
Why You Should Take NLP

— CS322, Spring 2019 —

But first.... What is Natural Language Processing?

$\text{AI} \cap \text{Linguistics} \cap \text{Computer Science} \approx \text{NLP}$

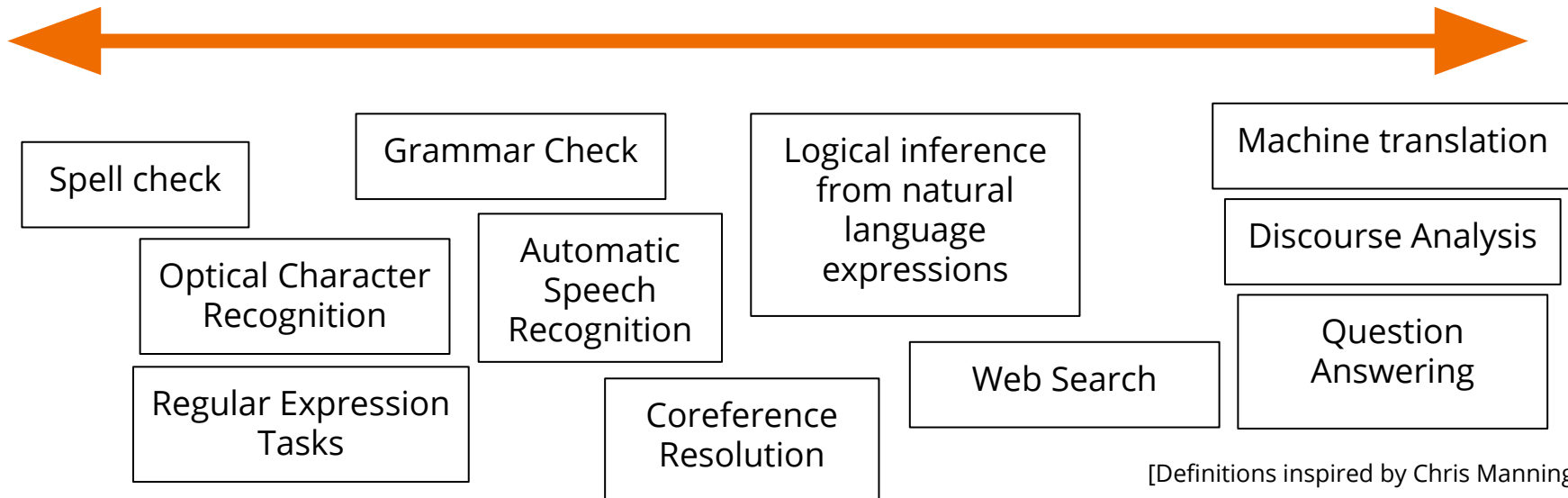
But first.... What is Natural Language Processing?

The goal of natural language processing is to build computers that "understand" natural languages (well enough to perform useful tasks)

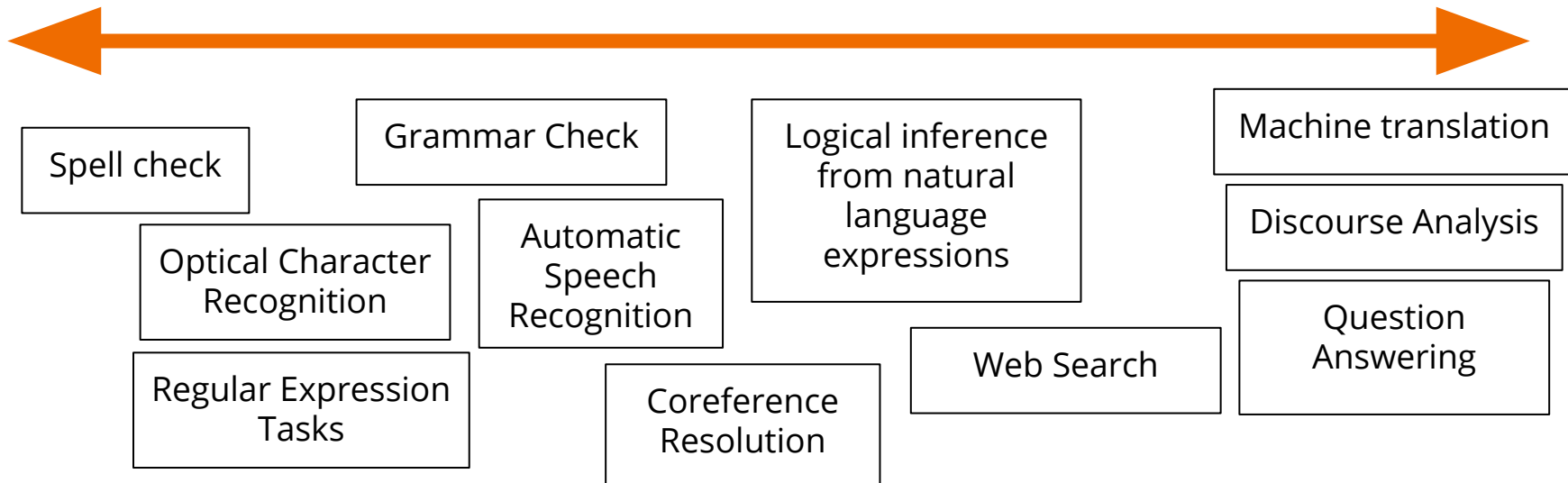
Hierarchy of abstraction...

Computers
can do these
pretty well
already

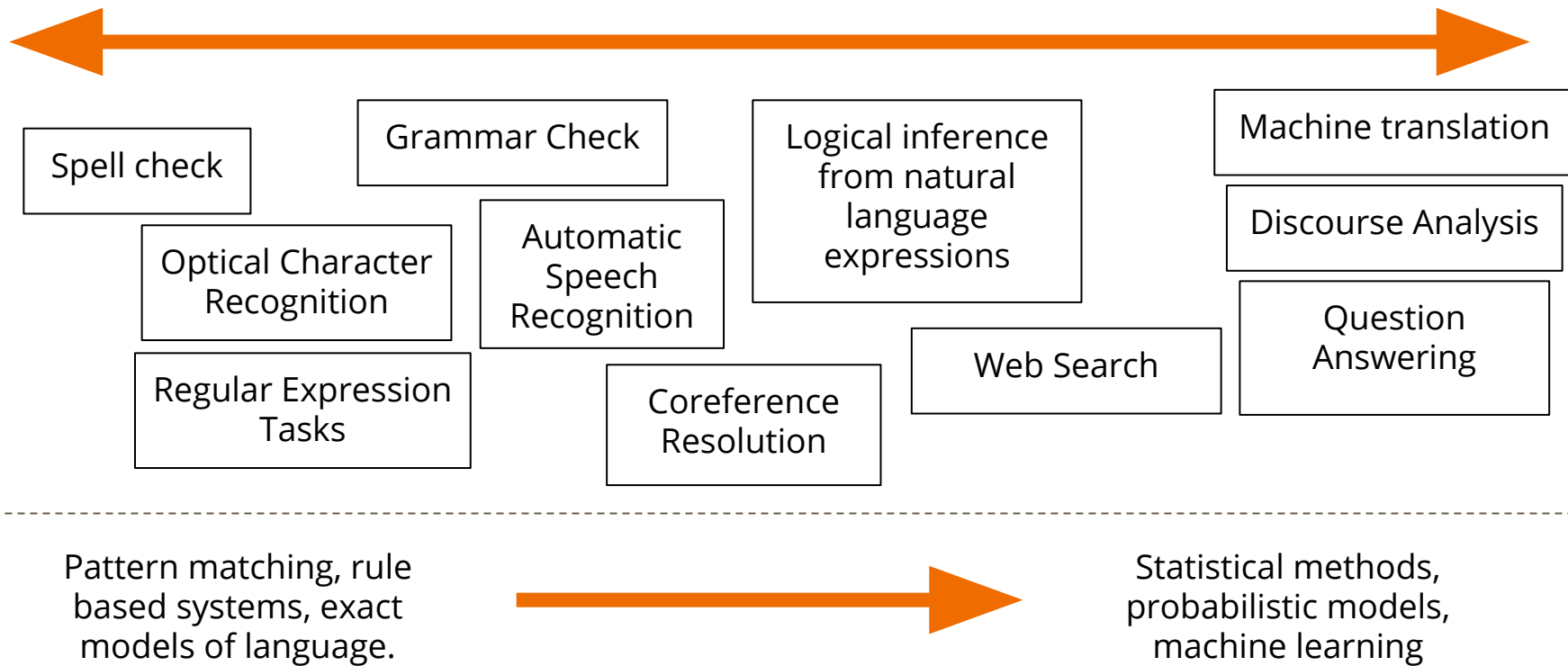
Requires
human-level
intelligence



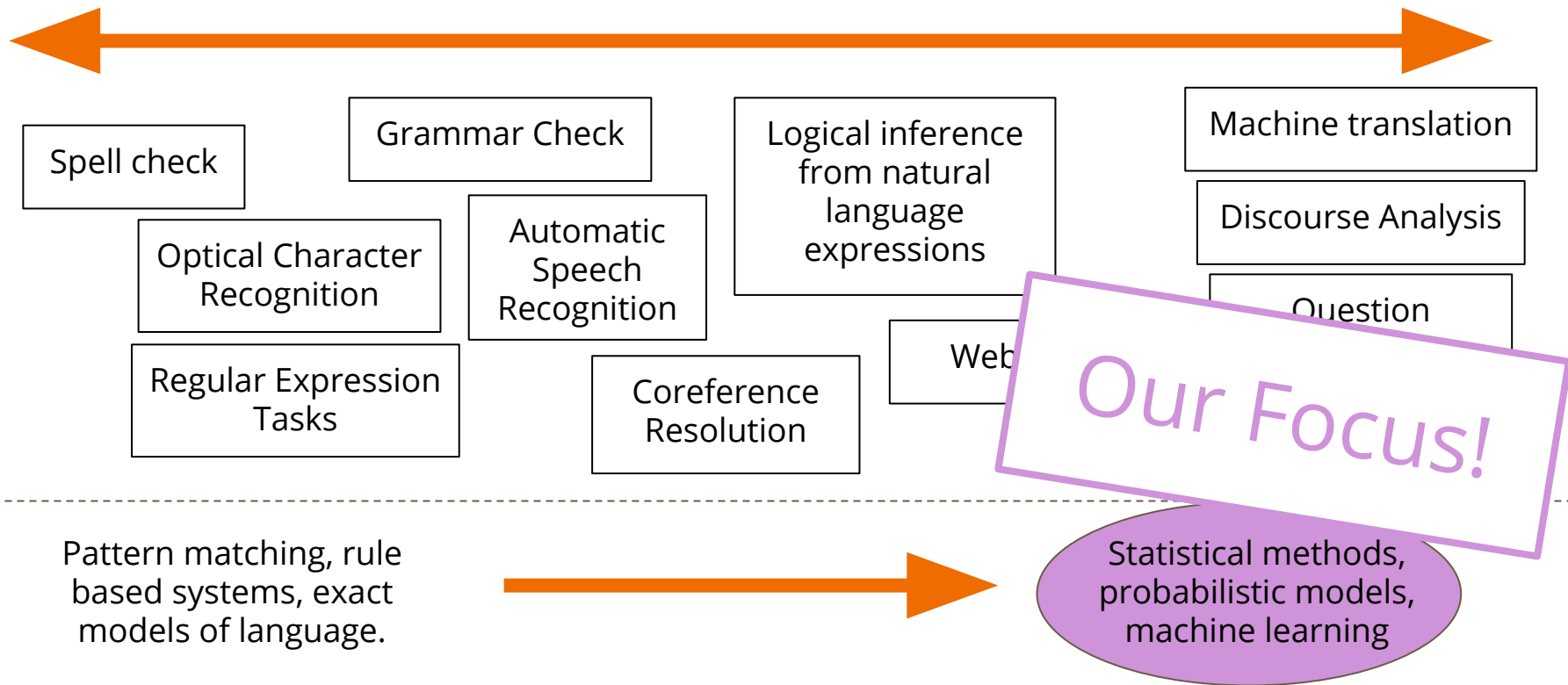
How to build natural language processing systems?



How to build natural language processing systems?



How to build natural language processing systems?



Why You Should Take NLP

— CS322, Spring 2019 —

Reason 1: NLP Enables Useful Applications

Machine Translation

English Spanish French Detect language ▾



English Spanish Arabic ▾

Translate

Even though automatic machine translation is high quality for some languages, the Carleton language requirement is still worthwhile.



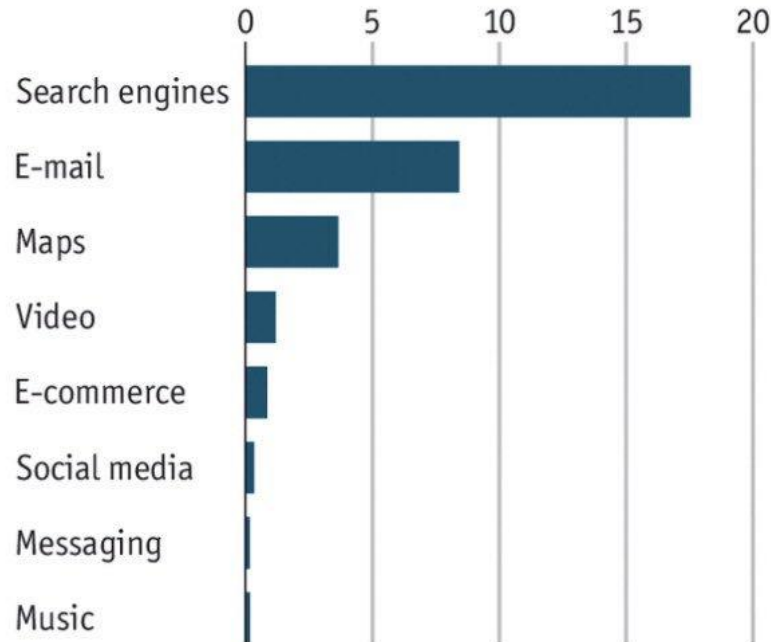
132/5000

Aunque la traducción automática de máquina es de alta calidad para algunos idiomas, el requisito de idioma de Carleton aún vale la pena.



Web Search

Median payment accepted to forgo use of internet service for a year, \$'000














[Brynjolfsson et al. 2018]

Evolving web search: websites → structured data!

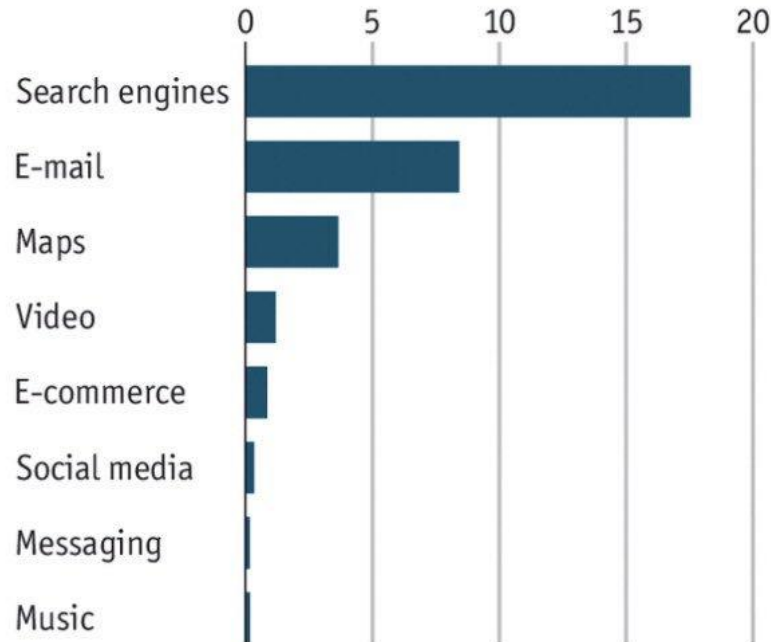
[All](#) [Videos](#) [Images](#) [News](#) [Maps](#) [More](#) [Settings](#) [Tools](#)

Roller Coasters (by Height)

										
Kingda Ka 456'	Top Thrill Dragster Coa... 420'	SUPERMAN: Escape From ... 415'	Tower of Terror II 377'	Red Force 367'	The Fury 325 325'	Steel Dragon 2000 318'	The Millennium F... 308'	Leviathan 306'	Intimidator 305 305'	Thunder Dolphin 262'

E-mail

Median payment accepted to forgo use of internet service for a year, \$'000

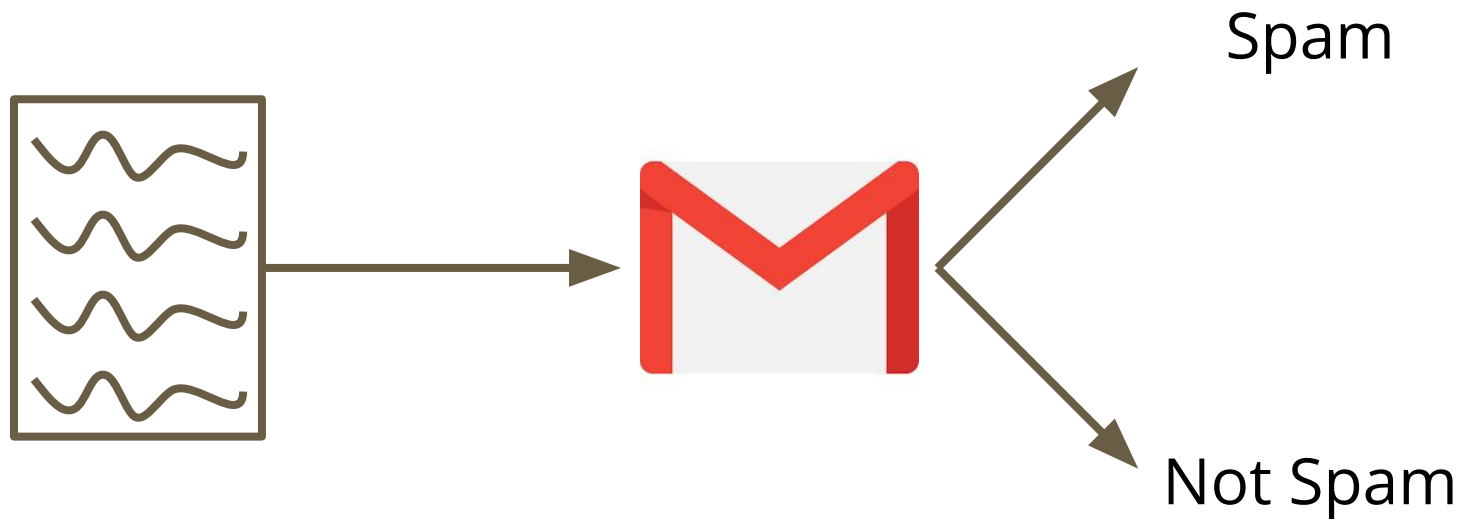


[Brynjolfsson et al. 2018]

NLP's Role in Saving E-mail...

Some estimates: 90% of all e-mail are spam!

And yet --- I don't really see any of it?



Vision and Language



"man in black shirt is playing guitar."

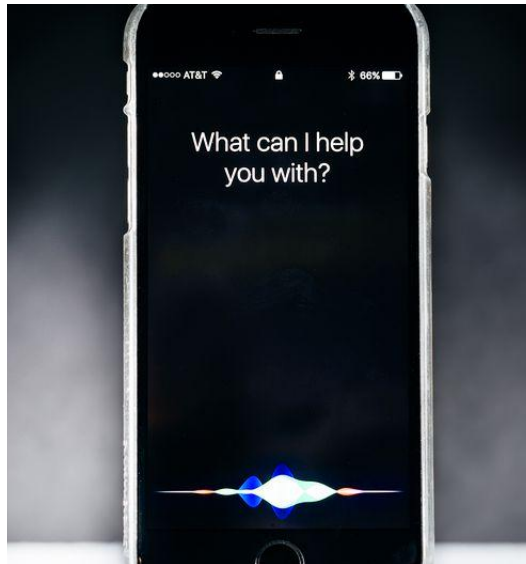
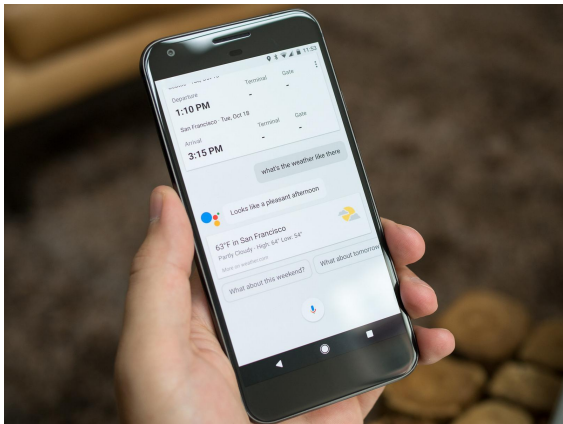


"construction worker in orange safety vest is working on road."



"two young girls are playing with lego toy."

Conversation assistants



Text is ubiquitous; thus, NLP is highly employable!

(if you'd like a job at a company someday)



Want their conversations to be "better" between
hosts and guests

(<https://medium.com/airbnb-engineering/>)



Want to recommend similar listings using text
content

(<https://www.zillow.com/data-science/>)

But, also --- NLP is used by non-profits!

(if you'd like a job at a non-profit someday)



CRISIS TEXT LINE |

Reason 2: It really works! And we're improving!

We've come a long way... !

Eliza (Weizenbaum, 1966) ([demo](#))

Google Assistant (for example) is a big improvement!

"How are you?"

"How many points did Michael Jordan score in his career?"

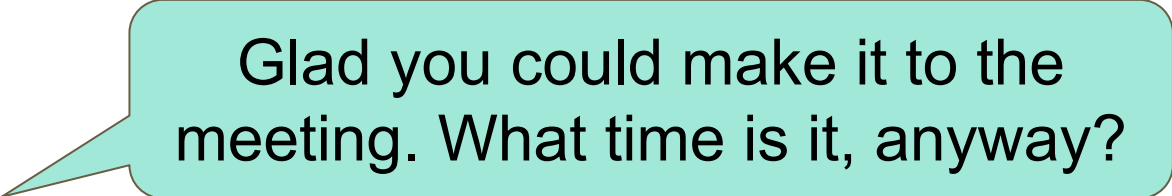
"What is the capital of the United States"

"What is the tallest building in MN?"

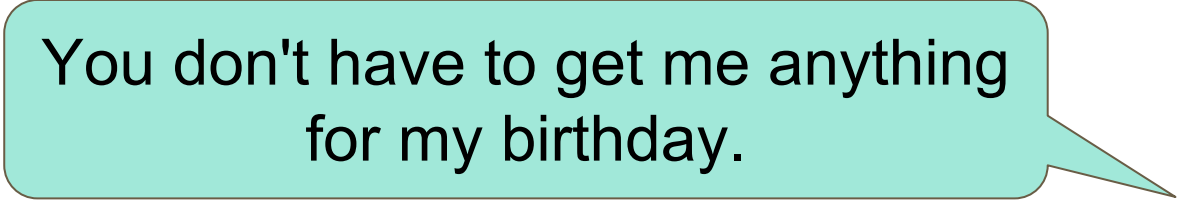
Reason 3: Language <--> Psychology

Communication and Pragmatics

Language is highly contextual! Reflects many psychological processes...



Glad you could make it to the meeting. What time is it, anyway?



You don't have to get me anything for my birthday.



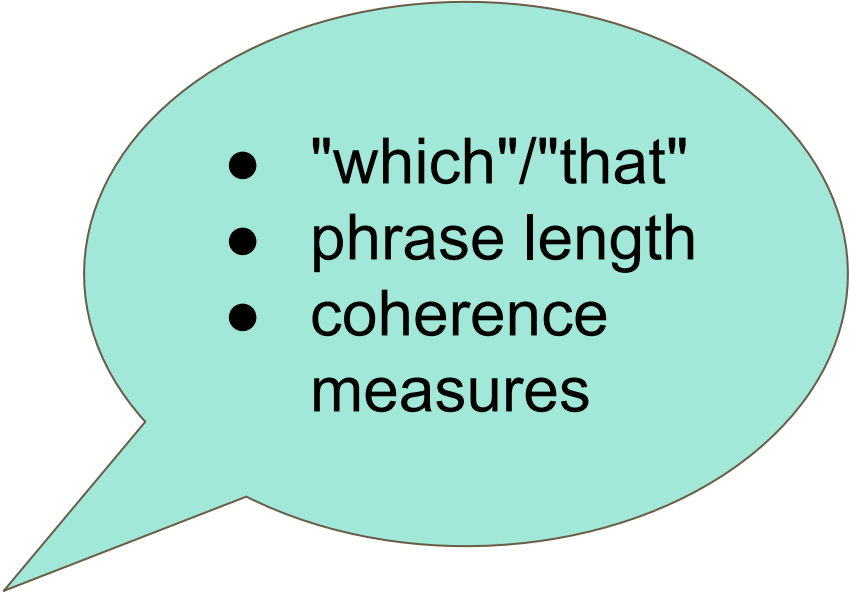
I like you a lot!

More recently... Mapping semantics in the brain

<https://www.youtube.com/watch?v=k61nJkx5aDQ>

(Huth et al. 2016; Nature)

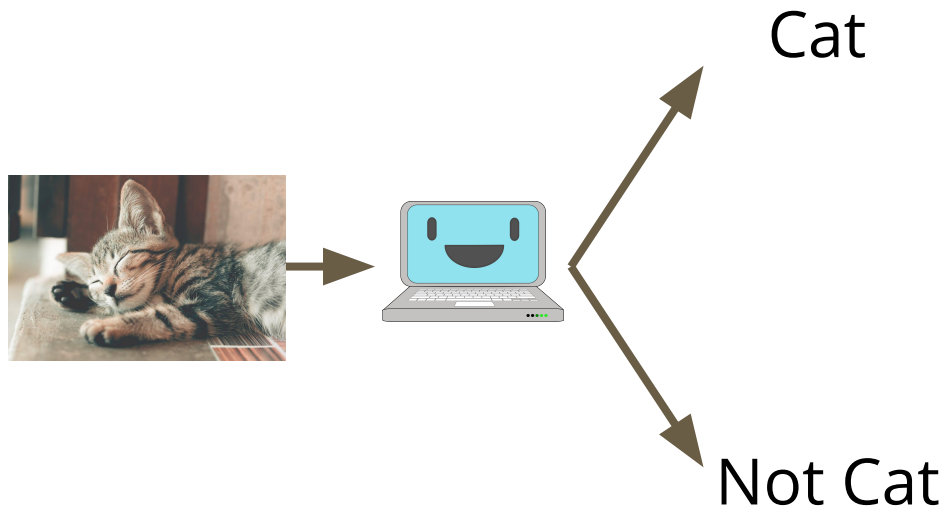
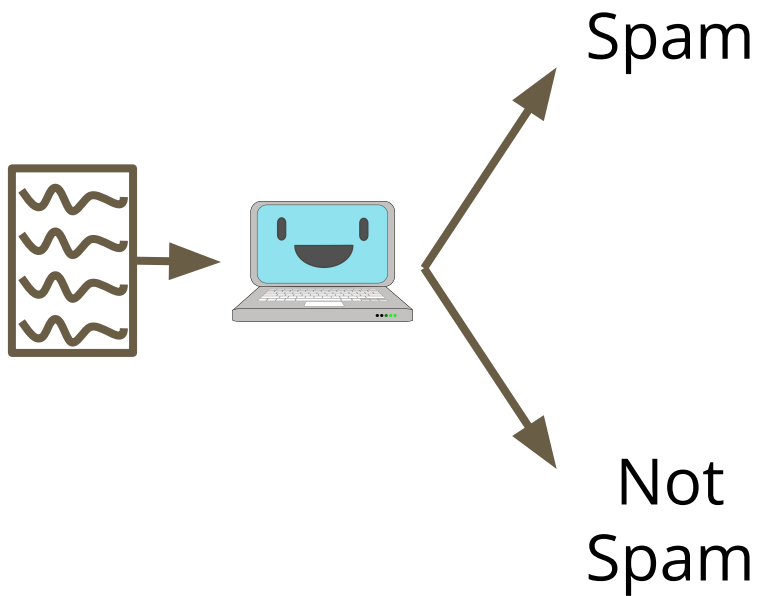
NLP for Healthcare: Objective Clinical Tests

- 
- "which"/"that"
 - phrase length
 - coherence measures

Reason 4: Transferable Skills

Machine learning with language data

can teach you about machine learning with other data!



Reason 5: NLP needs your help

Ambiguity (from Jurafsky and Martin)

I made her duck

How do you represent a sentence?

"an Italian restaurant near me??"

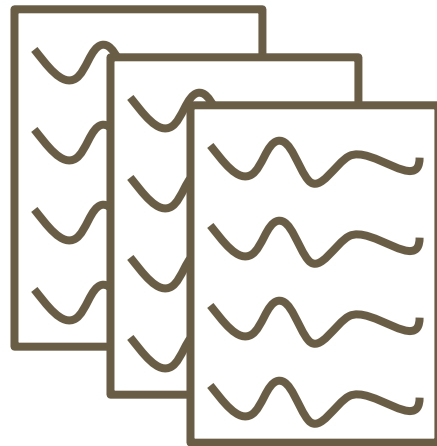


How do you represent a sentence?

"an Italian restaurant near me??"



{near, italian, restaurant} \cap



How do you represent a sentence?

"an Italian restaurant near me??"

→

$$\begin{aligned} \exists x \text{ s.t. } & \textit{restaurant}(x) \wedge \\ & \textit{serves}(x, \text{Italian food}) \wedge \\ & \textit{near}(x, \text{me}) \end{aligned}$$

Evaluation (what makes a good conversation anyway?)



Hey Alexa... I'm feeling sad.

Evaluation (what makes a good conversation anyway?)



Hey Alexa... I'm feeling sad.



I'm sorry you're feeling sad...

Evaluation (what makes a good conversation anyway?)

Hey Alexa... I'm feeling sad.

I'm sorry you're feeling sad...

Maybe you should call your friend!
Want me to connect you?

Evaluation (what makes a good conversation anyway?)

Hey Alexa... I'm feeling sad.

I'm sorry you're feeling sad...

Maybe you should call your friend!
Want me to connect you?

HERE ARE THE TOP WEB
RESULTS FOR "SAD"

Low Resource Languages

English <-> Spanish (Spanish has 480 million native speakers)

English <-> Urdu (Urdu has 70 million native speakers)

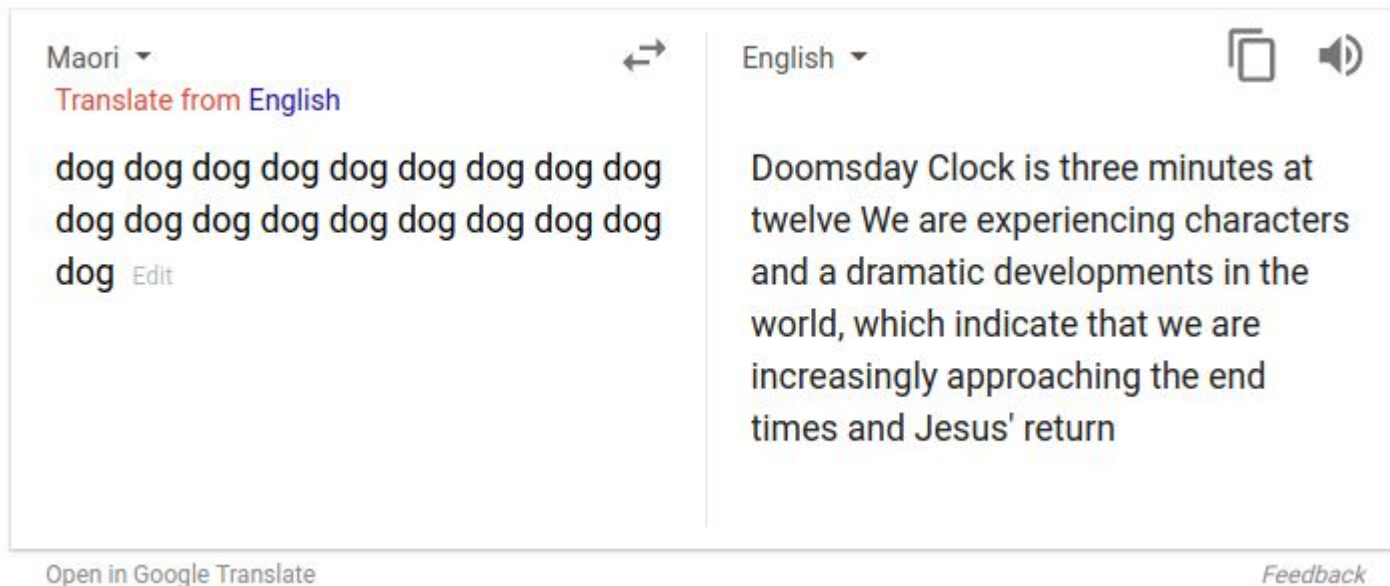
English <-> Hmong (Hmong has 12 million native speakers)

For fun: Fixedpoints of Google Translate?

I hopped off the plane at LAX with a dream and my cardigan. Welcome to the land of fame excess, whoa, am I gonna fit in? Jumped in the cab, here I am for the first time.

<https://www.translationparty.com/>

For fun 2: Google Translate Horror Movie?



We've come a long way... But we have a long path ahead...

Modified versions of the previous google assistant queries:

"What are Jordan's career point totals?" (ambiguity)

"What's not the capital of the United States" (negations)

"What's the tallest building in MN outside of minneapolis?" (representations)

and for fun "What are ... nevermind" (socially awkward!!)

What are we going to learn about?

Overview of topics

1. Language Modeling
2. Supervised Learning
3. Vector semantics
4. Feed-forward Deep Neural Networks
5. Recurrent Neural Networks
6. Parsing: Discovering the Latent Structures of Language
7. Ethics discussions
8. Computer Vision and Language (a topic close to my heart)

My goals for you skillswise

We'll get comfortable with...

1. Using python to turn text into data
2. Using numpy to manipulate vectors and matrices
3. Applied linear algebra (it's okay if you haven't taken linear algebra before!)
4. Python libraries for machine learning (sklearn, tensorflow, etc.)

Logistics