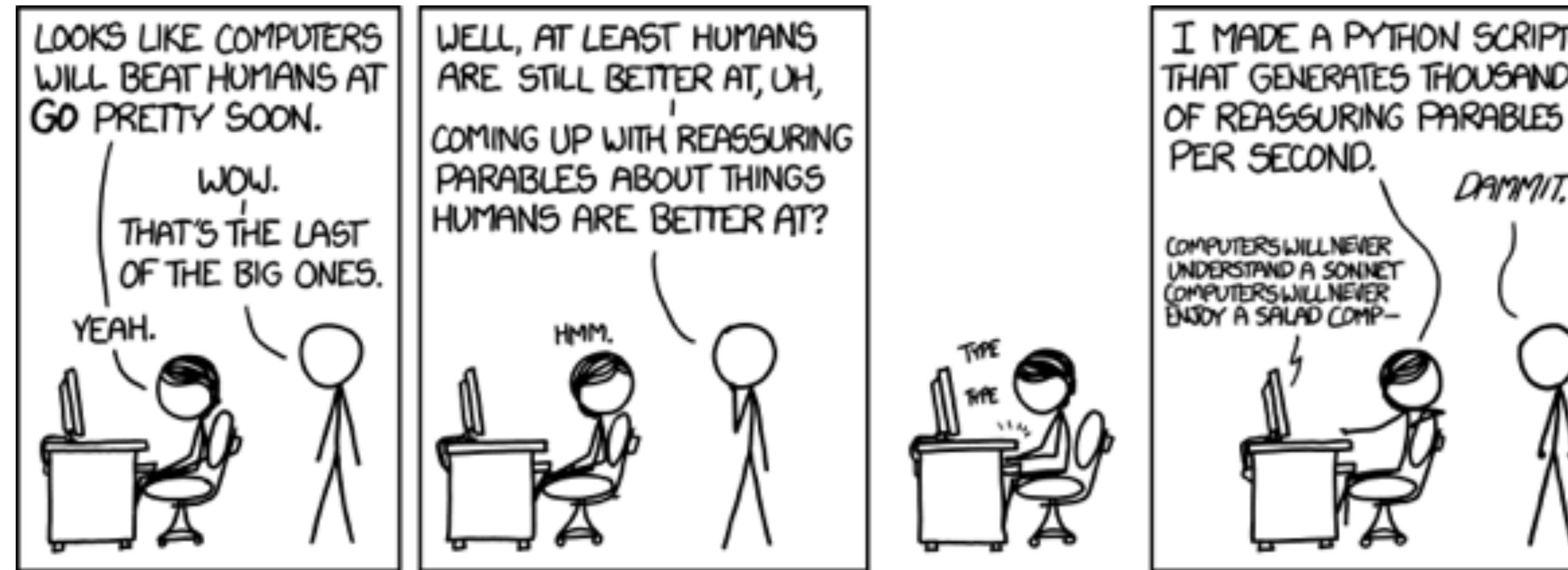


# 5525: Speech and Language Processing



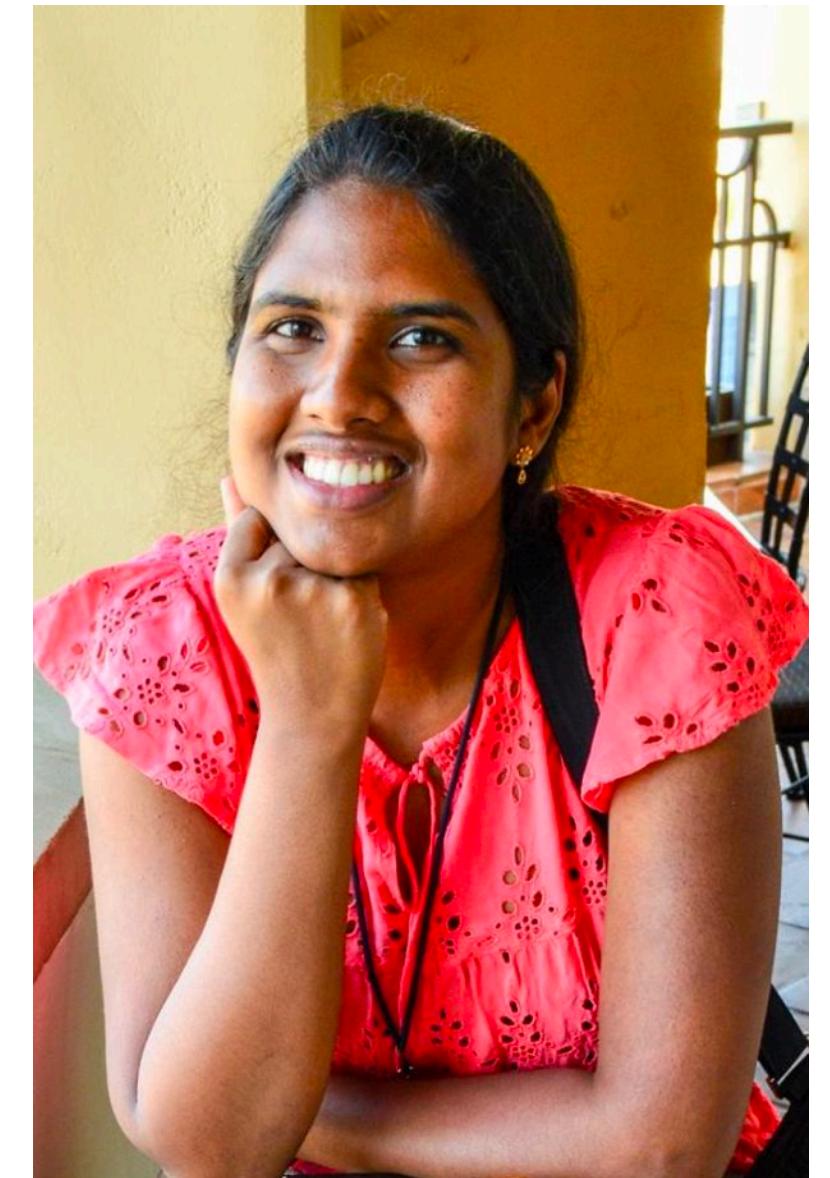
Alan Ritter

(many slides from Greg Durrett)

# Administrivia

---

- ▶ Course website:  
[http://aritter.github.io/courses/5525\\_spring19.html](http://aritter.github.io/courses/5525_spring19.html)
- ▶ Piazza: link on the course website
- ▶ My office hours: Friday 4-5pm DL 595
- ▶ TA: Mounica Maddela; Office hours:
  - ▶ Tuesday 2:30-3:30pm, Baker 480C



# Course Requirements

---

- ▶ Probability
- ▶ Linear Algebra
- ▶ Calculus
- ▶ Programming / Python experience
- ▶ Prior exposure to machine learning very helpful but not required

# Course Requirements

---

- ▶ Probability
- ▶ Linear Algebra
- ▶ Calculus
- ▶ Programming / Python experience
- ▶ Prior exposure to machine learning very helpful but not required

There will be a lot of math and programming!

# Enrollment

---

# Enrollment

---

- ▶ Homework 1 is out now (due in 2 weeks):

# Enrollment

---

- ▶ Homework 1 is out now (due in 2 weeks):
  - ▶ Please look at the assignment well before then

# Enrollment

---

- ▶ Homework 1 is out now (due in 2 weeks):
  - ▶ Please look at the assignment well before then
  - ▶ If this seems like it'll be challenging for you, come and talk to me (this is smaller-scale than the later assignments, which are smaller-scale than the final project)

# Texts

---

- ▶ 2 great textbooks for NLP
- ▶ There will be assigned readings from both
- ▶ Both freely available online

Natural Language Processing

**Speech and Language Processing** (3rd ed. draft)

[Dan Jurafsky](#) and [James H. Martin](#)

Jacob Eisenstein

# What's the goal of NLP?

---

# What's the goal of NLP?

---

- ▶ Be able to solve problems that require deep understanding of text

# What's the goal of NLP?

---

- ▶ Be able to solve problems that require deep understanding of text
- ▶ Example: dialogue systems

# What's the goal of NLP?

---

- ▶ Be able to solve problems that require deep understanding of text
- ▶ Example: dialogue systems



# What's the goal of NLP?

---

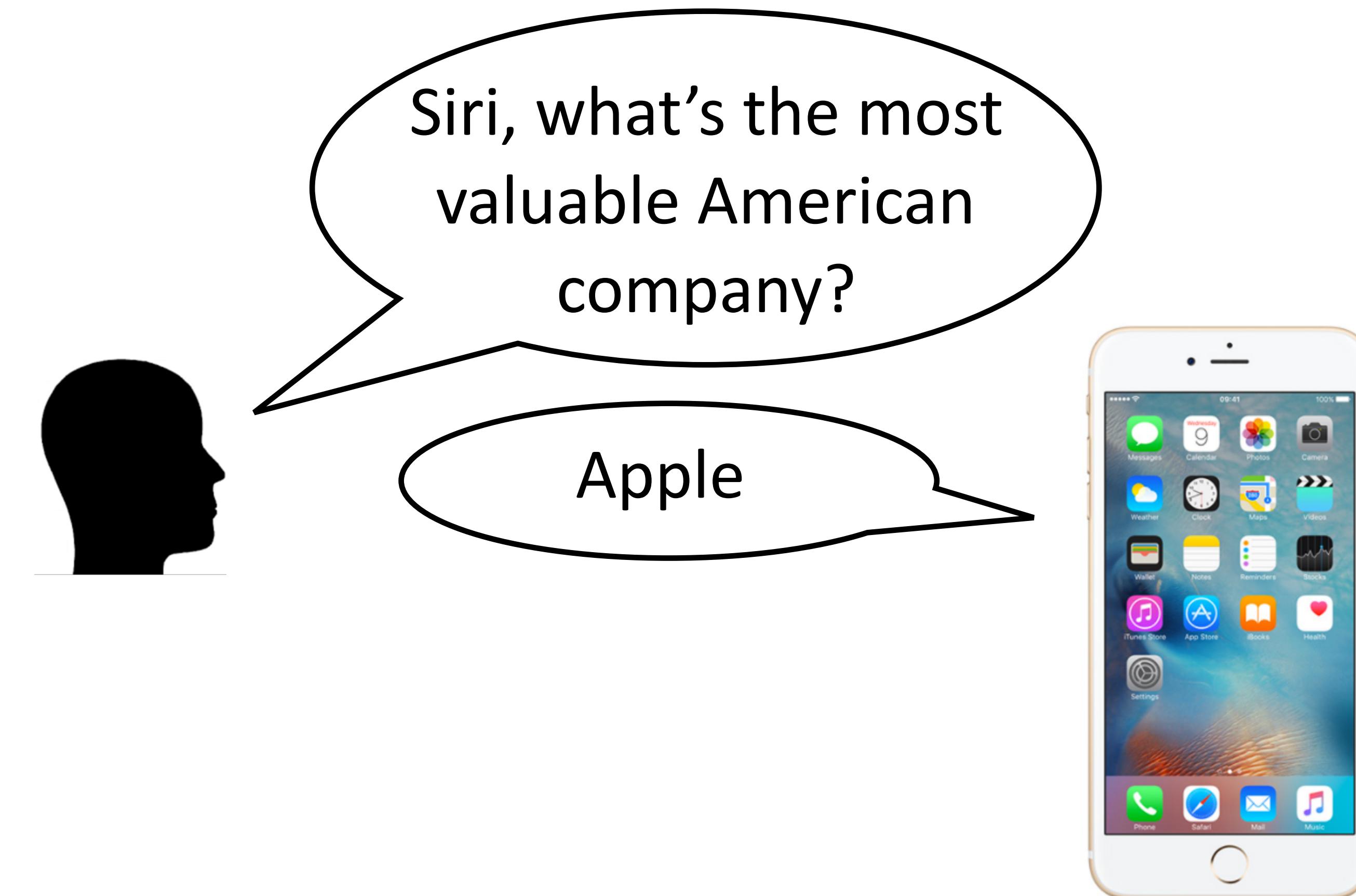
- ▶ Be able to solve problems that require deep understanding of text
- ▶ Example: dialogue systems



# What's the goal of NLP?

---

- ▶ Be able to solve problems that require deep understanding of text
- ▶ Example: dialogue systems



# What's the goal of NLP?

---

- ▶ Be able to solve problems that require deep understanding of text
- ▶ Example: dialogue systems



# What's the goal of NLP?

---

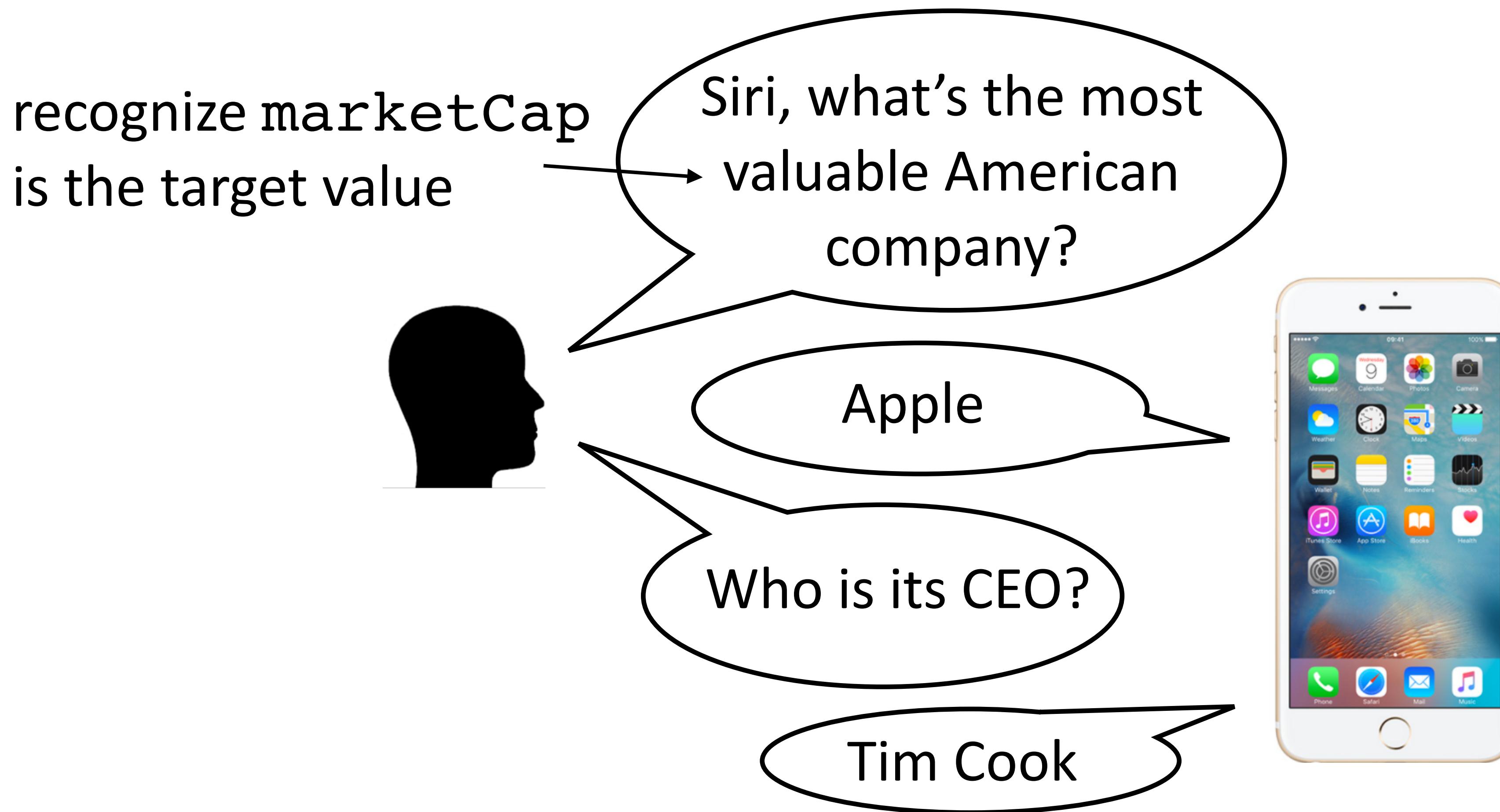
- ▶ Be able to solve problems that require deep understanding of text
- ▶ Example: dialogue systems



# What's the goal of NLP?

---

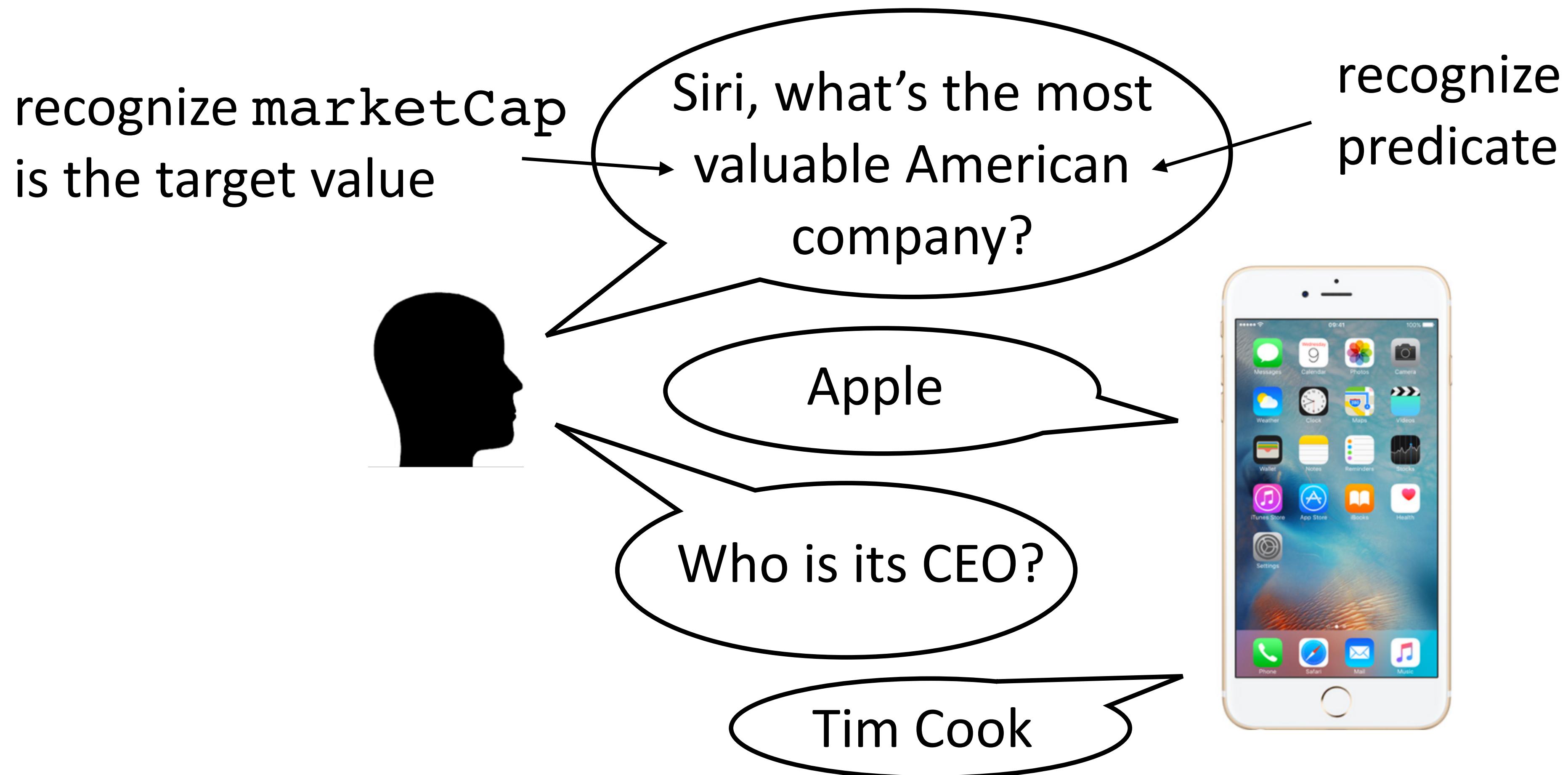
- ▶ Be able to solve problems that require deep understanding of text
- ▶ Example: dialogue systems



# What's the goal of NLP?

---

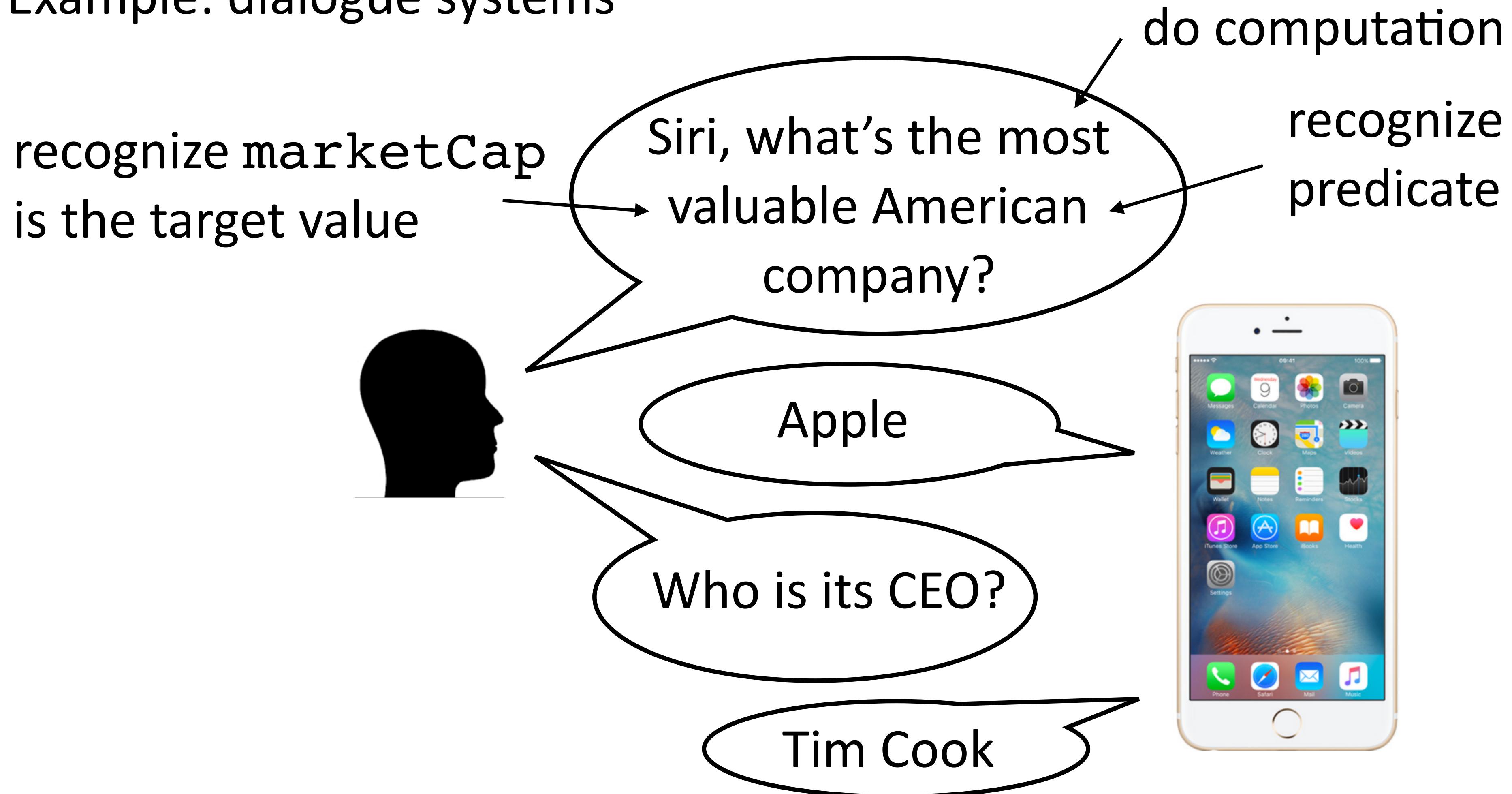
- ▶ Be able to solve problems that require deep understanding of text
- ▶ Example: dialogue systems



# What's the goal of NLP?

---

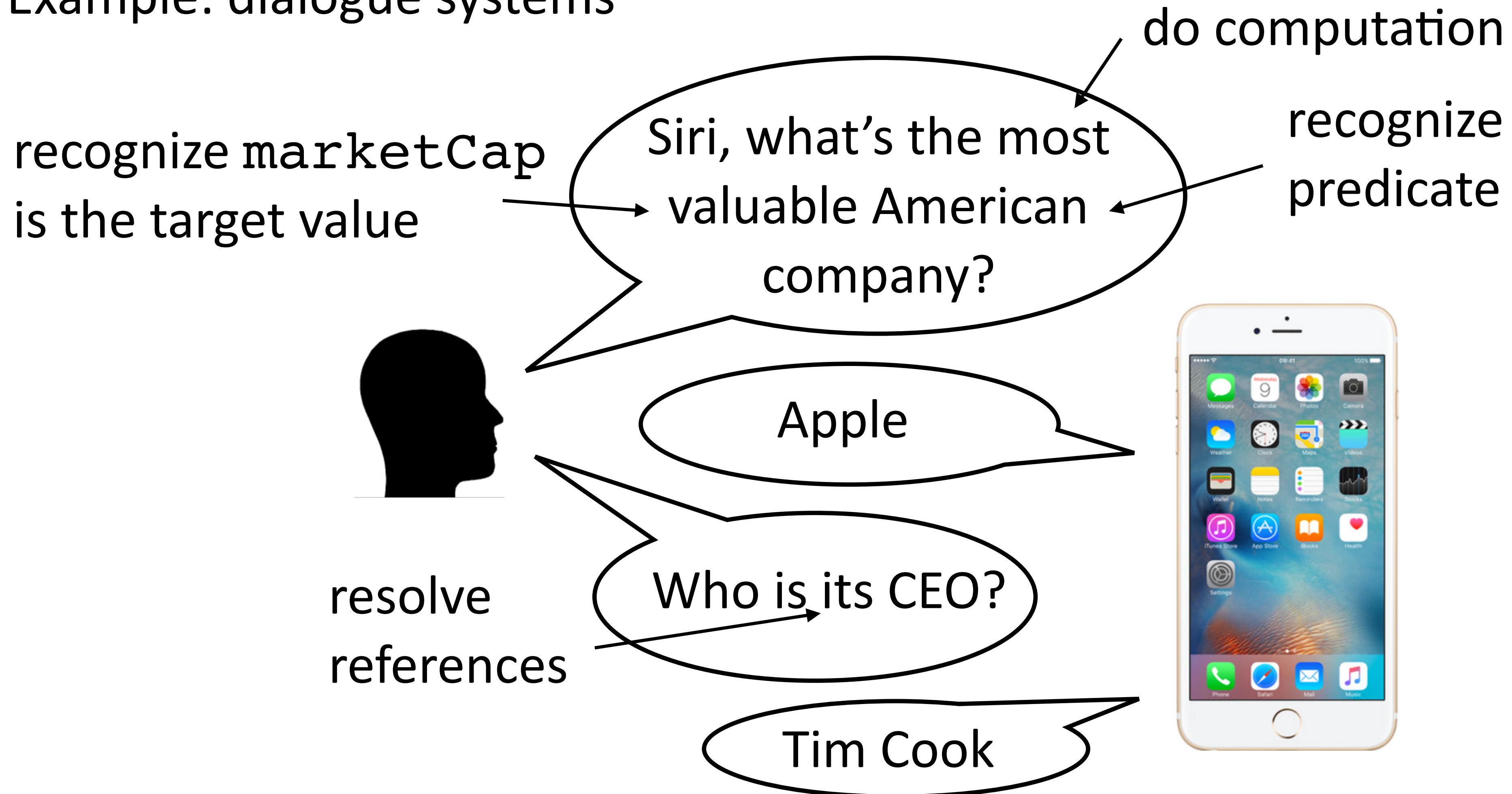
- ▶ Be able to solve problems that require deep understanding of text
- ▶ Example: dialogue systems



# What's the goal of NLP?

---

- ▶ Be able to solve problems that require deep understanding of text
- ▶ Example: dialogue systems



# Automatic Summarization

---

# Automatic Summarization

---

POLITICS

## ***Google Critic Ousted From Think Tank Funded by the Tech Giant***

WASHINGTON — In the hours after European antitrust regulators levied a record [\\$2.7 billion fine](#) against Google in late June, an influential Washington think tank learned what can happen when a tech giant that shapes public policy debates with its enormous wealth is criticized.

• • •

But not long after one of New America's scholars [posted a statement](#) on the think tank's website praising the European Union's penalty against Google, Mr. Schmidt, who had been chairman of New America until 2016, communicated his displeasure with the statement to the group's president, Anne-Marie Slaughter, according to the scholar.

• • •

Ms. Slaughter told Mr. Lynn that “the time has come for Open Markets and New America to part ways,” according to an email from Ms. Slaughter to Mr. Lynn. The email suggested that the entire Open Markets team — nearly 10 full-time employees and unpaid fellows — would be exiled from New America.

# Automatic Summarization

POLITICS

## ***Google Critic Ousted From Think Tank Funded by the Tech Giant***

WASHINGTON — In the hours after European antitrust regulators levied a record [\\$2.7 billion fine](#) against Google in late June, an influential Washington think tank learned what can happen when a tech giant that shapes public policy debates with its enormous wealth is criticized.

...

But not long after one of New America's scholars [posted a statement](#) on the think tank's website praising the European Union's penalty against Google, Mr. Schmidt, who had been chairman of New America until 2016, communicated his displeasure with the statement to the group's president, Anne-Marie Slaughter, according to the scholar.

...

Ms. Slaughter told Mr. Lynn that “the time has come for Open Markets and New America to part ways,” according to an email from Ms. Slaughter to Mr. Lynn. The email suggested that the entire Open Markets team — nearly 10 full-time employees and unpaid fellows — would be exiled from New America.

One of New America's writers posted a statement critical of Google. Eric Schmidt, Google's CEO, was displeased.

The writer and his team were dismissed.

# Automatic Summarization

POLITICS

## ***Google Critic Ousted From Think Tank Funded by the Tech Giant***

WASHINGTON — In the hours after European antitrust regulators levied a record [\\$2.7 billion fine](#) against Google in late June, an influential Washington think tank learned what can happen when a tech giant that shapes public policy debates with its enormous wealth is criticized.

...

But not long after one of New America's scholars [posted a statement](#) on the think tank's website praising the European Union's penalty against Google, Mr. Schmidt, who had been chairman of New America until 2016, communicated his displeasure with the statement to the group's president, Anne-Marie Slaughter, according to the scholar.

...

Ms. Slaughter told Mr. Lynn that “the time has come for Open Markets and New America to part ways,” according to an email from Ms. Slaughter to Mr. Lynn. The email suggested that the entire Open Markets team — nearly 10 full-time employees and unpaid fellows — would be exiled from New America.

compress  
text

One of New America's writers posted a statement critical of Google. Eric Schmidt, Google's CEO, was displeased.

The writer and his team were dismissed.

# Automatic Summarization

POLITICS

## ***Google Critic Ousted From Think Tank Funded by the Tech Giant***

WASHINGTON — In the hours after European antitrust regulators levied a record [\\$2.7 billion fine](#) against Google in late June, an influential Washington think tank learned what can happen when a tech giant that shapes public policy debates with its enormous wealth is criticized.

...

But not long after one of New America's scholars [posted a statement](#) on the think tank's website praising the European Union's penalty against Google, Mr. Schmidt, who had been chairman of New America until 2016, communicated his displeasure with the statement to the group's president, Anne-Marie Slaughter, according to the scholar.

...

Ms. Slaughter told Mr. Lynn that “the time has come for Open Markets and New America to part ways,” according to an email from Ms. Slaughter to Mr. Lynn. The email suggested that the entire Open Markets team — nearly 10 full-time employees and unpaid fellows — would be exiled from New America.

compress  
text

provide missing  
context

One of New America's writers posted a statement critical of Google. Eric Schmidt, [Google's CEO](#), was displeased.

The writer and his team were dismissed.

# Automatic Summarization

POLITICS

## ***Google Critic Ousted From Think Tank Funded by the Tech Giant***

WASHINGTON — In the hours after European antitrust regulators levied a record [\\$2.7 billion fine](#) against Google in late June, an influential Washington think tank learned what can happen when a tech giant that shapes public policy debates with its enormous wealth is criticized.

...

But not long after one of New America's scholars [posted a statement](#) on the think tank's website praising the European Union's penalty against Google, Mr. Schmidt, who had been chairman of New America until 2016, communicated his displeasure with the statement to the group's president, Anne-Marie Slaughter, according to the scholar.

...

Ms. Slaughter told Mr. Lynn that “the time has come for Open Markets and New America to part ways,” according to an email from Ms. Slaughter to Mr. Lynn. The email suggested that the entire Open Markets team — nearly 10 full-time employees and unpaid fellows — would be [exiled](#) from New America.

compress  
text

provide missing  
context

One of New America's writers posted a statement critical of Google. Eric Schmidt, [Google's CEO](#), was displeased.

The writer and his team were [dismissed](#).

paraphrase to provide clarity

# Machine Translation

---



< 2/8

特朗普偕家人在白宫阳台观看百年一遇日全食

>

People's Daily, August 30, 2017

# Machine Translation



A photograph of a woman with blonde hair, wearing a black sleeveless dress and 3D glasses, looking upwards. She is standing next to a man in a dark suit. The background is a plain wall.

Translate

English French Spanish Chinese - detected ▾

特朗普偕家人在白宫阳台观看百年一遇日全食

2/8 特朗普偕家人在白宫阳台观看百年一遇日全食

The image shows a woman wearing 3D glasses, looking up at a screen. A translation interface is overlaid on the image. The text "Translate" is in red. Below it is a row of language buttons: English, French, Spanish, Chinese - detected (which is highlighted), and a dropdown arrow. Below the buttons is a box containing the Chinese text "特朗普偕家人在白宫阳台观看百年一遇日全食". This text is enclosed in an orange border. At the bottom left, there is a navigation indicator showing "2/8" and the same Chinese text again. The overall context is a machine translation demonstration.

People's Daily, August 30, 2017

# Machine Translation



Trump Pope family watch a hundred years a year in the White House balcony

# Machine Translation



Trump Pope family watch a hundred years a year in the White House balcony

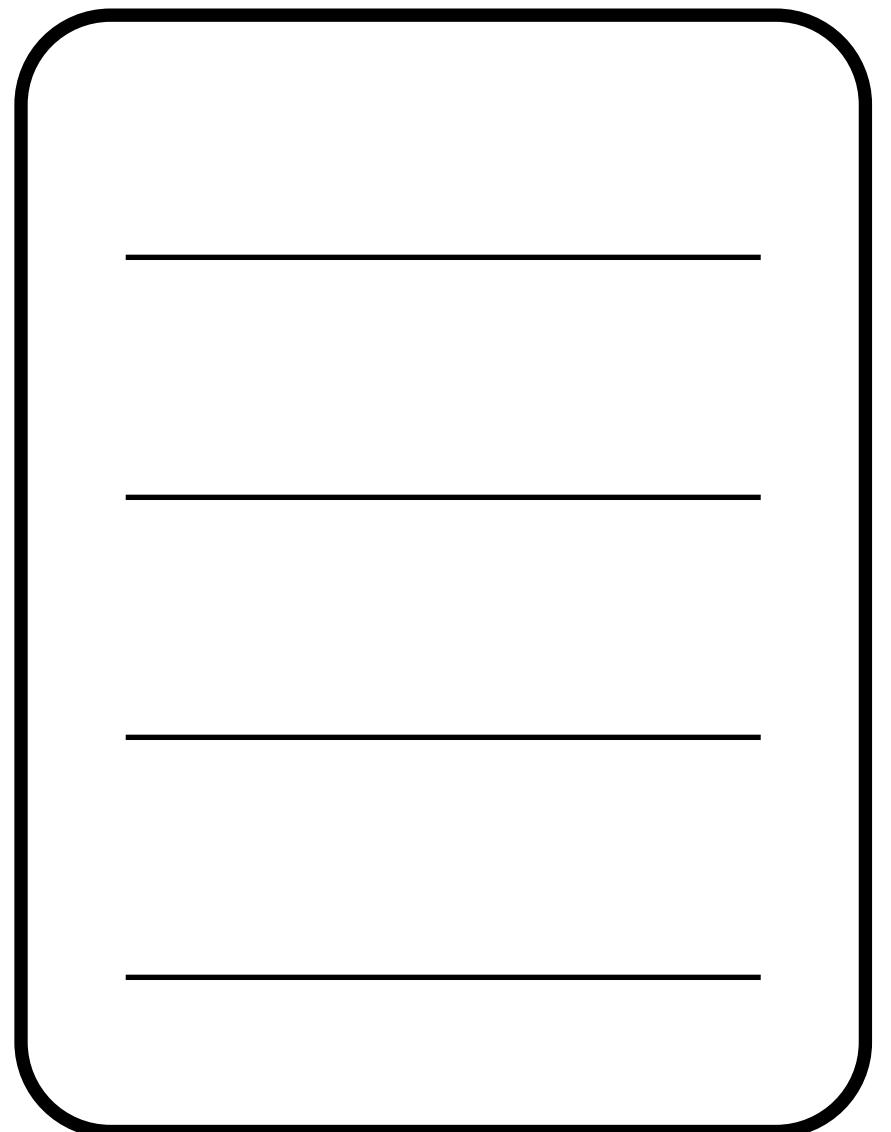
# NLP Analysis Pipeline

---

# NLP Analysis Pipeline

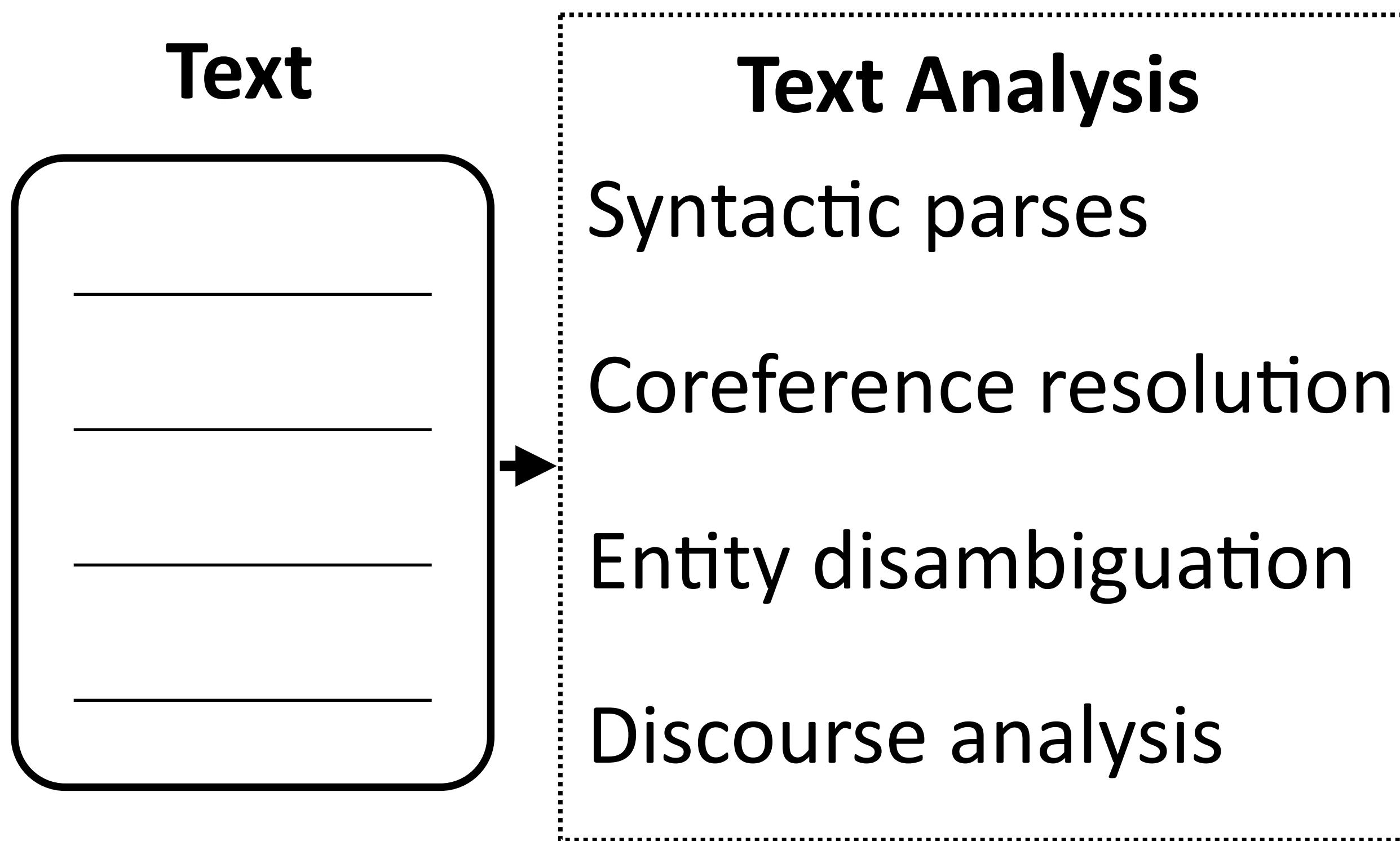
---

## Text

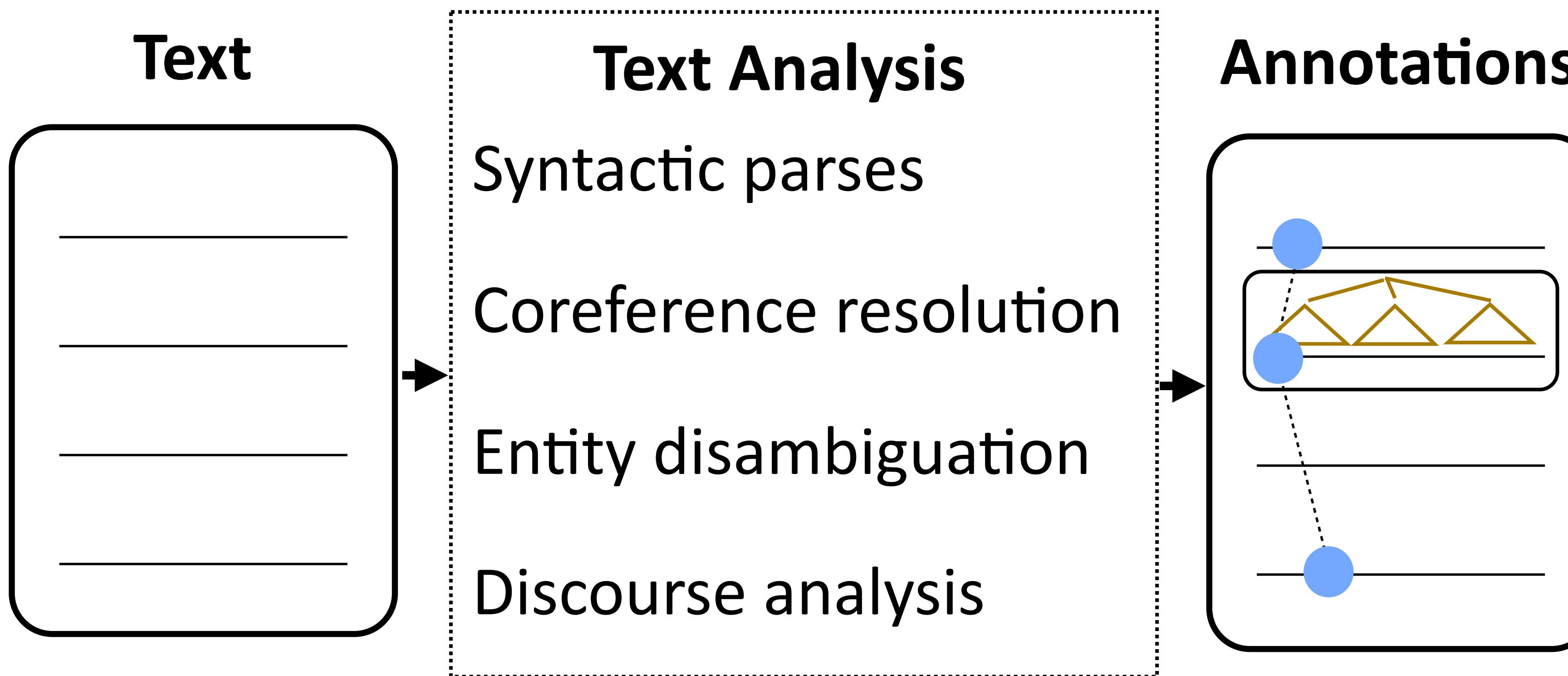


# NLP Analysis Pipeline

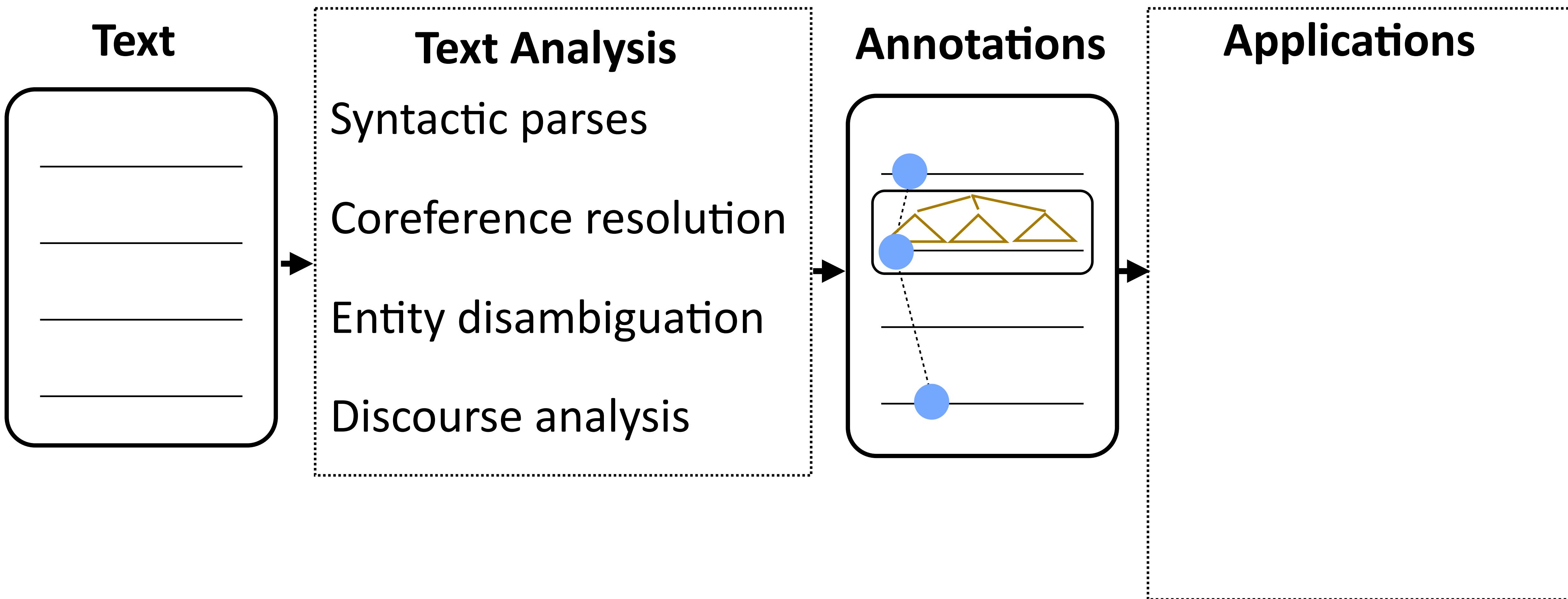
---



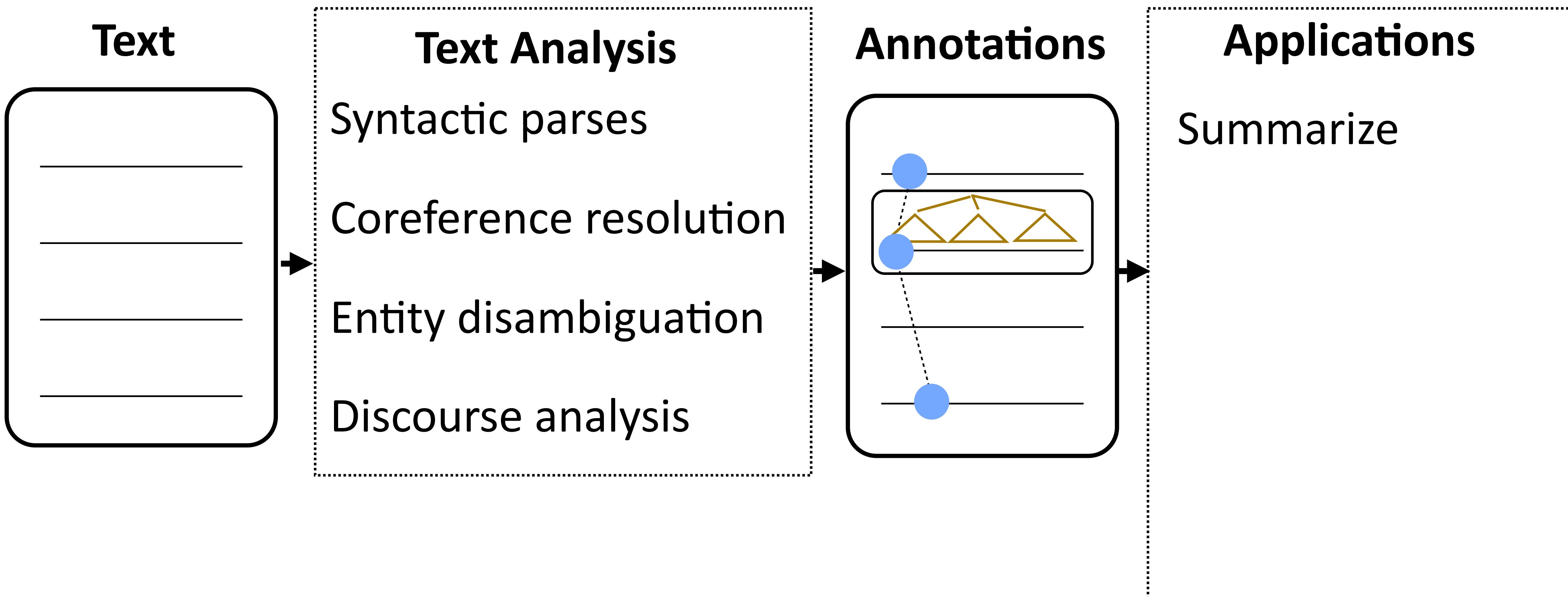
# NLP Analysis Pipeline



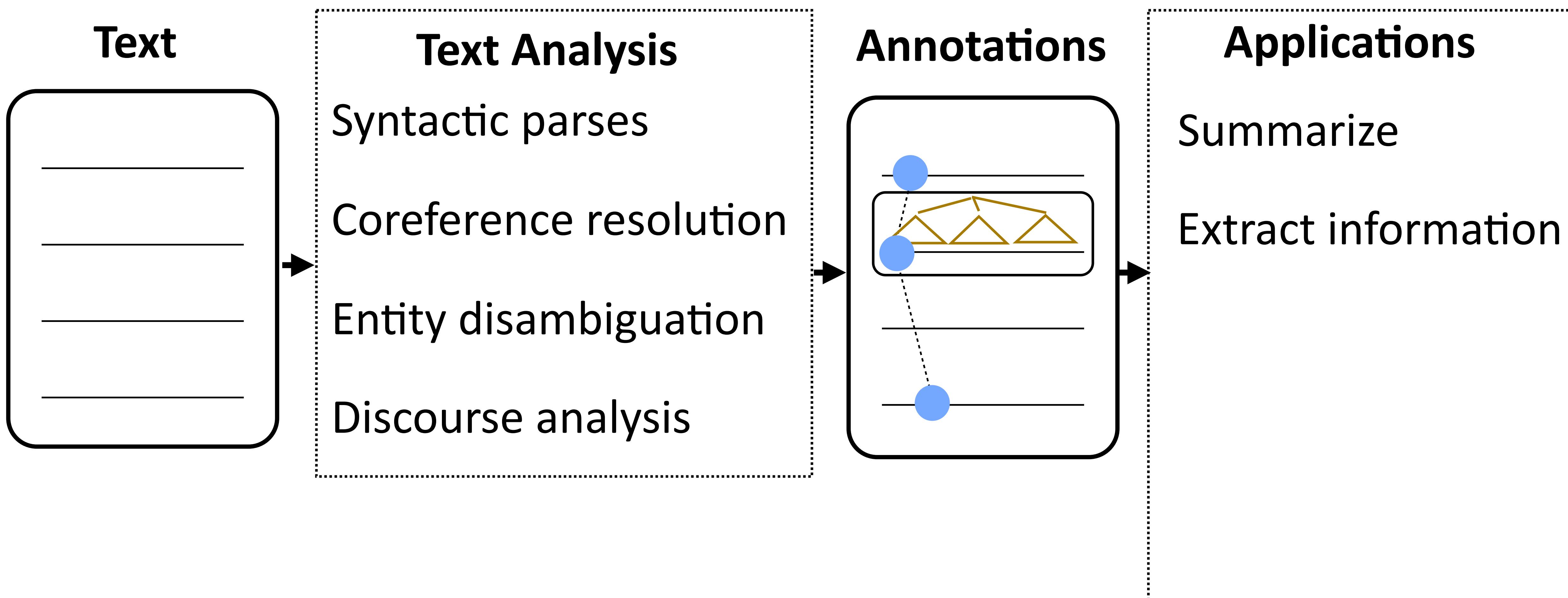
# NLP Analysis Pipeline



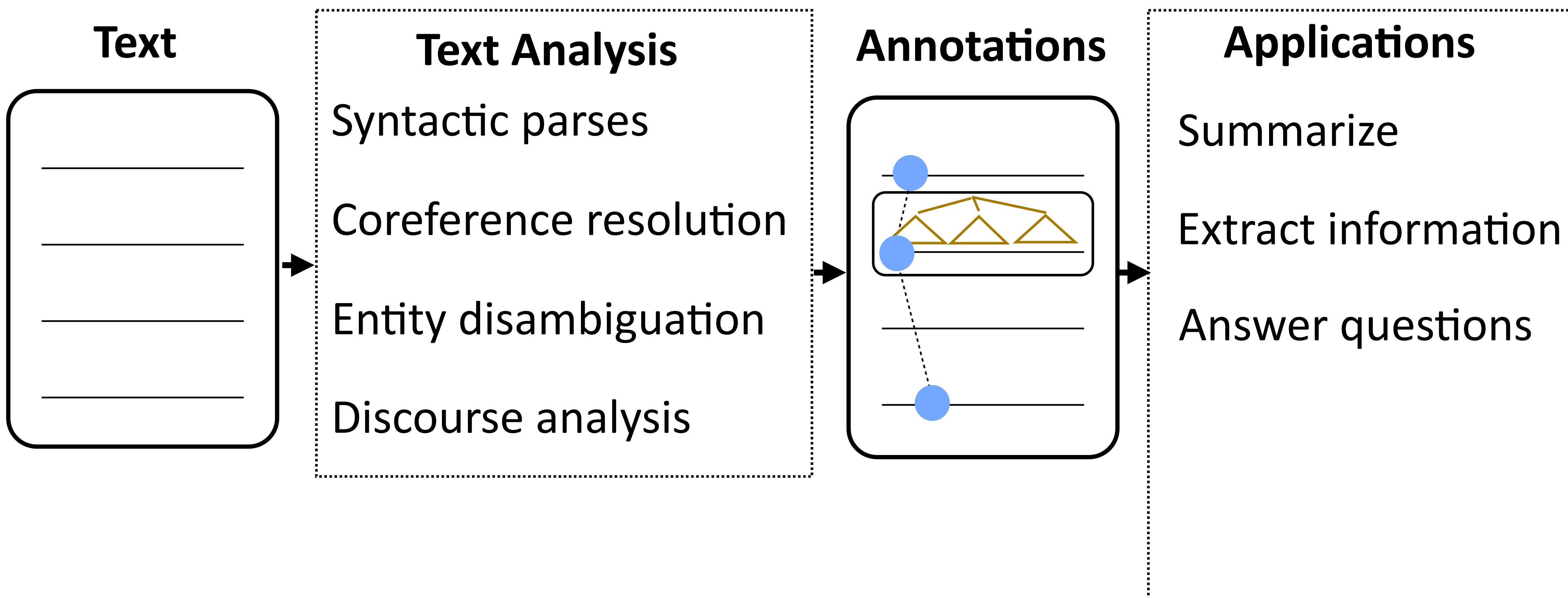
# NLP Analysis Pipeline



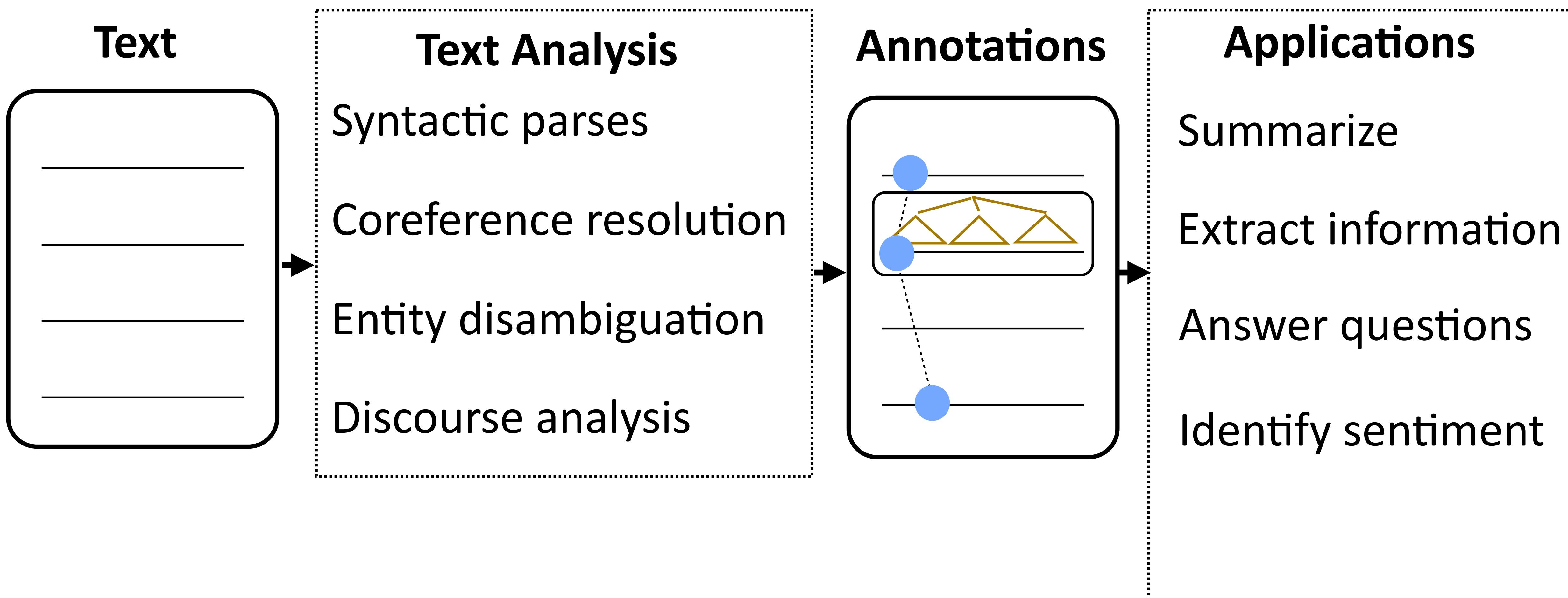
# NLP Analysis Pipeline



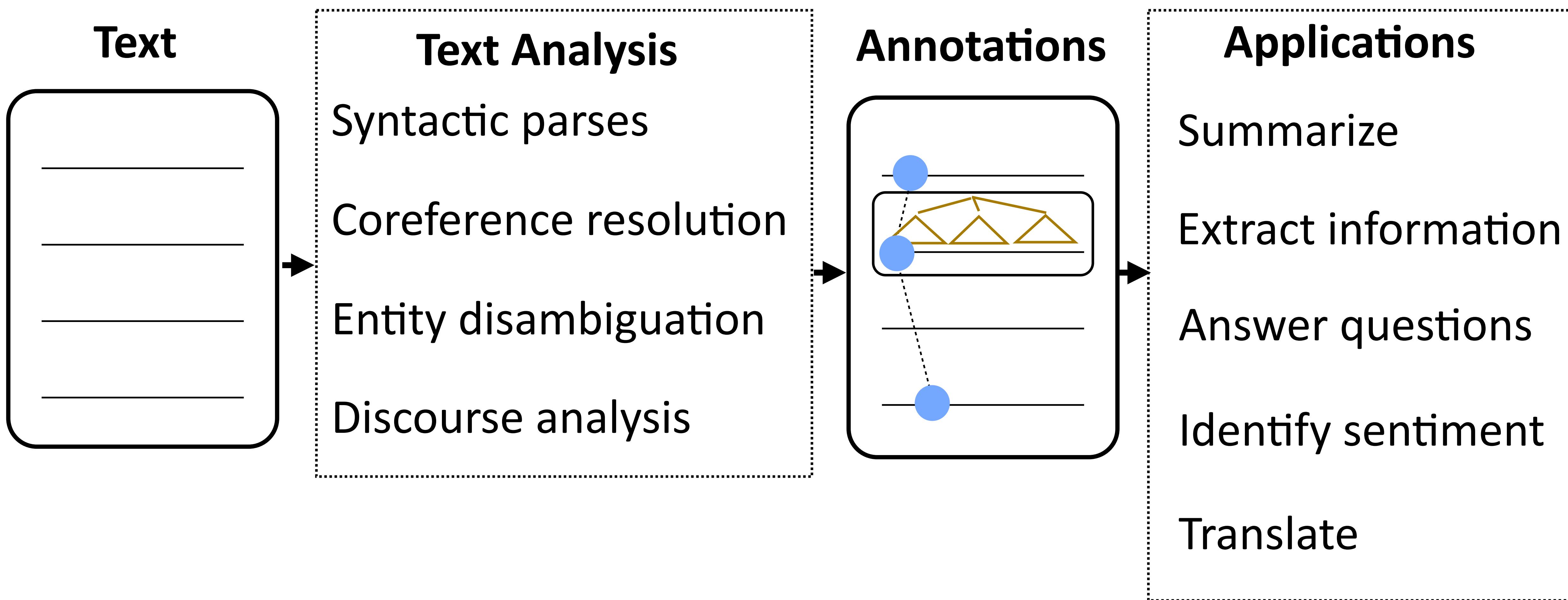
# NLP Analysis Pipeline



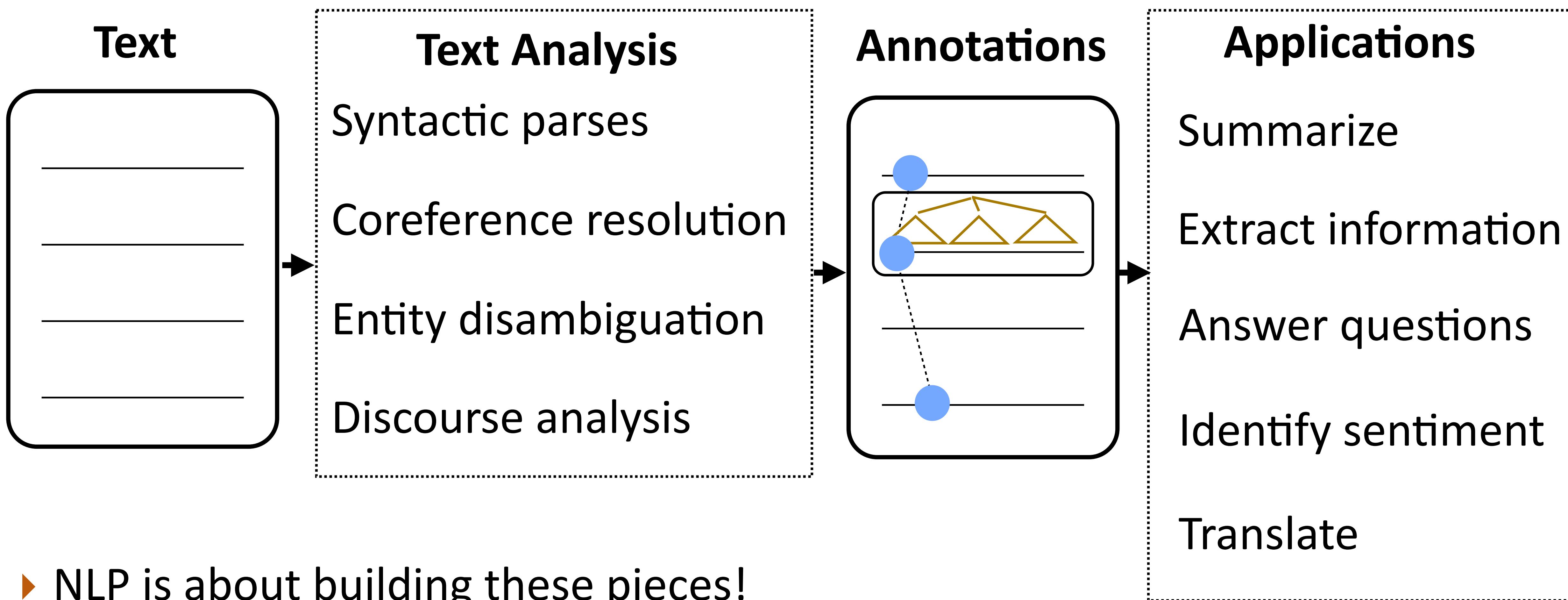
# NLP Analysis Pipeline



# NLP Analysis Pipeline

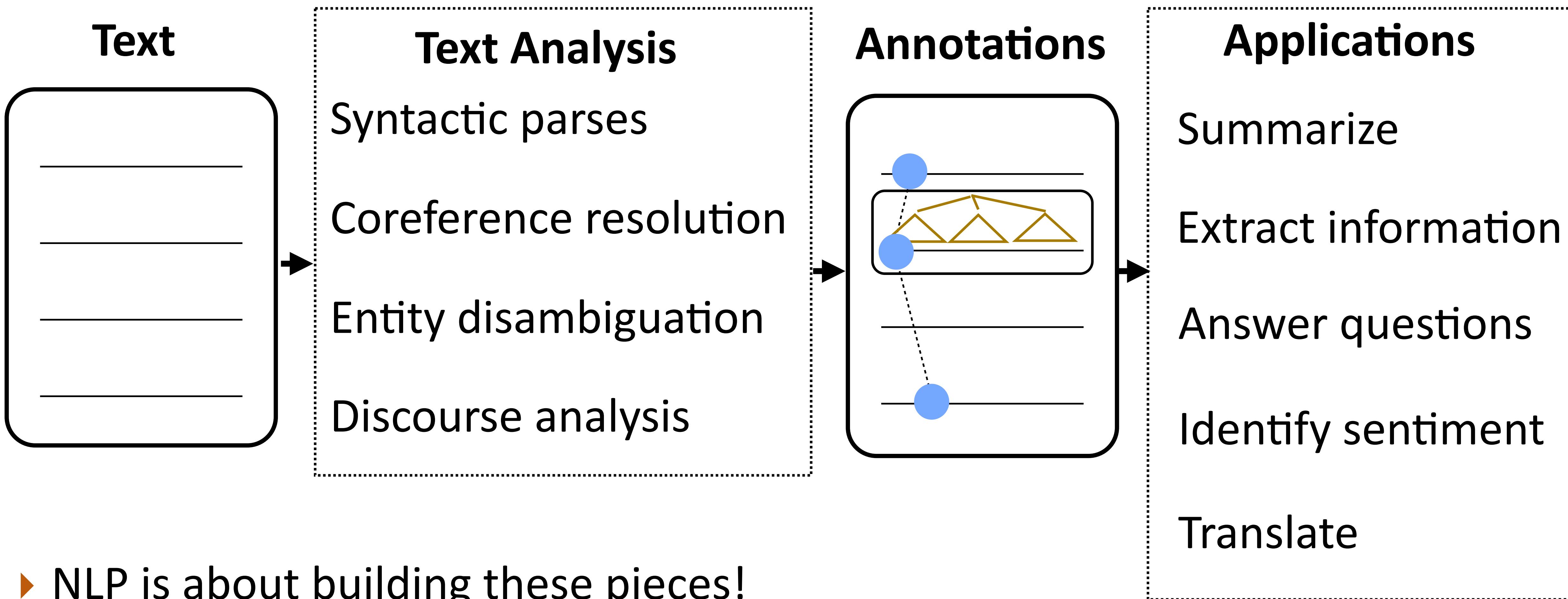


# NLP Analysis Pipeline



- ▶ NLP is about building these pieces!

# NLP Analysis Pipeline

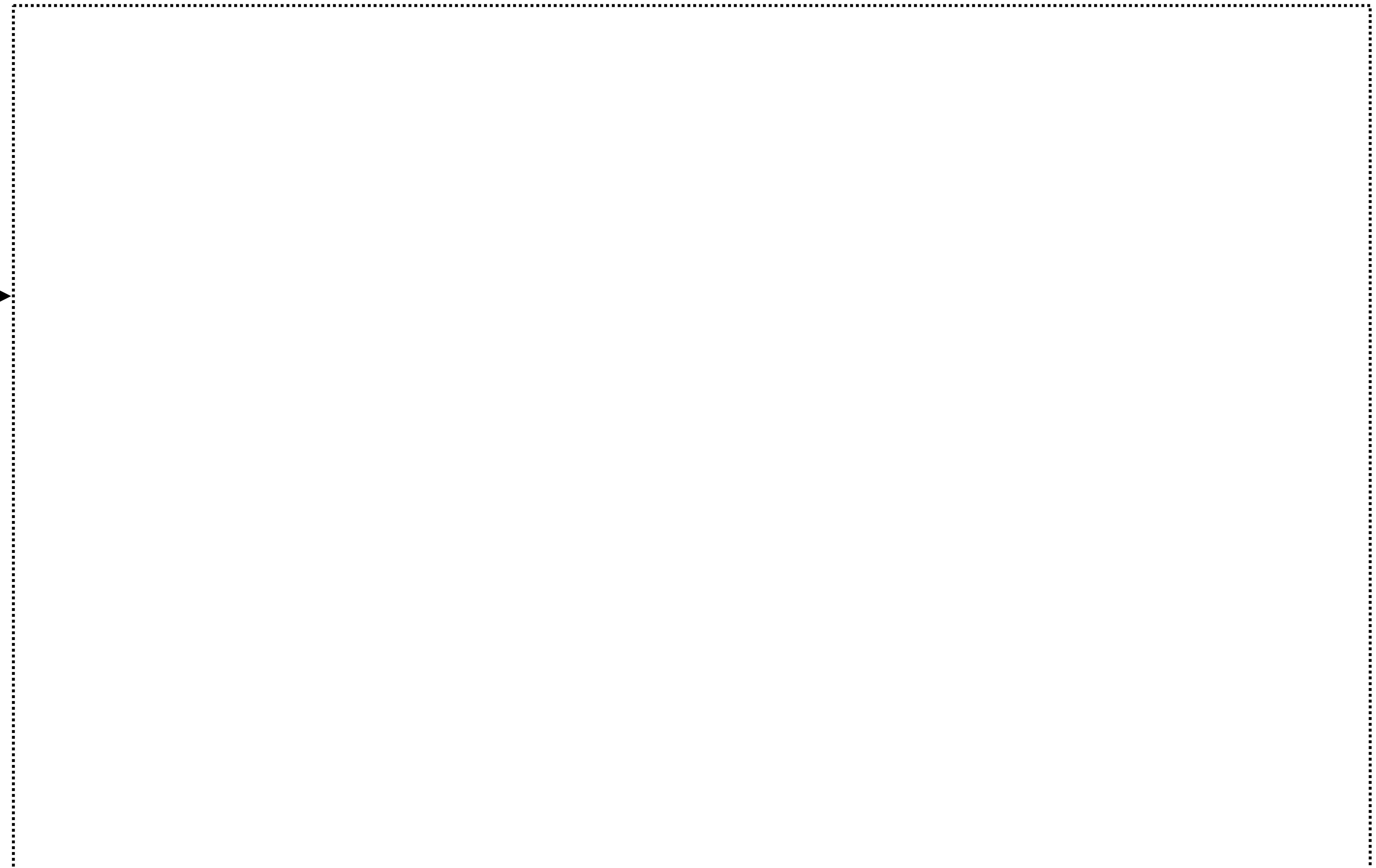
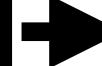
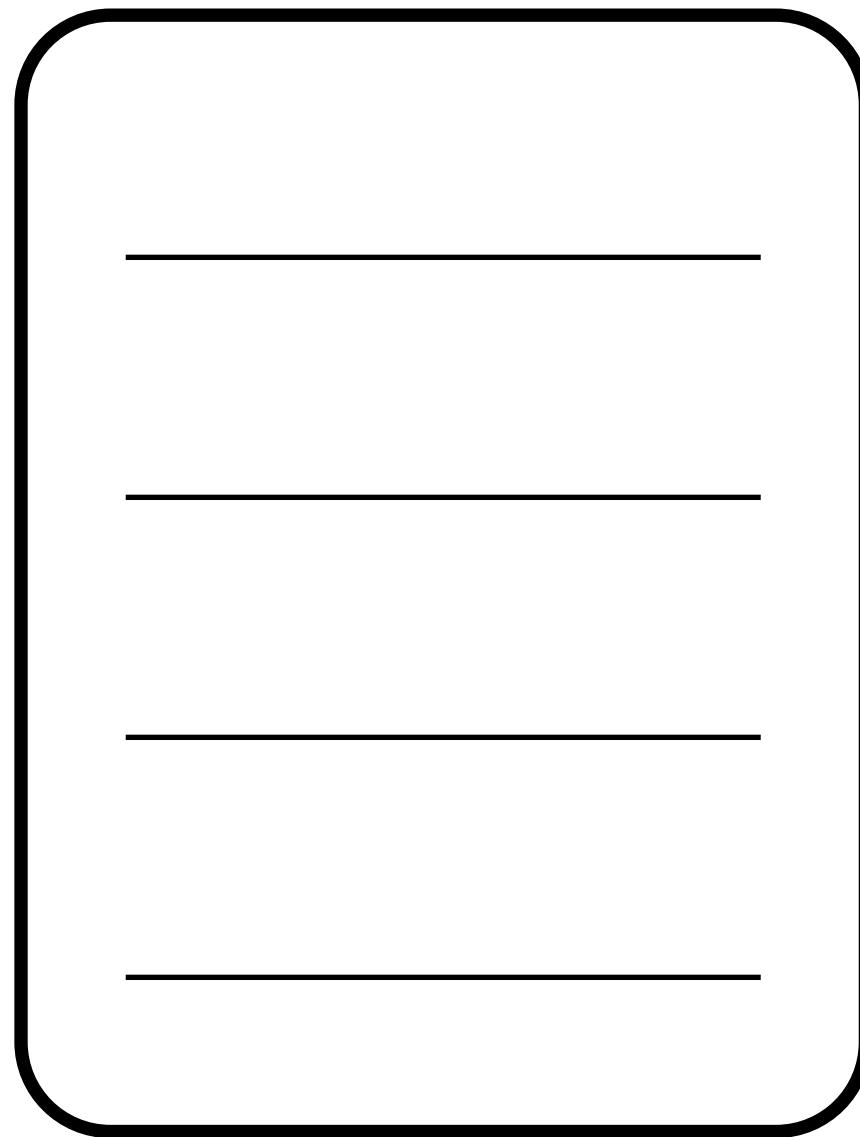


- ▶ NLP is about building these pieces!
- ▶ All of these components are modeled with statistical approaches trained with machine learning

# How do we represent language?

---

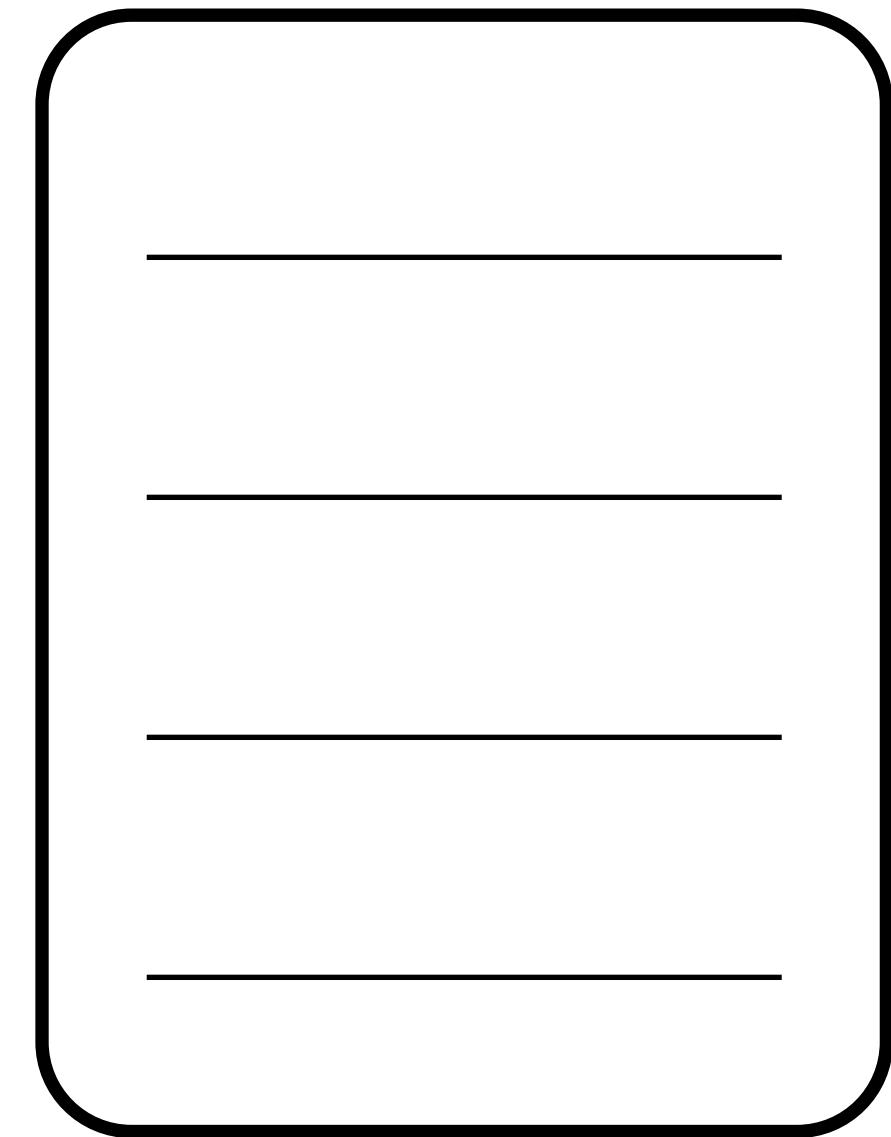
**Text**



# How do we represent language?

---

**Text**

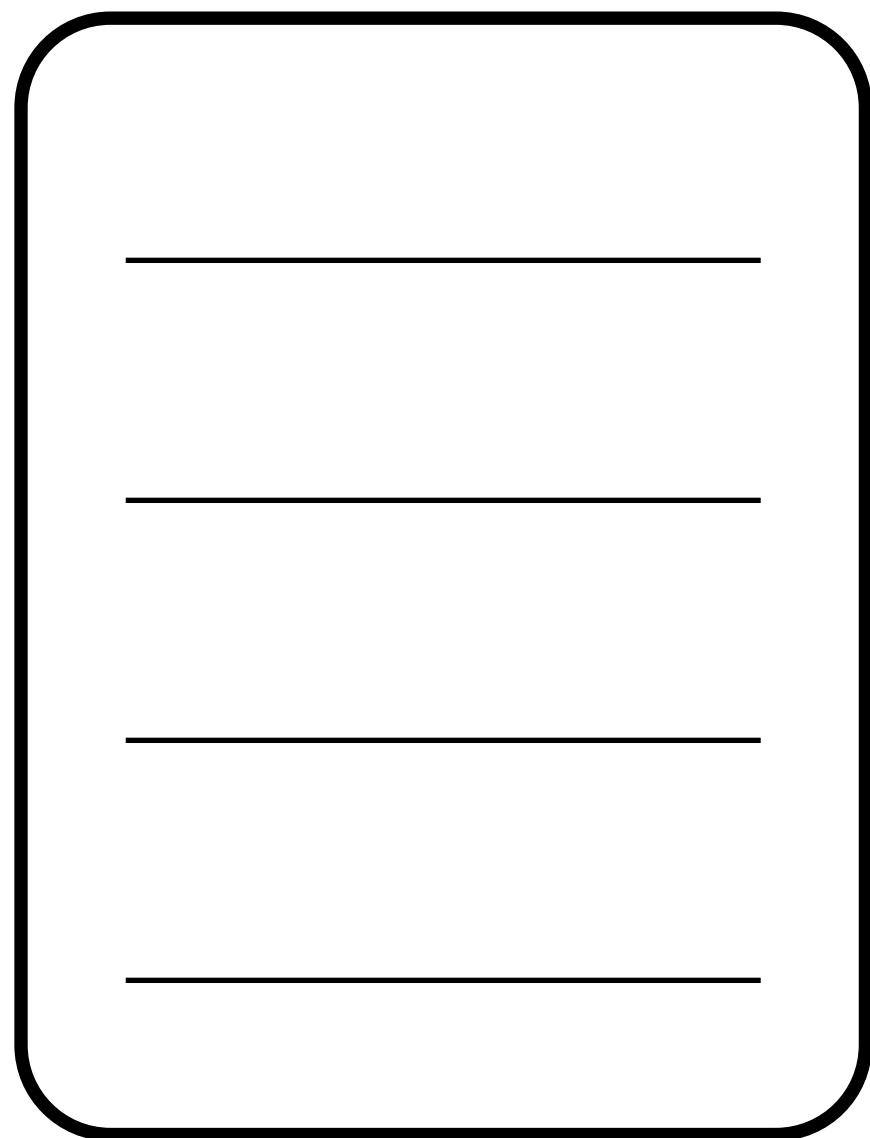


**Labels**



# How do we represent language?

**Text**

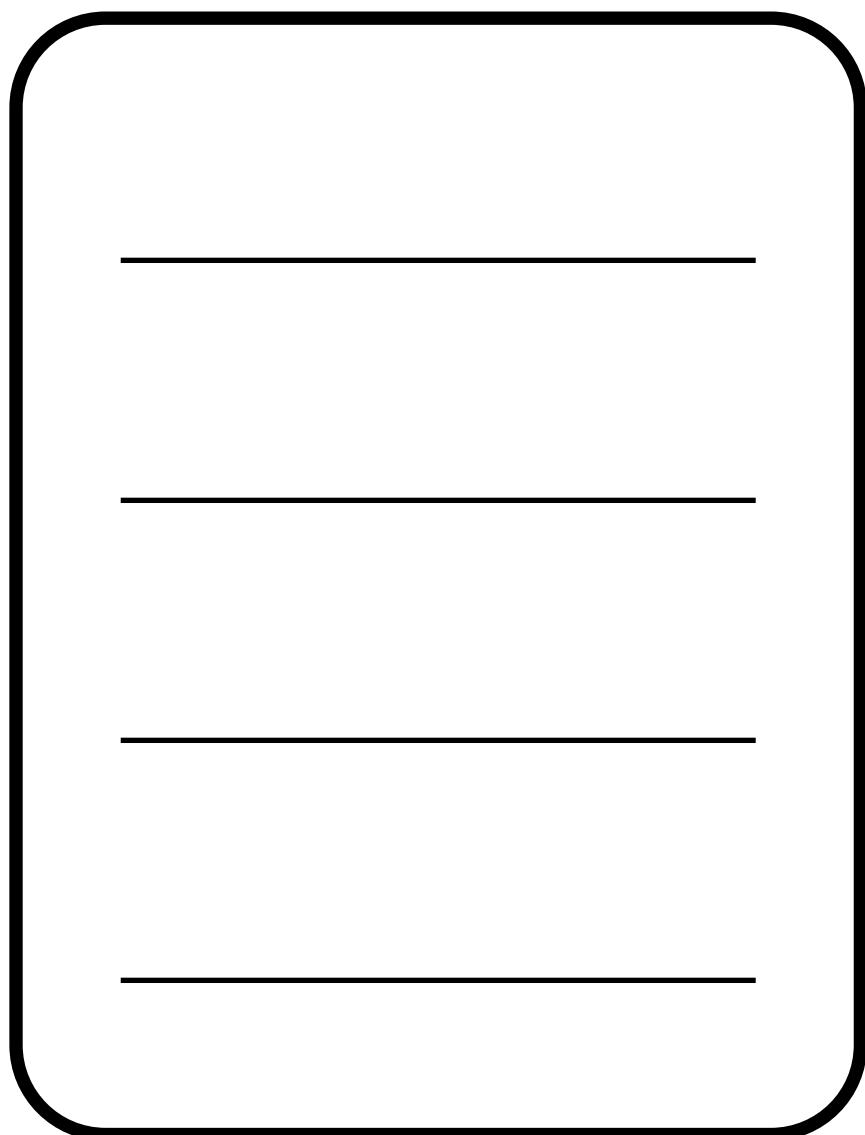


**Labels**

*the movie was good* +

# How do we represent language?

Text



Labels

*the movie was good* +

*Beyoncé had one of the best videos of all time* subjective

# How do we represent language?

Text

Labels

*the movie was good* +

*Beyoncé had one of the best videos of all time* subjective

Sequences/tags

PERSON

*Tom Cruise* stars in the new *Mission Impossible* film

WORK\_OF\_ART

# How do we represent language?

Text

Labels

*the movie was good* +

*Beyoncé had one of the best videos of all time* subjective

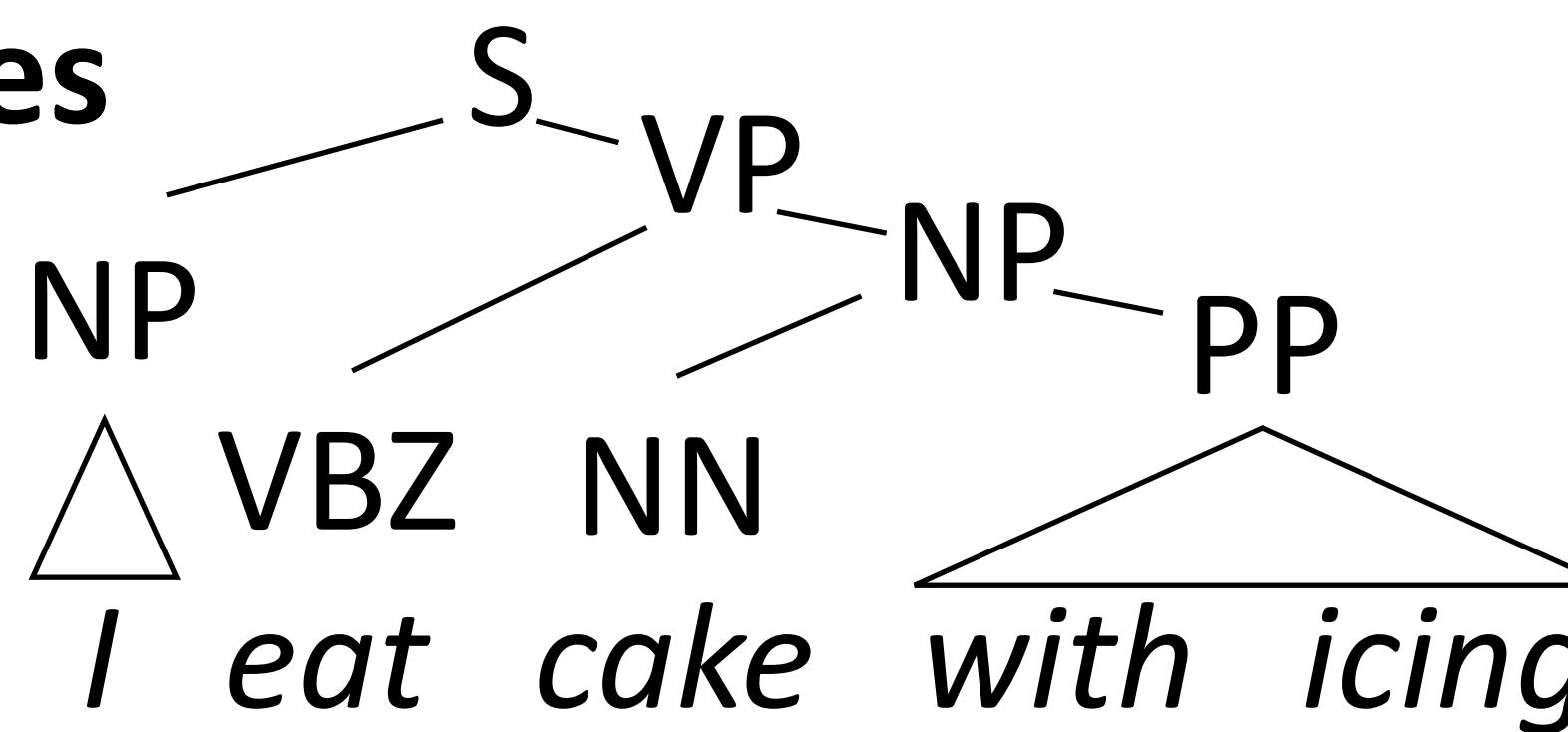
Sequences/tags

PERSON

*Tom Cruise stars in the new Mission Impossible film*

WORK\_OF\_ART

Trees



# How do we represent language?

Text

Labels

*the movie was good* +

*Beyoncé had one of the best videos of all time* subjective

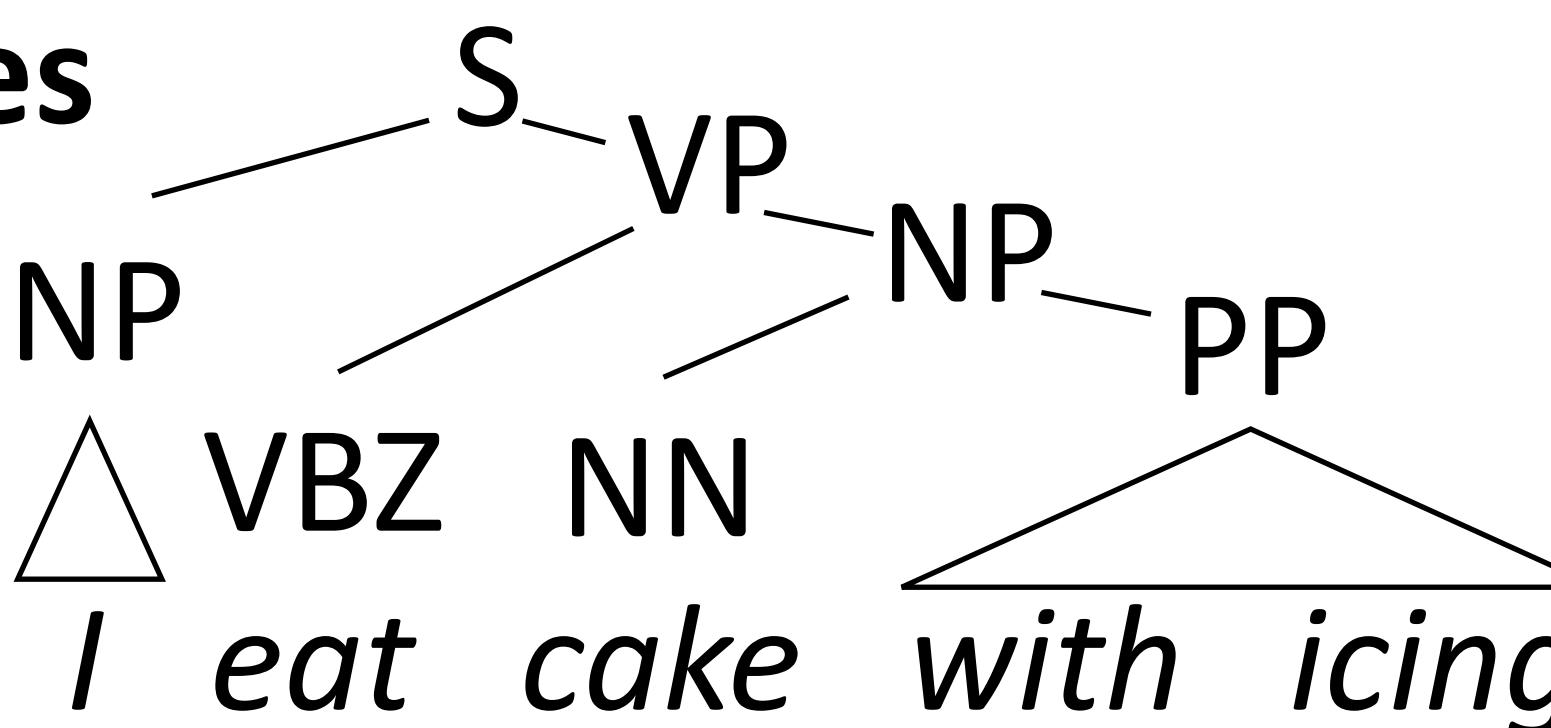
Sequences/tags

PERSON

*Tom Cruise stars in the new Mission Impossible film*

WORK\_OF\_ART

Trees

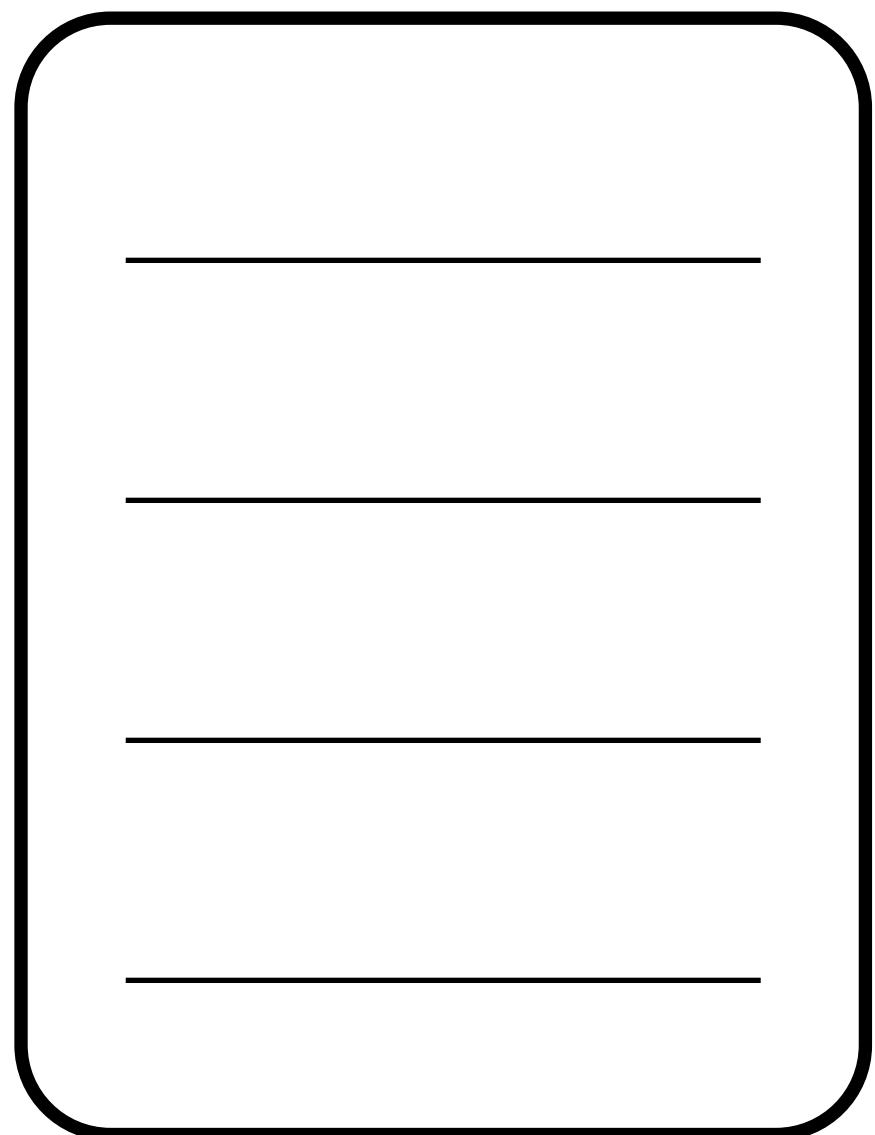


$\lambda x. \text{flight}(x) \wedge \text{dest}(x) = \text{Miami}$   
*flights to Miami*

# How do we use these representations?

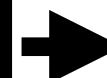
---

**Text**

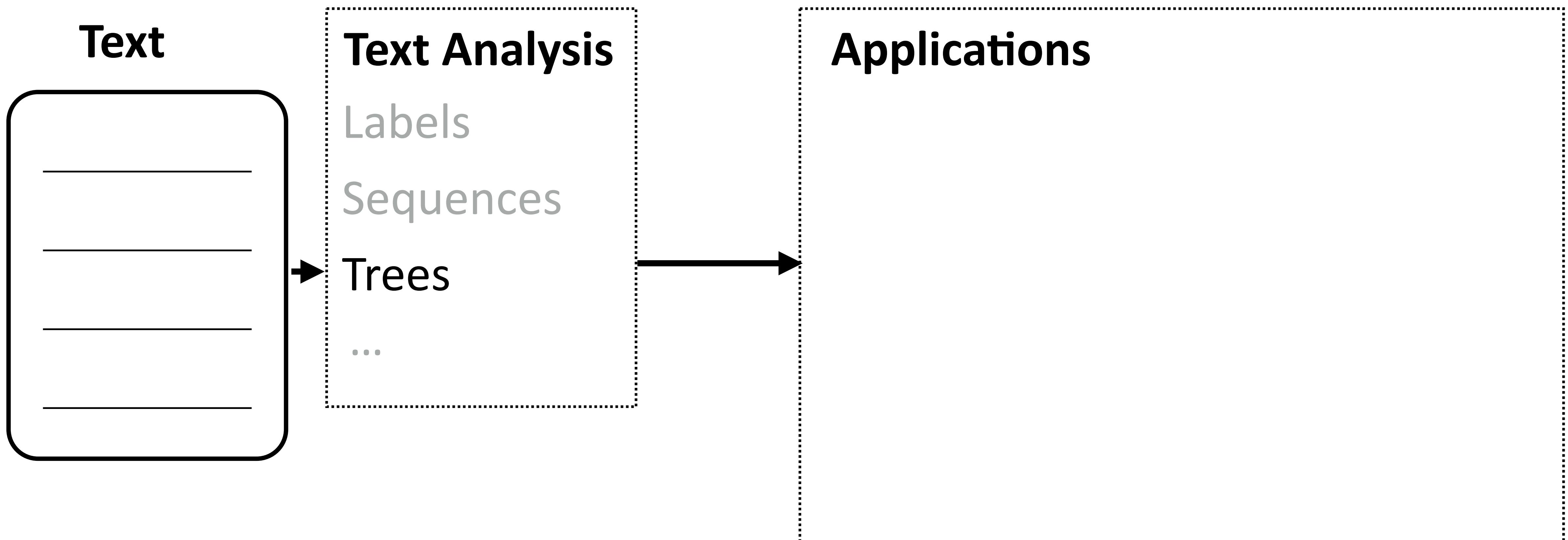


**Text Analysis**

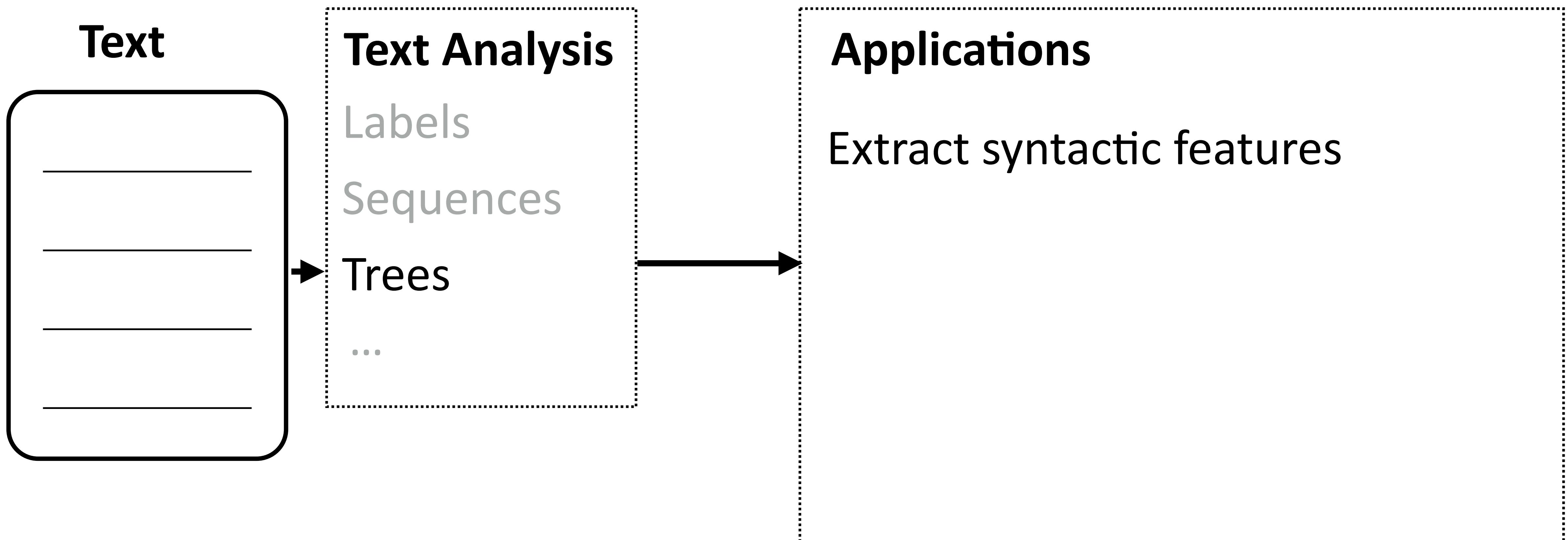
Labels  
Sequences  
**Trees**  
...



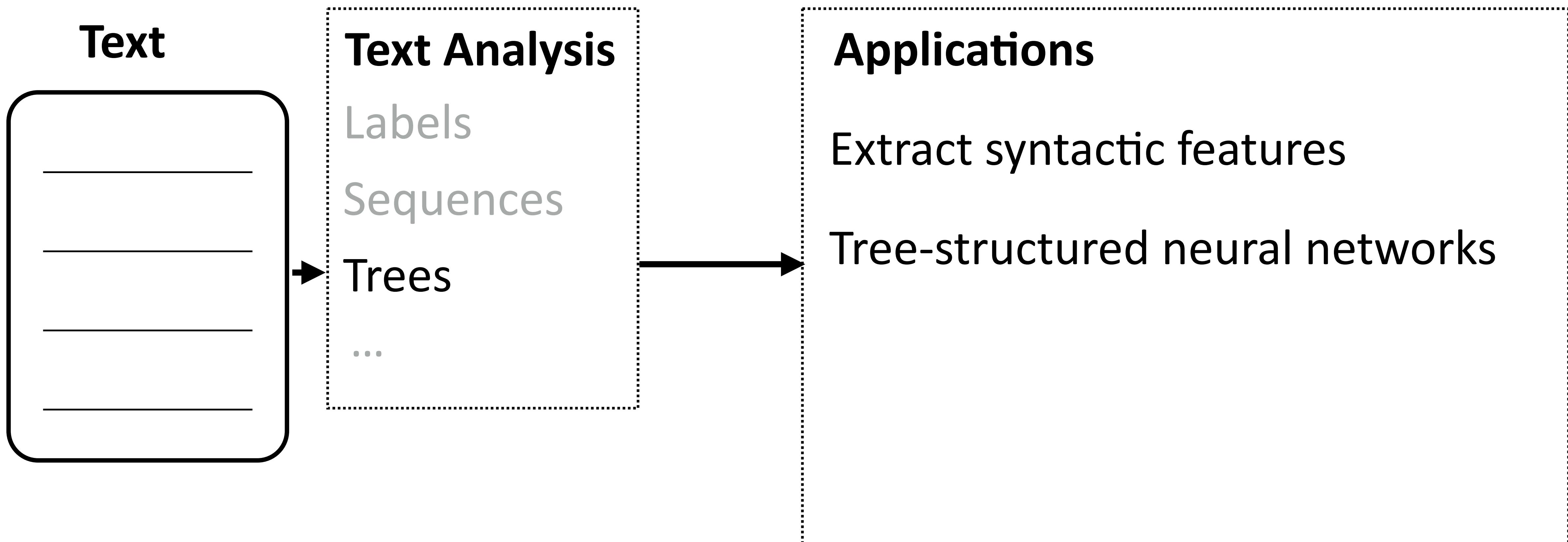
# How do we use these representations?



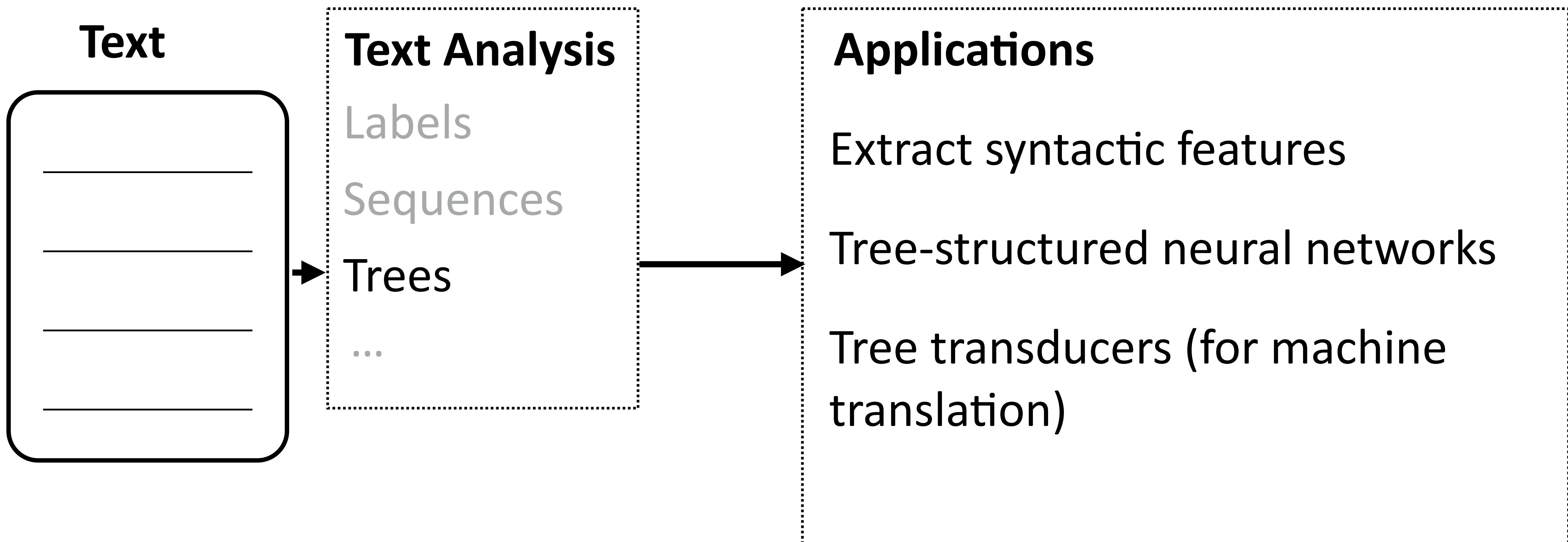
# How do we use these representations?



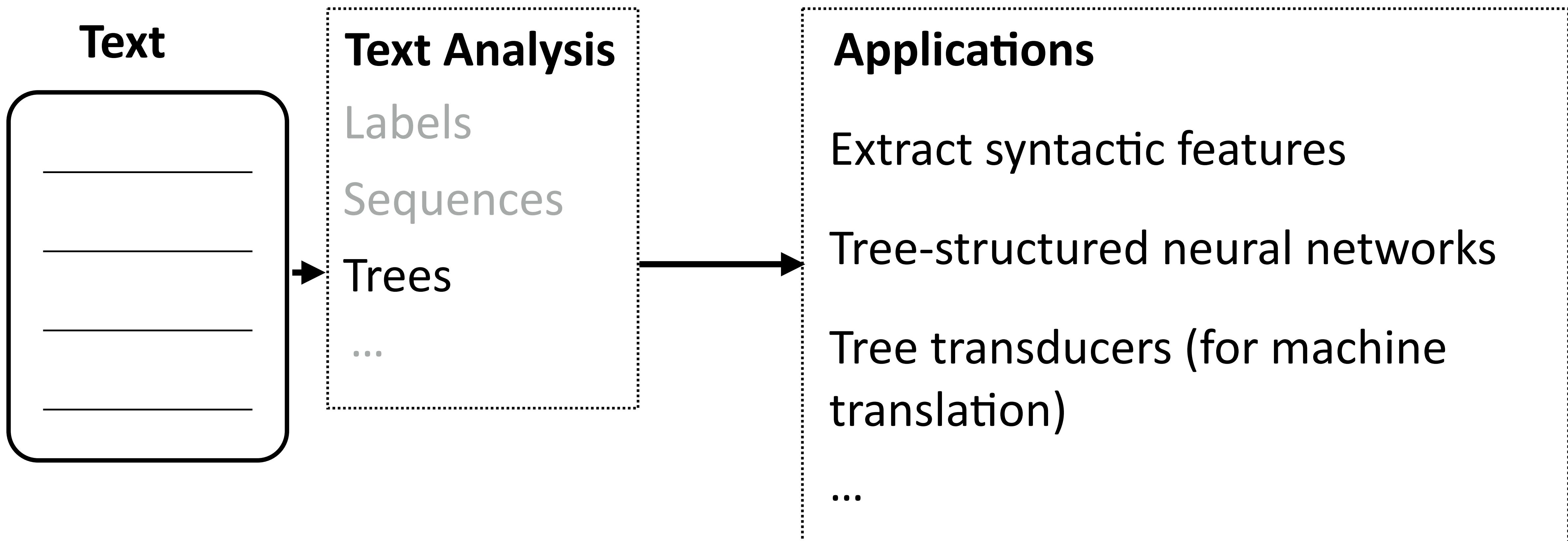
# How do we use these representations?



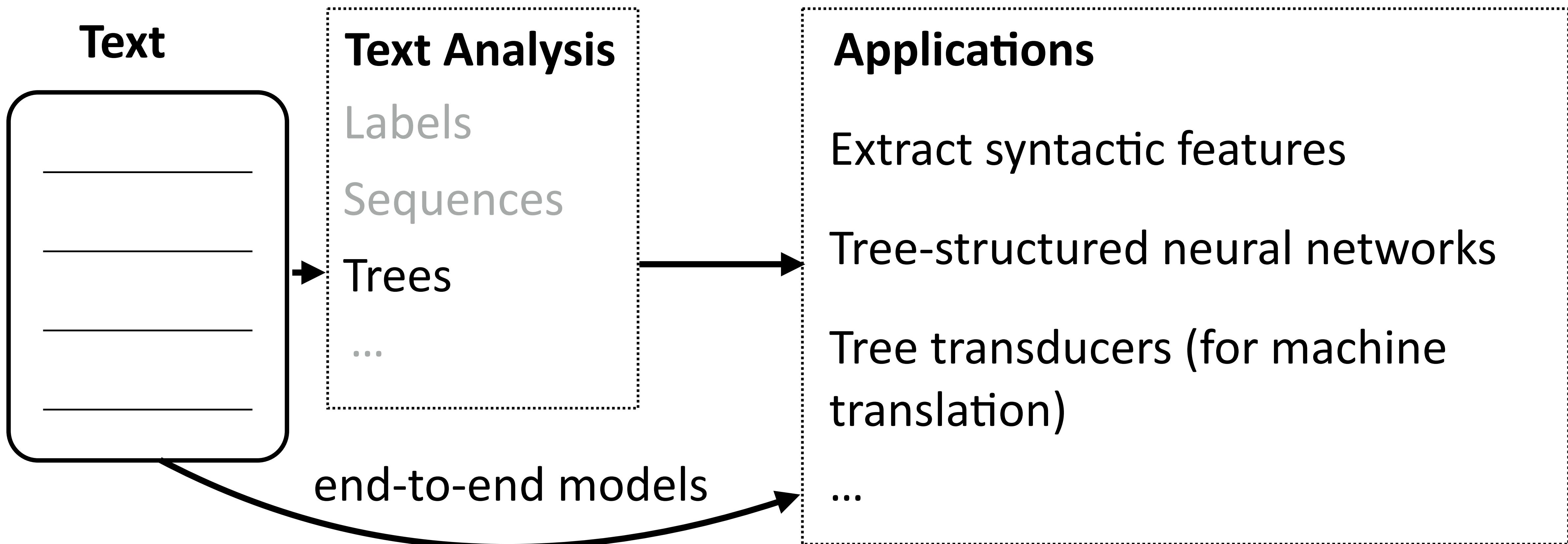
# How do we use these representations?



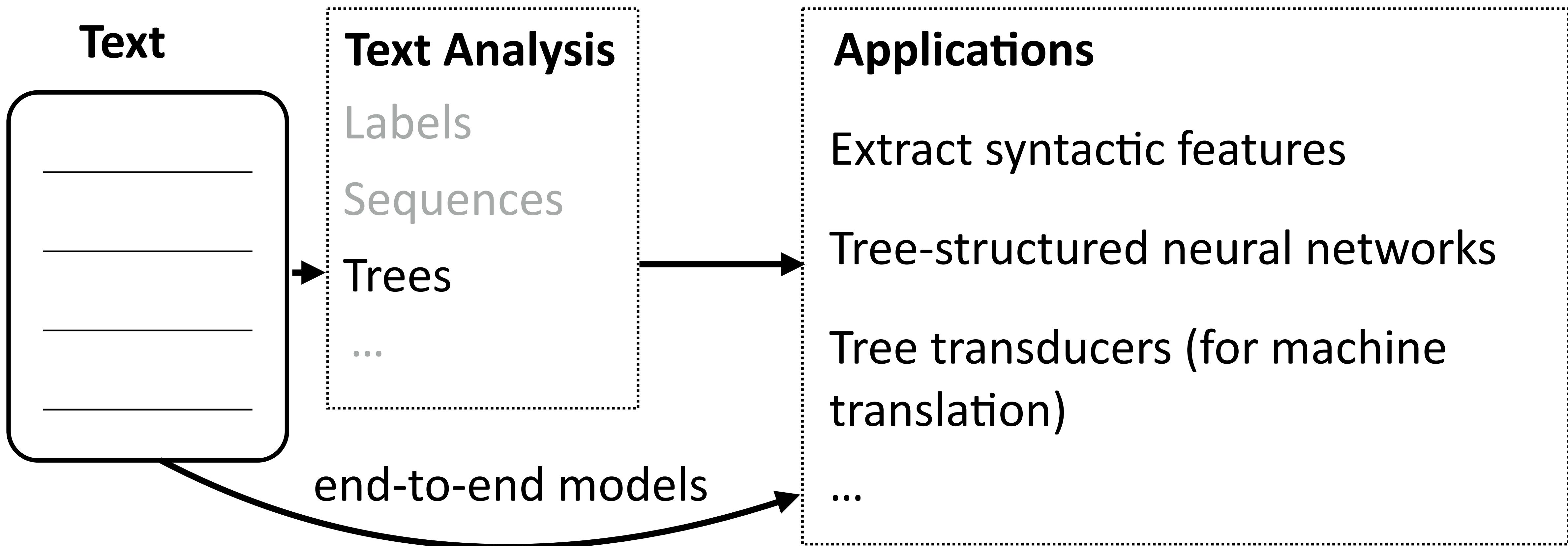
# How do we use these representations?



# How do we use these representations?

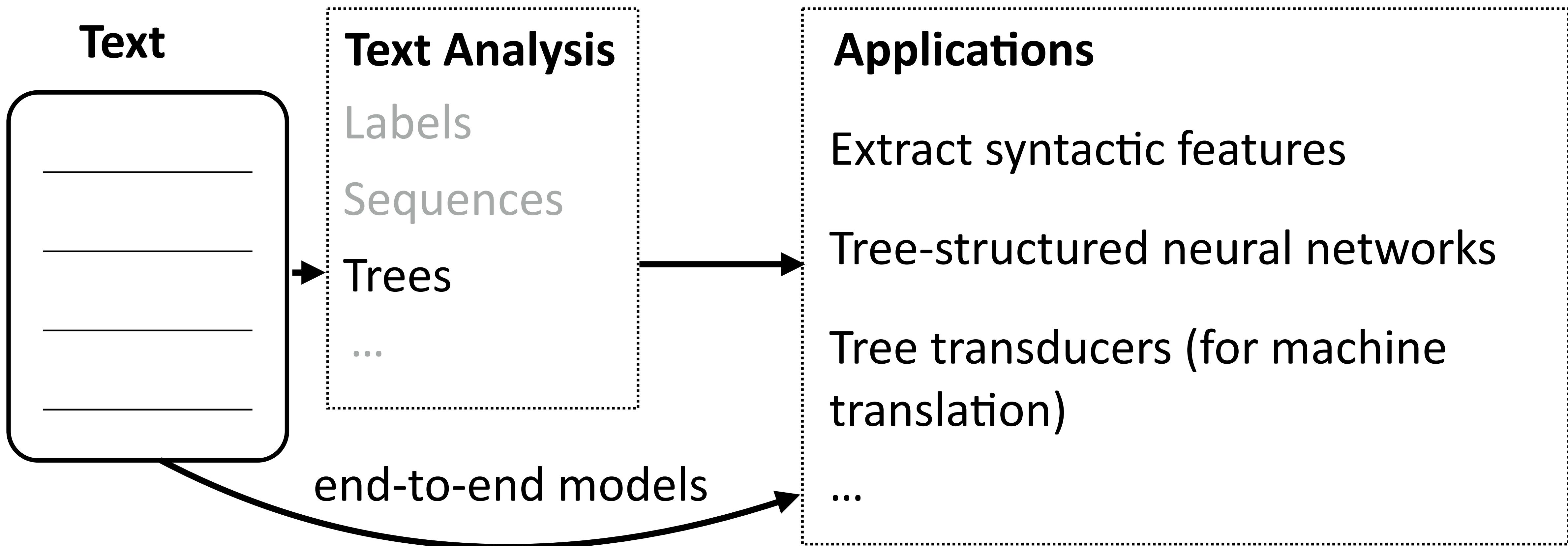


# How do we use these representations?



- ▶ Main question: What representations do we need for language? What do we want to know about it?

# How do we use these representations?



- ▶ Main question: What representations do we need for language? What do we want to know about it?
- ▶ Boils down to: what ambiguities do we need to resolve?

Why is language hard?  
(and how can we handle that?)

# Language is Ambiguous!

---

- ▶ Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

# Language is Ambiguous!

---

- ▶ Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because they \_\_\_\_\_ violence

# Language is Ambiguous!

---

- ▶ Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because they \_\_\_\_\_ violence

# Language is Ambiguous!

---

- ▶ Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because they \_\_\_\_\_ violence

they advocated

# Language is Ambiguous!

---

- ▶ Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because they \_\_\_\_\_ violence

they advocated

# Language is Ambiguous!

---

- ▶ Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because they \_\_\_\_\_ violence

they advocated

they feared

# Language is Ambiguous!

---

- ▶ Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because they \_\_\_\_\_ violence

they advocated

they feared

# Language is Ambiguous!

---

- ▶ Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because they \_\_\_\_\_ violence

they advocated

they feared

- ▶ This is so complicated that it’s an AI challenge problem! (AI-complete)

# Language is Ambiguous!

---

- ▶ Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because they \_\_\_\_\_ violence

they advocated

they feared

- ▶ This is so complicated that it’s an AI challenge problem! (AI-complete)
- ▶ Referential/semantic ambiguity

# Language is Ambiguous!

---

# Language is Ambiguous!

---

- ▶ Headlines

# Language is Ambiguous!

---

- ▶ Headlines
- ▶ Teacher Strikes Idle Kids

# Language is Ambiguous!

---

- ▶ Headlines
  - ▶ Teacher Strikes Idle Kids
  - ▶ Hospitals Sued by 7 Foot Doctors

# Language is Ambiguous!

---

- ▶ Headlines
  - ▶ Teacher Strikes Idle Kids
  - ▶ Hospitals Sued by 7 Foot Doctors
  - ▶ Ban on Nude Dancing on Governor's Desk

# Language is Ambiguous!

---

- ▶ Headlines
  - ▶ Teacher Strikes Idle Kids
  - ▶ Hospitals Sued by 7 Foot Doctors
  - ▶ Ban on Nude Dancing on Governor's Desk
  - ▶ Iraqi Head Seeks Arms

# Language is Ambiguous!

---

- ▶ Headlines
  - ▶ Teacher Strikes Idle Kids
  - ▶ Hospitals Sued by 7 Foot Doctors
  - ▶ Ban on Nude Dancing on Governor's Desk
  - ▶ Iraqi Head Seeks Arms
  - ▶ Stolen Painting Found by Tree

# Language is Ambiguous!

---

- ▶ Headlines
  - ▶ Teacher Strikes Idle Kids
  - ▶ Hospitals Sued by 7 Foot Doctors
  - ▶ Ban on Nude Dancing on Governor's Desk
  - ▶ Iraqi Head Seeks Arms
  - ▶ Stolen Painting Found by Tree
  - ▶ Kids Make Nutritious Snacks

# Language is Ambiguous!

---

- ▶ Headlines
  - ▶ Teacher Strikes Idle Kids
  - ▶ Hospitals Sued by 7 Foot Doctors
  - ▶ Ban on Nude Dancing on Governor's Desk
  - ▶ Iraqi Head Seeks Arms
  - ▶ Stolen Painting Found by Tree
  - ▶ Kids Make Nutritious Snacks
  - ▶ Local HS Dropouts Cut in Half

# Language is Ambiguous!

---

- ▶ Headlines
  - ▶ Teacher Strikes Idle Kids
  - ▶ Hospitals Sued by 7 Foot Doctors
  - ▶ Ban on Nude Dancing on Governor's Desk
  - ▶ Iraqi Head Seeks Arms
  - ▶ Stolen Painting Found by Tree
  - ▶ Kids Make Nutritious Snacks
  - ▶ Local HS Dropouts Cut in Half
- ▶ Syntactic/semantic ambiguity: parsing needed to resolve these, but need context to figure out which parse is correct

# Language is Really Ambiguous!

---

- ▶ There aren't just one or two possibilities which are resolved pragmatically

# Language is Really Ambiguous!

---

- ▶ There aren't just one or two possibilities which are resolved pragmatically

*il fait vraiment beau* —————→

# Language is Really Ambiguous!

---

- ▶ There aren't just one or two possibilities which are resolved pragmatically

It is really nice out

*il fait vraiment beau* →

# Language is Really Ambiguous!

---

- ▶ There aren't just one or two possibilities which are resolved pragmatically

*il fait vraiment beau*  It is really nice out  
It's really nice

# Language is Really Ambiguous!

---

- ▶ There aren't just one or two possibilities which are resolved pragmatically

*il fait vraiment beau*



It is really nice out

It's really nice

The weather is beautiful

# Language is Really Ambiguous!

---

- ▶ There aren't just one or two possibilities which are resolved pragmatically

*il fait vraiment beau*



It is really nice out

It's really nice

The weather is beautiful

It is really beautiful outside

# Language is Really Ambiguous!

---

- ▶ There aren't just one or two possibilities which are resolved pragmatically

*il fait vraiment beau*



It is really nice out

It's really nice

The weather is beautiful

It is really beautiful outside

He makes truly beautiful

# Language is Really Ambiguous!

---

- ▶ There aren't just one or two possibilities which are resolved pragmatically

*il fait vraiment beau*



It is really nice out

It's really nice

The weather is beautiful

It is really beautiful outside

He makes truly beautiful

He makes truly boyfriend

# Language is Really Ambiguous!

---

- ▶ There aren't just one or two possibilities which are resolved pragmatically

*il fait vraiment beau*



It is really nice out

It's really nice

The weather is beautiful

It is really beautiful outside

He makes truly beautiful

He makes truly boyfriend

It fact actually handsome

# Language is Really Ambiguous!

---

- ▶ There aren't just one or two possibilities which are resolved pragmatically

*il fait vraiment beau*



It is really nice out

It's really nice

The weather is beautiful

It is really beautiful outside

He makes truly beautiful

He makes truly boyfriend

It fact actually handsome

- ▶ Combinatorially many possibilities, many you won't even register as ambiguities, but systems still have to resolve them

# What do we need to understand language?

## ► Lots of data!

SOURCE	Cela constituerait une solution transitoire qui permettrait de conduire à terme à une charte à valeur contraignante.
HUMAN	That would be an interim solution which would make it possible to work towards a binding charter in the long term .
1x DATA	[this] [constituerait] [assistance] [transitoire] [who] [permettrait] [licences] [to] [terme] [to] [a] [charter] [to] [value] [contraignante] [.]
10x DATA	[it] [would] [a solution] [transitional] [which] [would] [of] [lead] [to] [term] [to a] [charter] [to] [value] [binding] [.]
100x DATA	[this] [would be] [a transitional solution] [which would] [lead to] [a charter] [legally binding] [.]
1000x DATA	[that would be] [a transitional solution] [which would] [eventually lead to] [a binding charter] [.]

# What do we need to understand language?

---

- ▶ World knowledge: have access to information beyond the training data

# What do we need to understand language?

---

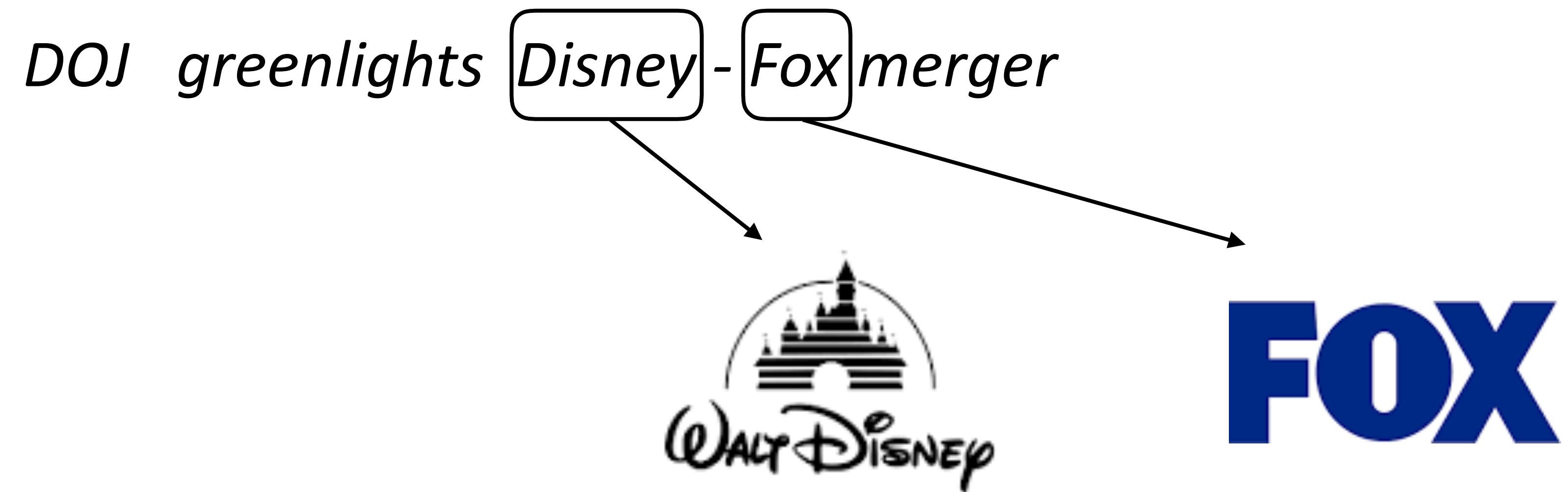
- ▶ World knowledge: have access to information beyond the training data

*DOJ greenlights Disney - Fox merger*

# What do we need to understand language?

---

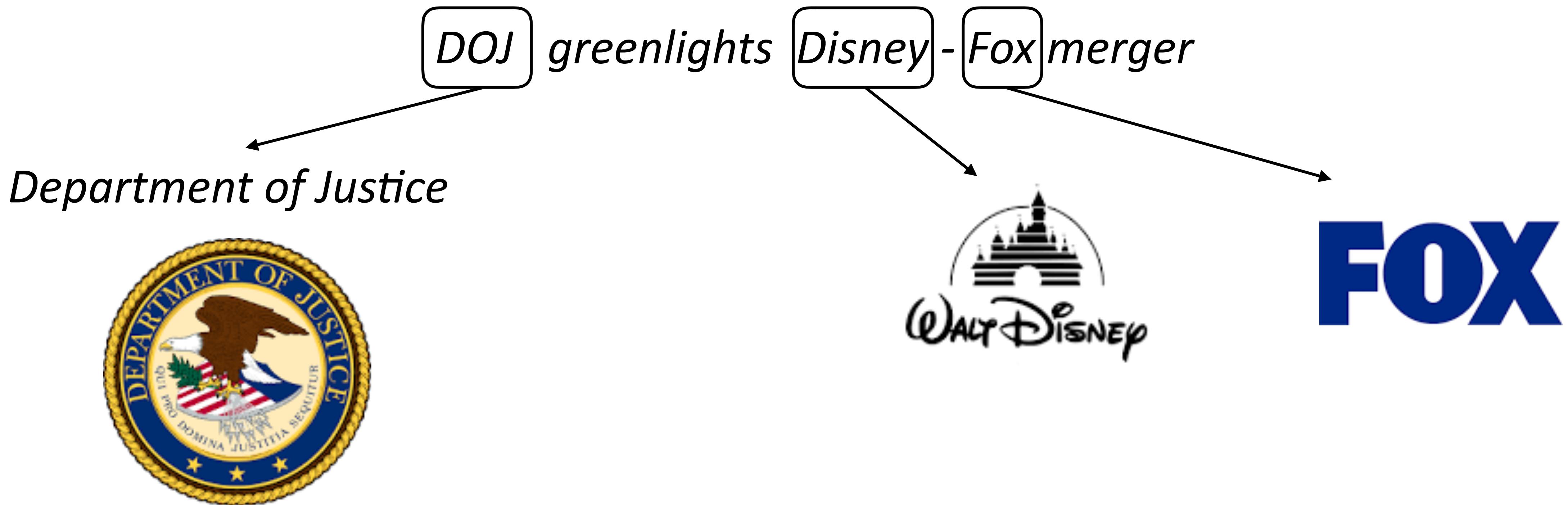
- ▶ World knowledge: have access to information beyond the training data



# What do we need to understand language?

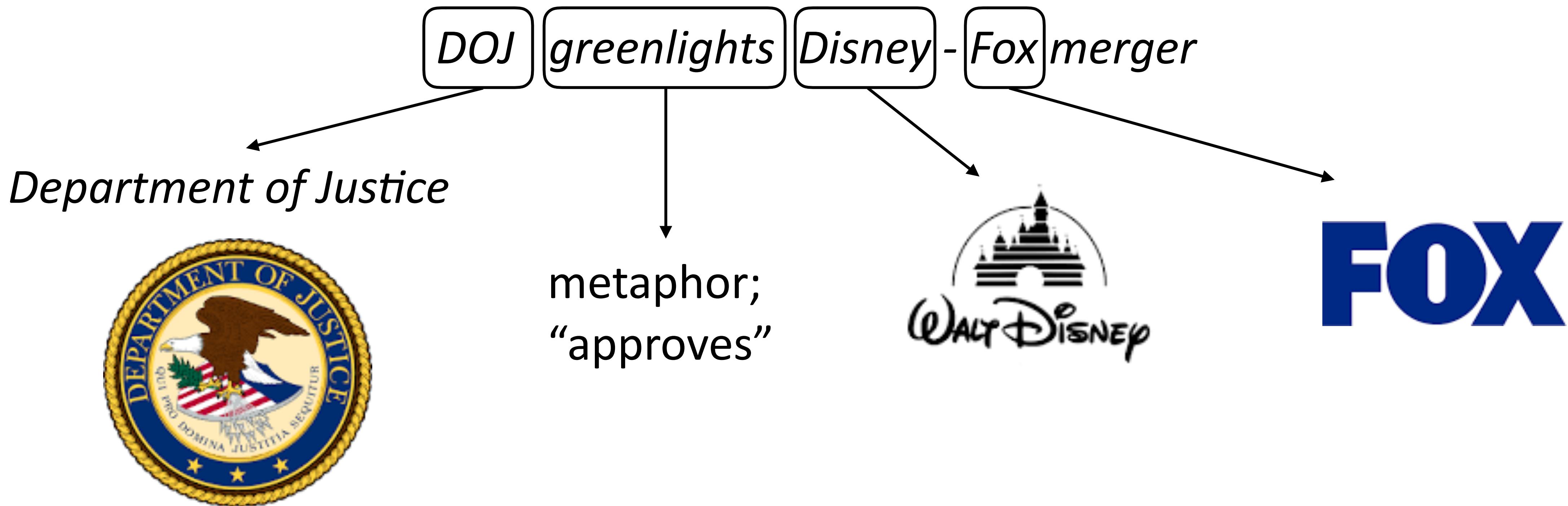
---

- ▶ World knowledge: have access to information beyond the training data



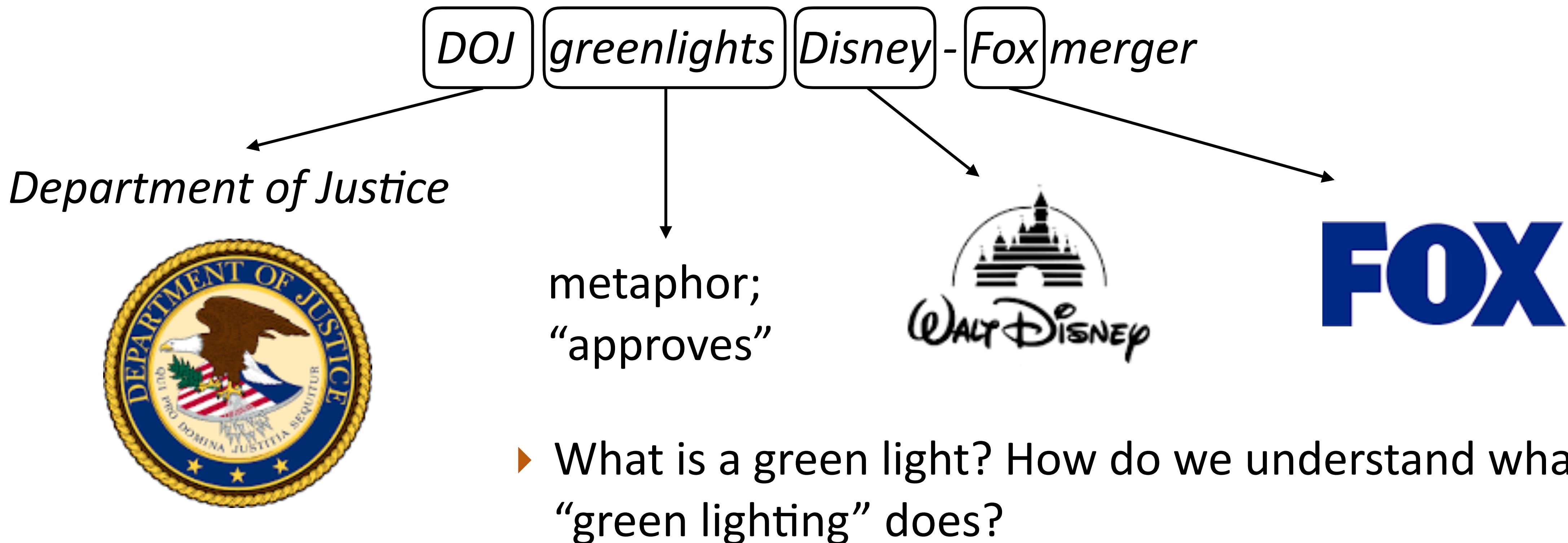
# What do we need to understand language?

- ▶ World knowledge: have access to information beyond the training data



# What do we need to understand language?

- ▶ World knowledge: have access to information beyond the training data



# What do we need to understand language?

---

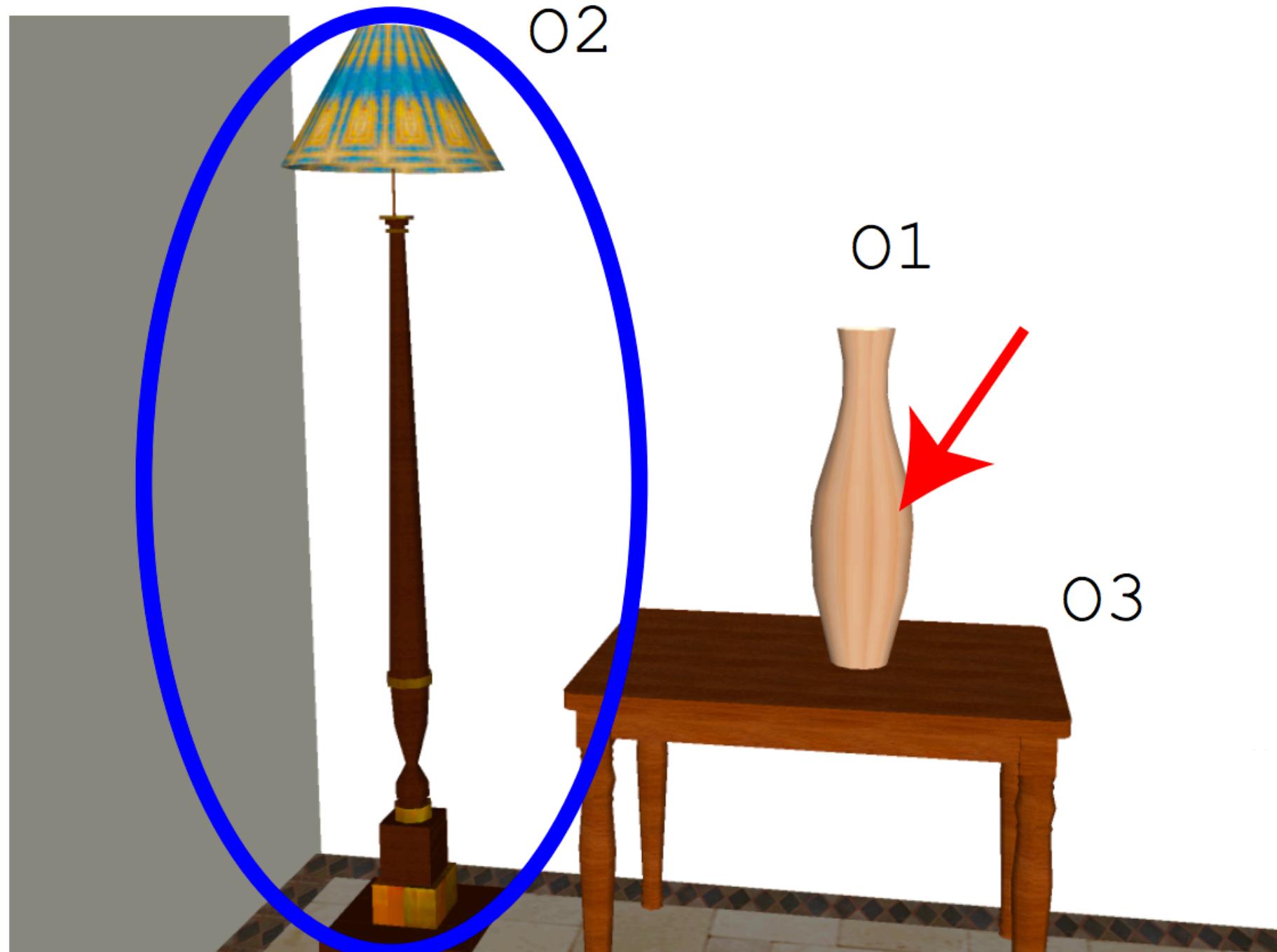
- ▶ Grounding: learn what fundamental concepts actually mean in a data-driven way

# What do we need to understand language?

---

- ▶ Grounding: learn what fundamental concepts actually mean in a data-driven way

**Question:** What object is **right of** **02** ?

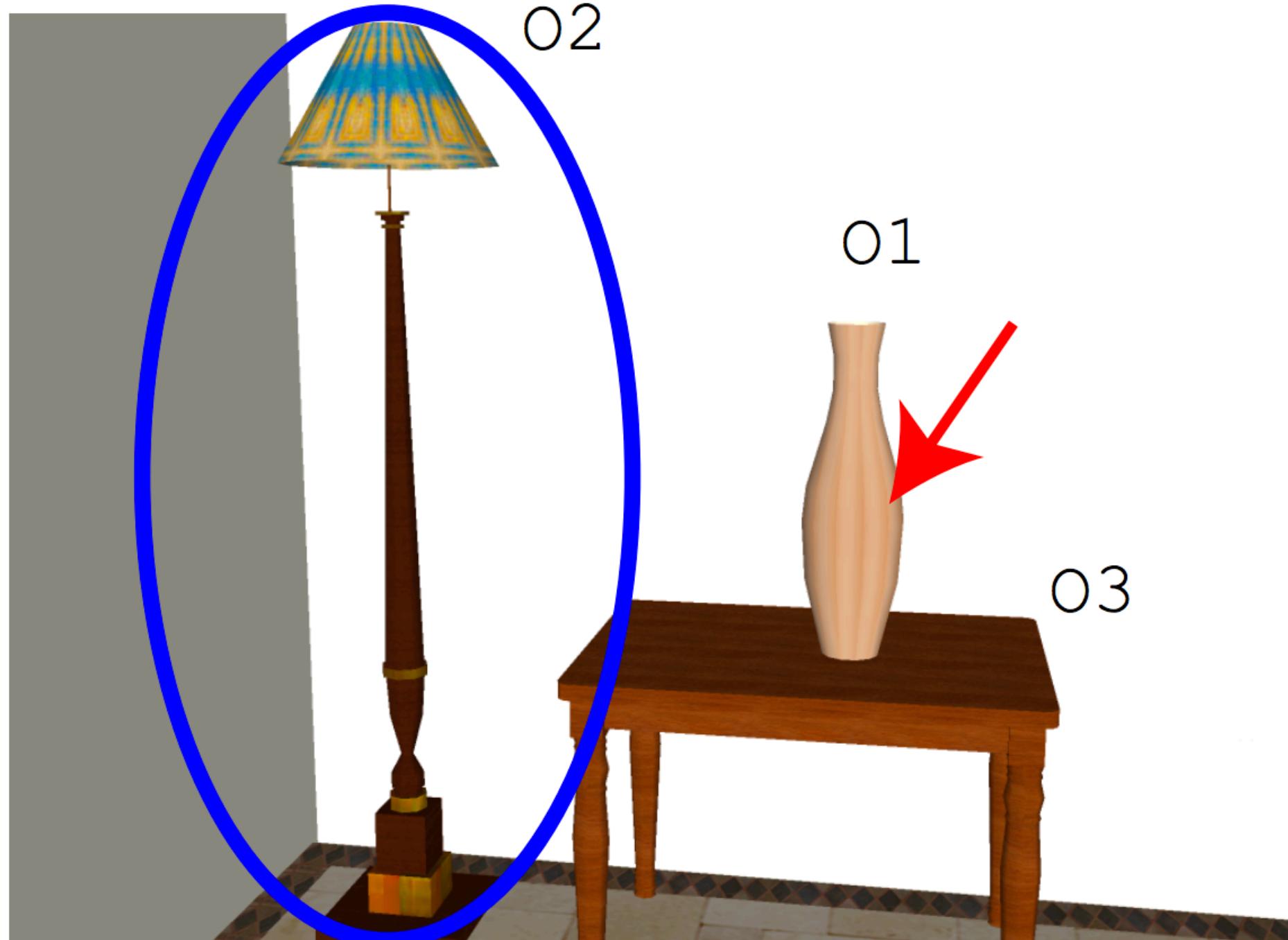


Golland et al. (2010)

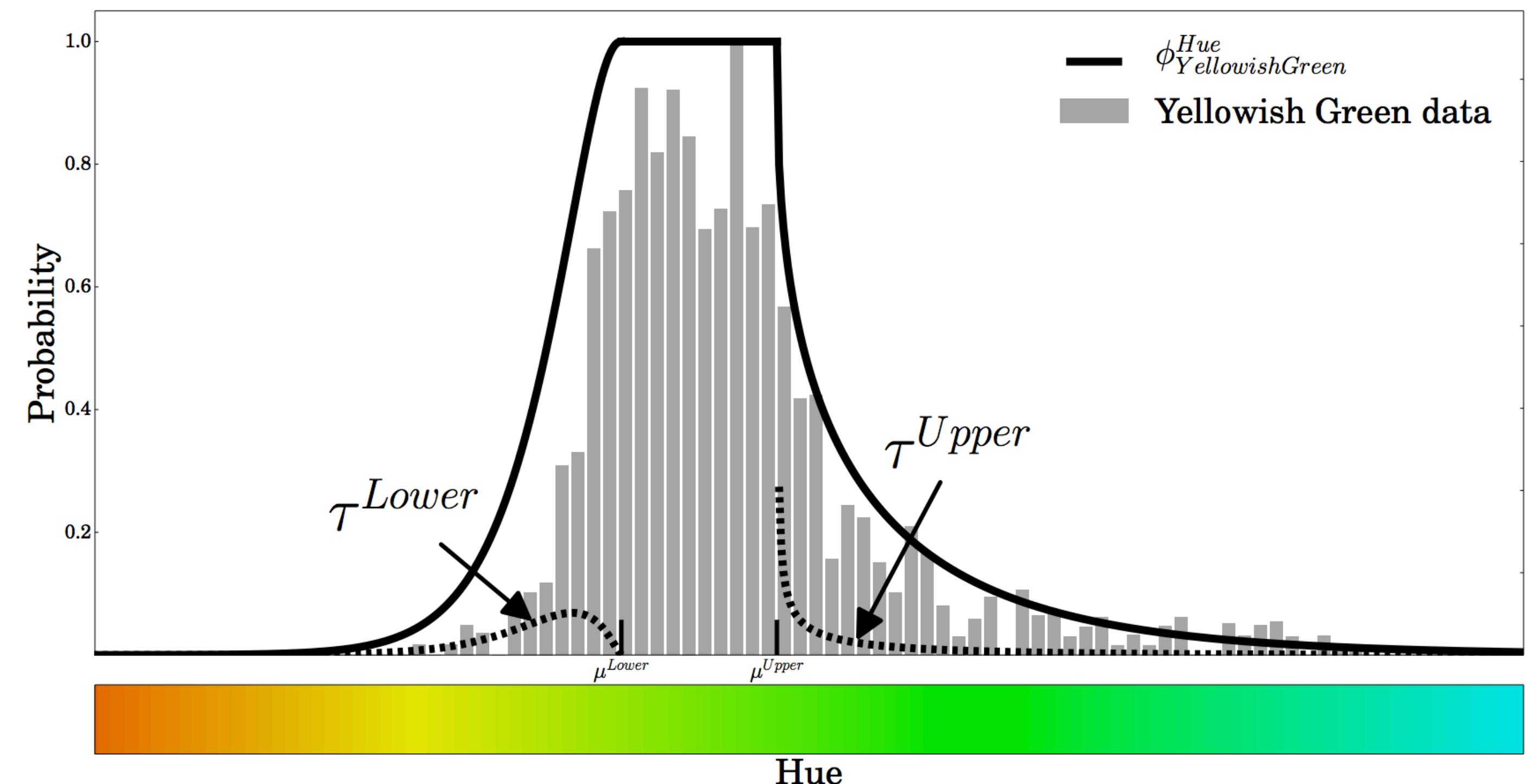
# What do we need to understand language?

- Grounding: learn what fundamental concepts actually mean in a data-driven way

**Question:** What object is right of **O2** ?



Golland et al. (2010)



McMahan and Stone (2015)

# What do we need to understand language?

---

- ▶ Linguistic structure

# What do we need to understand language?

---

- ▶ Linguistic structure
- ▶ ...but computers probably won't understand language the same way humans do

# What do we need to understand language?

---

- ▶ Linguistic structure
- ▶ ...but computers probably won't understand language the same way humans do
- ▶ However, linguistics tells us what phenomena we need to be able to deal with and gives us hints about how language works

# What do we need to understand language?

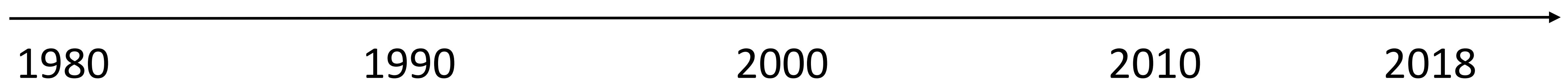
---

- ▶ Linguistic structure
- ▶ ...but computers probably won't understand language the same way humans do
- ▶ However, linguistics tells us what phenomena we need to be able to deal with and gives us hints about how language works
  - a. John has been having a lot of trouble arranging his vacation.
  - b. He cannot find anyone to take over his responsibilities. (he = John)  
 $C_b = \text{John}; C_f = \{\text{John}\}$
  - c. He called up Mike yesterday to work out a plan. (he = John)  
 $C_b = \text{John}; C_f = \{\text{John}, \text{Mike}\}$  (CONTINUE)
  - d. Mike has annoyed him a lot recently.  
 $C_b = \text{John}; C_f = \{\text{Mike}, \text{John}\}$  (RETAIN)
  - e. He called John at 5 AM on Friday last week. (he = Mike)  
 $C_b = \text{Mike}; C_f = \{\text{Mike}, \text{John}\}$  (SHIFT)

What techniques do we use?  
(to combine data, knowledge, linguistics, etc.)

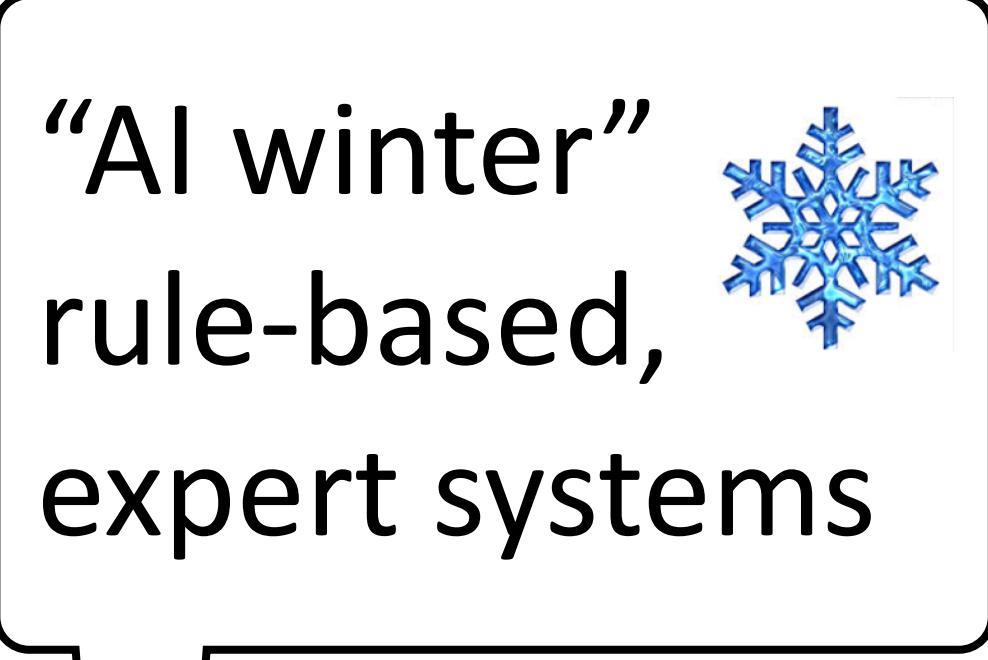
# A brief history of (modern) NLP

---



# A brief history of (modern) NLP

---



1980

1990

2000

2010

2018

# A brief history of (modern) NLP

---

“AI winter”  
rule-based,  
expert systems

earliest stat MT  
work at IBM



1980

1990

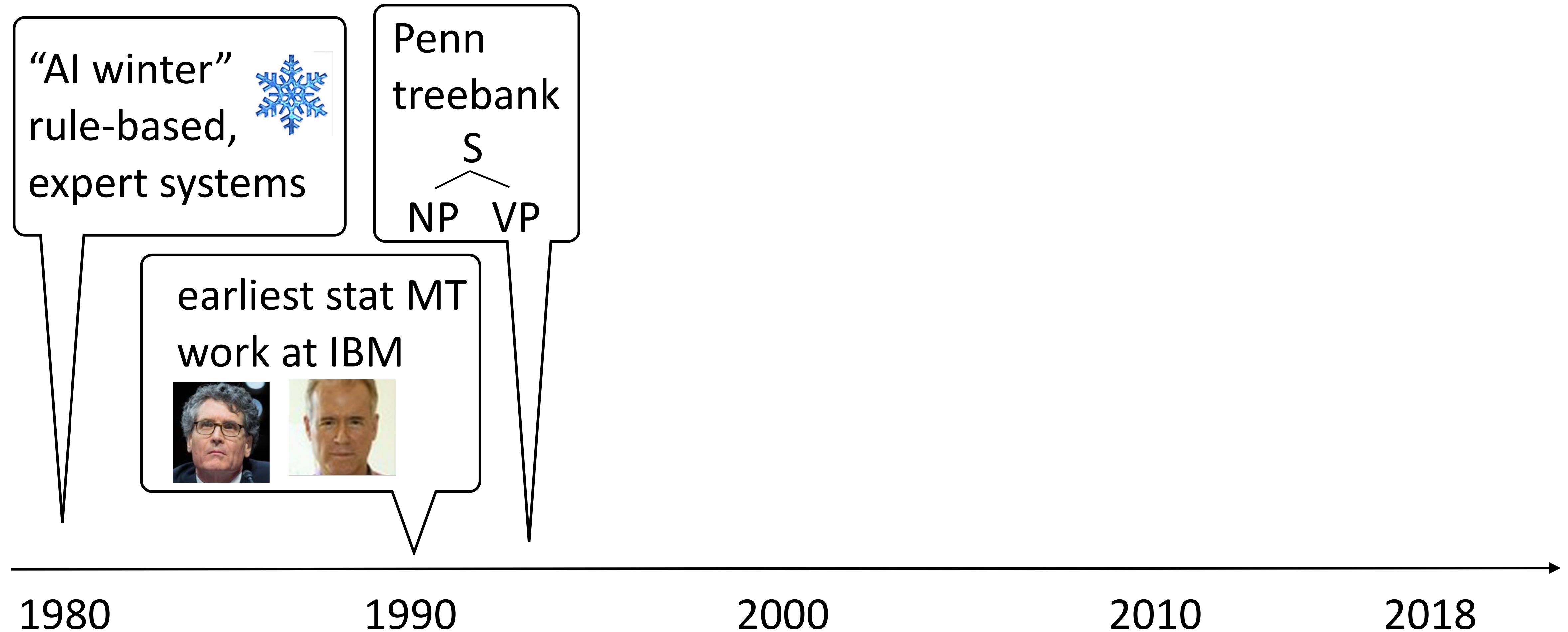
2000

2010

2018

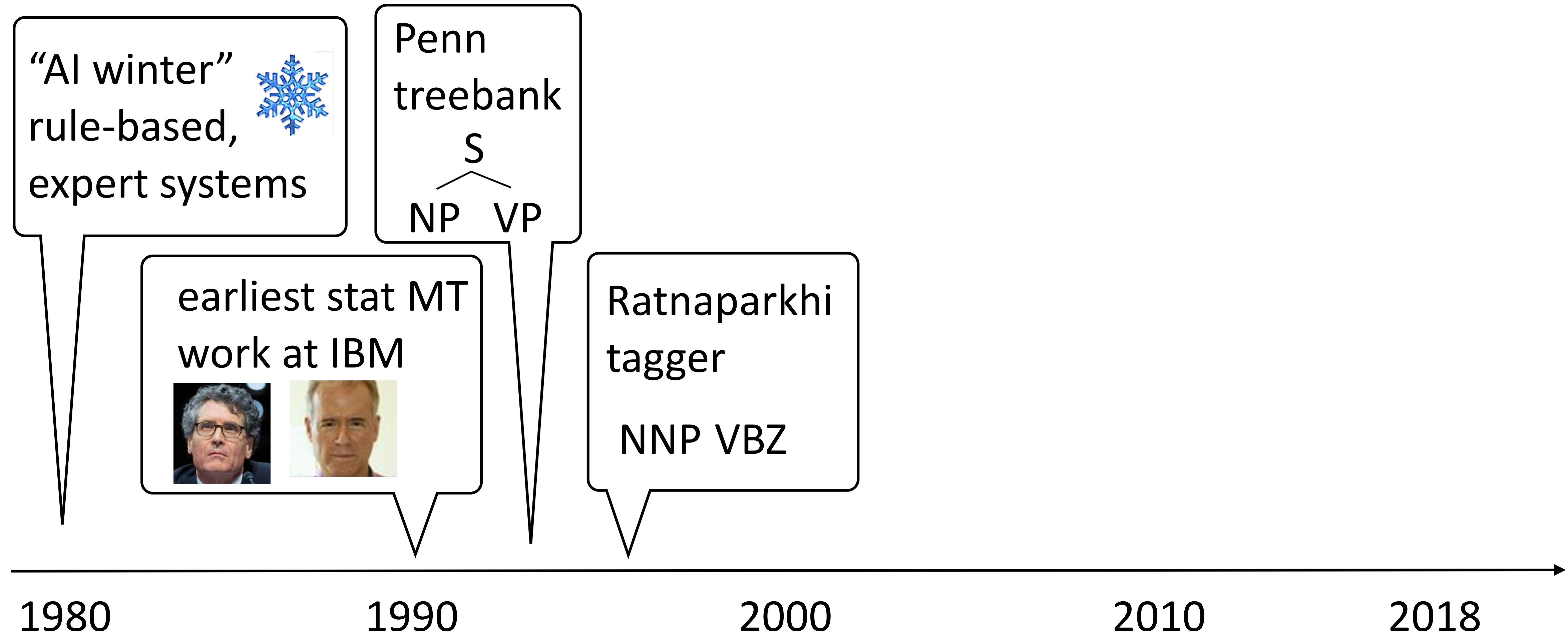
# A brief history of (modern) NLP

---



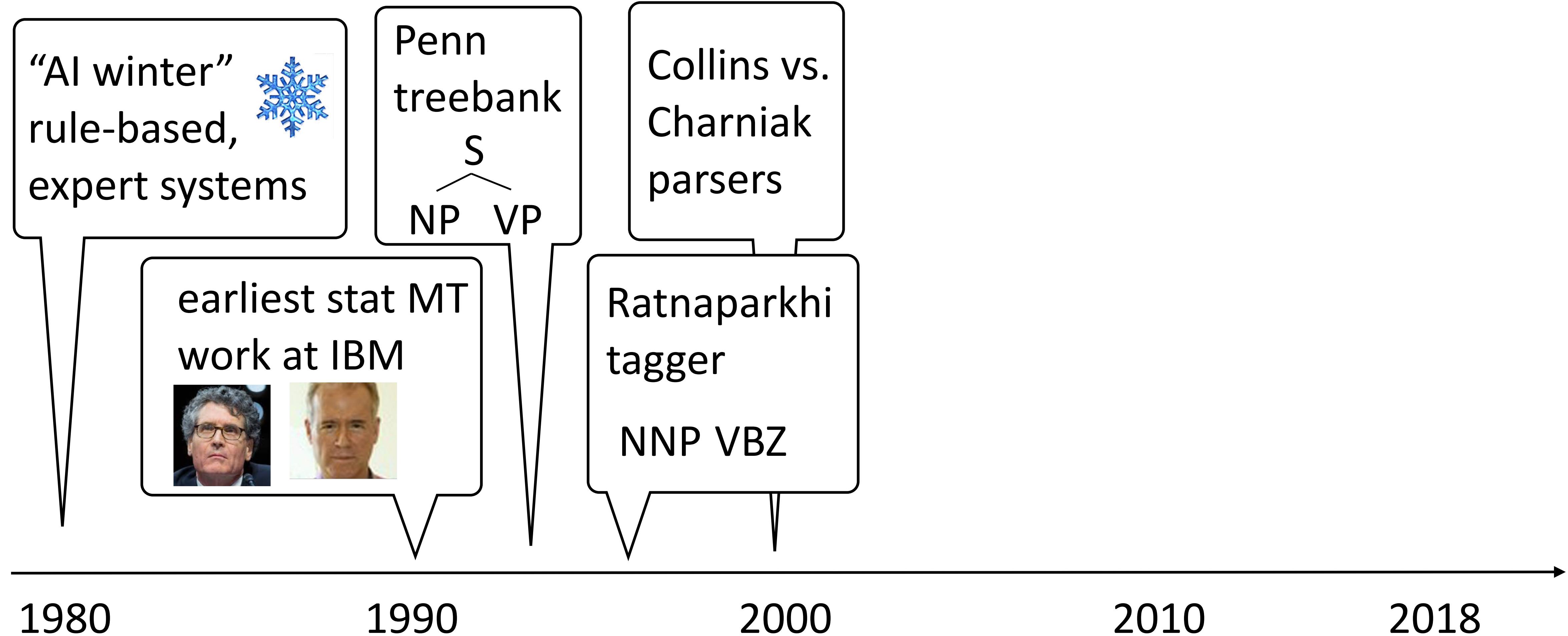
# A brief history of (modern) NLP

---



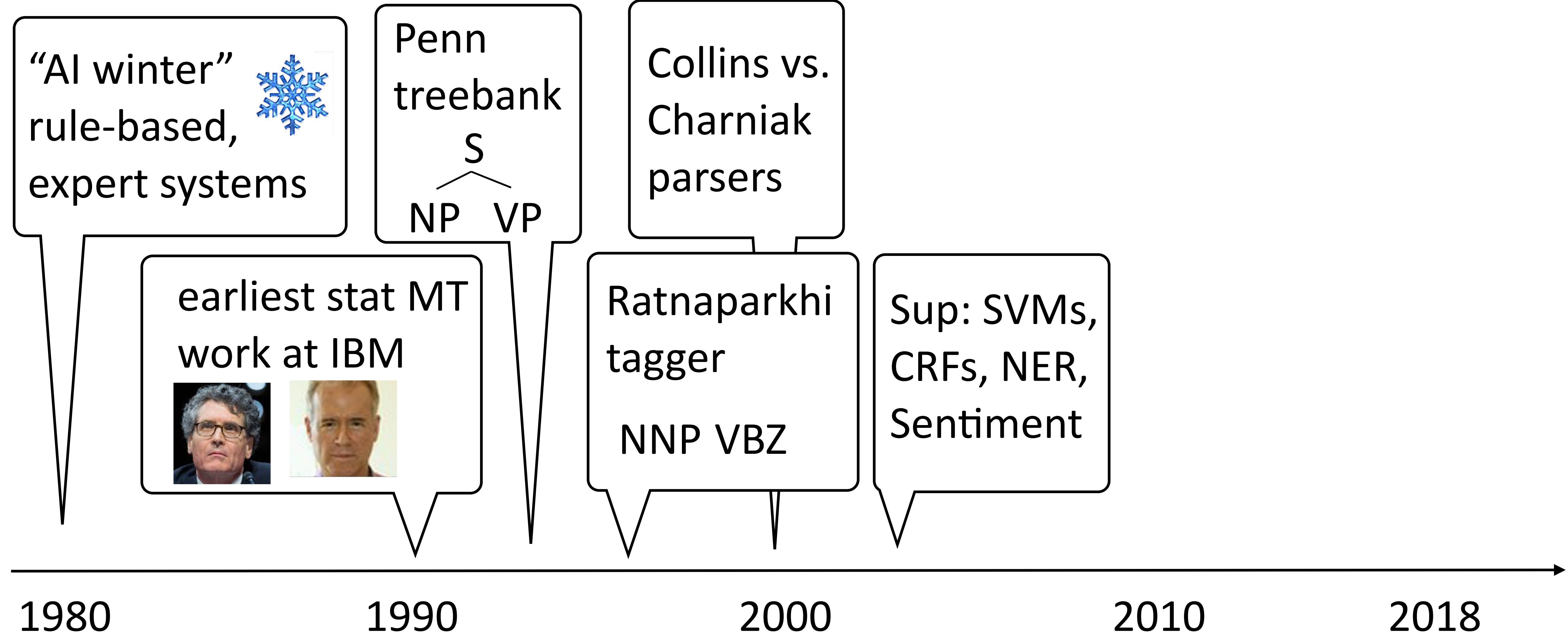
# A brief history of (modern) NLP

---

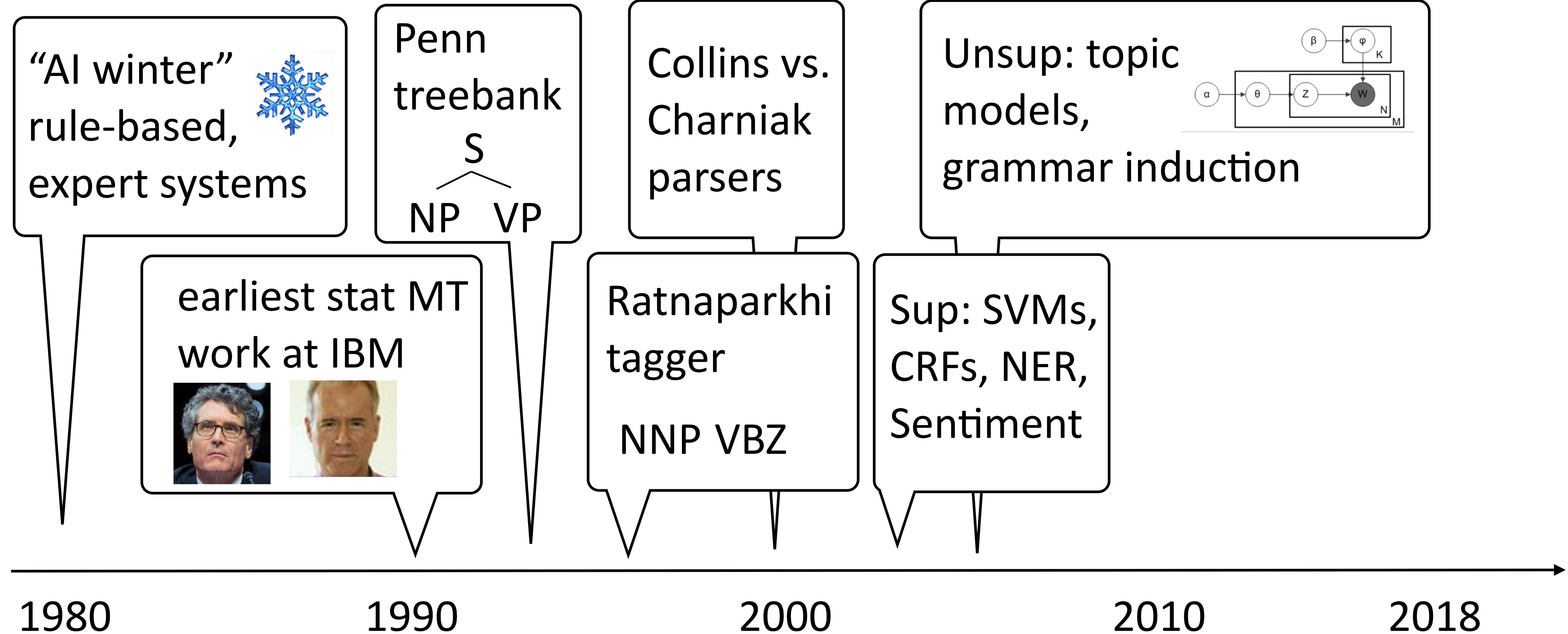


# A brief history of (modern) NLP

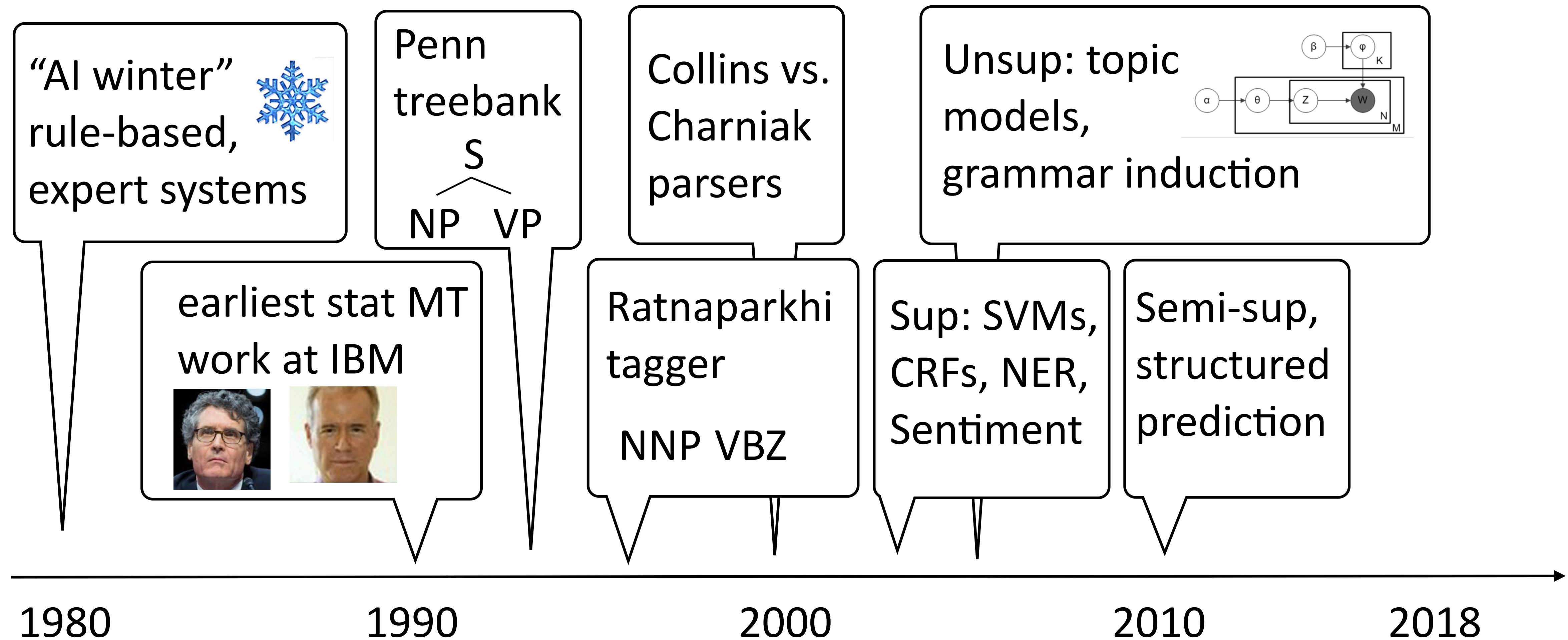
---



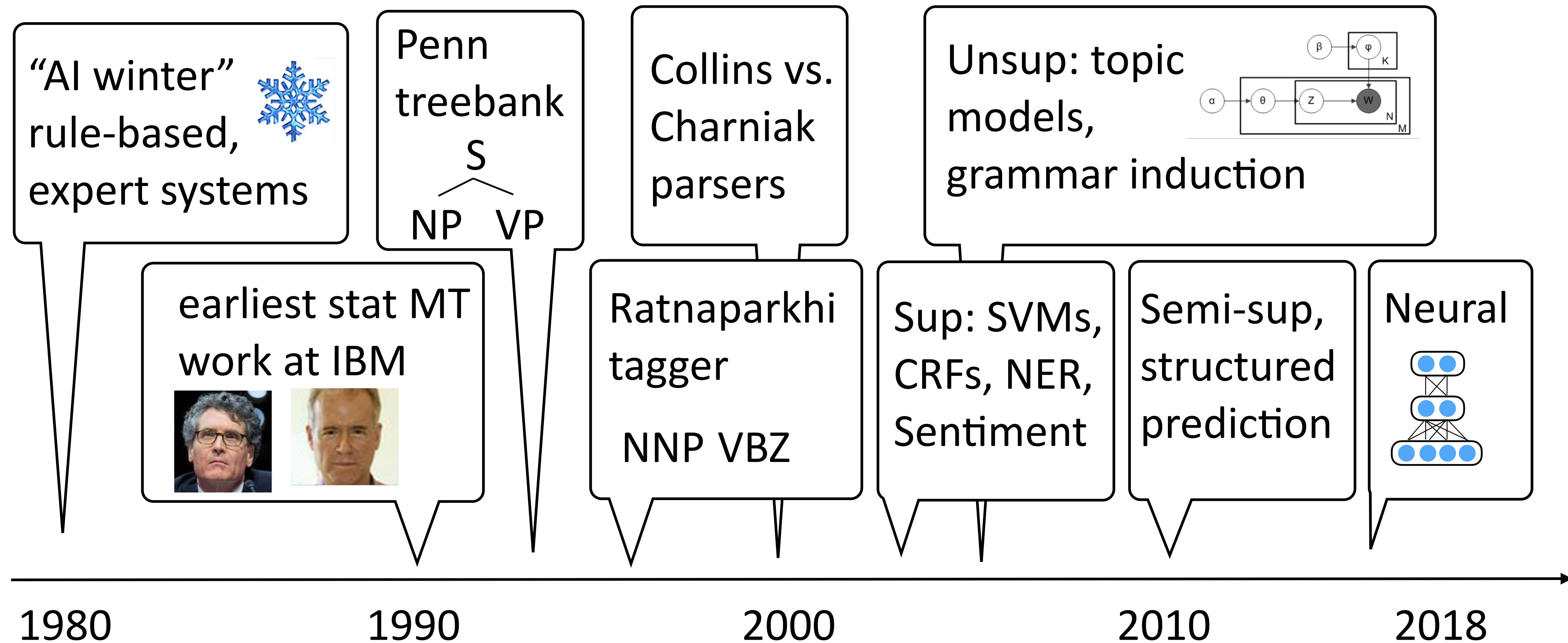
# A brief history of (modern) NLP



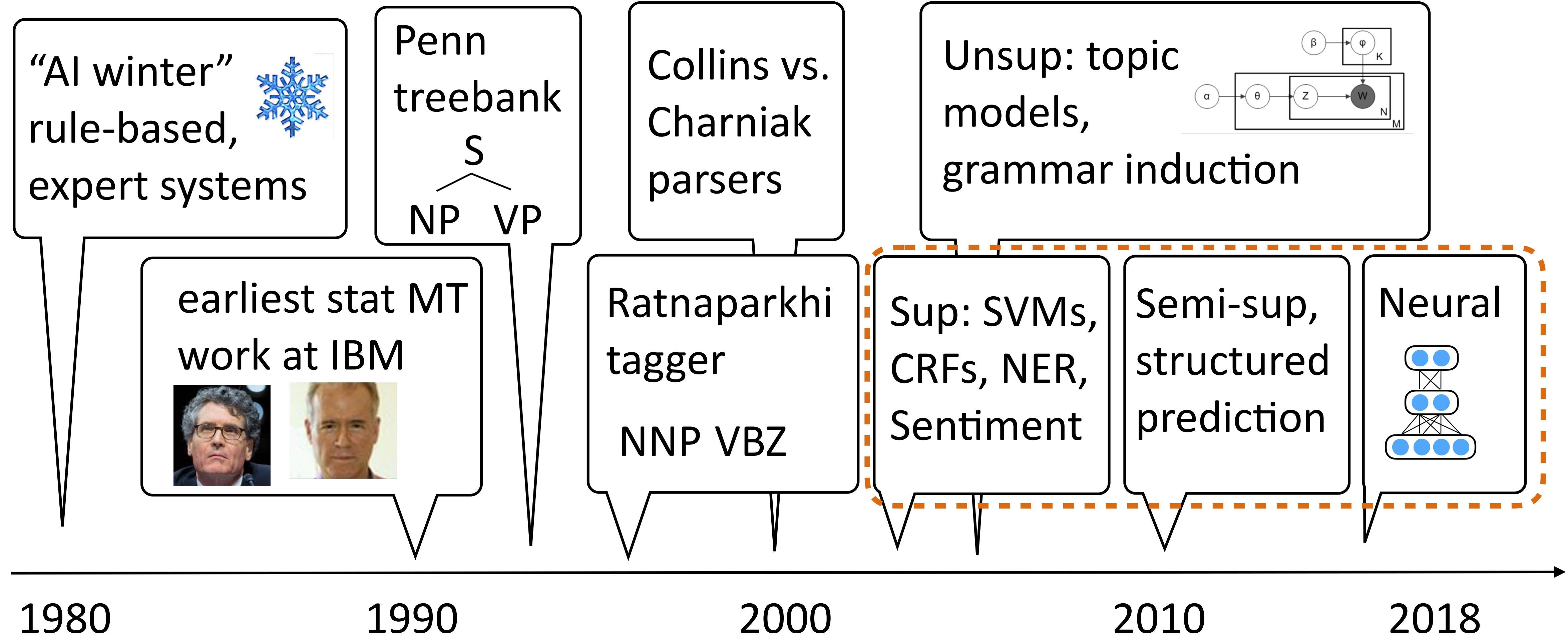
# A brief history of (modern) NLP



# A brief history of (modern) NLP



# A brief history of (modern) NLP



# Structured Prediction

---

“Learning a Part-of-Speech Tagger from Two Hours of Annotation”  
Garrette and Baldridge (2013)

# Structured Prediction

---

- ▶ All of these techniques are data-driven! Some data is naturally occurring, but may need to label

“Learning a Part-of-Speech Tagger from Two Hours of Annotation”  
Garrette and Baldridge (2013)

# Structured Prediction

---

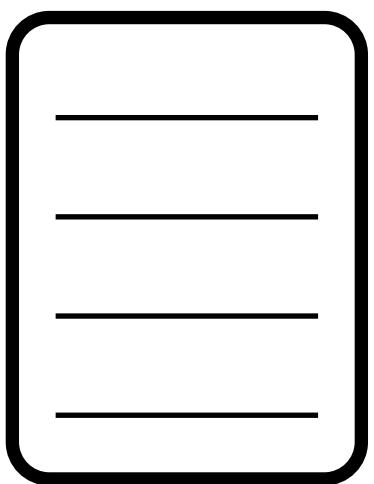
- ▶ All of these techniques are data-driven! Some data is naturally occurring, but may need to label
- ▶ Supervised techniques work well on very little data

“Learning a Part-of-Speech Tagger from Two Hours of Annotation”  
Garrette and Baldridge (2013)

# Structured Prediction

---

- ▶ All of these techniques are data-driven! Some data is naturally occurring, but may need to label
- ▶ Supervised techniques work well on very little data

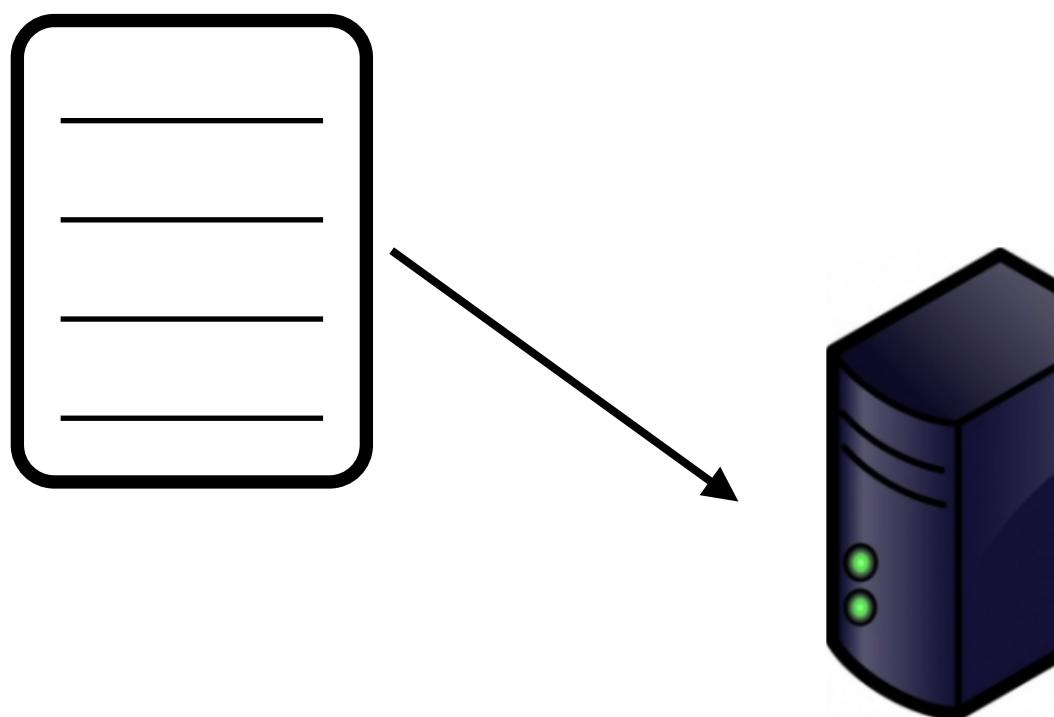


“Learning a Part-of-Speech Tagger from Two Hours of Annotation”  
Garrette and Baldridge (2013)

# Structured Prediction

---

- ▶ All of these techniques are data-driven! Some data is naturally occurring, but may need to label
- ▶ Supervised techniques work well on very little data



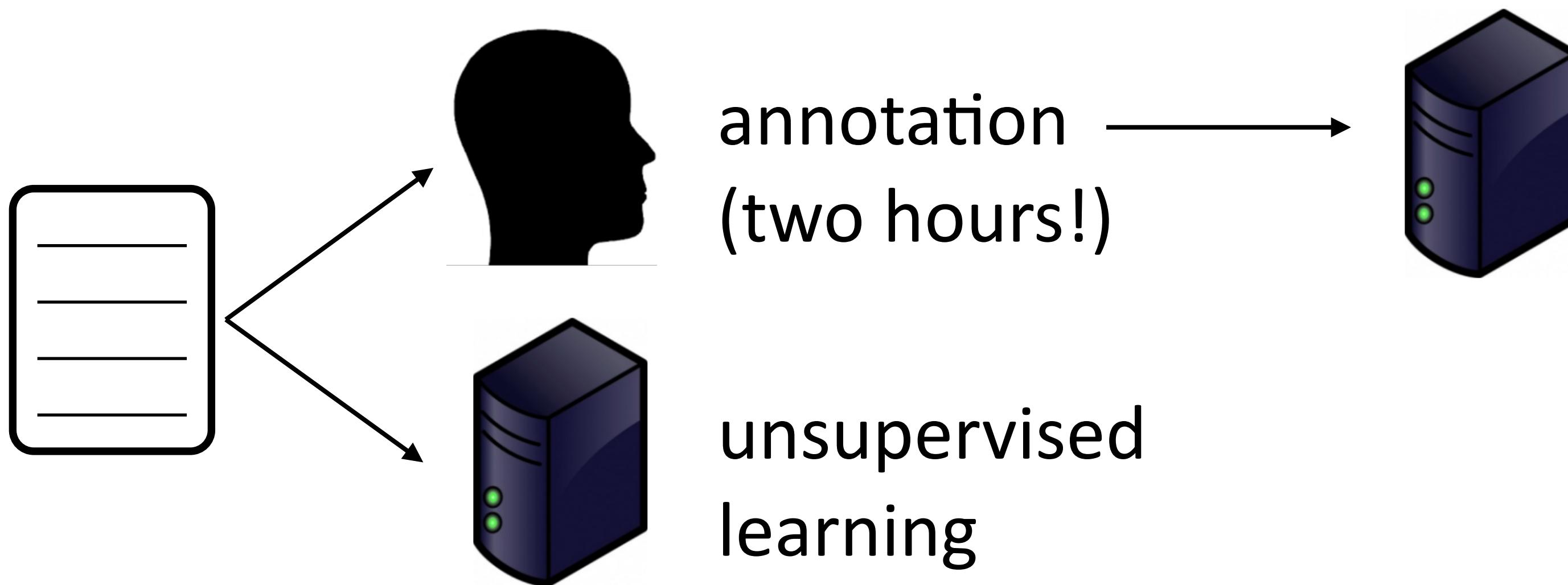
unsupervised  
learning

“Learning a Part-of-Speech Tagger from Two Hours of Annotation”  
Garrette and Baldridge (2013)

# Structured Prediction

---

- ▶ All of these techniques are data-driven! Some data is naturally occurring, but may need to label
- ▶ Supervised techniques work well on very little data

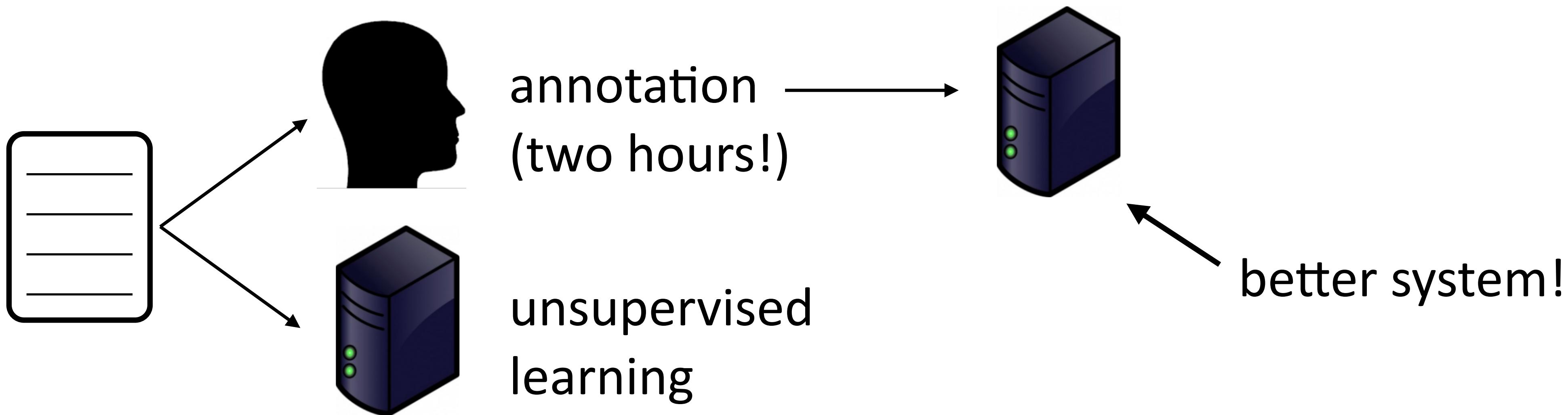


“Learning a Part-of-Speech Tagger from Two Hours of Annotation”  
Garrette and Baldridge (2013)

# Structured Prediction

---

- ▶ All of these techniques are data-driven! Some data is naturally occurring, but may need to label
- ▶ Supervised techniques work well on very little data

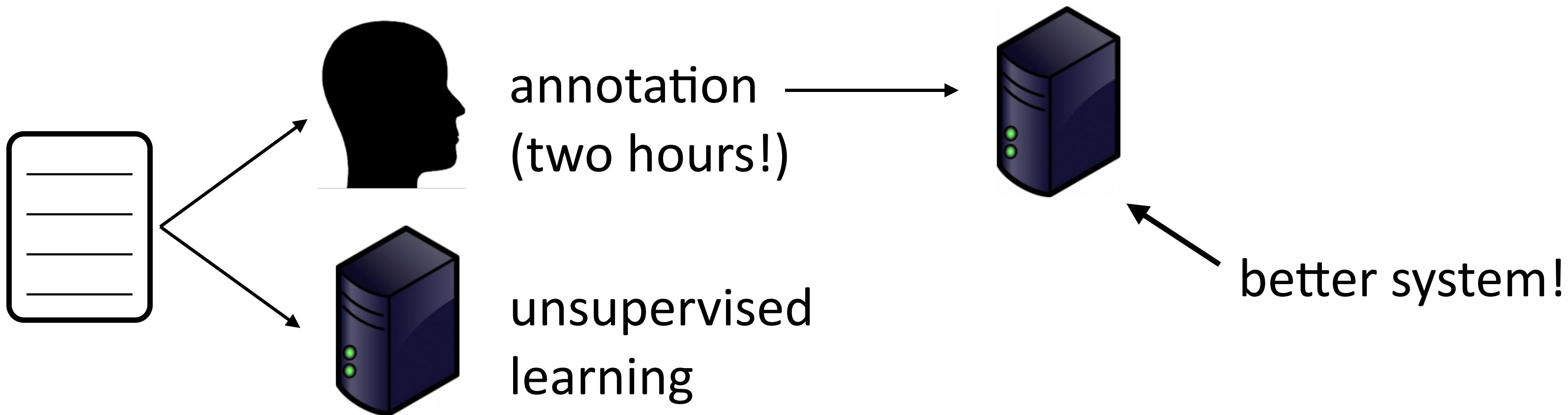


“Learning a Part-of-Speech Tagger from Two Hours of Annotation”  
Garrette and Baldridge (2013)

# Structured Prediction

---

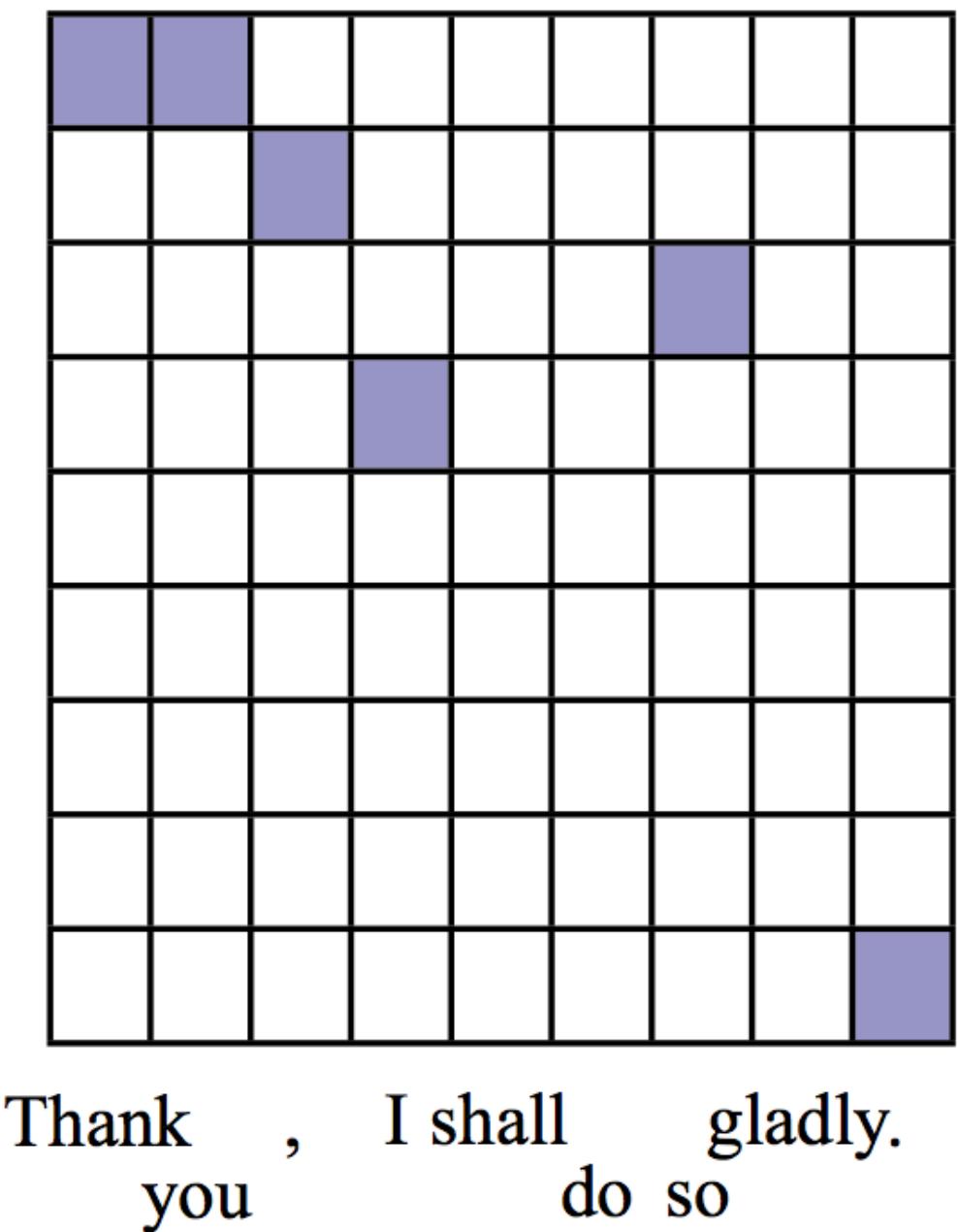
- ▶ All of these techniques are data-driven! Some data is naturally occurring, but may need to label
- ▶ Supervised techniques work well on very little data



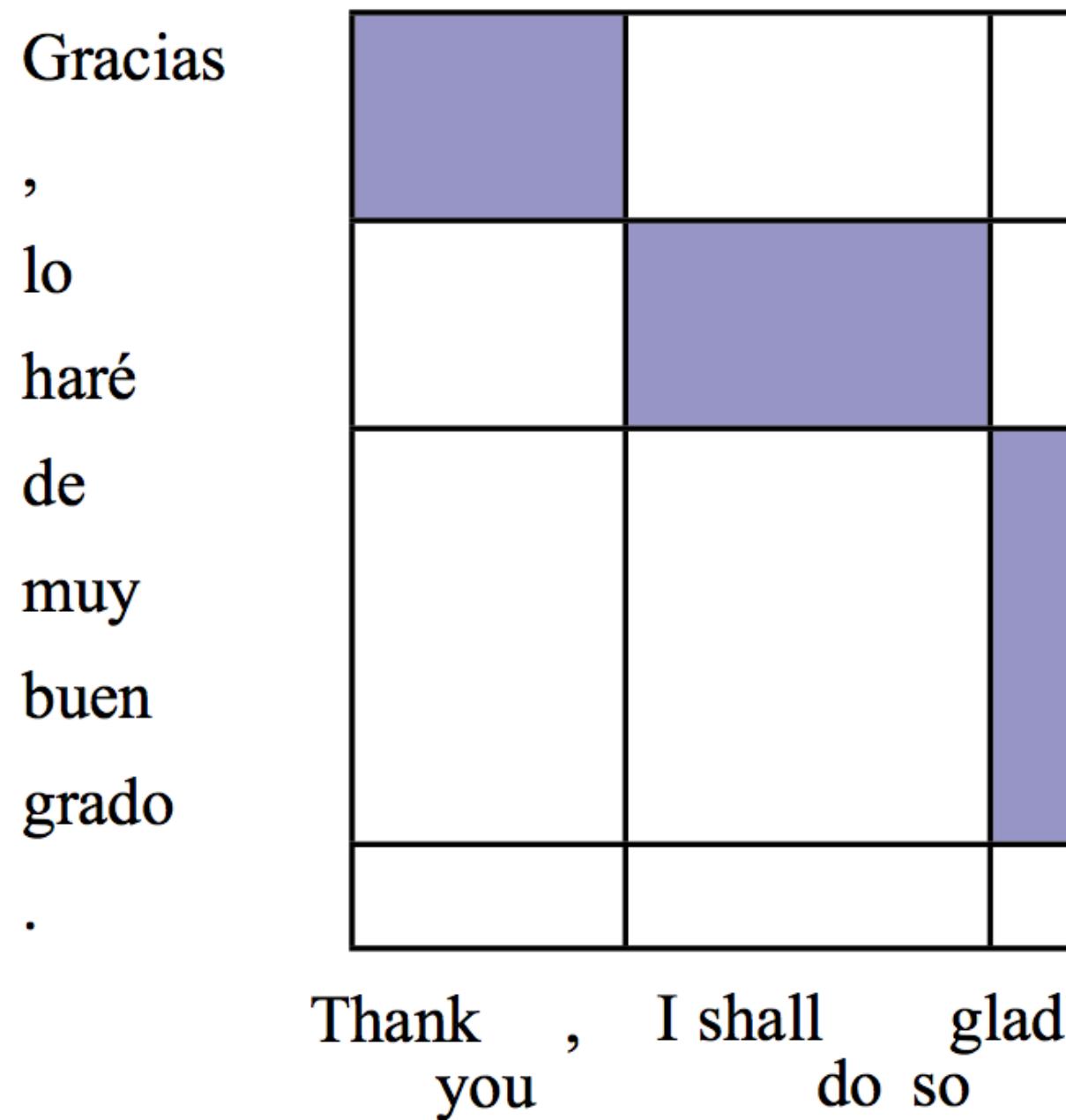
- ▶ Even neural nets can do pretty well!

“Learning a Part-of-Speech Tagger from Two Hours of Annotation”  
Garrette and Baldridge (2013)

# Less Manual Structure?



### (a) example word alignment



**(b) example phrase alignment**

The figure displays a 2D word cloud with two columns of text. The left column is in French and the right column is in English. The size of each word's pixel block corresponds to its frequency.

**French Text:**

La destruction de l'équipement signifie que la Syrie ne peut plus produire de nouvelles armes chimiques.

**English Text:**

Destruction of the equipment means that Syria can no longer produce new chemical weapons.

. <end>

DeNero et al. (2008)

Bahdanau et al. (2014)

# Does manual structure have a place?

---

# Does manual structure have a place?

---

- ▶ Neural nets don't always work out of domain!

# Does manual structure have a place?

---

- ▶ Neural nets don't always work out of domain!
- ▶ Coreference: rule-based systems are still about as good as deep learning out-of-domain

# Does manual structure have a place?

- ▶ Neural nets don't always work out of domain!
- ▶ Coreference: rule-based systems are still about as good as deep learning out-of-domain

	CoNLL Avg. F <sub>1</sub>
<b>Newswire</b>	
rule-based	55.60
berkeley	61.24
cort	63.37
deep-coref [conll]	65.39
deep-coref [lea]	65.60
<b>Wikipedia</b>	
rule-based	51.77
berkeley	51.01
cort	49.94
deep-coref [conll]	52.65
deep-coref [lea]	53.14
deep-coref <sup>-</sup>	51.01

Moosavi and Strube (2017)

# Does manual structure have a place?

- ▶ Neural nets don't always work out of domain!
- ▶ Coreference: rule-based systems are still about as good as deep learning out-of-domain
- ▶ LORELEI: transition point below which phrase-based systems are better

	CoNLL Avg. F <sub>1</sub>
<b>Newswire</b>	
rule-based	55.60
berkeley	61.24
cort	63.37
deep-coref [conll]	65.39
deep-coref [lea]	65.60
<b>Wikipedia</b>	
rule-based	51.77
berkeley	51.01
cort	49.94
deep-coref [conll]	52.65
deep-coref [lea]	53.14
deep-coref <sup>-</sup>	51.01

Moosavi and Strube (2017)

# Does manual structure have a place?

- ▶ Neural nets don't always work out of domain!
- ▶ Coreference: rule-based systems are still about as good as deep learning out-of-domain
- ▶ LORELEI: transition point below which phrase-based systems are better
- ▶ Why is this? Inductive bias!

	CoNLL
	Avg. F <sub>1</sub>
Newswire	
rule-based	55.60
berkeley	61.24
cort	63.37
deep-coref [conll]	65.39
deep-coref [lea]	65.60

	Wikipedia
rule-based	51.77
berkeley	51.01
cort	49.94
deep-coref [conll]	52.65
deep-coref [lea]	53.14
deep-coref-	51.01

Moosavi and Strube (2017)

# Does manual structure have a place?

- ▶ Neural nets don't always work out of domain!
- ▶ Coreference: rule-based systems are still about as good as deep learning out-of-domain
- ▶ LORELEI: transition point below which phrase-based systems are better
- ▶ Why is this? Inductive bias!
- ▶ Can multi-task learning help?

	CoNLL Avg. F <sub>1</sub>
<b>Newswire</b>	
rule-based	55.60
berkeley	61.24
cort	63.37
deep-coref [conll]	65.39
deep-coref [lea]	65.60
<b>Wikipedia</b>	
rule-based	51.77
berkeley	51.01
cort	49.94
deep-coref [conll]	52.65
deep-coref [lea]	53.14
deep-coref <sup>-</sup>	51.01

Moosavi and Strube (2017)

# Does manual structure have a place?

Translate

English French Spanish Chinese - detected ▾

特朗普偕家人在白宫阳台观看百年一遇日全食 X

Trump Pope family watch a hundred years a year in the White House balcony

# Does manual structure have a place?



Trump **Pope** family watch a hundred years a year in the White House balcony

- ▶ Maybe manual structure would help...

# Where are we?

---

# Where are we?

---

- ▶ NLP consists of: analyzing and building representations for text, solving problems involving text

# Where are we?

---

- ▶ NLP consists of: analyzing and building representations for text, solving problems involving text
- ▶ These problems are hard because language is ambiguous, requires drawing on data, knowledge, and linguistics to solve

# Where are we?

---

- ▶ NLP consists of: analyzing and building representations for text, solving problems involving text
- ▶ These problems are hard because language is ambiguous, requires drawing on data, knowledge, and linguistics to solve
- ▶ Knowing which techniques use requires understanding dataset size, problem complexity, and a lot of tricks!

# Where are we?

---

- ▶ NLP consists of: analyzing and building representations for text, solving problems involving text
- ▶ These problems are hard because language is ambiguous, requires drawing on data, knowledge, and linguistics to solve
- ▶ Knowing which techniques use requires understanding dataset size, problem complexity, and a lot of tricks!
- ▶ NLP encompasses all of these things

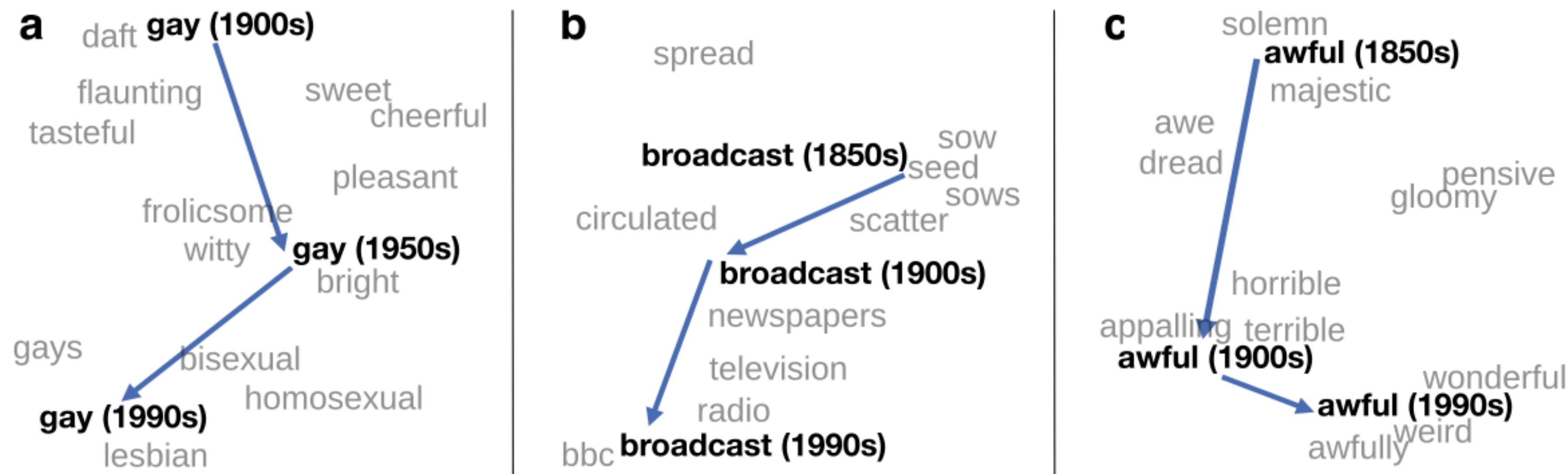
# NLP vs. Computational Linguistics

---

- ▶ NLP: build systems that deal with language data
- ▶ CL: use computational tools to study language

# NLP vs. Computational Linguistics

- ▶ NLP: build systems that deal with language data
- ▶ CL: use computational tools to study language



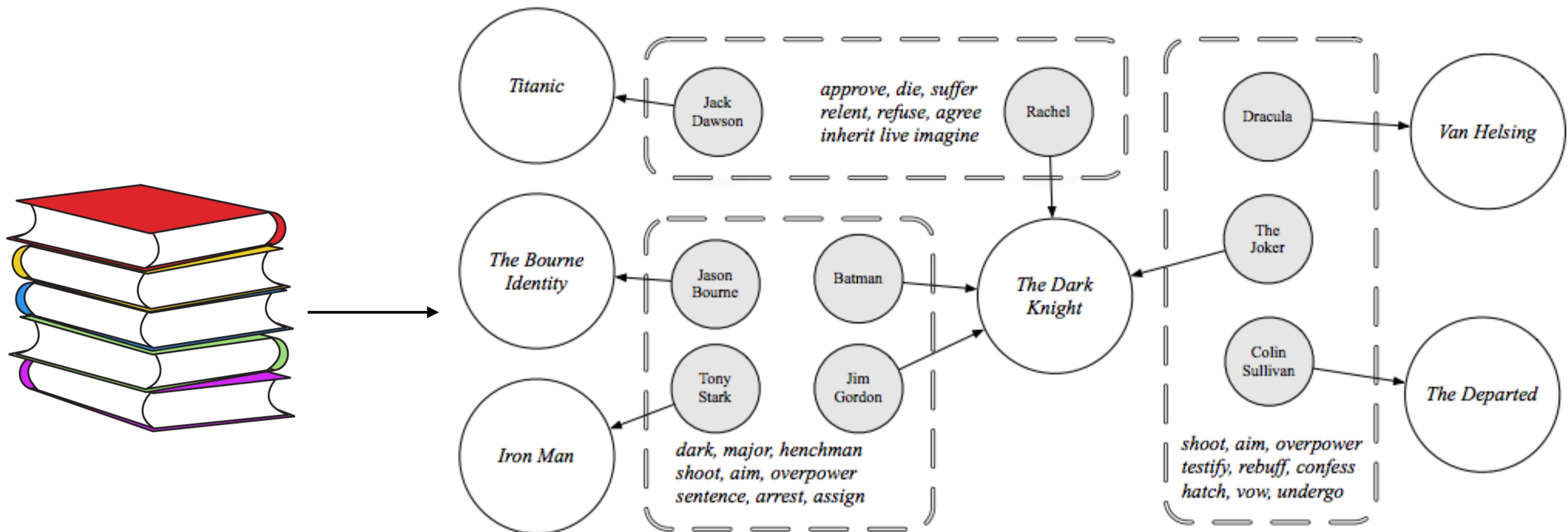
# NLP vs. Computational Linguistics

---

- ▶ Computational tools for other purposes: literary theory, political science...

# NLP vs. Computational Linguistics

- ▶ Computational tools for other purposes: literary theory, political science...



# Outline of the Course

---

Date	Topics	Readings
Aug 30	Introduction [4pp]	
Sept 4	Binary classification	JM 6.1-6.3
Sept 6	Multiclass classification	JM 7, Structured SVM secs 1-2
Sept 11	Sequence models I: HMMs	JM 9, JM 10.4, Manning POS
Sept 13	Sequence models II: CRFs	Sutton CRFs 2.3, 2.6.1, Wallach CRFs tutorial, Illinois NER
Sept 18	Neural Nets I: FFNNs	Goldberg 1-4, 6, NLP with FFNNs, DANs
Sept 20	Neural Nets II: NN impl / word embeddings	Goldberg 5, word2vec, GloVe, Dropout
Sept 25	Neural Nets III: RNN and CNN encoders	Goldberg 9-11, Kim
Sept 27	Neural Nets IV: Neural CRFs	Collobert and Weston, Neural NER, Neural CRF parsing
Oct 2	Trees I: Constituency, PCFGs	JM 13.1-13.7, Structural, Lexicalized, State-split
Oct 4	Trees II: Dependency I	JM 14.1-14.4, Huang 1-2
Oct 9	Trees III: Dependency II	Parsey, Huang 2
Oct 11	Semantics I	
Oct 16	Semantics II / Seq2seq I	
Oct 18	Seq2seq II: Beam search, attention	Seq2seq, Attention, Luong Attention
Oct 23	Information Extraction / SRL	Distant supervision, RL for slot filling, TextRunner, ReVerb, NELL
Oct 25	Discourse and Coreference	
Oct 30	Machine Translation I: Phrase-based	HMM alignment, Pharaoh
Nov 1	Machine Translation II: Neural	
Nov 6	Applications I: Reading comprehension / MemNets	E2E Memory Networks, CBT, SQuAD, BiDAF
Nov 8	Applications II: Language grounding	
Nov 13	Applications III: Summarization	MMR, Gillick, Sentence compression, SummaRuNNER, Pointer
Nov 15	Applications IV: Dialogue	RNN chatbots, Diversity, Goal-oriented, Latent Intention, QA-as-dialogue
Nov 20	Unsupervised Learning	
Nov 22	NO CLASS (Thanksgiving)	
Nov 27	Multilinguality and morphology	
Nov 29	Wrapup	
Dec 4	Project presentations I	
Dec 6	Project presentations II	
Dec 15		

# Outline of the Course

## ML and structured prediction for NLP

Date	Topics	Readings
Aug 30	Introduction [4pp]	
Sept 4	Binary classification	JM 6.1-6.3
Sept 6	Multiclass classification	JM 7, Structured SVM secs 1-2
Sept 11	Sequence models I: HMMs	JM 9, JM 10.4, Manning POS
Sept 13	Sequence models II: CRFs	Sutton CRFs 2.3, 2.6.1, Wallach CRFs tutorial, Illinois NER
Sept 18	Neural Nets I: FFNNs	Goldberg 1-4, 6, NLP with FFNNs, DANs
Sept 20	Neural Nets II: NN impl / word embeddings	Goldberg 5, word2vec, GloVe, Dropout
Sept 25	Neural Nets III: RNN and CNN encoders	Goldberg 9-11, Kim
Sept 27	Neural Nets IV: Neural CRFs	Collobert and Weston, Neural NER, Neural CRF parsing
Oct 2	Trees I: Constituency, PCFGs	JM 13.1-13.7, Structural, Lexicalized, State-split
Oct 4	Trees II: Dependency I	JM 14.1-14.4, Huang 1-2
Oct 9	Trees III: Dependency II	Parsey, Huang 2
Oct 11	Semantics I	
Oct 16	Semantics II / Seq2seq I	
Oct 18	Seq2seq II: Beam search, attention	Seq2seq, Attention, Luong Attention
Oct 23	Information Extraction / SRL	Distant supervision, RL for slot filling, TextRunner, ReVerb, NELL
Oct 25	Discourse and Coreference	
Oct 30	Machine Translation I: Phrase-based	HMM alignment, Pharaoh
Nov 1	Machine Translation II: Neural	
Nov 6	Applications I: Reading comprehension / MemNets	E2E Memory Networks, CBT, SQuAD, BiDAF
Nov 8	Applications II: Language grounding	
Nov 13	Applications III: Summarization	MMR, Gillick, Sentence compression, SummaRuNNER, Pointer
Nov 15	Applications IV: Dialogue	RNN chatbots, Diversity, Goal-oriented, Latent Intention, QA-as-dialogue
Nov 20	Unsupervised Learning	
Nov 22	NO CLASS (Thanksgiving)	
Nov 27	Multilinguality and morphology	
Nov 29	Wrapup	
Dec 4	Project presentations I	
Dec 6	Project presentations II	
Dec 15		

# Outline of the Course

## ML and structured prediction for NLP

### Neural nets

Date	Topics	Readings
Aug 30	Introduction [4pp]	
Sept 4	Binary classification	JM 6.1-6.3
Sept 6	Multiclass classification	JM 7, Structured SVM secs 1-2
Sept 11	Sequence models I: HMMs	JM 9, JM 10.4, Manning POS
Sept 13	Sequence models II: CRFs	Sutton CRFs 2.3, 2.6.1, Wallach CRFs tutorial, Illinois NER
Sept 18	Neural Nets I: FFNNs	Goldberg 1-4, 6, NLP with FFNNs, DANs
Sept 20	Neural Nets II: NN impl / word embeddings	Goldberg 5, word2vec, GloVe, Dropout
Sept 25	Neural Nets III: RNN and CNN encoders	Goldberg 9-11, Kim
Sept 27	Neural Nets IV: Neural CRFs	Collobert and Weston, Neural NER, Neural CRF parsing
Oct 2	Trees I: Constituency, PCFGs	JM 13.1-13.7, Structural, Lexicalized, State-split
Oct 4	Trees II: Dependency I	JM 14.1-14.4, Huang 1-2
Oct 9	Trees III: Dependency II	Parsey, Huang 2
Oct 11	Semantics I	
Oct 16	Semantics II / Seq2seq I	
Oct 18	Seq2seq II: Beam search, attention	Seq2seq, Attention, Luong Attention
Oct 23	Information Extraction / SRL	Distant supervision, RL for slot filling, TextRunner, ReVerb, NELL
Oct 25	Discourse and Coreference	
Oct 30	Machine Translation I: Phrase-based	HMM alignment, Pharaoh
Nov 1	Machine Translation II: Neural	
Nov 6	Applications I: Reading comprehension / MemNets	E2E Memory Networks, CBT, SQuAD, BiDAF
Nov 8	Applications II: Language grounding	
Nov 13	Applications III: Summarization	MMR, Gillick, Sentence compression, SummaRuNNER, Pointer
Nov 15	Applications IV: Dialogue	RNN chatbots, Diversity, Goal-oriented, Latent Intention, QA-as-dialogue
Nov 20	Unsupervised Learning	
Nov 22	NO CLASS (Thanksgiving)	
Nov 27	Multilinguality and morphology	
Nov 29	Wrapup	
Dec 4	Project presentations I	
Dec 6	Project presentations II	
Dec 15		

# Outline of the Course

ML and structured prediction for NLP

Neural nets

Syntax/  
semantics

Date	Topics	Readings
Aug 30	Introduction [4pp]	
Sept 4	Binary classification	JM 6.1-6.3
Sept 6	Multiclass classification	JM 7, Structured SVM secs 1-2
Sept 11	Sequence models I: HMMs	JM 9, JM 10.4, Manning POS
Sept 13	Sequence models II: CRFs	Sutton CRFs 2.3, 2.6.1, Wallach CRFs tutorial, Illinois NER
Sept 18	Neural Nets I: FFNNs	Goldberg 1-4, 6, NLP with FFNNs, DANs
Sept 20	Neural Nets II: NN impl / word embeddings	Goldberg 5, word2vec, GloVe, Dropout
Sept 25	Neural Nets III: RNN and CNN encoders	Goldberg 9-11, Kim
Sept 27	Neural Nets IV: Neural CRFs	Collobert and Weston, Neural NER, Neural CRF parsing
Oct 2	Trees I: Constituency, PCFGs	JM 13.1-13.7, Structural, Lexicalized, State-split
Oct 4	Trees II: Dependency I	JM 14.1-14.4, Huang 1-2
Oct 9	Trees III: Dependency II	Parsey, Huang 2
Oct 11	Semantics I	
Oct 16	Semantics II / Seq2seq I	
Oct 18	Seq2seq II: Beam search, attention	Seq2seq, Attention, Luong Attention
Oct 23	Information Extraction / SRL	Distant supervision, RL for slot filling, TextRunner, ReVerb, NELL
Oct 25	Discourse and Coreference	
Oct 30	Machine Translation I: Phrase-based	HMM alignment, Pharaoh
Nov 1	Machine Translation II: Neural	
Nov 6	Applications I: Reading comprehension / MemNets	E2E Memory Networks, CBT, SQuAD, BiDAF
Nov 8	Applications II: Language grounding	
Nov 13	Applications III: Summarization	MMR, Gillick, Sentence compression, SummaRuNNER, Pointer
Nov 15	Applications IV: Dialogue	RNN chatbots, Diversity, Goal-oriented, Latent Intention, QA-as-dialogue
Nov 20	Unsupervised Learning	
Nov 22	NO CLASS (Thanksgiving)	
Nov 27	Multilinguality and morphology	
Nov 29	Wrapup	
Dec 4	Project presentations I	
Dec 6	Project presentations II	
Dec 15		

# Outline of the Course

ML and structured prediction for NLP  
Neural nets  
Syntax/  
semantics  
Applications:  
MT, IE,  
summarization,  
dialogue, etc.

Date	Topics	Readings
Aug 30	Introduction [4pp]	
Sept 4	Binary classification	JM 6.1-6.3
Sept 6	Multiclass classification	JM 7, Structured SVM secs 1-2
Sept 11	Sequence models I: HMMs	JM 9, JM 10.4, Manning POS
Sept 13	Sequence models II: CRFs	Sutton CRFs 2.3, 2.6.1, Wallach CRFs tutorial, Illinois NER
Sept 18	Neural Nets I: FFNNs	Goldberg 1-4, 6, NLP with FFNNs, DANs
Sept 20	Neural Nets II: NN impl / word embeddings	Goldberg 5, word2vec, GloVe, Dropout
Sept 25	Neural Nets III: RNN and CNN encoders	Goldberg 9-11, Kim
Sept 27	Neural Nets IV: Neural CRFs	Collobert and Weston, Neural NER, Neural CRF parsing
Oct 2	Trees I: Constituency, PCFGs	JM 13.1-13.7, Structural, Lexicalized, State-split
Oct 4	Trees II: Dependency I	JM 14.1-14.4, Huang 1-2
Oct 9	Trees III: Dependency II	Parsey, Huang 2
Oct 11	Semantics I	
Oct 16	Semantics II / Seq2seq I	
Oct 18	Seq2seq II: Beam search, attention	Seq2seq, Attention, Luong Attention
Oct 23	Information Extraction / SRL	Distant supervision, RL for slot filling, TextRunner, ReVerb, NELL
Oct 25	Discourse and Coreference	
Oct 30	Machine Translation I: Phrase-based	HMM alignment, Pharaoh
Nov 1	Machine Translation II: Neural	
Nov 6	Applications I: Reading comprehension / MemNets	E2E Memory Networks, CBT, SQuAD, BiDAF
Nov 8	Applications II: Language grounding	
Nov 13	Applications III: Summarization	MMR, Gillick, Sentence compression, SummaRuNNER, Pointer
Nov 15	Applications IV: Dialogue	RNN chatbots, Diversity, Goal-oriented, Latent Intention, QA-as-dialogue
Nov 20	Unsupervised Learning	
Nov 22	NO CLASS (Thanksgiving)	
Nov 27	Multilinguality and morphology	
Nov 29	Wrapup	
Dec 4	Project presentations I	
Dec 6	Project presentations II	
Dec 15		

# Course Goals

---

# Course Goals

---

- ▶ Cover fundamental machine learning techniques used in NLP

# Course Goals

---

- ▶ Cover fundamental machine learning techniques used in NLP
- ▶ Understand how to look at language data and approach linguistic phenomena

# Course Goals

---

- ▶ Cover fundamental machine learning techniques used in NLP
- ▶ Understand how to look at language data and approach linguistic phenomena
- ▶ Cover modern NLP problems encountered in the literature: what are the active research topics in 2018?

# Course Goals

---

- ▶ Cover fundamental machine learning techniques used in NLP
- ▶ Understand how to look at language data and approach linguistic phenomena
- ▶ Cover modern NLP problems encountered in the literature: what are the active research topics in 2018?
- ▶ Make you a “producer” rather than a “consumer” of NLP tools

# Course Goals

---

- ▶ Cover fundamental machine learning techniques used in NLP
- ▶ Understand how to look at language data and approach linguistic phenomena
- ▶ Cover modern NLP problems encountered in the literature: what are the active research topics in 2018?
- ▶ Make you a “producer” rather than a “consumer” of NLP tools
  - ▶ The four assignments should teach you what you need to know to understand nearly any system in the literature

# Assignments

---

- ▶ 4 Homework Assignments
  - ▶ Implementation-oriented, with an open-ended component to each
  - ▶ Homework 1 (Naive Bayes for sentiment classification) is out NOW
  - ▶ ~2 weeks per assignment, 5 “slip days” for automatic extensions

# Assignments

---

- ▶ 4 Homework Assignments
  - ▶ Implementation-oriented, with an open-ended component to each
  - ▶ Homework 1 (Naive Bayes for sentiment classification) is out NOW
  - ▶ ~2 weeks per assignment, 5 “slip days” for automatic extensions

These projects require understanding of the concepts, ability to write performant code, and ability to think about how to debug complex systems. **They are challenging, so start early!**

# Final Project

---

- ▶ Final project (20%)
  - ▶ Groups of 3-4 preferred, 1 is possible.
  - ▶ Good idea to talk to run your project idea by me in office hours or email.
  - ▶ 4 page report + final project presentation.

# Survey

---

1. Fill in: I am a [CS / \_\_\_\_] [PhD / masters / undergrad] in year [1 2 3 4 5+]
2. Which of the following have you learned in a class?
  1. Bayes' Rule
  2. SVMs
  3. Expectation maximization
  4. RNNs
3. Which of the following have you used?
  1. Python
  2. numpy/scipy/scikit-learn
  3. Tensorflow/(Py)Torch/Theano
4. One interesting fact about yourself, or what you like to do in your spare time