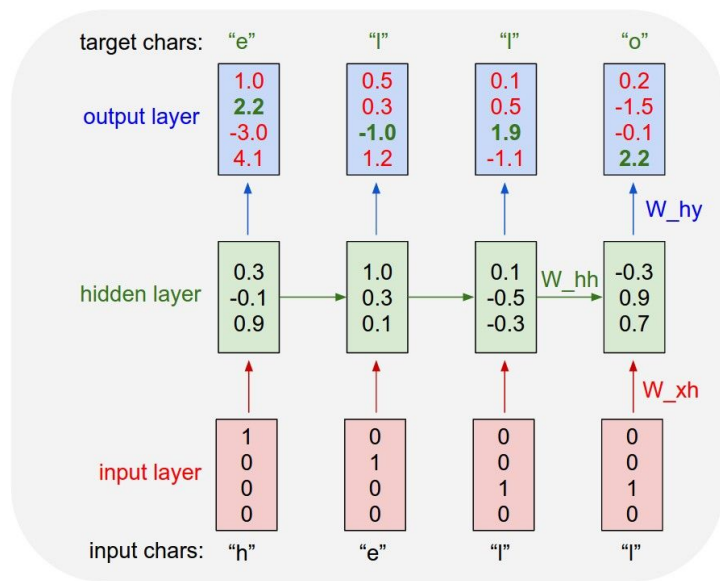
[https://www.reddit.com/r/MachineLearning/comments/2lmo0l/ama\\_geoffrey\\_hinton/](https://www.reddit.com/r/MachineLearning/comments/2lmo0l/ama_geoffrey_hinton/)

(3 years ago)

# Themes of deep learning

- **Use very general models and algorithms**
- Distributed representations
- Large design space of potential models
  - Engineering model architectures rather than features
  
- Leverage available compute and data
  - Learn as much as possible from data rather than design for domains

# The hammer: deep learning



For  $\bigoplus_{n=1, \dots, m}$  where  $\mathcal{L}_{m, \bullet} = 0$ , hence we can find a closed subset  $\mathcal{H}$  in  $\mathcal{H}$  and any sets  $\mathcal{F}$  on  $X$ ,  $U$  is a closed immersion of  $S$ , then  $U \rightarrow T$  is a separated algebraic space.

*Proof.* Proof of (1). It also start we get

$$S = \text{Spec}(R) = U \times_X U \times_X U$$

and the comparicoly in the fibre product covering we have to prove the lemma generated by  $\coprod Z \times_U U \rightarrow V$ . Consider the maps  $M$  along the set of points  $\text{Sch}_{f_{ppf}}$  and  $U \rightarrow U$  is the fibre category of  $S$  in  $U$  in Section, ?? and the fact that any  $U$  affine, see Morphisms, Lemma ?? . Hence we obtain a scheme  $S$  and any open subset  $W \subset U$  in  $\text{Sh}(G)$  such that  $\text{Spec}(R') \rightarrow S$  is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

which has a nonzero morphism we may assume that  $f_i$  is of finite presentation over  $S$ . We claim that  $\mathcal{O}_{X,x}$  is a scheme where  $x, x', s'' \in S'$  such that  $\mathcal{O}_{X,x'} \rightarrow \mathcal{O}_{X',x'}$  is separated. By Algebra, Lemma ?? we can define a map of complexes  $\text{GL}_{S'}(x'/S'')$  and we win.  $\square$

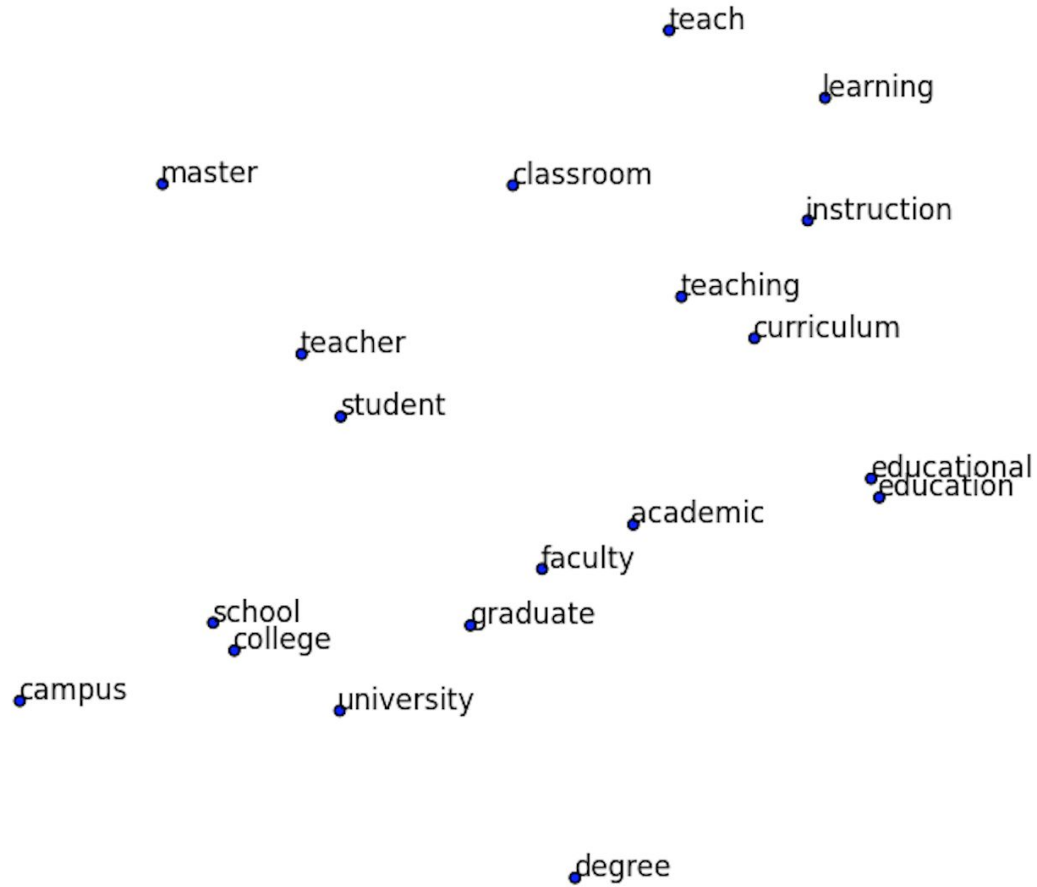
To prove study we see that  $\mathcal{F}|_U$  is a covering of  $\mathcal{X}'$ , and  $\mathcal{T}_i$  is an object of  $\mathcal{F}_{X/S}$  for  $i > 0$  and  $\mathcal{F}_p$  exists and let  $\mathcal{F}_i$  be a presheaf of  $\mathcal{O}_X$ -modules on  $\mathcal{C}$  as a  $\mathcal{F}$ -module. In particular  $\mathcal{F} = U/\mathcal{F}$  we have to show that

$$\widetilde{M}^\bullet = \mathcal{I}^\bullet \otimes_{\text{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1} \mathcal{F}$$

# Themes of deep learning

- Use very general models and algorithms
- **Distributed representations**
- Large design space of potential models
  - Engineering model architectures rather than features
  
- Leverage available compute and data
  - Learn as much as possible from data rather than design for domains

# Representations

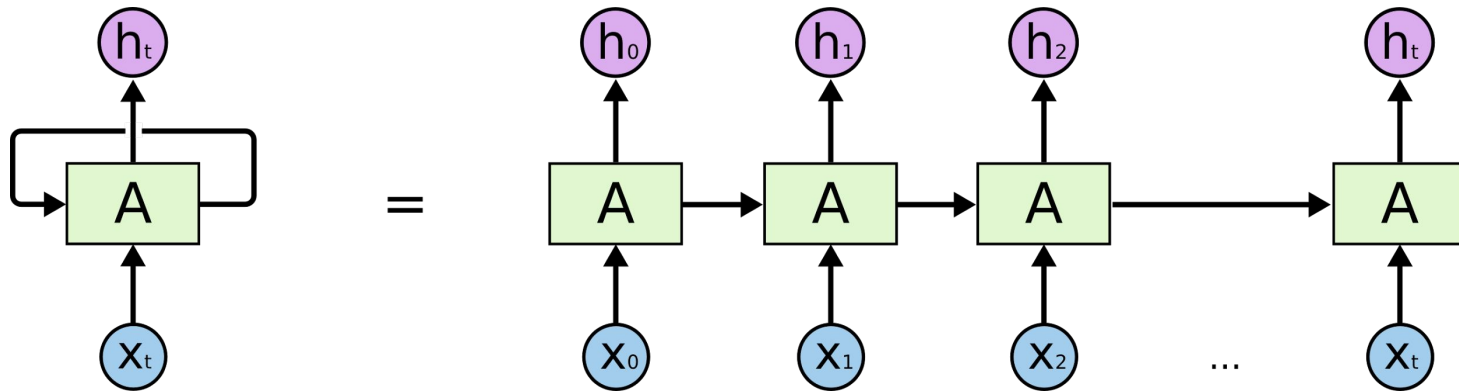




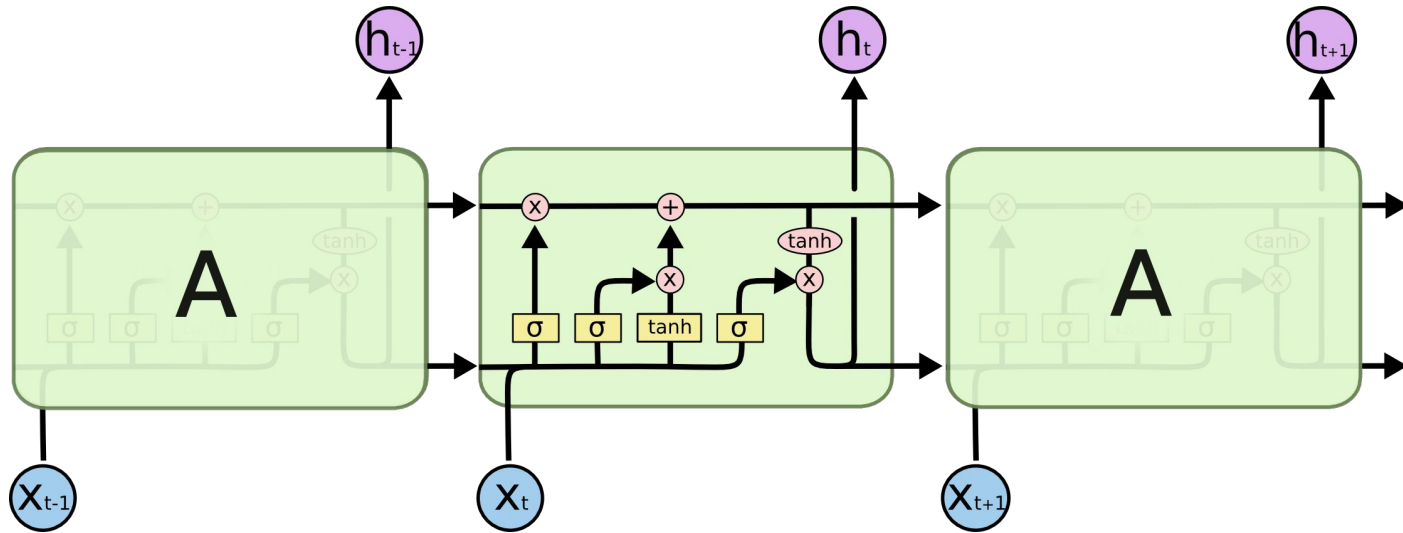
# Themes of deep learning

- Use very general models and algorithms
- Distributed representations
- **Large design space of potential models**
  - Engineering model architectures rather than features
  
- Leverage available compute and data
  - Learn as much as possible from data rather than design for domains

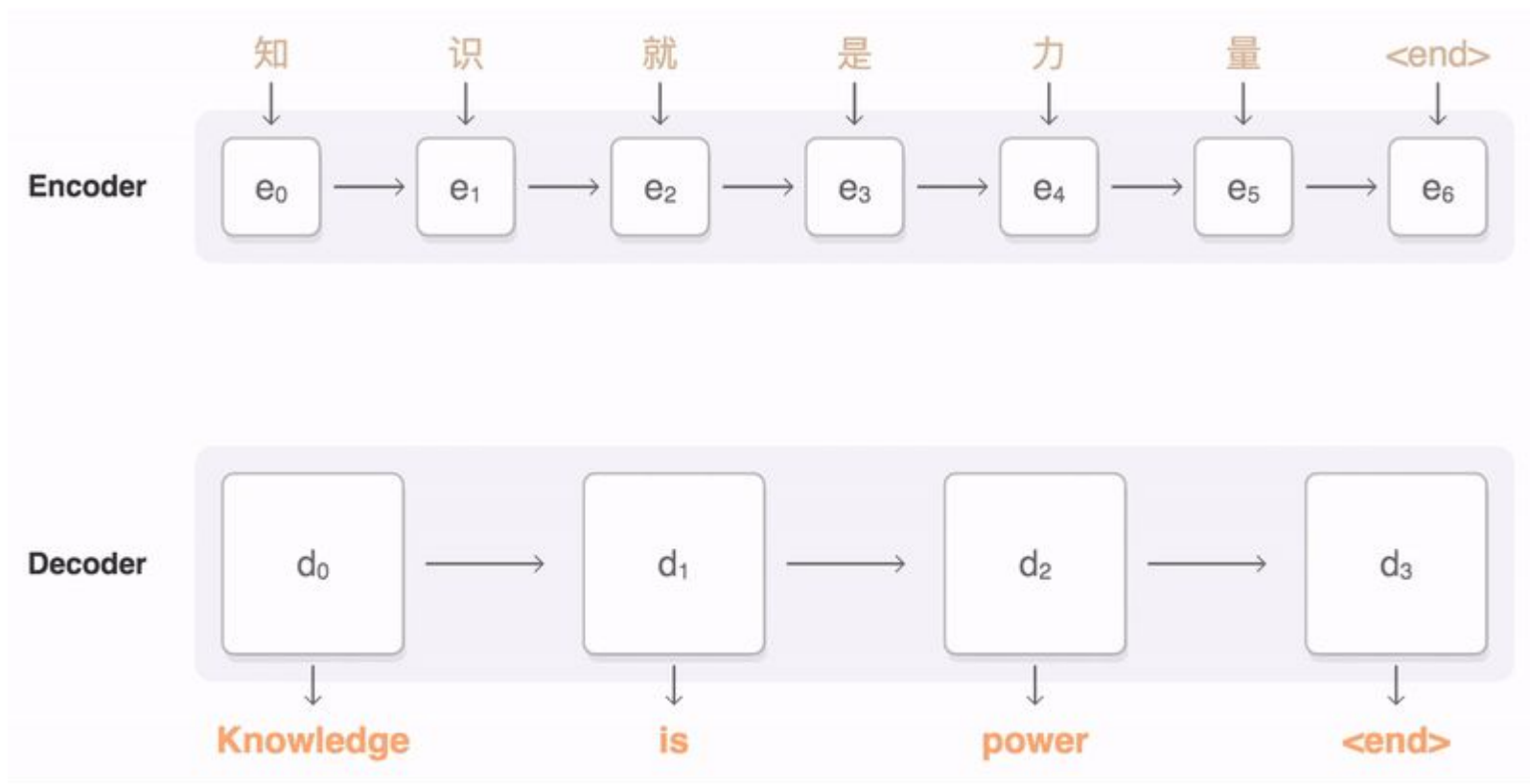
# Models



# LSTM cell

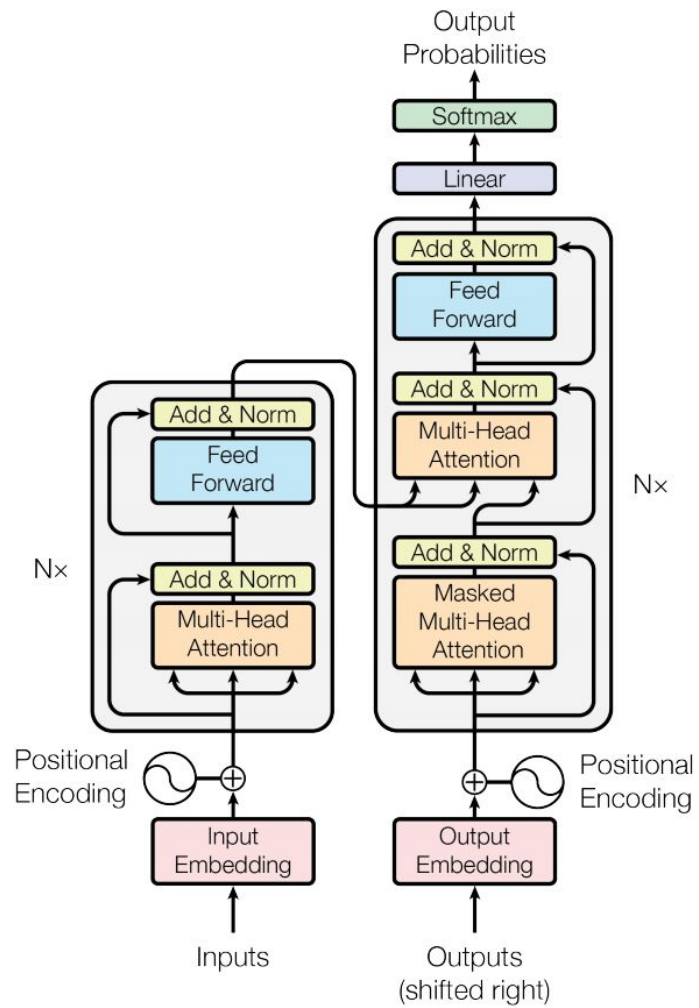


# Seq2seq with attention



<https://research.googleblog.com/2016/09/a-neural-network-for-machine.html>

# Transformer



The nails in NLP

## Question

Which parts of the shuttle are reused

A reusable launch system (RLS, or reusable launch vehicle, RLV) is a launch system which is capable of launching a payload into space more than once. This contrasts with expendable launch systems, where each launch vehicle is launched once and then discarded. No completely reusable orbital launch system has ever been created. Two partially reusable launch systems were developed, the Space Shuttle and Falcon 9. The Space Shuttle was partially reusable: the orbiter (which included the **Space Shuttle main engines and the Orbital Maneuvering System engines**), and the **two solid rocket boosters** were reused after several months of refitting work for each launch. The external tank was discarded after each flight.

## Question

which part of the shuttle are not reused?

## Passage Context

A reusable launch system (RLS, or reusable launch vehicle, RLV) is a launch system which is capable of launching a payload into space more than once. This contrasts with expendable launch systems, where each launch vehicle is launched once and then discarded. No completely reusable orbital launch system has ever been created. Two partially reusable launch systems were developed, the **Space Shuttle** and Falcon 9. The Space Shuttle was partially reusable: the orbiter (which included the Space Shuttle main engines and the Orbital Maneuvering System engines), and the two solid rocket boosters were reused after several months of refitting work for each launch. The external tank was discarded after each flight.





Where is **The Hobbit** playing in **Mountain View**?

The Hobbit is playing at the Century 16 Theater.

When is **it** playing **there**?

It's playing at 2pm, 5pm, and 8pm.

OK. I'd like 1 **adult** and 2 **children** for **the first show**.  
How much would **that** cost?



Need **domain knowledge**, **discourse knowledge**, **world knowledge**

[MacCartney and Potts cs224u]

# Reality

“ Where is the hobbit playing in Mountain View ”

OK, here's 'The Hobbit: An Unexpected Journey' playing in Mountain View today:

THE HOBBIT: AN UNE...	
Playing at 1 Theaters	
<b>CENTURY CINEMA 16</b>	5.6 MI
11:10	14:05 3D 14:50 17:50 3D 21:55 3D ...

“ When is it playing there ”

OK, here's 'The Hobbit: An Unexpected Journey' playing a bit far from Palo Alto today:

Here's 'The Hobbit: An Unexpected Journey' playing a bit far from Palo Alto today:

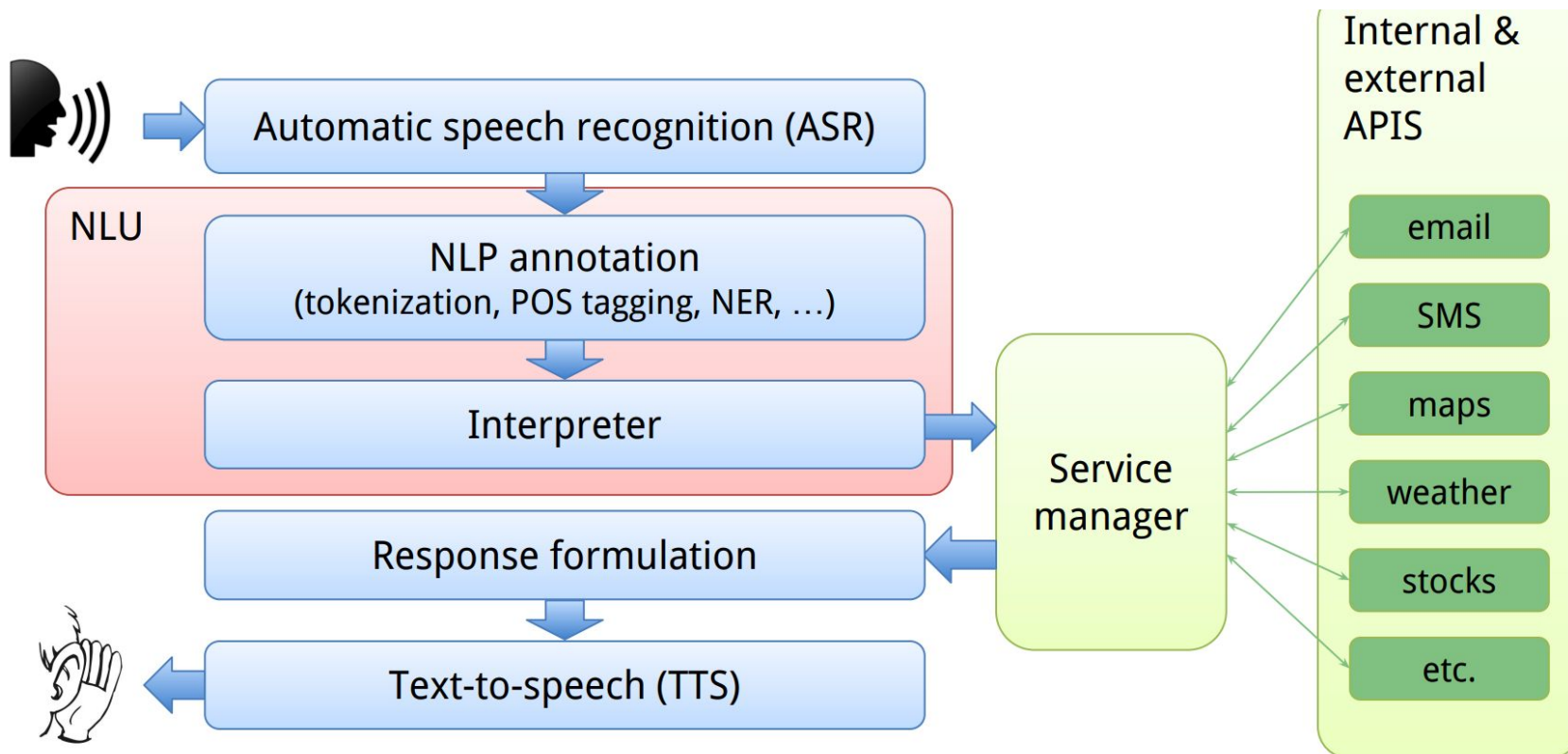
THE HOBBIT: AN UNE...	
Playing at 2 Theaters	
<b>CENTURY 20 DOWNTOWN RED...</b>	4.2 MI
10:50	12:30 3D 14:35 16:10 3D 18:15 ...
<b>CENTURY CINEMA 16</b>	5.9 MI
11:10	14:05 3D 14:50 17:50 3D 21:55 3D ...

“ Okay I like one adult and two children for the first show how much would that cost ”

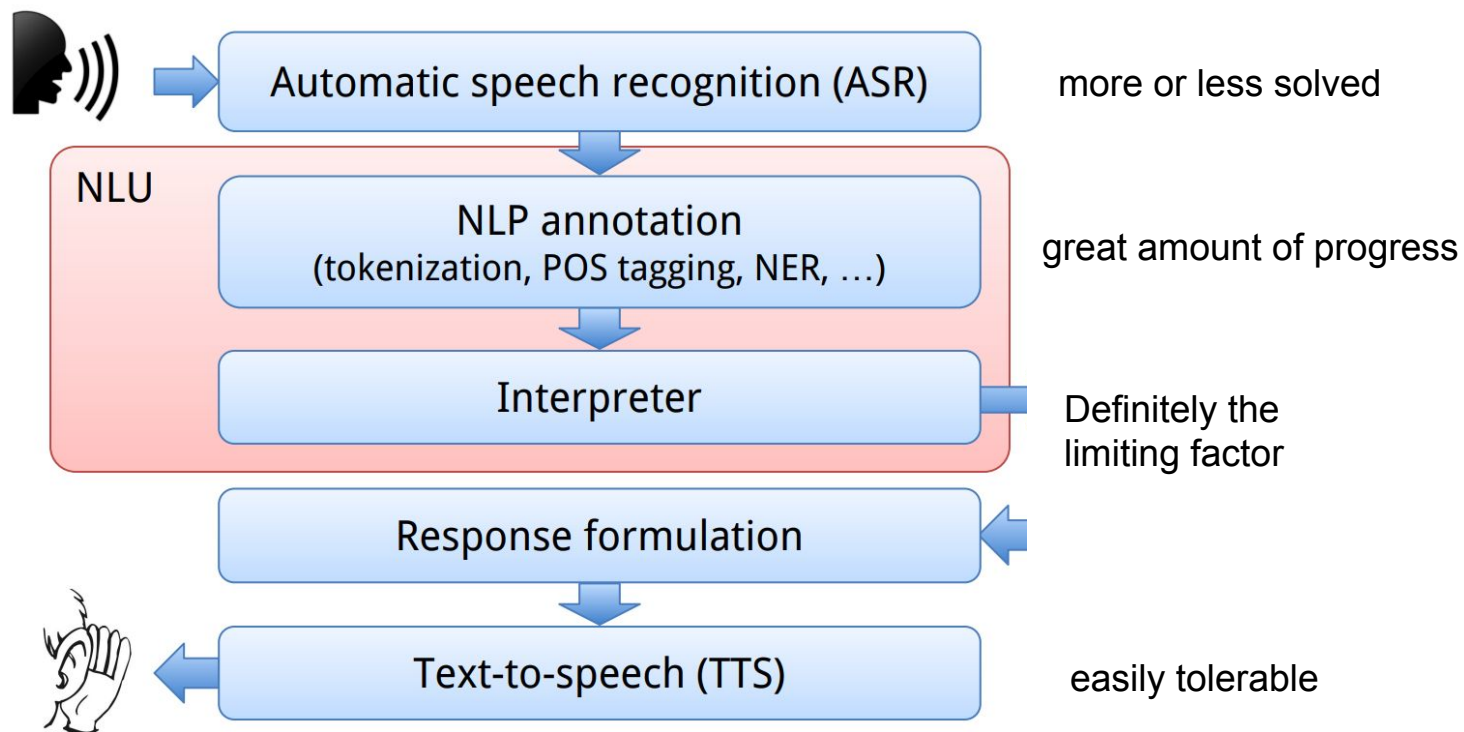
Searching the web for 'Okay I like one adult and two children for the first show how much would that cost'...

[MacCartney and Potts cs224u]

# Conversational interfaces



# Conversational interfaces



# More conversational search

how many legs does the zebra have

Here's what I found

<p>How many legs does the Zebra have? - Album on Imgur Imgur &gt; gallery</p> <p>Post with 16 votes and 370 views. Tagged with funny, puppy, creepy, baby animals. Shared by Eesobuhiri350</p>	<p>How many legs does a zebra have? - Quora &gt; Home</p> <p>There is no question. I have different questions.</p>
--	--

how many legs does a goat have


Here's the top Quora result

You could argue only the cats are specified as on the bed (0 cats, 4 bed legs), birds are flying (0), 4 monkeys (**8 legs** (assuming monkeys have **2 legs** each and not 4)), 5 dogs (**20 legs**), 3 goats (**12 legs**), and you (**2 legs**).

You walk into your room and see 4 monkeys, 3 **goats**, 5 dogs, 2 cats ...  
Quora

Tested by Sida on 02/05/2018

which states border new jersey?




New Jersey is bordered by New York, **Pennsylvania**, and **Delaware**. Although, **Connecticut** (in the north) & **Maryland** (in the south) are very close to the border, but these states don't actually touch with New Jersey.

**What states border New Jersey?** - Quora  
Quora

how many eyes do the chicken have

Here are some images



# Conversational interfaces

## Works great for a number of useful queries

- which states borders New Jersey?
- who is Donald Trump married to?
- what is 15% of \$120?

## Fails to work (tested on Google assistant 02/05/2018)

- how many legs does the goat have?
- how many eyes does the chicken have?
- is boiling water hotter than iced water?
- is the day brighter than the night?

Why?

# Realities of NLI, sometimes



# Ambiguity

## **Ambiguity**

The Pope's baby steps on gays

Boy paralyzed after tumor fights back to gain black belt

Enraged cow injures farmer with axe

Juvenile Court to Try Shooting Defendant



# World knowledge

## **Arbitrary world knowledge required for full understanding**

The city councilmen refused the demonstrators a permit because they [feared/advocated] violence.

What does the pronoun "they" refer to in each case?

# Common sense

- how many eyes does the chicken have?
- is boiling water hotter than iced water?
- is the day brighter than the night?

Often missing from the data

# Pragmatics

When a diplomat says *yes*, he means 'perhaps';  
When he says *perhaps*, he means 'no';  
When he says *no*, he is not a diplomat.—*Voltaire*  
(Quoted, in Spanish, in Escandell 1993.)



# Goals of this class

- Learn the fundamentals
  - Machine learning / deep learning focused
  - Broadly applicable methods and abstractions
  - Identify important problems
- Gain practical experience
  - Complete assignments with significant implementation components
  - To support you in completing a project that is worthy of a top NLP conference
- Learn the language of the field
  - Read research papers, and gain the ability to learn more on your own

# Diagnosis

- In this class:
  - Methods and approaches of applied machine learning
  - Build NLP systems and learn from empirical feedback
  - Understanding NLP problems and issues
  - Critically read NLP papers
  
- Not in this class
  - details of deep learning CoreNLP, NLTK, deep learning software
    - but the ability to learn such things is expected
  - background in calculus, linear algebra, etc.
  - theory of machine learning / linguistics

# Class plan

- First half
  - the machine learning tools
  - distributed representations: word vectors etc.
  - deep learning models: RNN, seq2seq, attention
  - test barebone models, and compare against baselines
- Midterm
- Second half (not too sure yet)
  - study/implementation of representative/important/cutting-edge papers
  - more careful study of semantic parsing and natural language interfaces
- Project

Warm up with least squares

# Logistics

- Class website
  - Piazza
  - Schedule office hours
    - Misha: 10am on Monday
  - Waitlist and enrollment problems
- 
- Sections next week
    - review of python, numpy
    - assignment setup
    - review of probability, optimization etc.?