What is the goal of NLP?

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

- Systems that talk to us: dialogue systems, machine translation, summarization

Siri, what's your favorite kind of movie?

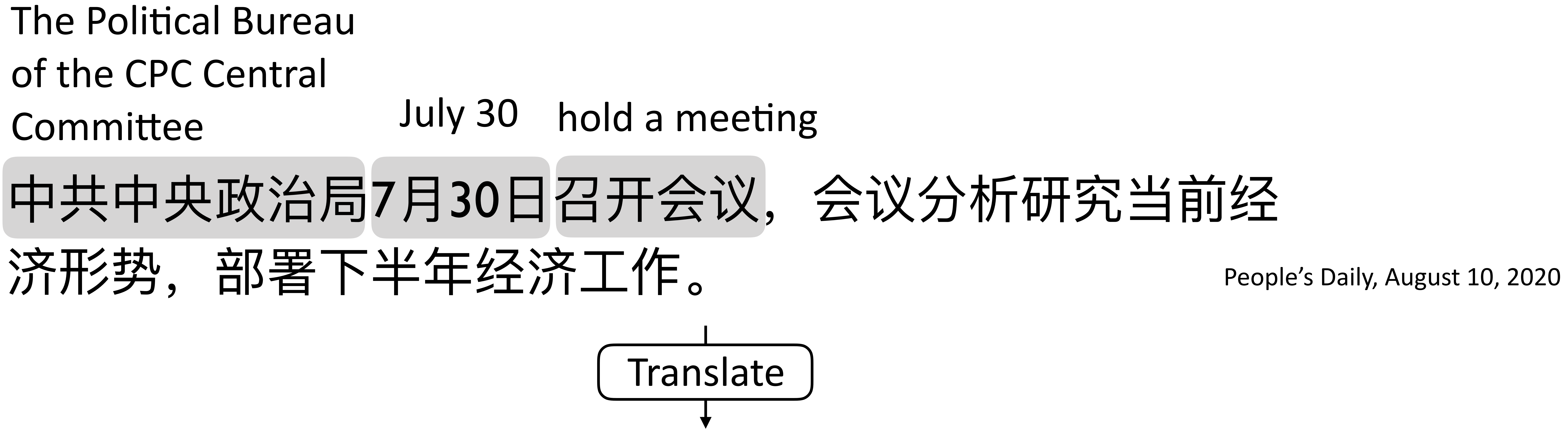
I like superhero movies!

What's come out recently?

The Avengers

What is the goal of NLP?

- Be able to solve problems that require **deep understanding** of text
- Systems that talk to us: dialogue systems, machine translation, summarization



The Political Bureau of the CPC Central Committee

July 30

hold a meeting

中共中央政治局7月30日召开会议，会议分析研究当前经济形势，部署下半年经济工作。

People’s Daily, August 10, 2020

Translate

The Political Bureau of the CPC Central Committee held a meeting on July 30 to analyze and study the current economic situation and plan economic work in the second half of the year.

What is the goal of NLP?

- Build systems that extract information from text and answer questions

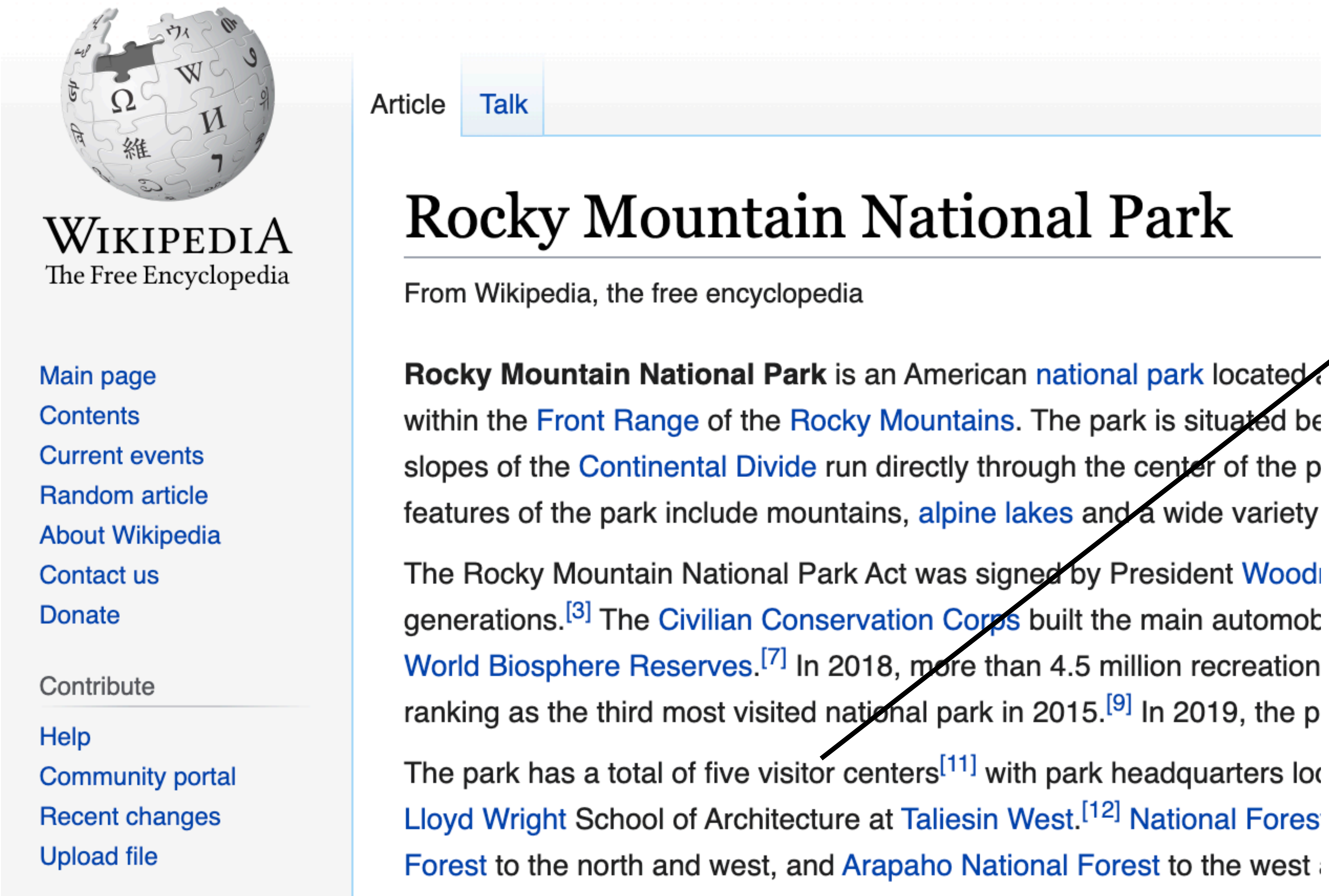
When was Abraham Lincoln born?

map to Birthday field

Name	Birthday
Lincoln, Abraham	2/12/1809
Washington, George	2/22/1732
Adams, John	10/30/1735

→ **February 12, 1809**

How many visitors centers are there in Rocky Mountain National Park?



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute

Help
Community portal
Recent changes
Upload file

Article Talk

Rocky Mountain National Park

From Wikipedia, the free encyclopedia

Rocky Mountain National Park is an American national park located within the Front Range of the Rocky Mountains. The park is situated be slopes of the Continental Divide run directly through the center of the p features of the park include mountains, alpine lakes and a wide variety

The Rocky Mountain National Park Act was signed by President Wood generations.[3] The Civilian Conservation Corps built the main automob World Biosphere Reserves.[7] In 2018, more than 4.5 million recreation ranking as the third most visited national park in 2015.[9] In 2019, the p

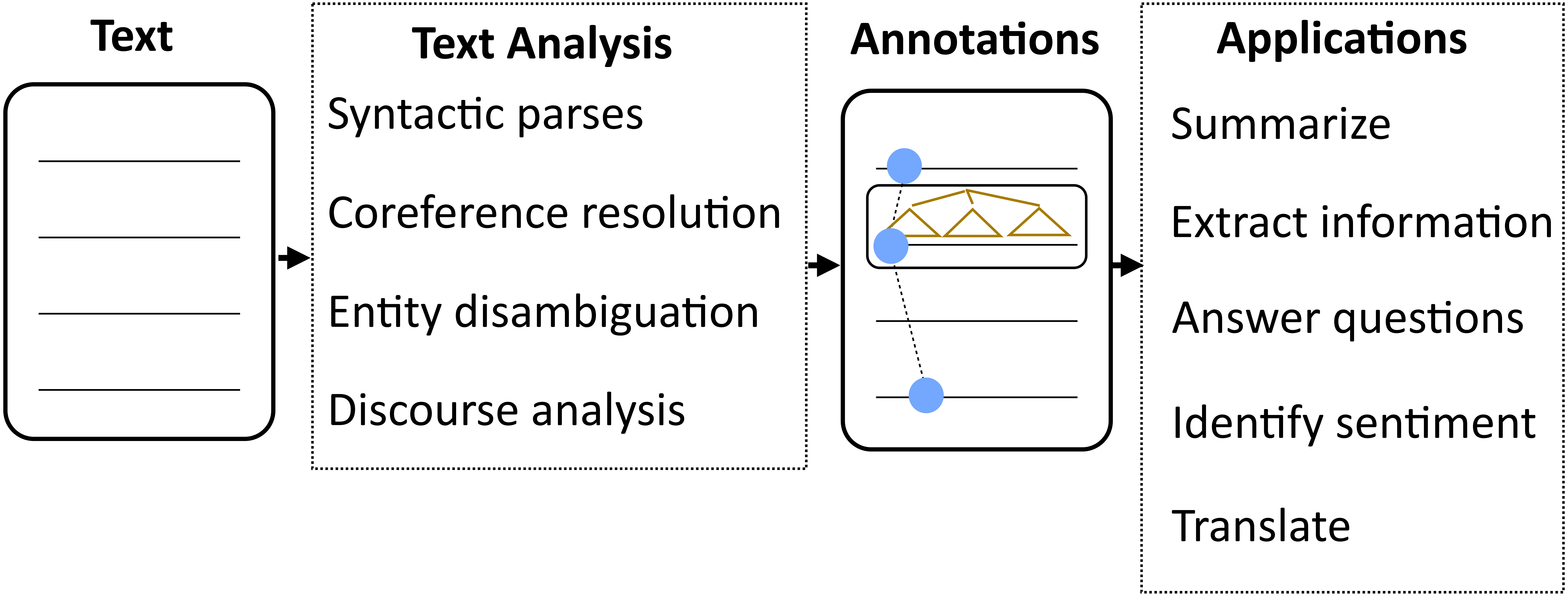
The park has a total of five visitor centers[11] with park headquarters lo Lloyd Wright School of Architecture at Taliesin West.[12] National Fores Forest to the north and west, and Arapaho National Forest to the west

The park has a total of five visitor centers

↓

five

“Standard” NLP Pipeline



Text

Text Analysis

Syntactic parses

Coreference resolution

Entity disambiguation

Discourse analysis

Annotations

Applications

Summarize

Extract information

Answer questions

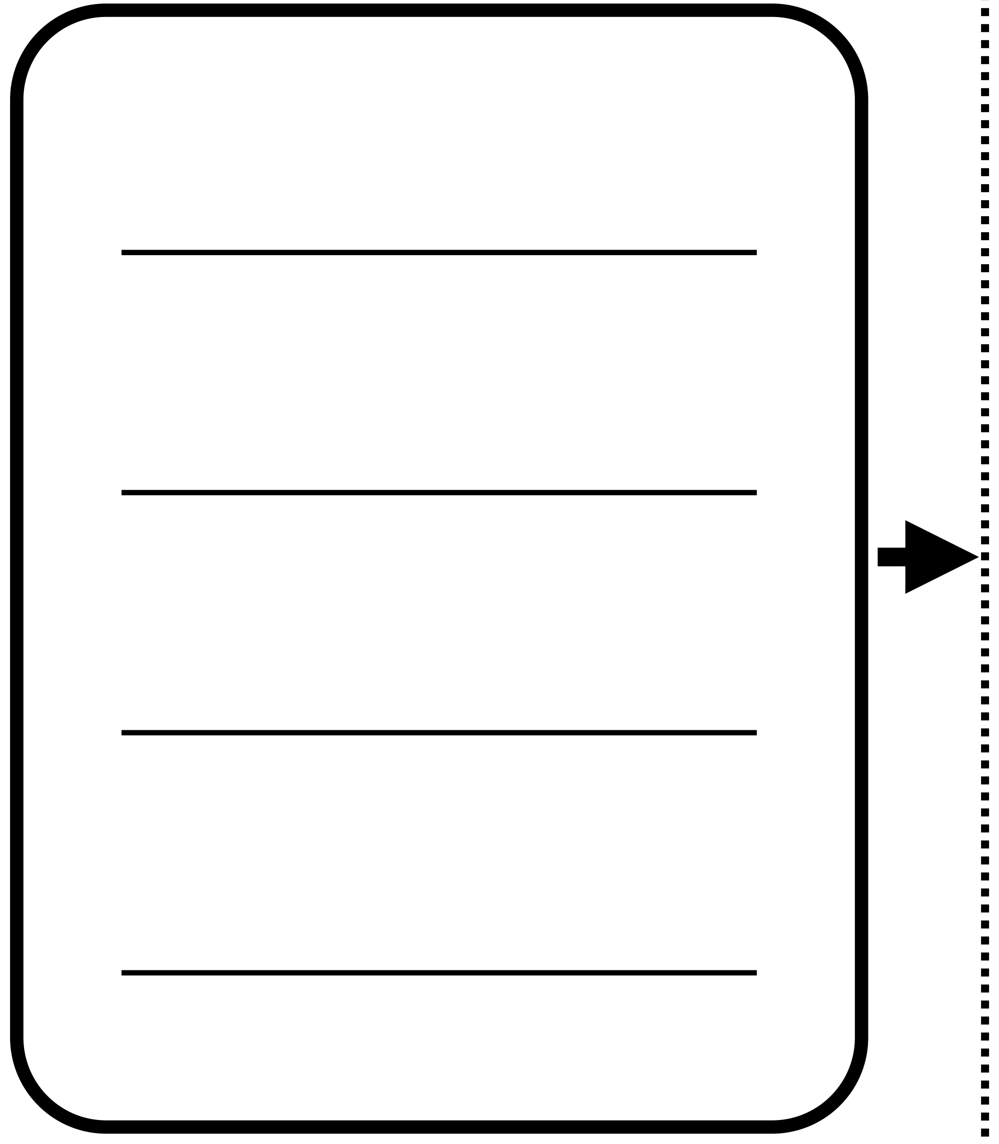
Identify sentiment

Translate

- All of these components are modeled with statistical approaches using machine learning

How do we represent language?

Text



Labels

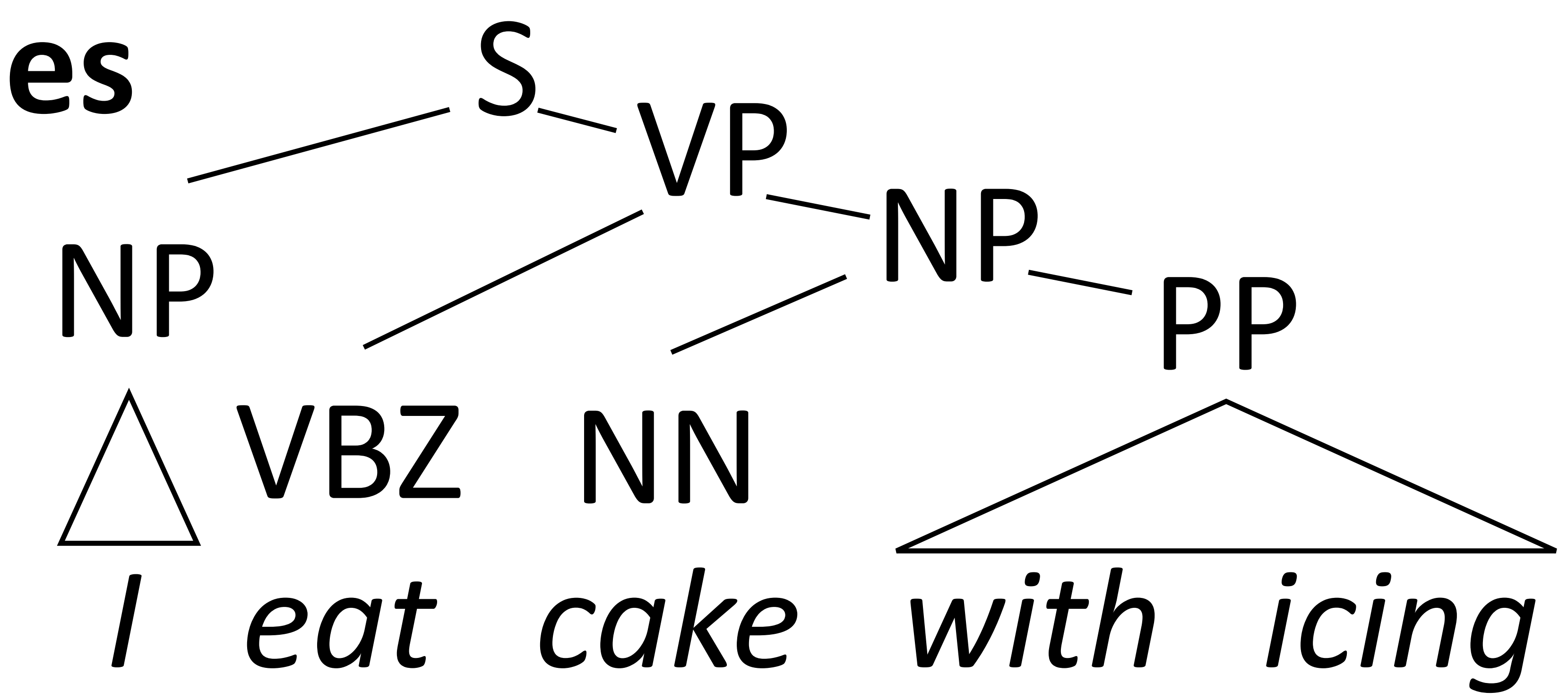
the movie was good +

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

Sequences/tags

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

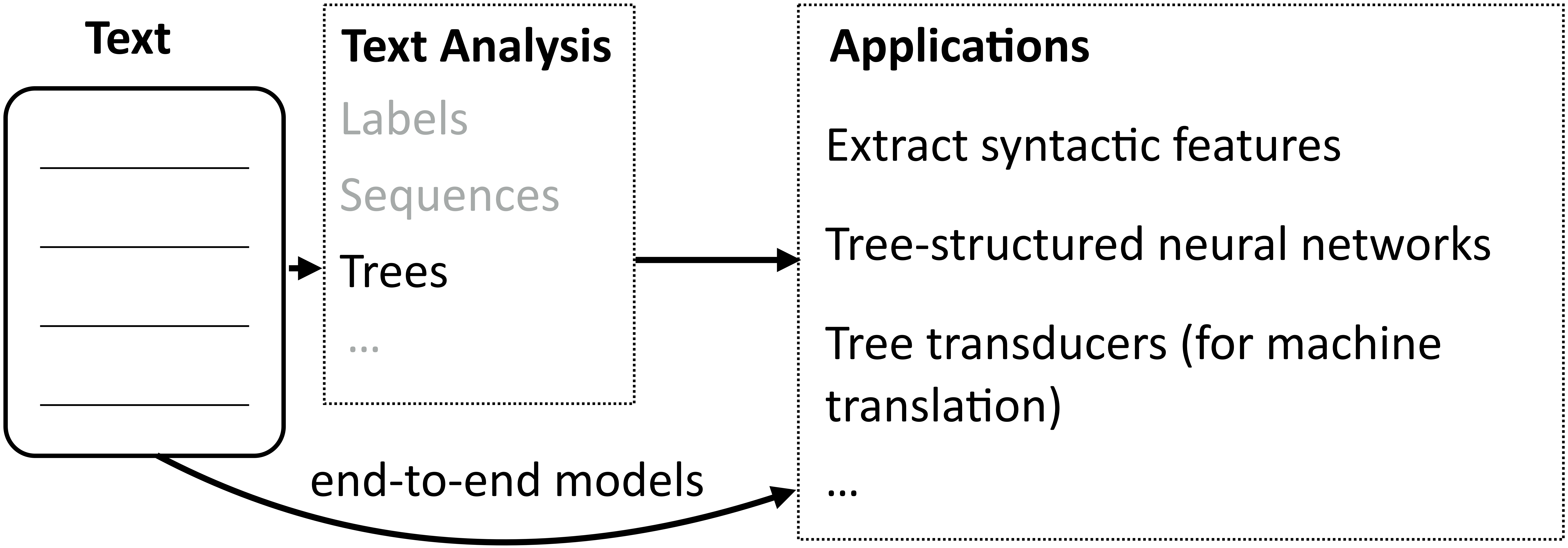
Trees



λx. flight(x) ∧ dest(x)=Miami

flights to Miami

How do we use these representations?



Text

Text Analysis

Labels

Sequences

Trees

…

Applications

Extract syntactic features

Tree-structured neural networks

Tree transducers (for machine translation)

…

end-to-end models

- Why is this prediction hard? Because language is complex and ambiguous!
- What ambiguities do we need to resolve?

What makes language hard?

Teacher Strikes Idle Kids

Ban on Nude Dancing on Governor’s Desk

Iraqi Head Seeks Arms

What makes language hard?

- There aren't just one or two possibilities, but many!

il fait vraiment beau →

It is really nice out

It's really nice

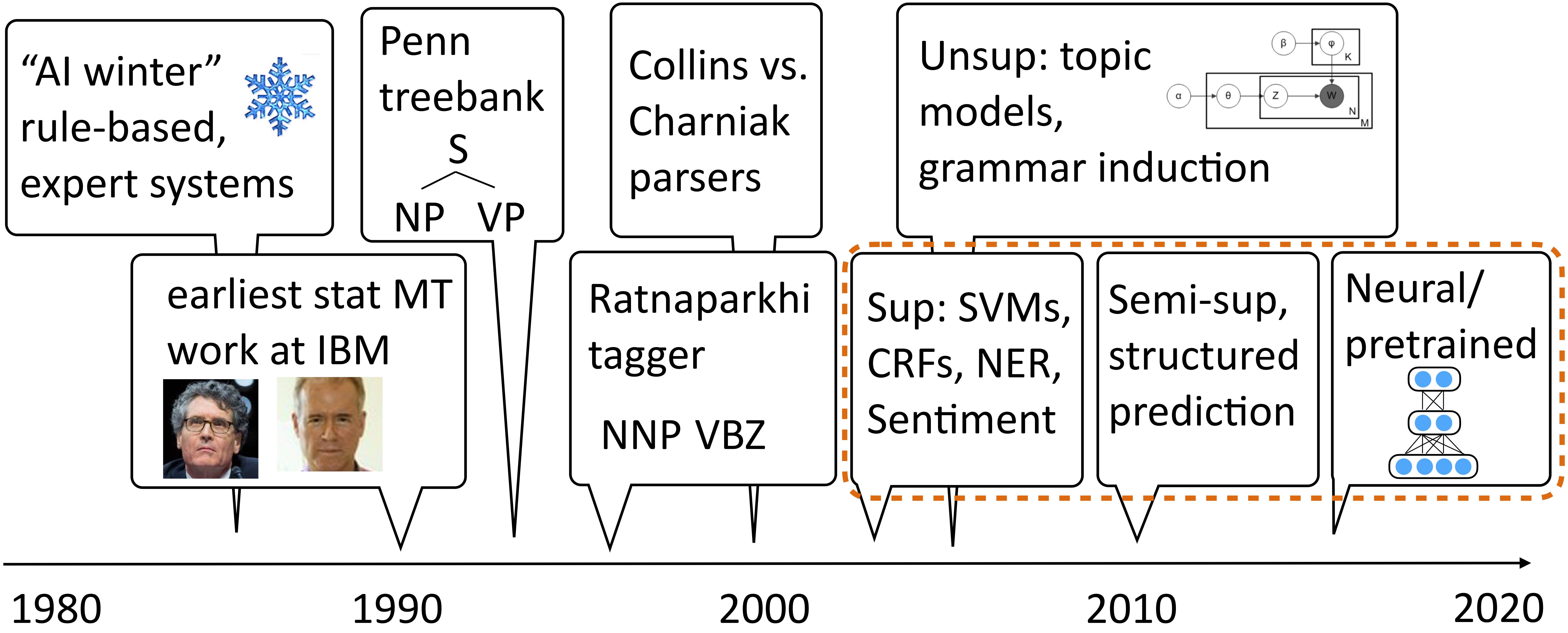
The weather is beautiful

It is really beautiful outside

He makes truly beautiful

It fact actually handsome

A brief history of NLP techniques



"AI winter" rule-based, expert systems

earliest stat MT work at IBM

Penn treebank

S

NP VP

Collins vs. Charniak parsers

Ratnaparkhi tagger

NNP VBZ

Unsup: topic models, grammar induction

Sup: SVMs, CRFs, NER, Sentiment

Semi-sup, structured prediction

Neural/ pretrained

- This course focuses on **supervised learning**, **semi-supervised methods**, and **neural models (including pre-training)**
- All of these models can handle ambiguity by learning how to map text into linguistic representations using data

Outline of the course

- Classification: linear and neural, word representations
- Text analysis: tagging (HMMs, CRFs) and parsing (PCFGs)
- Language modeling and pre-training
- Question answering and semantics
- Machine translation
- Applications