CS 4650: Natural Language Processing

Alan Ritter

Administrivia

- Course website: https://aritter.github.io/CS-4650-au24/
- Piazza and Gradescope: links on the course website
  - We will do our best to answer questions within 24 hours (or Monday for questions asked over the weekend).
  - **Please Sign Yourself up for Piazza**
- TA Office hours:
  - See spreadsheet

https://tinyurl.com/4650-TAs

Instructor

Alan Ritter
alan.ritter@cc.gatech.edu

Teaching Assistants

Duong Le
dminh6@gatech.edu

Jeongrok Yu
jyu673@gatech.edu

Junmo Kang
junmo.kang@gatech.edu

Prerequisites

- Probability/Statistics
- Linear Algebra
- Multivariable Calculus
- Programming / Python experience
- A prior Machine Learning course will be very helpful

Prerequisites

- Probability/Statistics
- Linear Algebra
- Multivariable Calculus
- Programming / Python experience
- A prior Machine Learning course will be very helpful

There will be a lot of math and programming!

Coursework

- 4 Programming Projects (fairly substantial implementation effort)
  - Logistic Regression
  - Text classification
  - Sequence Labeling
  - Neural chatbot

Coursework

- 4 Programming Projects (fairly substantial implementation effort)
  - Logistic Regression
  - Text classification
  - Sequence Labeling
  - Neural chatbot
- 3 written assignments + midterm exam
  - Mostly math problems related to ML / NLP

Coursework

- 4 Programming Projects (fairly substantial implementation effort)
  - Logistic Regression
  - Text classification
  - Sequence Labeling
  - Neural chatbot
- 3 written assignments + midterm exam
  - Mostly math problems related to ML / NLP
- Final project

Coursework

- 4 Programming Projects (fairly substantial implementation effort)
  - Logistic Regression
  - Text classification
  - Sequence Labeling
  - Neural chatbot
- 3 written assignments + midterm exam
  - Mostly math problems related to ML / NLP
- Final project
- Problem Set 0 (background review) is out now and **due Thursday**.

Final Project

- Final project (20%)
  - Groups of 3-4.
    - 1 is possible, but will require more work. Grading requirements are the same for individual projects.
  - 4 page report.
  - We will discuss more and have a “project kickoff” later in the semester.

Problem Set 0 (Background Review)

- Due this Thursday.
- Background review on probability, linear algebra, calculus.
- **Waitlisted students:** please submit PS0 by Thursday if you plan to enroll in the course.
    - We can’t predict whether or not you will get in, as this depends on other students dropping the class…
- Submit on Gradescope
    - If you don’t have access to Gradescope, send the course staff a private message on Piazza, and we will give you the access code.

Project 0 is also out (please look!)

Logistic Regression

CS 4650 "Natural Language Processing" Project 0
Georgia Tech, Fall 2023 (Instructor: Alan Ritter)

In this assignment, we will walk you through the process of implementing logistic regression from scratch. You will also apply your implemented logistic regression model to a small dataset and predict whether a student will be admitted to a university. This dataset will allow you to visualize the data and debug more easily. You may find this documentation very helpful (although it uses Octave rather than Python).

This assignment also serves as a programming preparation test. We will use NumPy -- a popular Python package for scientific computing and implementing machine learning algorithms. It provides very good support for matrix and vector operations. You need to feel comfortable working with matrics, vectors, and tensors in order to complete all the programming projects in CS 4650.

To start, first make a copy of this notebook to your local drive, so you can edit it.

4. Gradient Computation [5 points]

Implement the gradient computations for logistic regression.

```
def gradient_update(theta, X, y):
    """ The gradient update for logistic regression"""
    ###############################
    # Compute the gradient update #
    ###############################
```

Free Textbooks! 😊

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

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

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

Natural Language Processing

Jacob Eisenstein

Not free: GPUs 😢

- Modern NLP methods require non-trivial computation
  - Training neural networks with many parameters can take a long time (it is a very good idea to start working on the assignments early!)
    - This is a big part of modern NLP methods. It is important to get experience training these networks.
  - You will need to use GPUs to complete the programming assignments.
  - Google Colab: has free GPUs, but with some big limitations that will make the assignments very difficult to complete.
  - The programming projects are designed with Colab in mind
  - Colab Pro subscription ($10/month). This is highly recommended once we start working with PyTorch.

Outline of the Course

- Machine Learning Review (Naive Bayes, Log. Reg. SVMs, Neural Nets)
- Sequence Models (HMMs, CRFs)
- Word Embeddings
- Neural Networks in NLP (NBOW, RNNs, CNNs, Transformers)
- Pre-trained Transformers (e.g. BERT, BART, T5, GPT)
- Machine Translation
- Dialogue
- Question Answering

Course Goals

Course Goals

- Cover fundamental machine learning techniques used in NLP

Course Goals

- Cover fundamental machine learning techniques used in NLP
    - Deeper understanding of algorithms beyond “how to use ML/NLP libraries”.

Course Goals

- Cover fundamental machine learning techniques used in NLP
  - Deeper understanding of algorithms beyond “how to use ML/NLP libraries”.
- Know about modern NLP methods: what is the state-of-the-art in 2024?

Course Goals

- Cover fundamental machine learning techniques used in NLP
    - Deeper understanding of algorithms beyond “how to use ML/NLP libraries”.
- Know about modern NLP methods: what is the state-of-the-art in 2024?
- Make you a “producer” rather than a “consumer” of NLP tools

Programming Assignments

- 3 Programming Assignments
  - Implementation-oriented
  - ~2 weeks per assignment

Programming Assignments

- 3 Programming Assignments
  - Implementation-oriented
  - ~2 weeks per assignment

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!**

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?

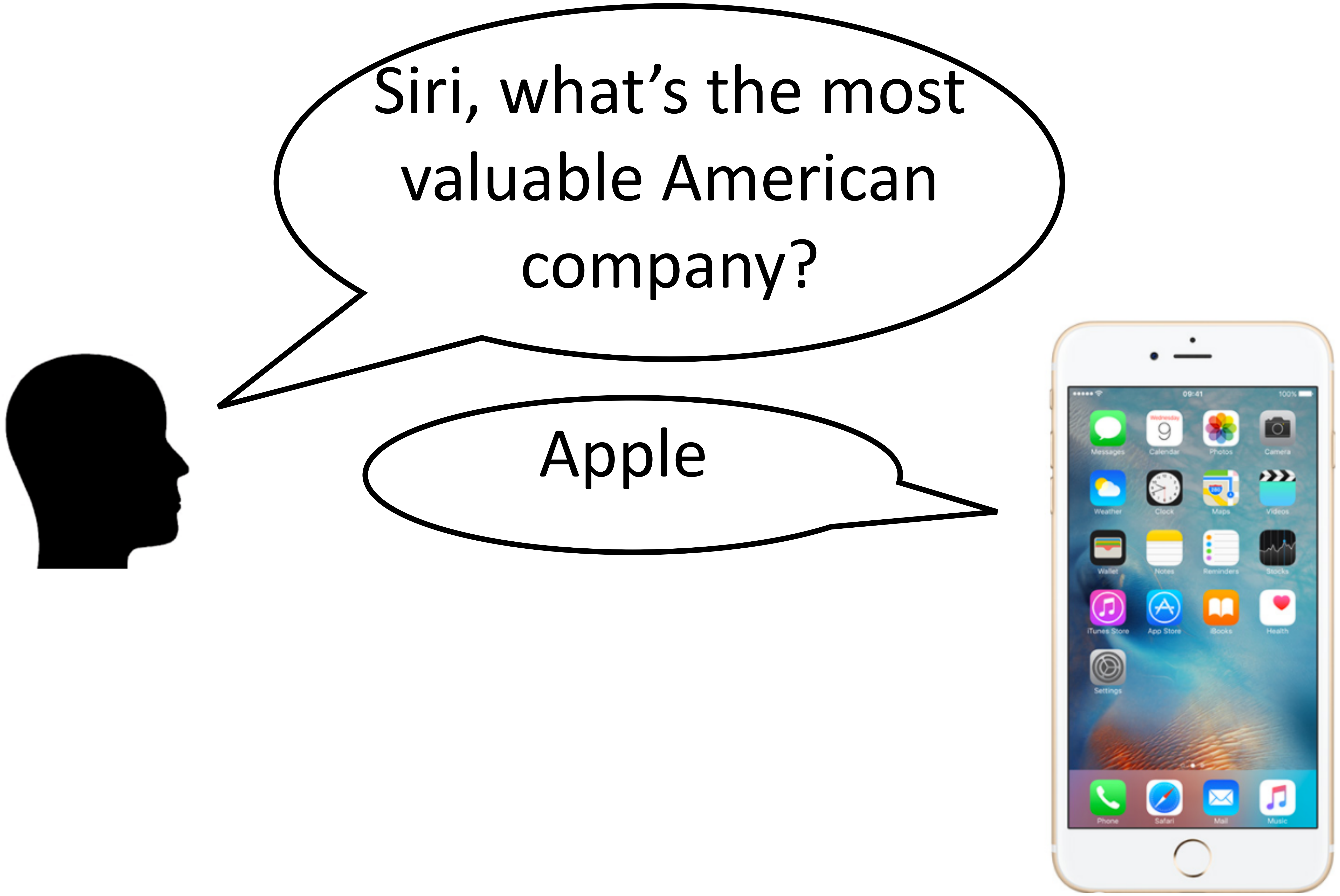
- Be able to solve problems that require deep understanding of text
- Example: dialogue systems

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

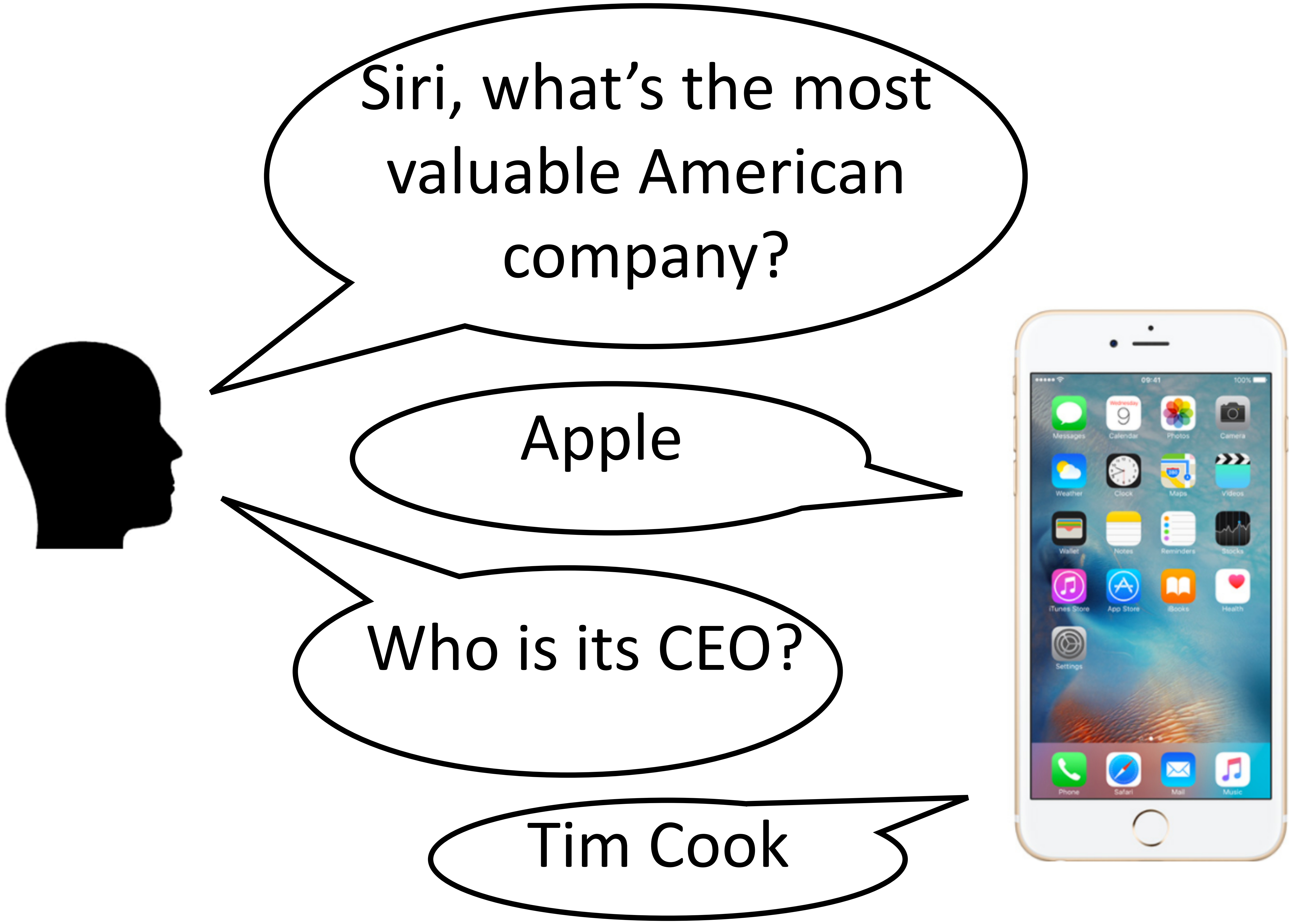


What's the goal of NLP?

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

What's the goal of NLP?

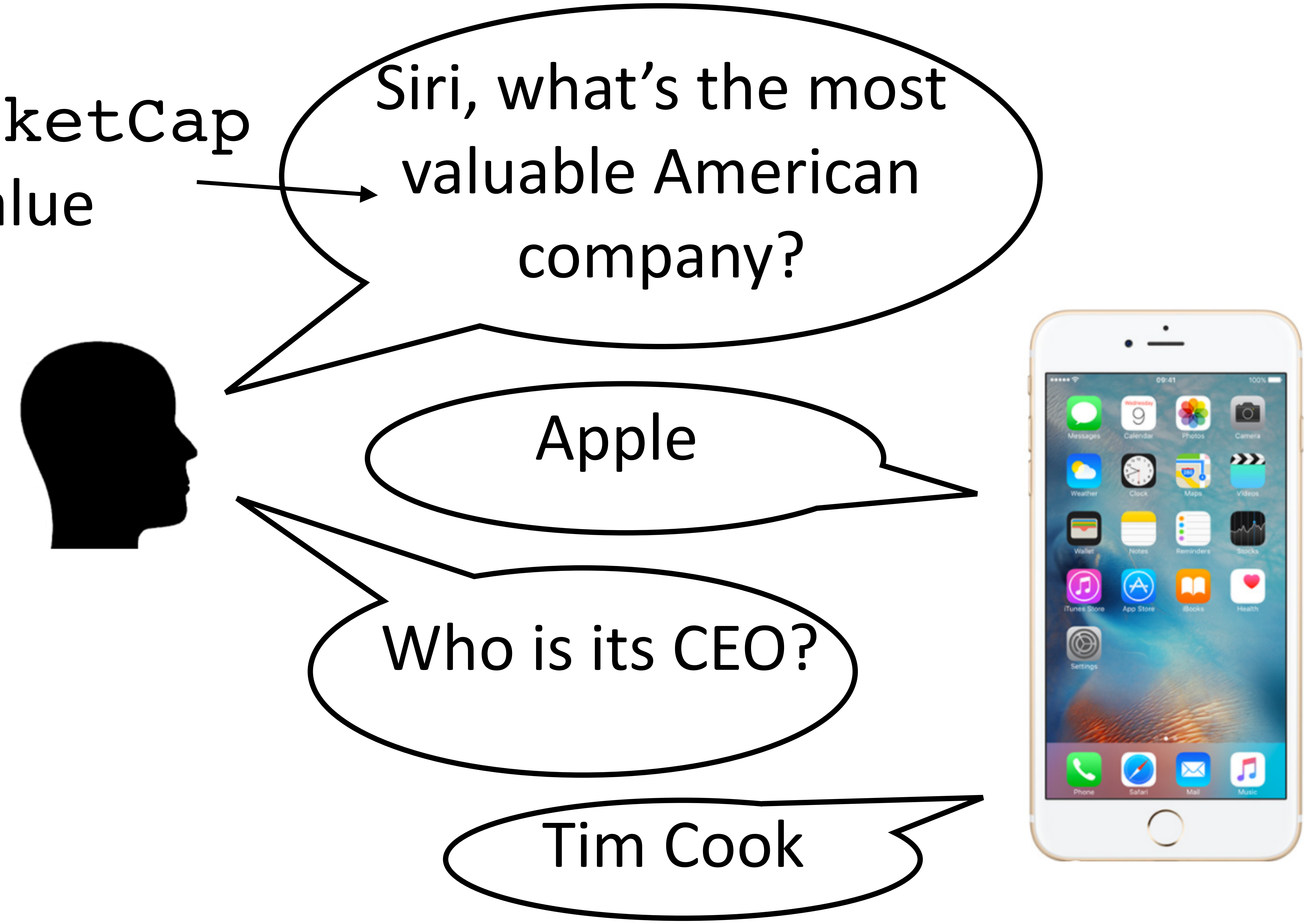
- Be able to solve problems that require deep understanding of text
- Example: dialogue systems



What’s the goal of NLP?

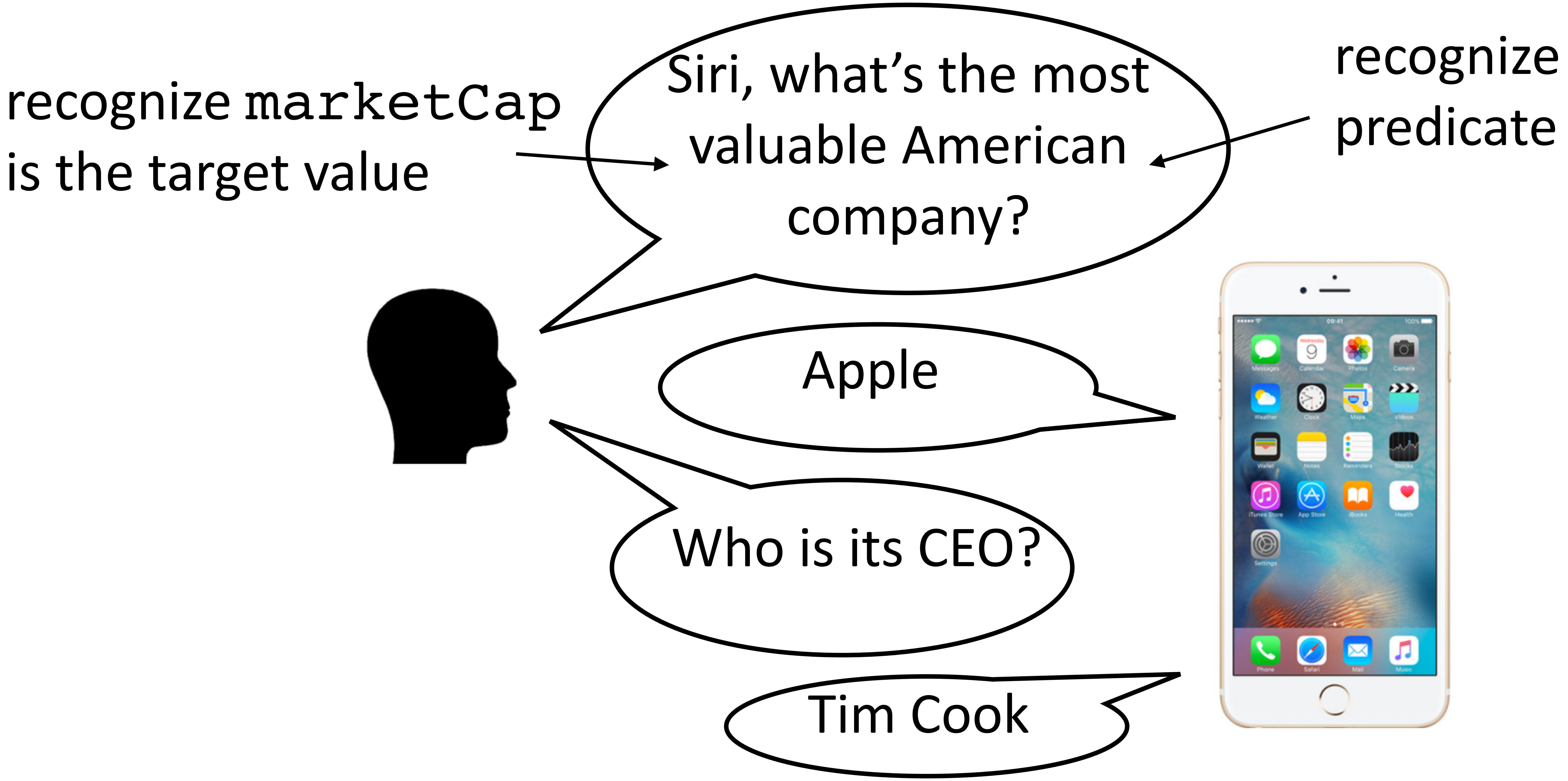
- Be able to solve problems that require deep understanding of text
- Example: dialogue systems

recognize `marketCap` is the target value



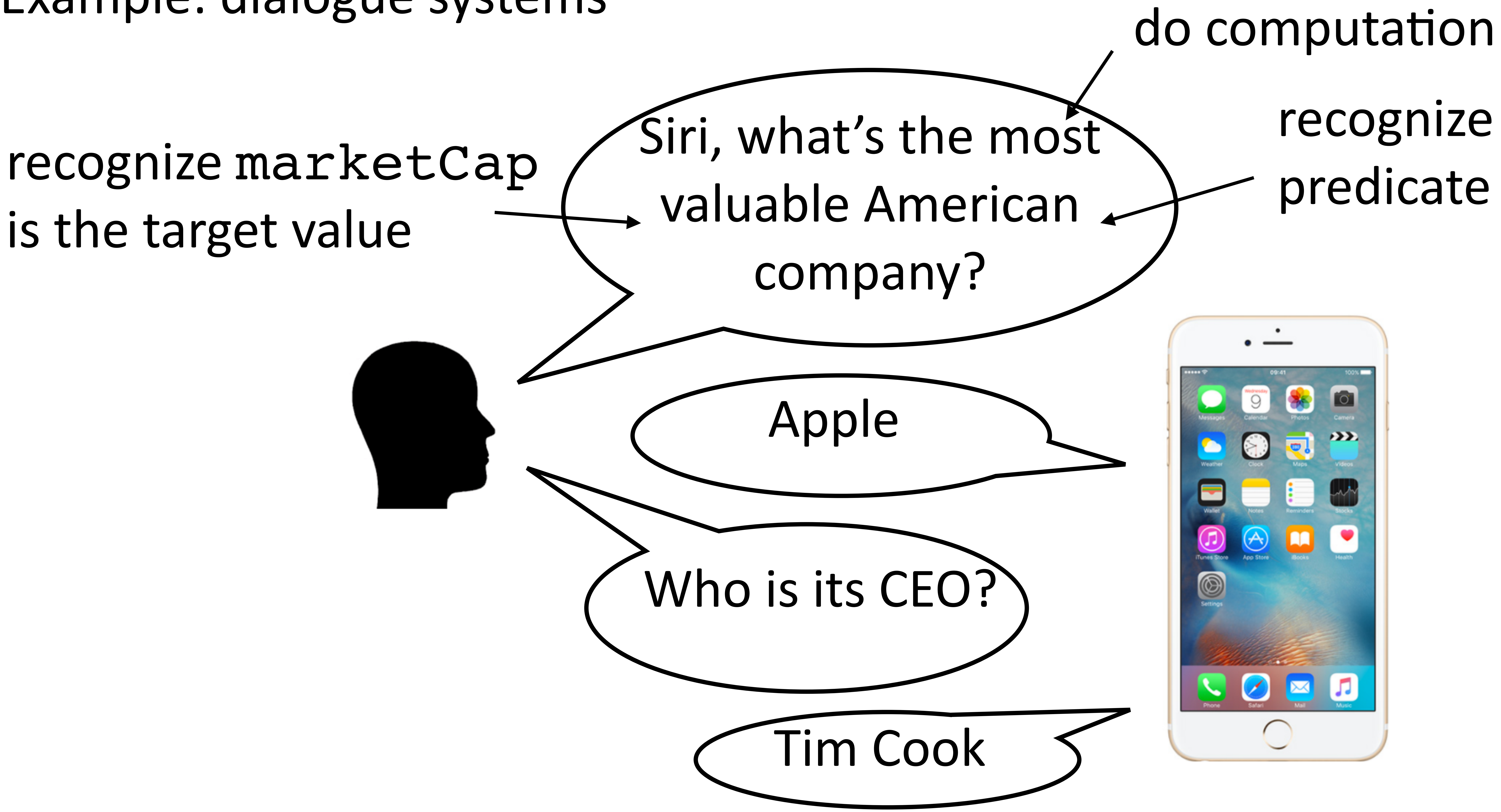
What's the goal of NLP?

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



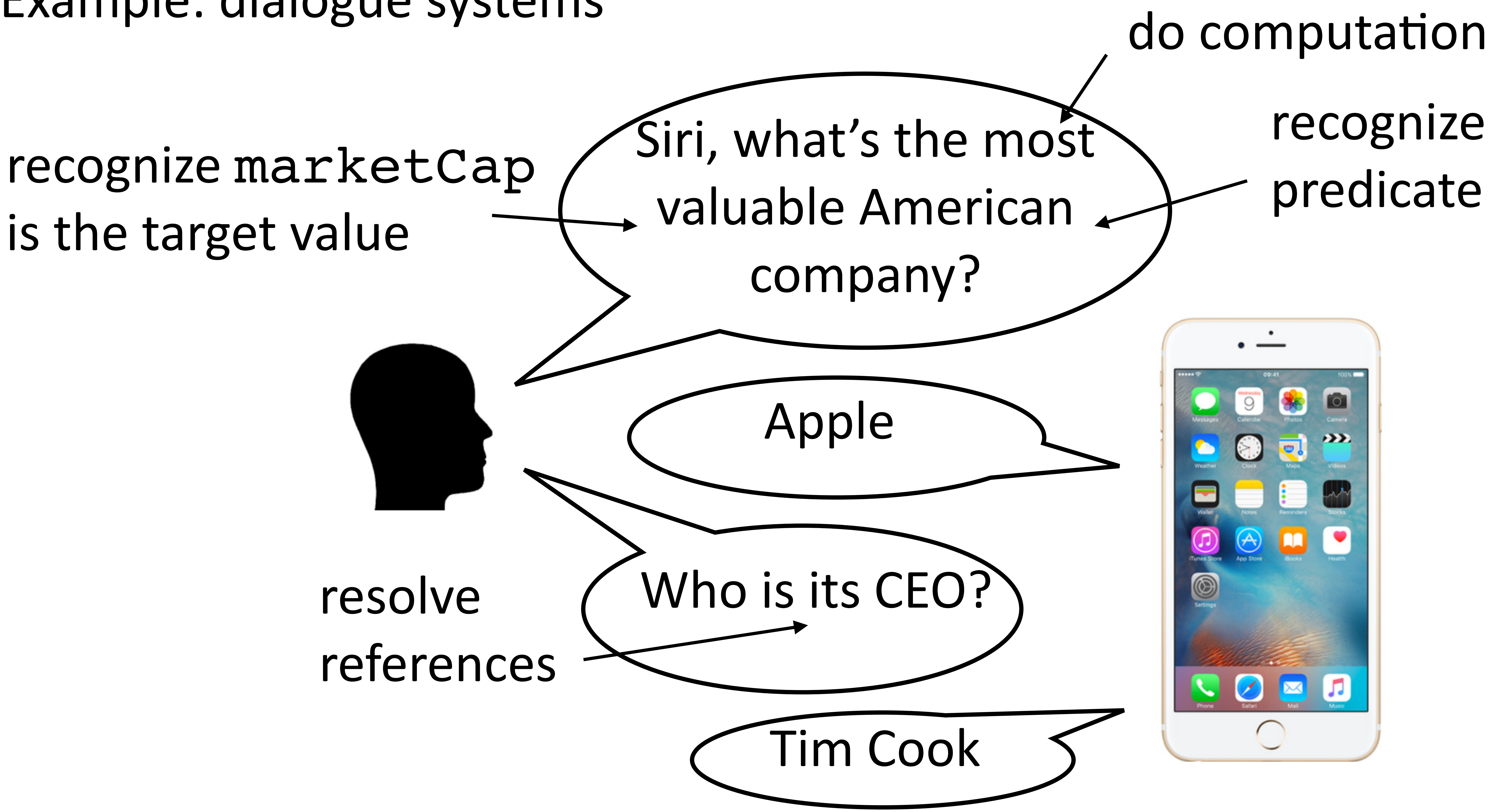
What's the goal of NLP?

- 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

THE WALL STREET JOURNAL.

美国众议院议长选举大戏落幕，共和党议员重点转向支出及中国问题

7 小时前

Machine Translation

THE WALL STREET JOURNAL.

美国众议院议长选举大戏落幕，共和党议员重点转向支出及中国问题

7 小时前

CHINESE (SIMPLIFIED) - DETECTED | CHINESE (SIMPLIFIED) | HINDI | FRE | ENGLISH | SPANISH | ARABIC

美国众议院议长选举大戏落幕，共和党议员重点转向支出及中国问题

U.S. House speaker race ends as Republican lawmakers focus on spending, China

Machine Translation

THE WALL STREET JOURNAL.

美国众议院议长选举大戏落幕，共和党议员重点转向支出及中国问题

7 小时前

- Working very well now for high-resource languages.
- Some language pairs more difficult (e.g. English-Chinese)
- Still a number of challenges (scaling up to thousands of languages, etc.)

CHINESE (SIMPLIFIED) - DETECTED | CHINESE (SIMPLIFIED) | HINDI | FRE | ENGLISH | SPANISH | ARABIC

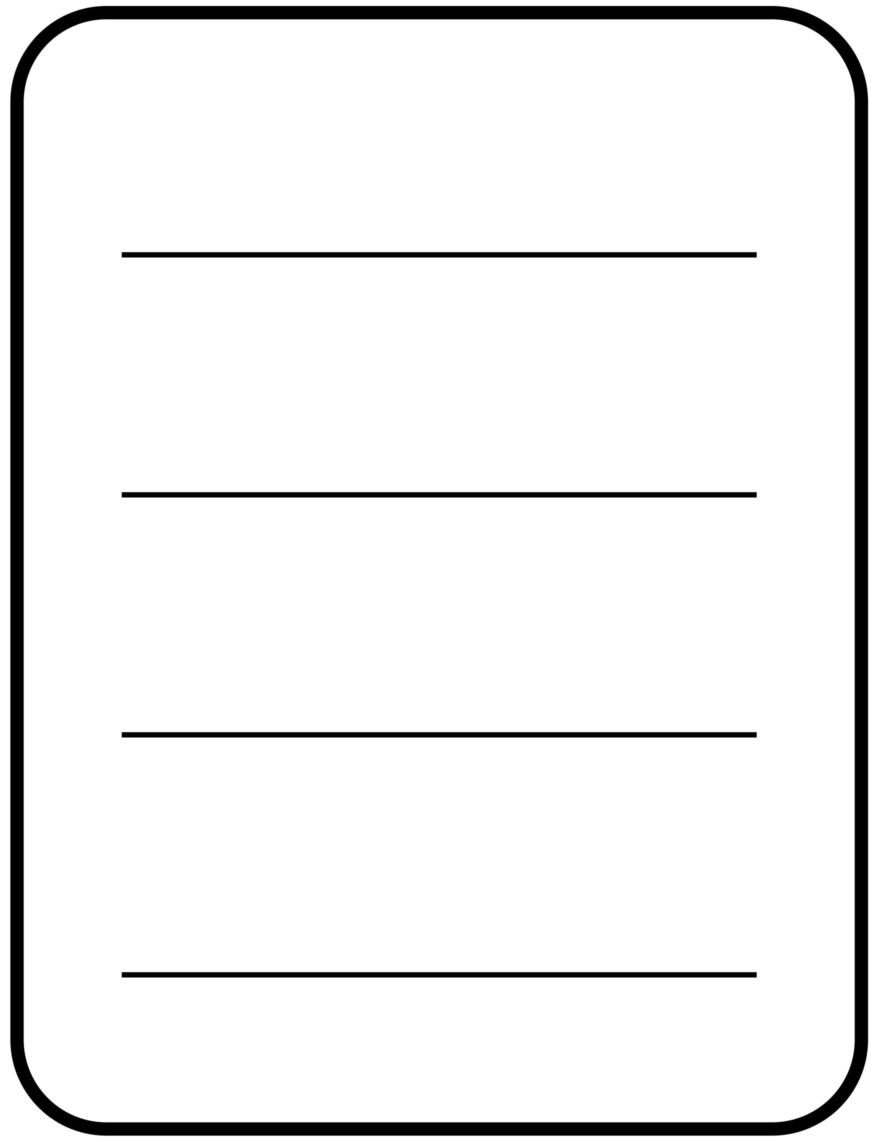
美国众议院议长选举大戏落幕，共和党议员重点转向支出及中国问题

U.S. House speaker race ends as Republican lawmakers focus on spending, China

Traditional NLP Analysis Pipeline

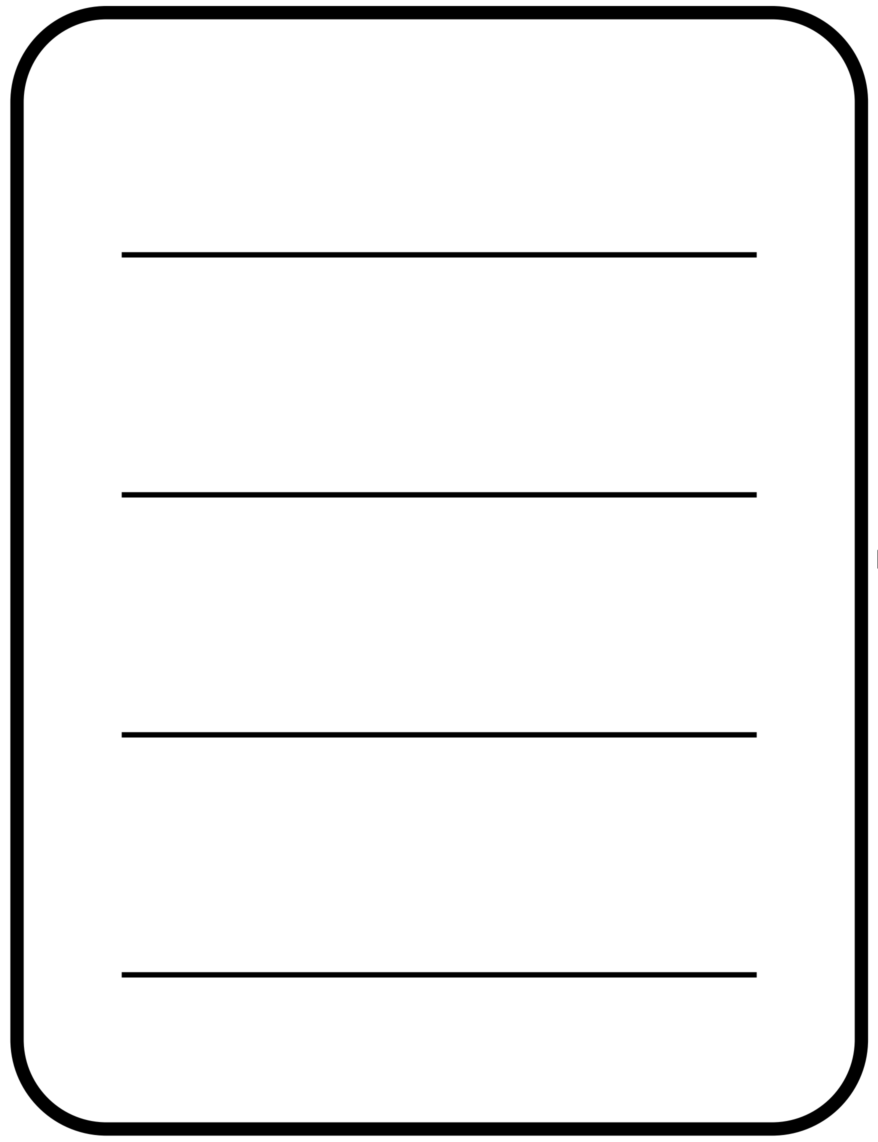
Traditional NLP Analysis Pipeline

Text



Traditional NLP Analysis Pipeline

Text



Text Analysis

Syntactic parses

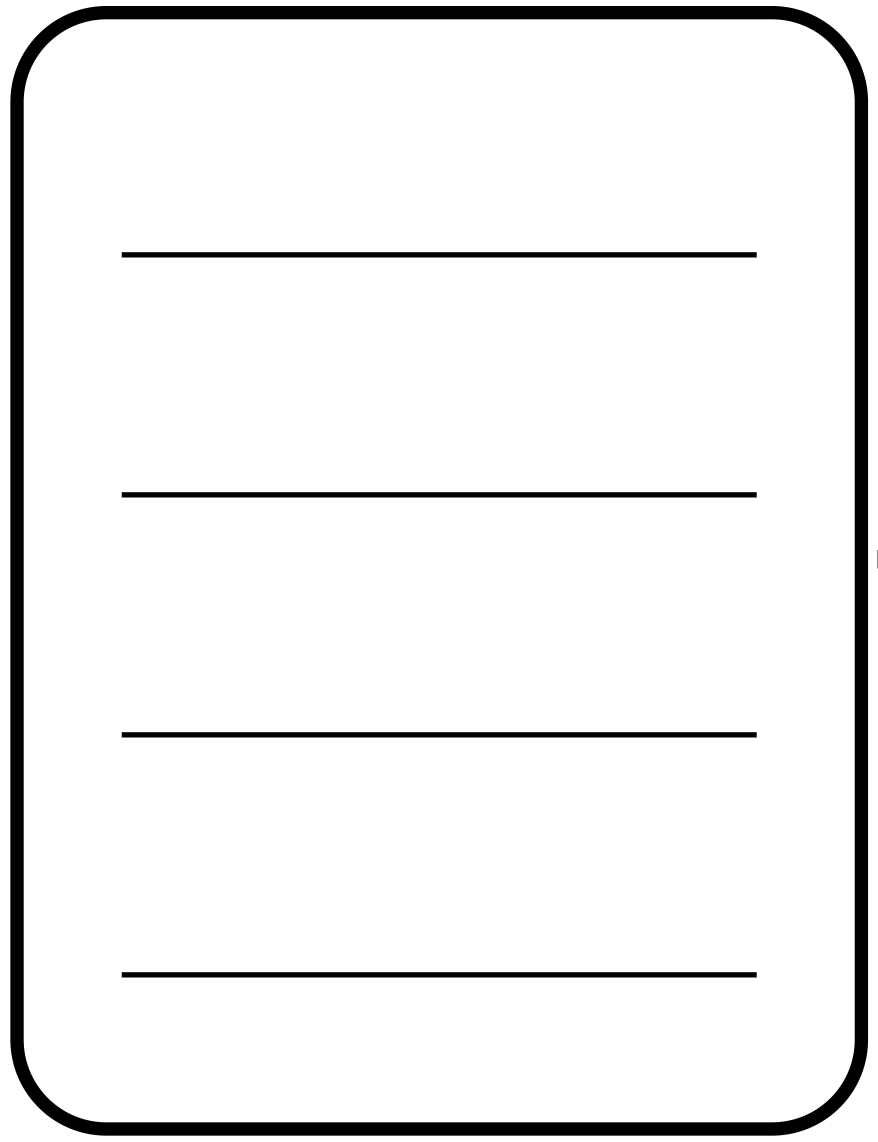
Coreference resolution

Entity disambiguation

Discourse analysis

Traditional NLP Analysis Pipeline

Text



Text Analysis

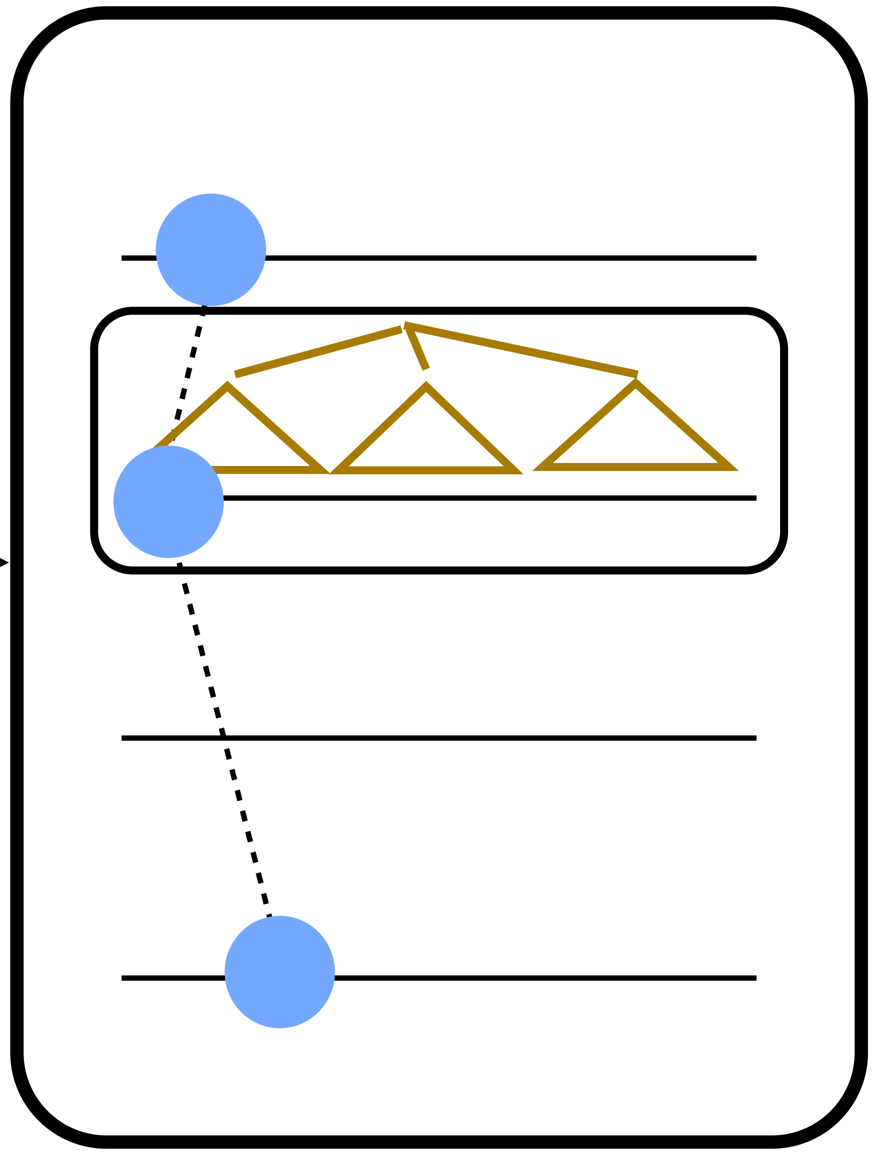
Syntactic parses

Coreference resolution

Entity disambiguation

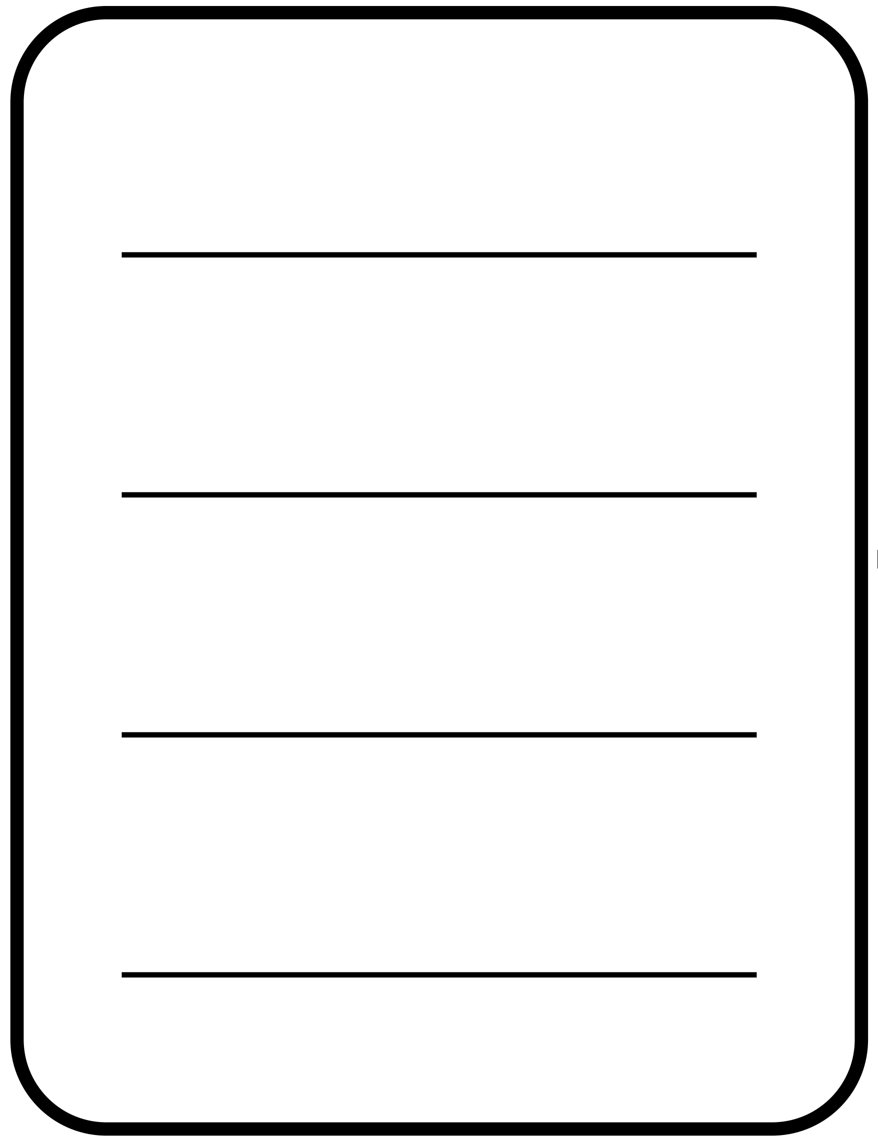
Discourse analysis

Annotations



Traditional NLP Analysis Pipeline

Text



Text Analysis

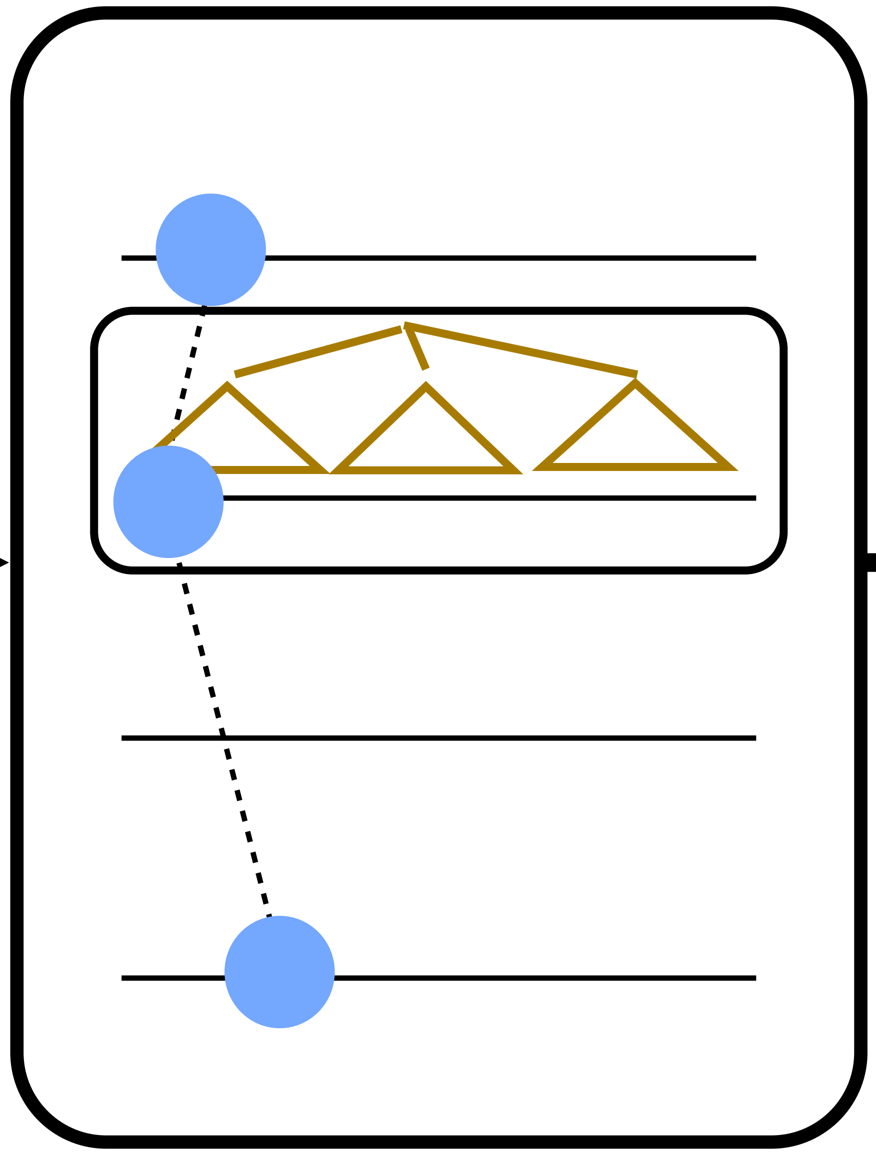
Syntactic parses

Coreference resolution

Entity disambiguation

Discourse analysis

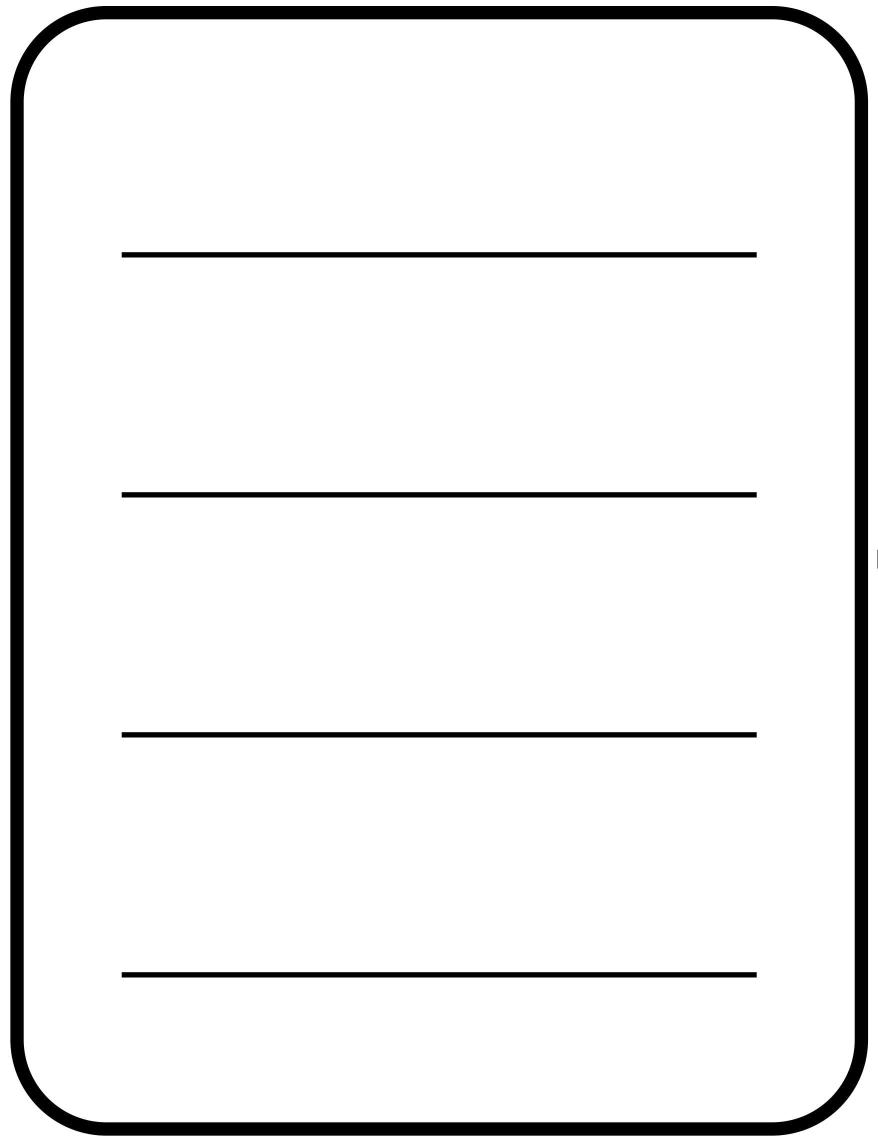
Annotations



Applications

Traditional NLP Analysis Pipeline

Text



Text Analysis

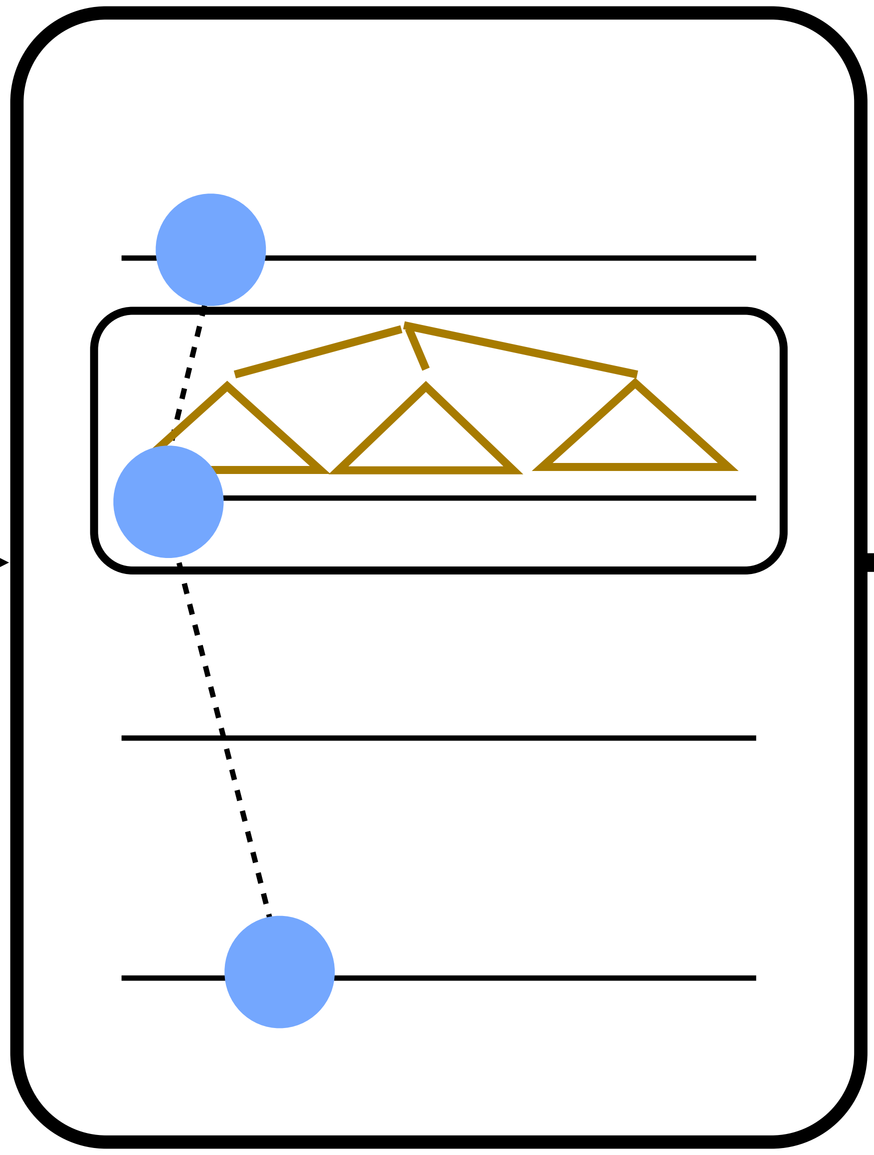
Syntactic parses

Coreference resolution

Entity disambiguation

Discourse analysis

Annotations

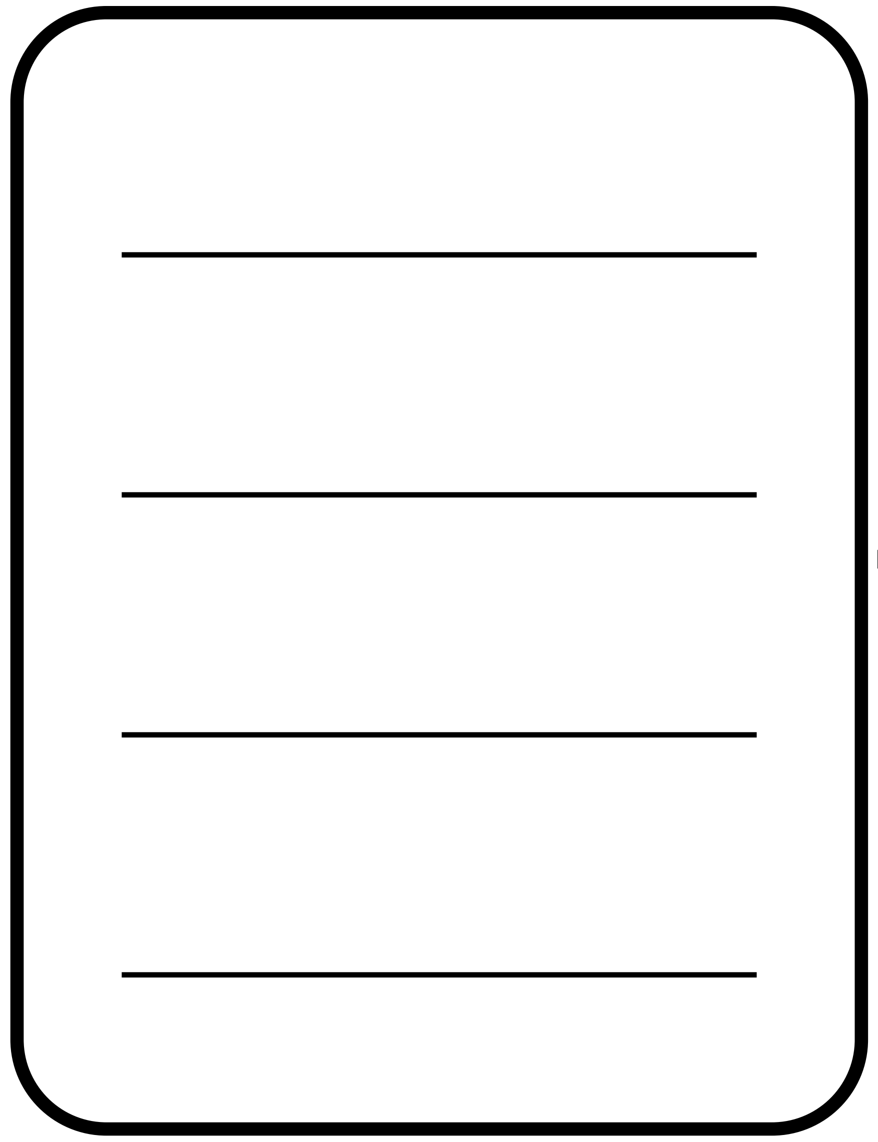


Applications

Summarize

Traditional NLP Analysis Pipeline

Text



Text Analysis

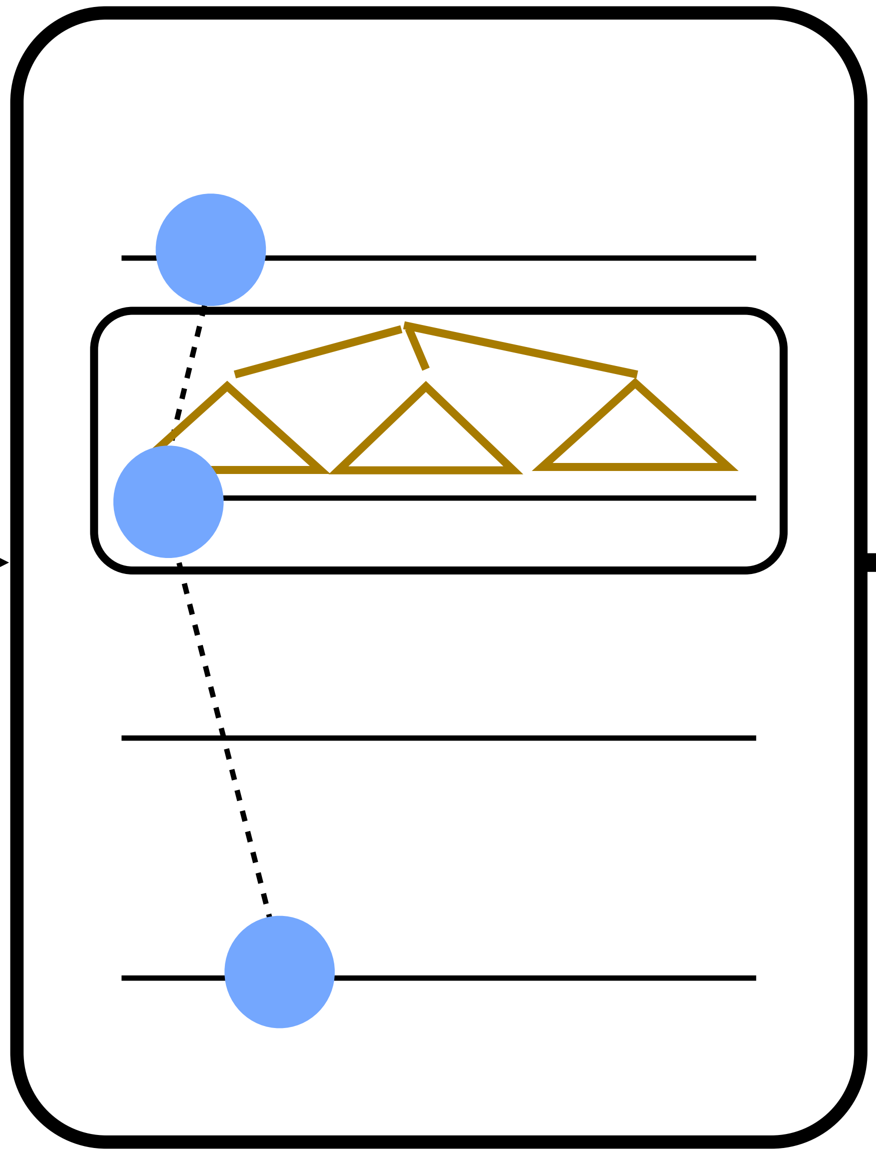
Syntactic parses

Coreference resolution

Entity disambiguation

Discourse analysis

Annotations



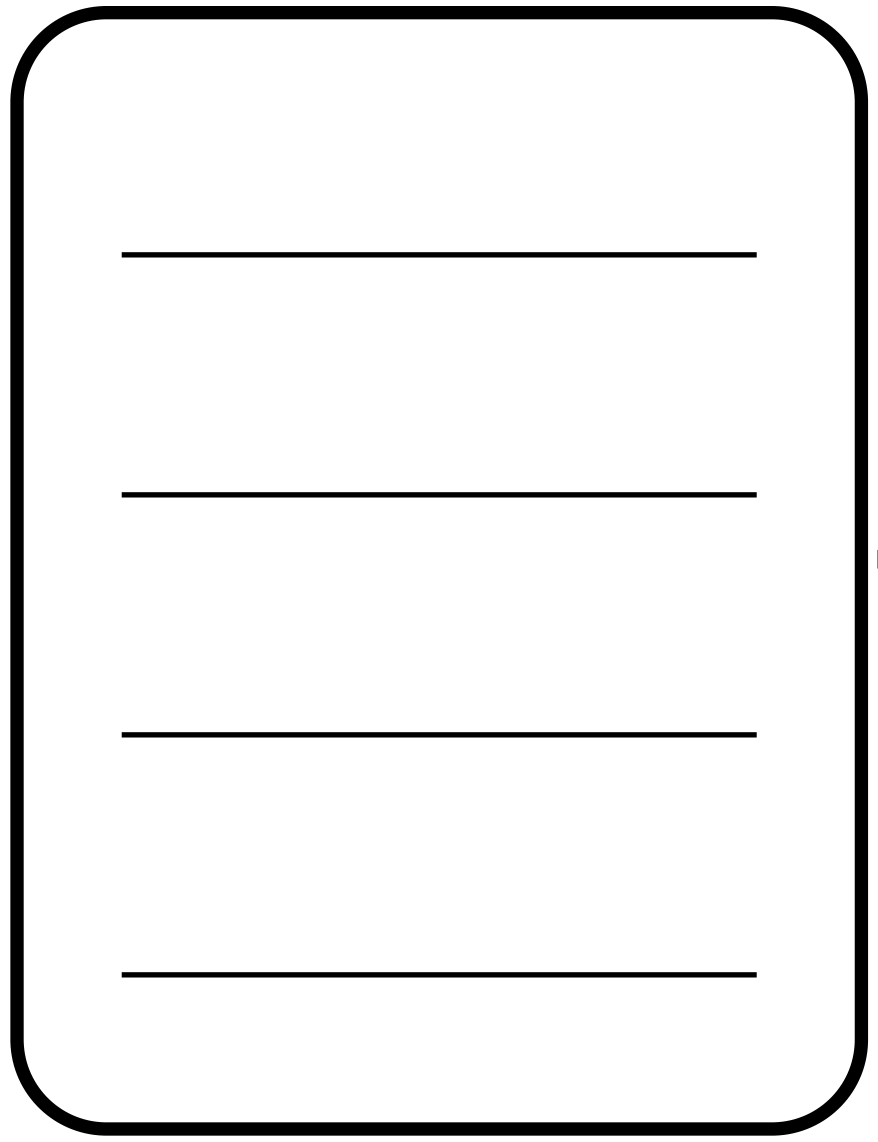
Applications

Summarize

Extract information

Traditional NLP Analysis Pipeline

Text



Text Analysis

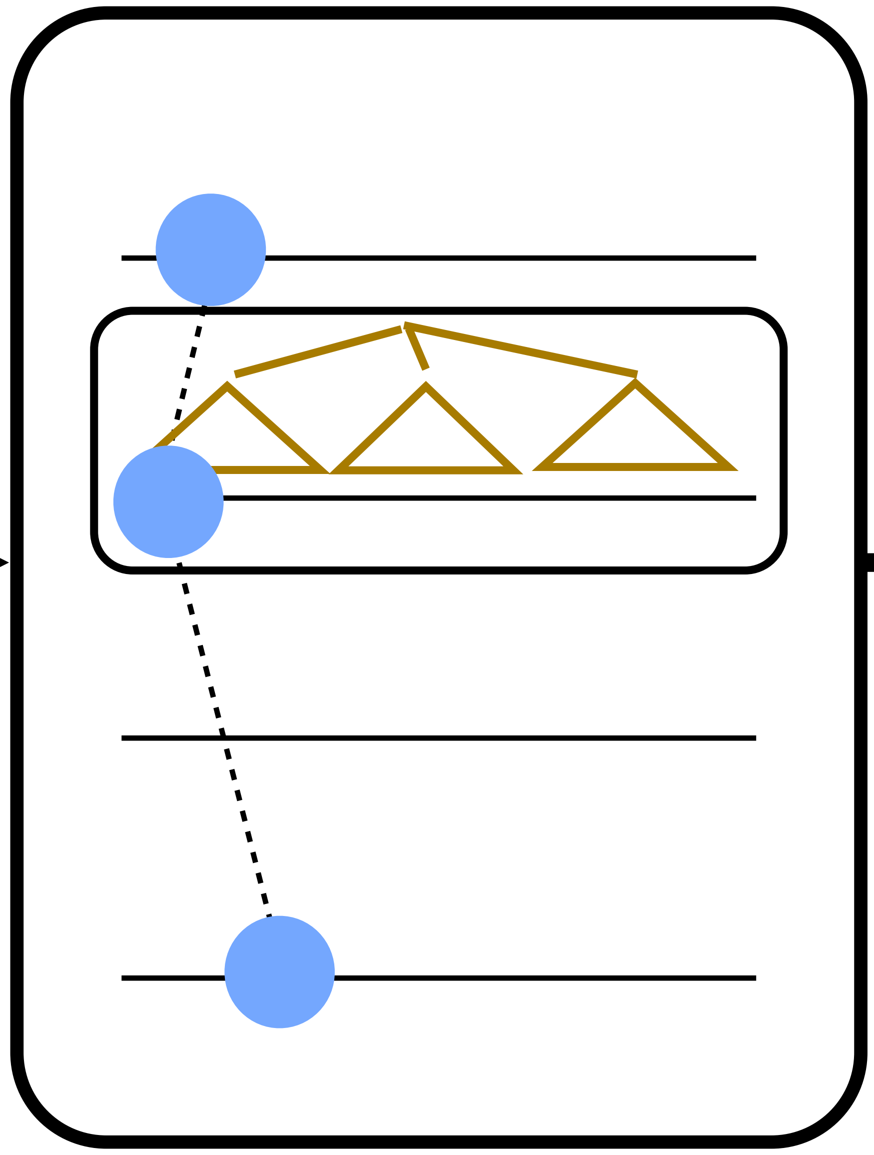
Syntactic parses

Coreference resolution

Entity disambiguation

Discourse analysis

Annotations



Applications

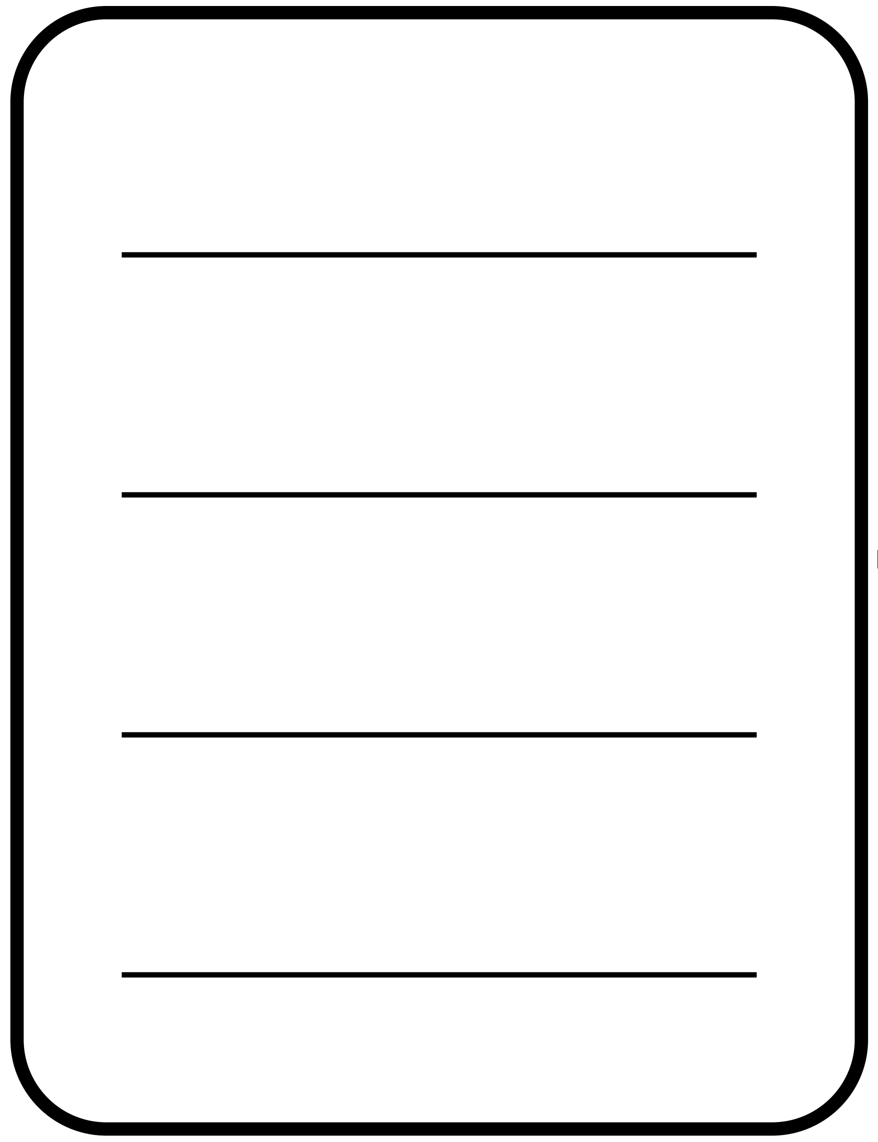
Summarize

Extract information

Answer questions

Traditional NLP Analysis Pipeline

Text



Text Analysis

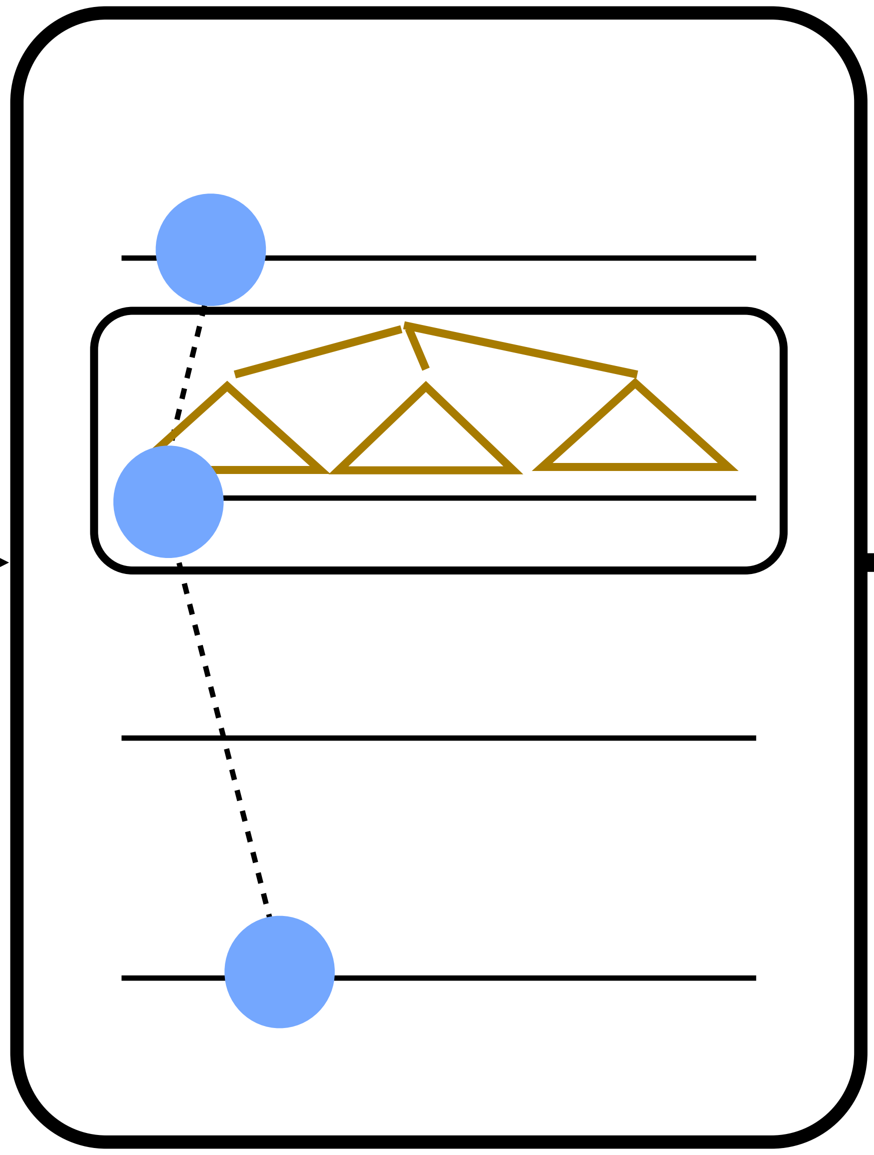
Syntactic parses

Coreference resolution

Entity disambiguation

Discourse analysis

Annotations



Applications

Summarize

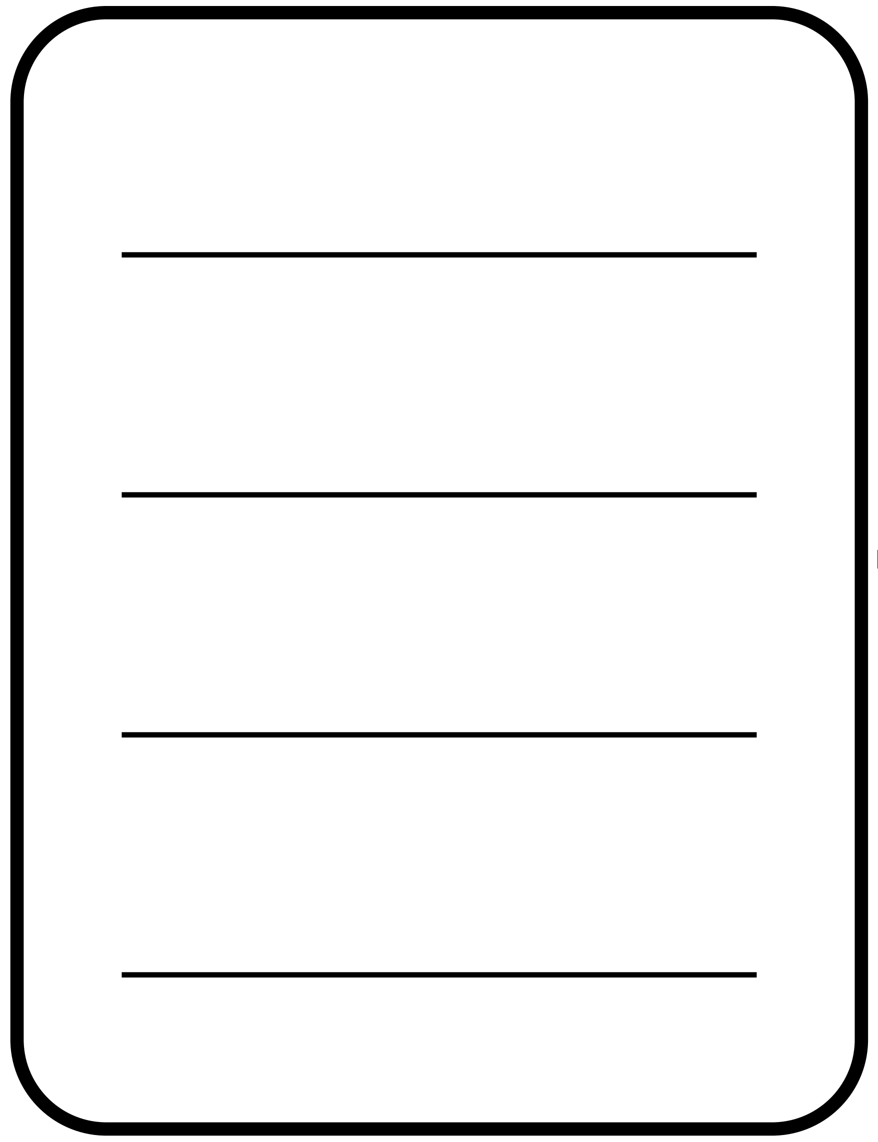
Extract information

Answer questions

Identify sentiment

Traditional NLP Analysis Pipeline

Text



Text Analysis

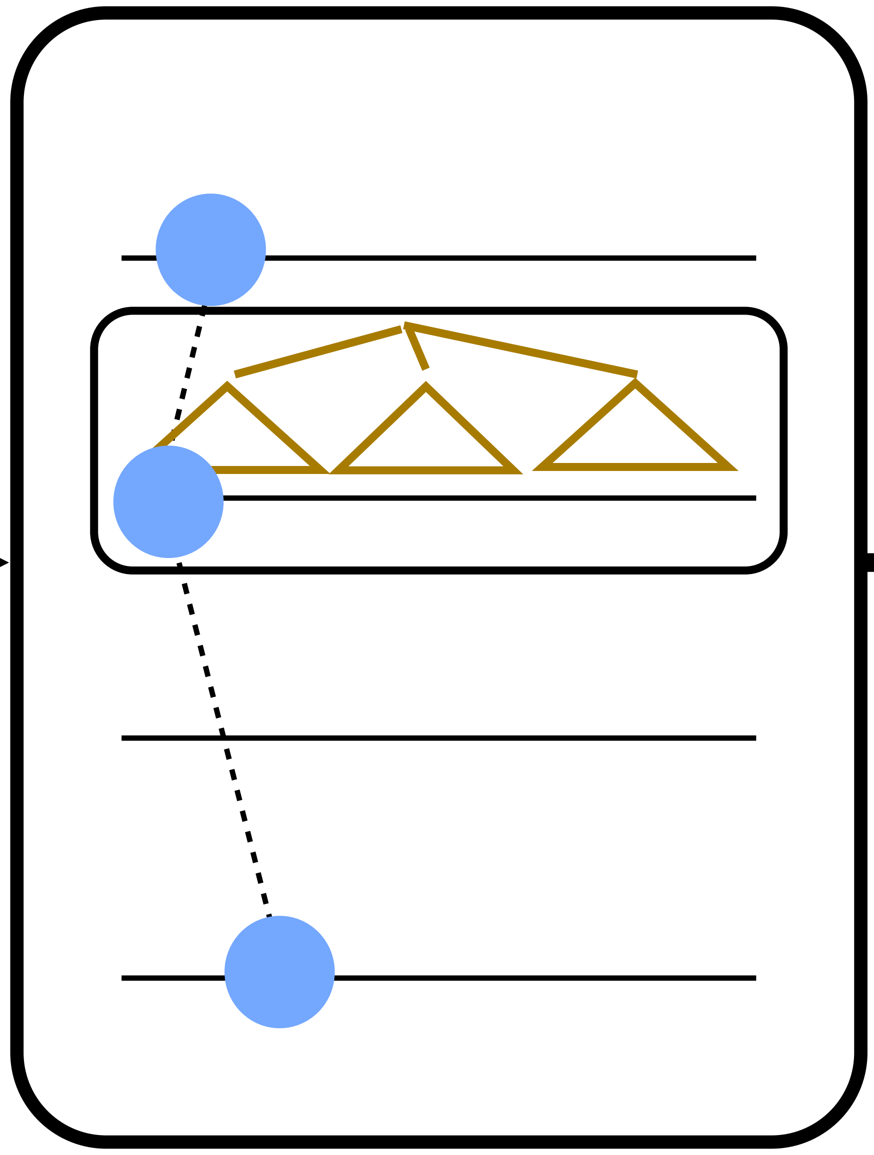
Syntactic parses

Coreference resolution

Entity disambiguation

Discourse analysis

Annotations



Applications

Summarize

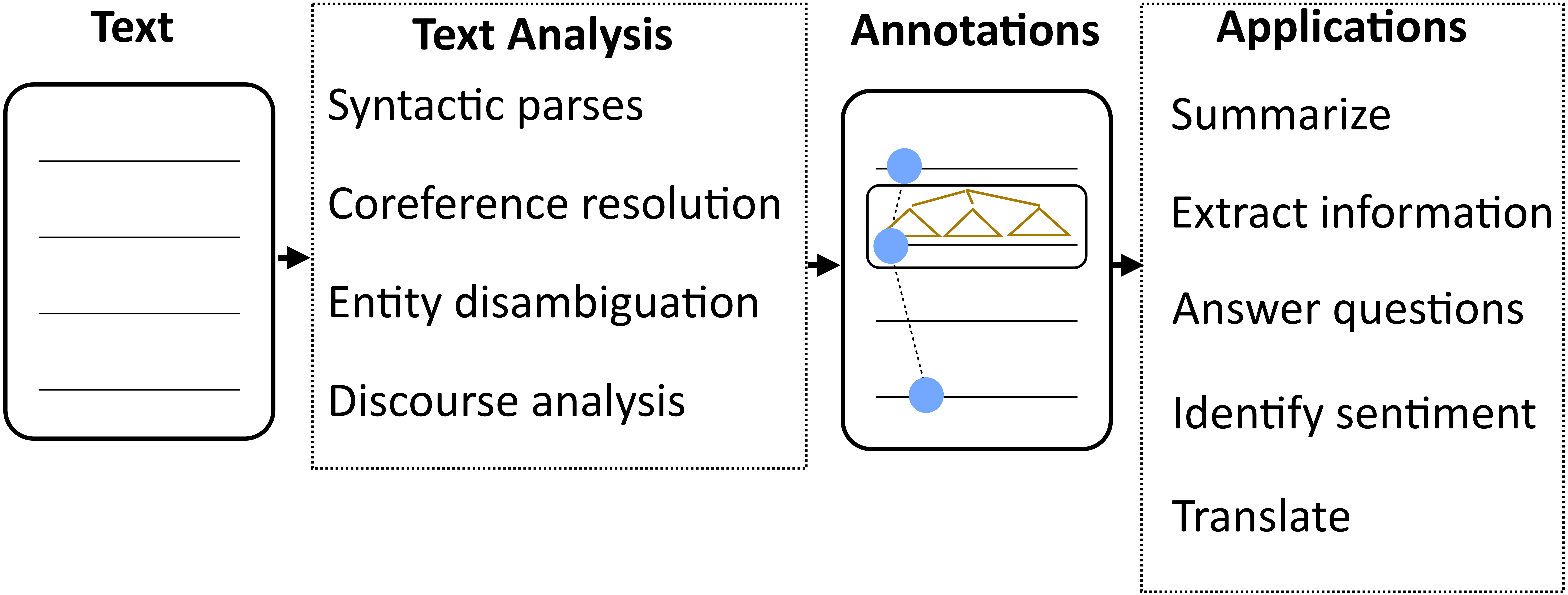
Extract information

Answer questions

Identify sentiment

Translate

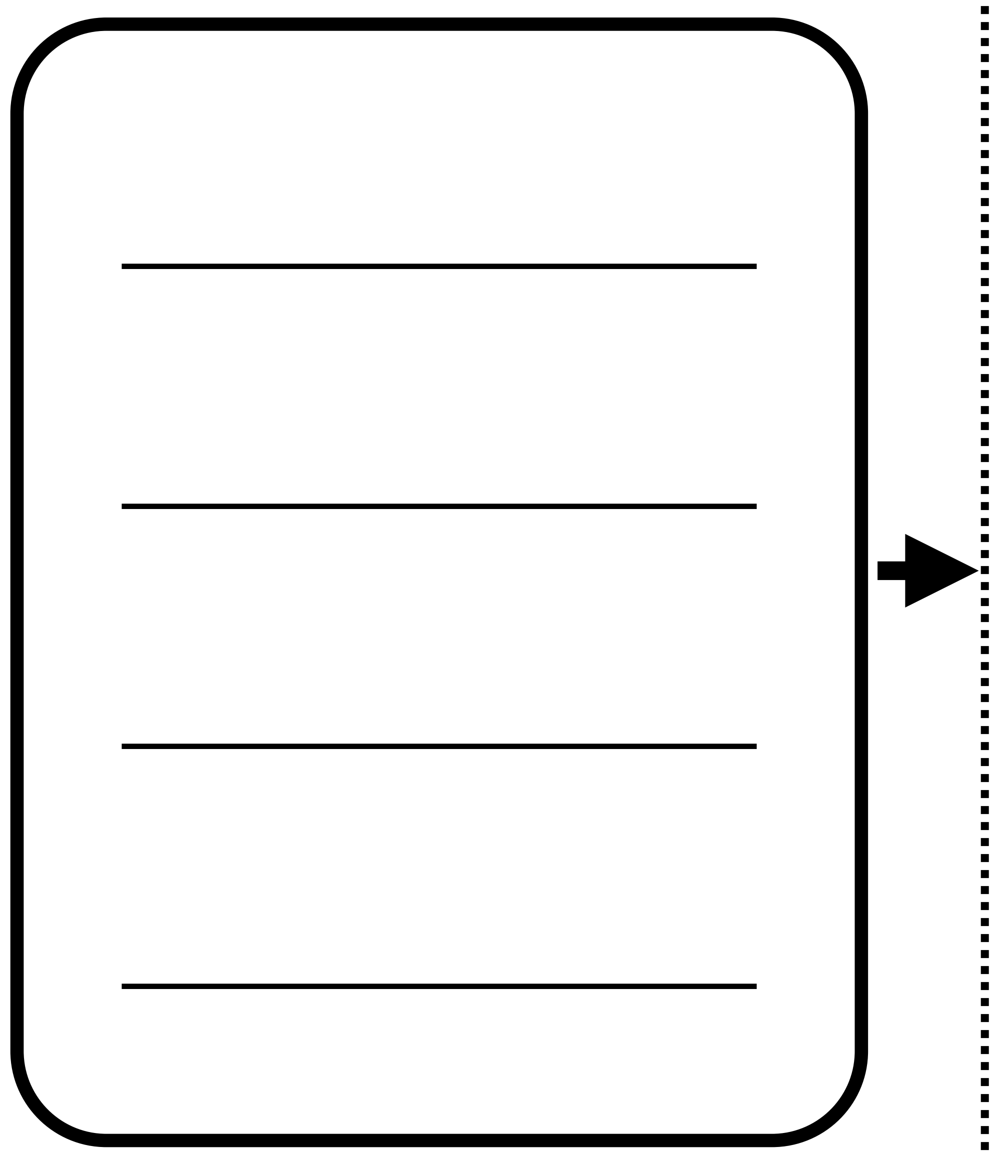
Traditional NLP Analysis Pipeline



- All of these components were modeled with statistical approaches trained with machine learning

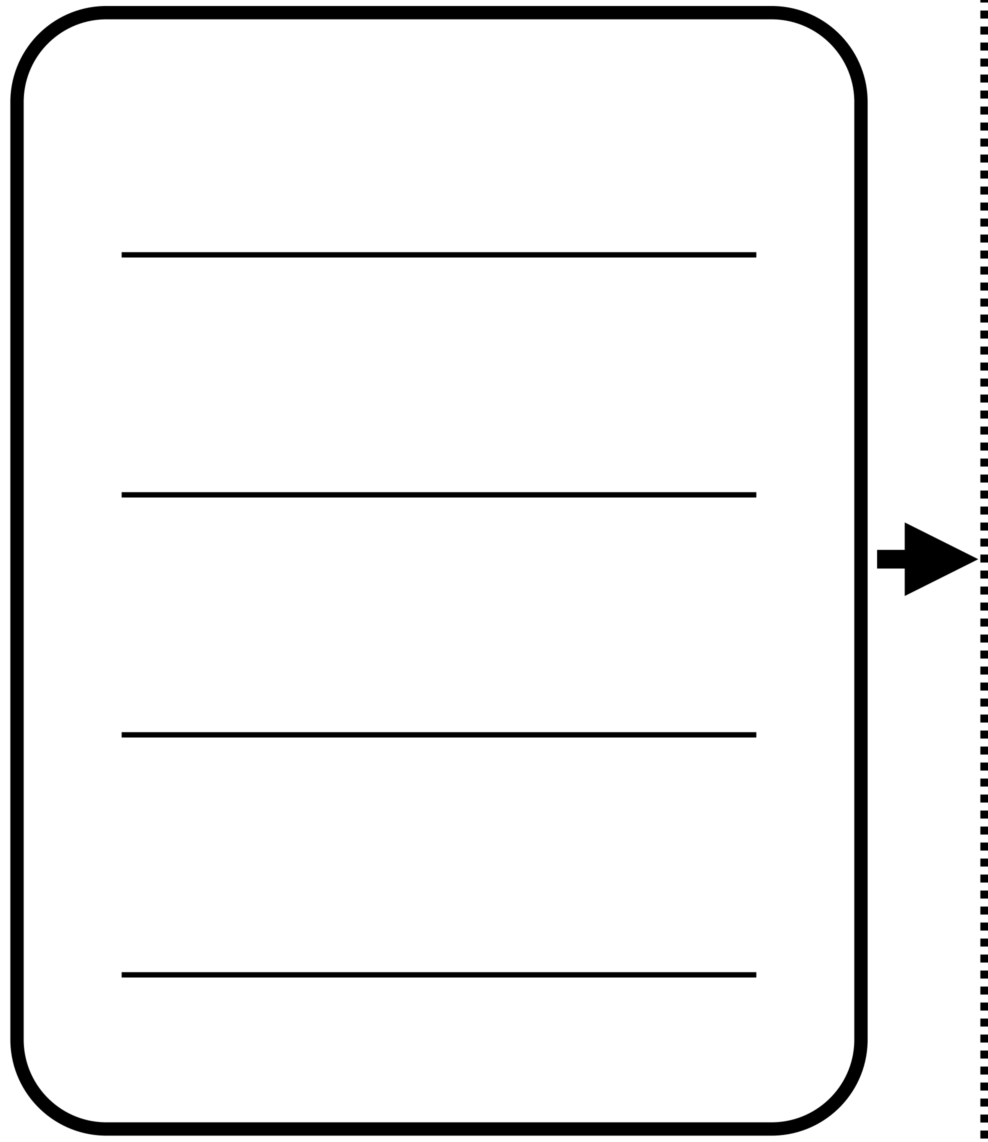
How do we represent language?

Text



How do we represent language?

Text



Labels

How do we represent language?

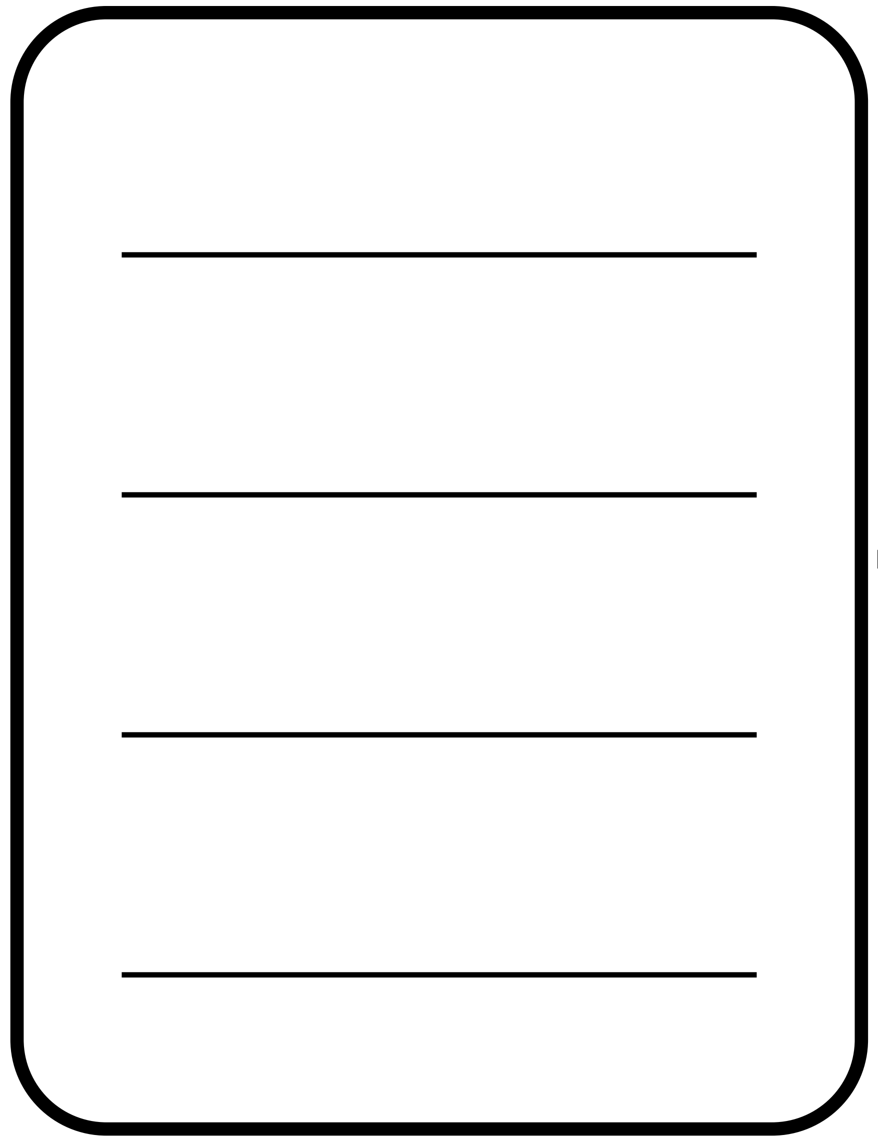
Text

Labels

the movie was good +

How do we represent language?

Text



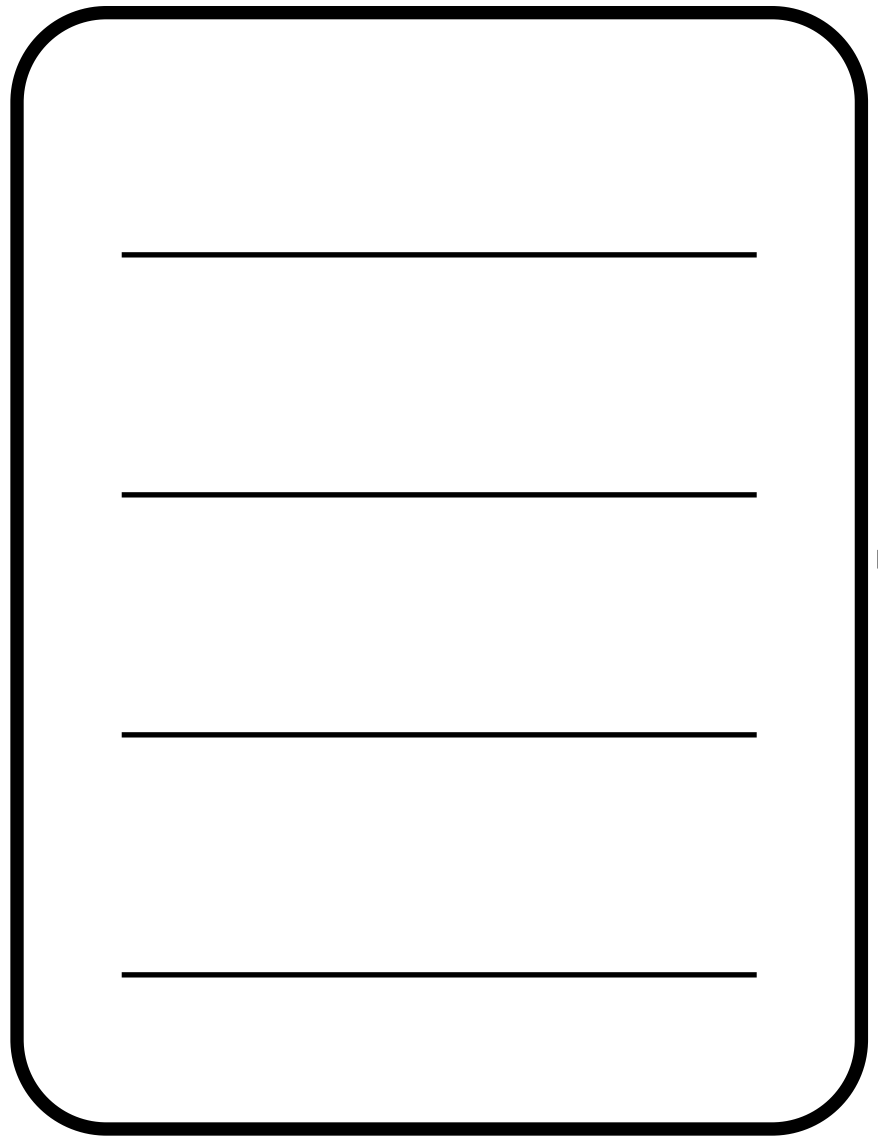
Labels

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 **MOVIE**

Tom Cruise stars in the new Mission Impossible film

How do we represent language?

Text

Labels

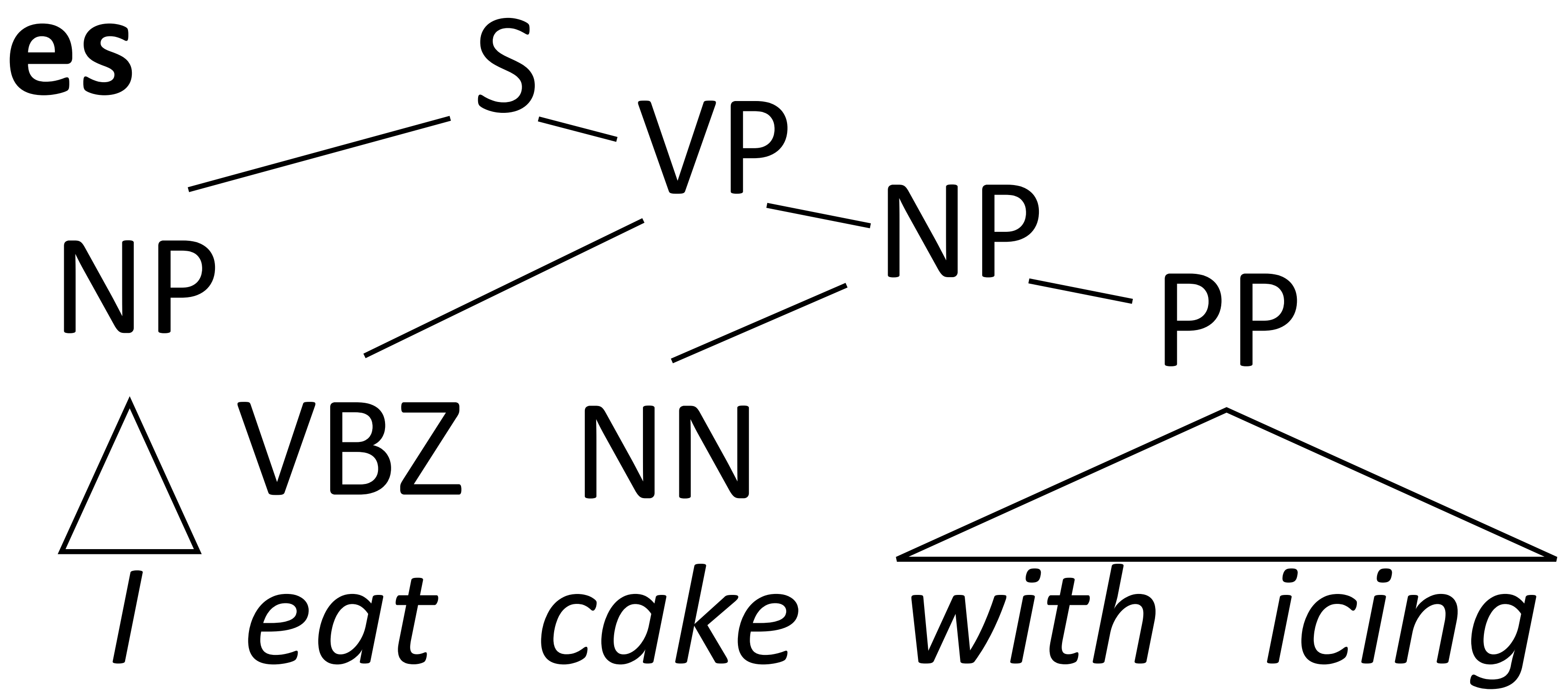
the movie was good +

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

Sequences/tags

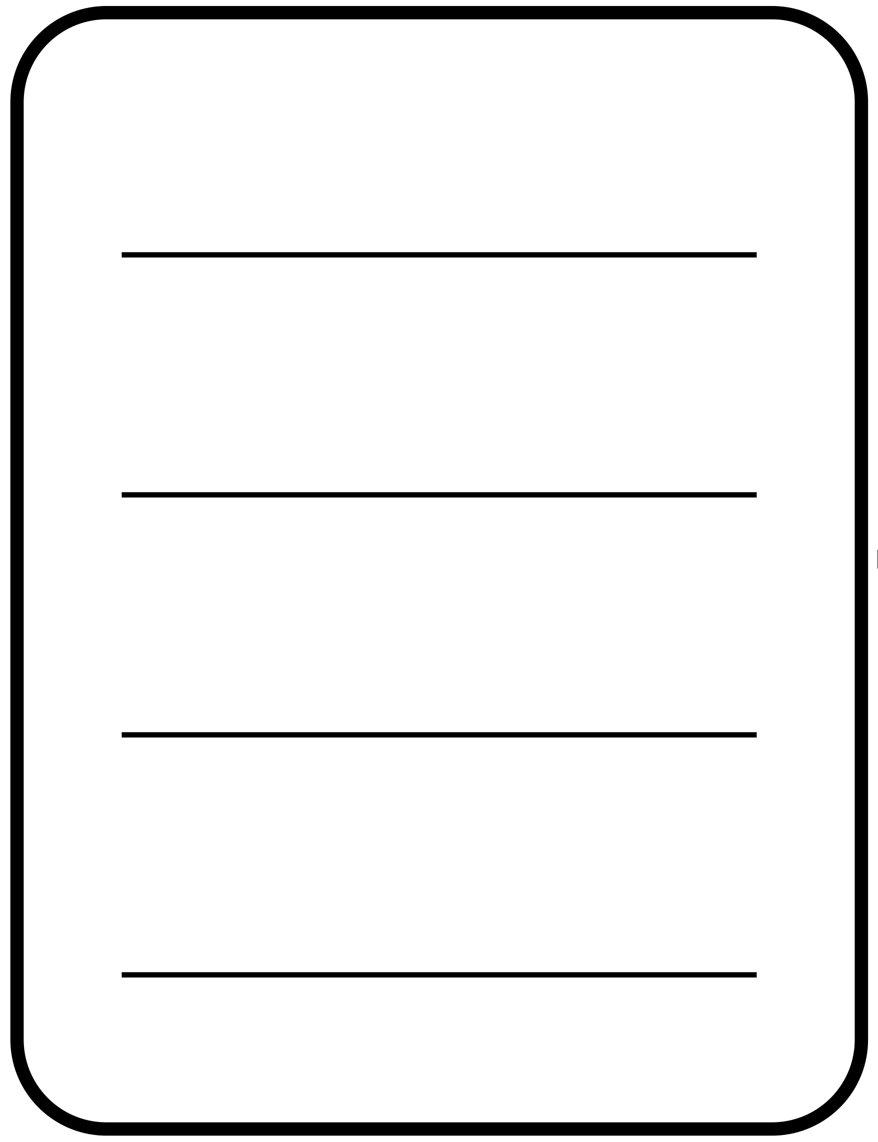
PERSON *Tom Cruise* *stars in the new* **MOVIE** *Mission Impossible* *film*

Trees



How do we represent language?

Text



Labels

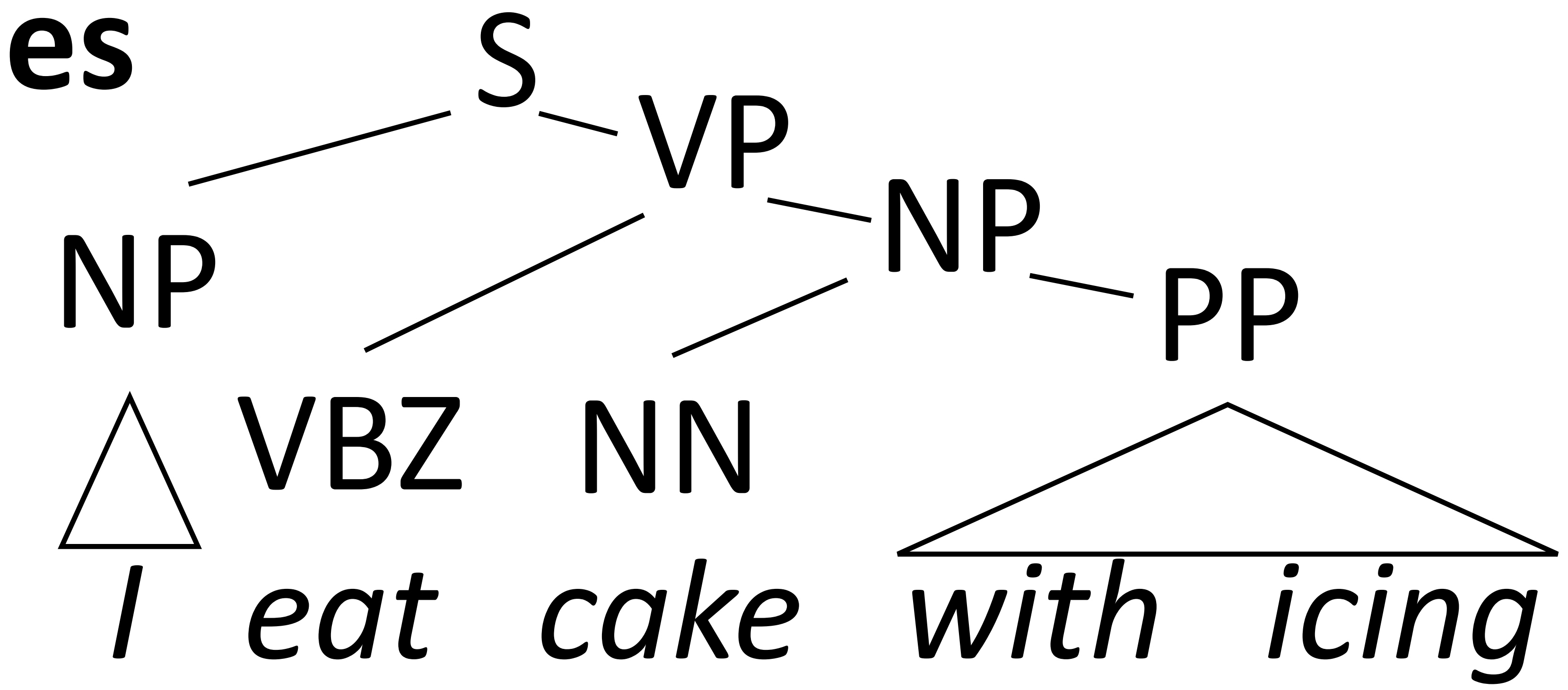
the movie was good +

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

Sequences/tags

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

Trees

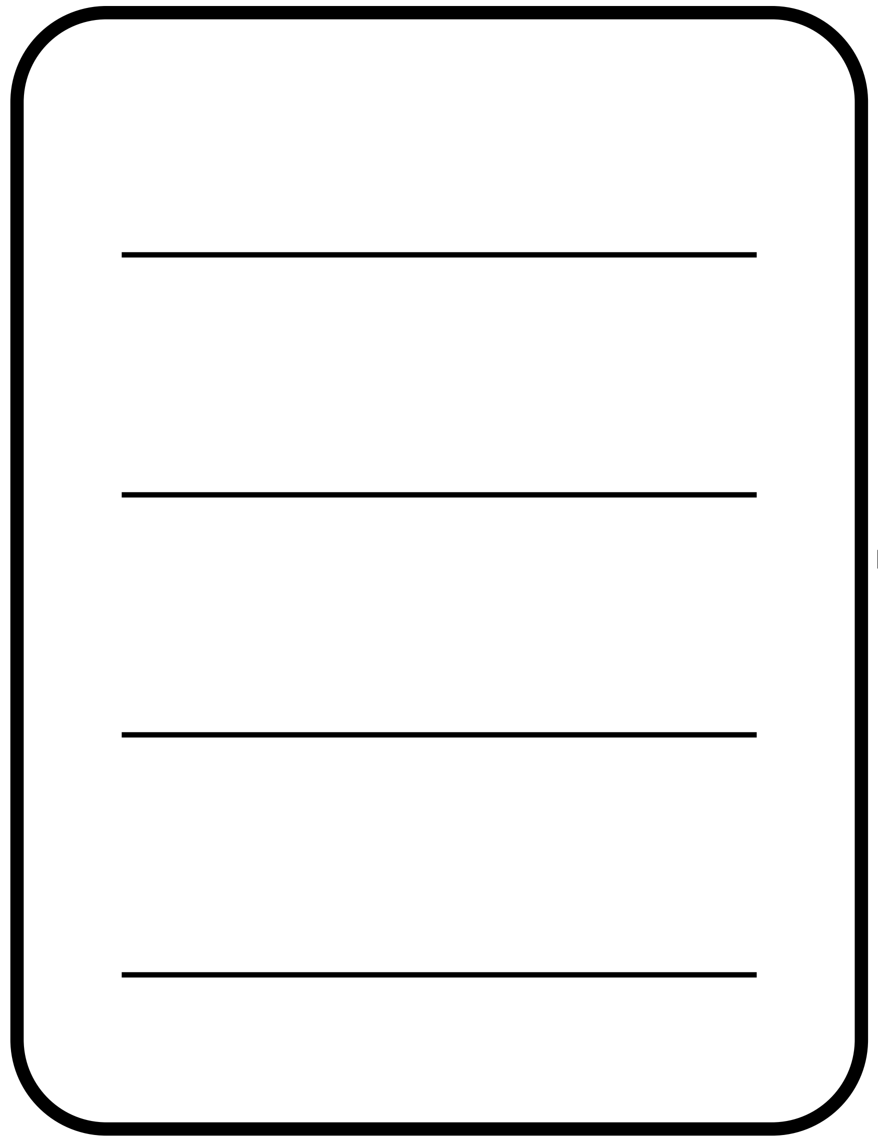


λx. flight(x) ∧ dest(x)=Miami

flights to Miami

How do we use these representations?

Text



Text Analysis

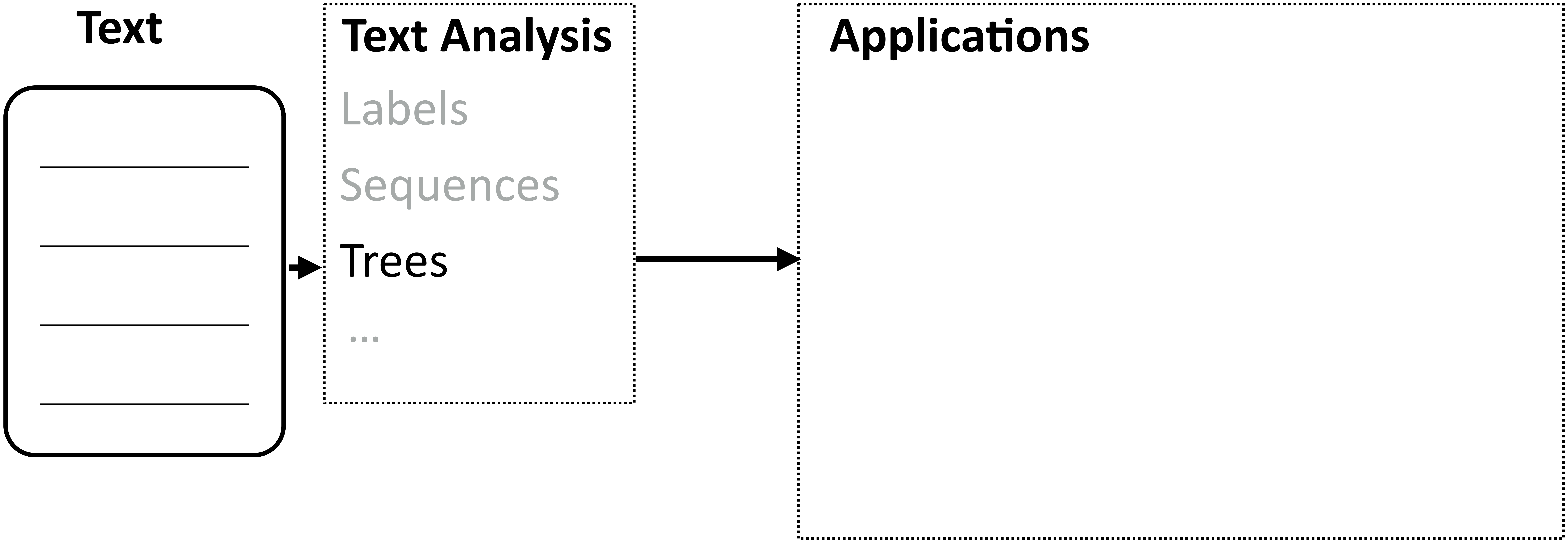
Labels

Sequences

Trees

…

How do we use these representations?



How do we use these representations?

Text

Text Analysis

Labels

Sequences

Trees

…

Applications

Extract syntactic features

How do we use these representations?

Text

Text Analysis

Labels

Sequences

Trees

…

Applications

Extract syntactic features

Tree-structured neural networks

How do we use these representations?

Text

Text Analysis

Labels

Sequences

Trees

…

Applications

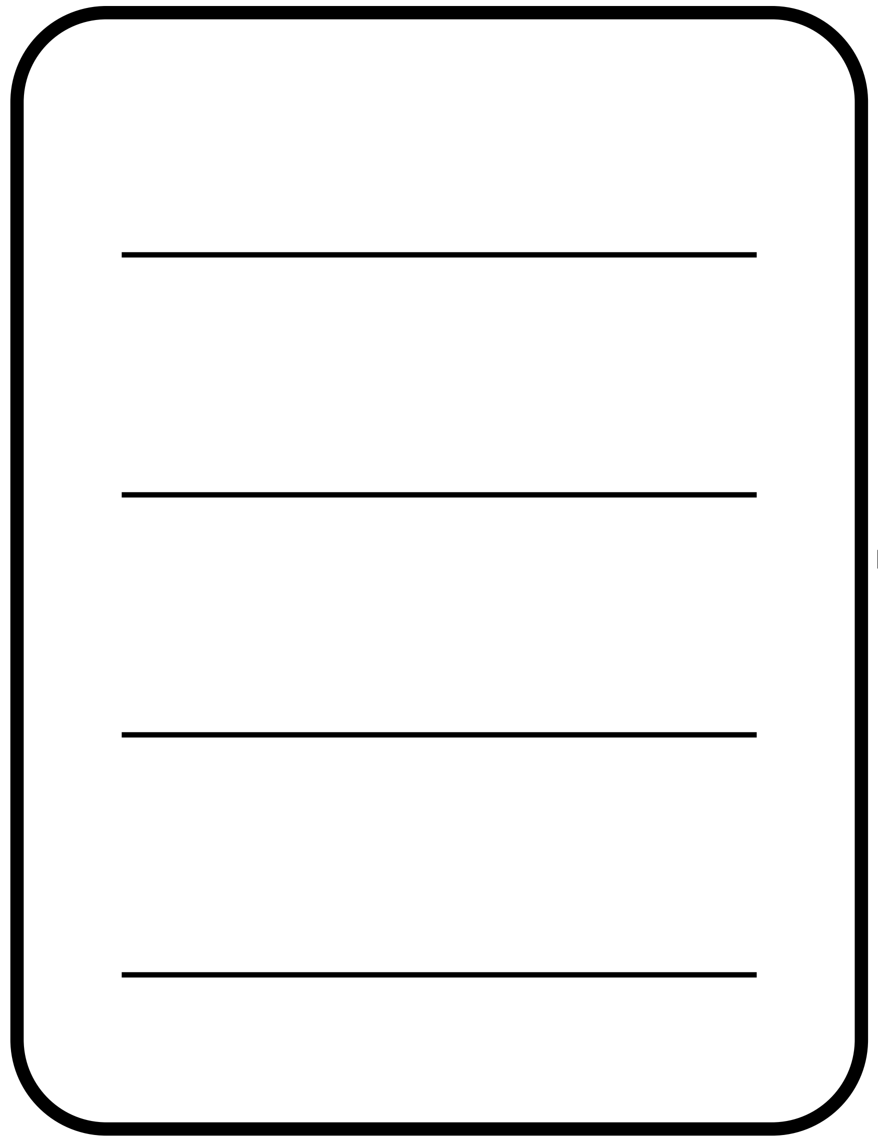
Extract syntactic features

Tree-structured neural networks

Tree transducers (for machine translation)

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)

…

How do we use these representations?

Text

Text Analysis

Labels

Sequences

Trees

…

Applications

Extract syntactic features

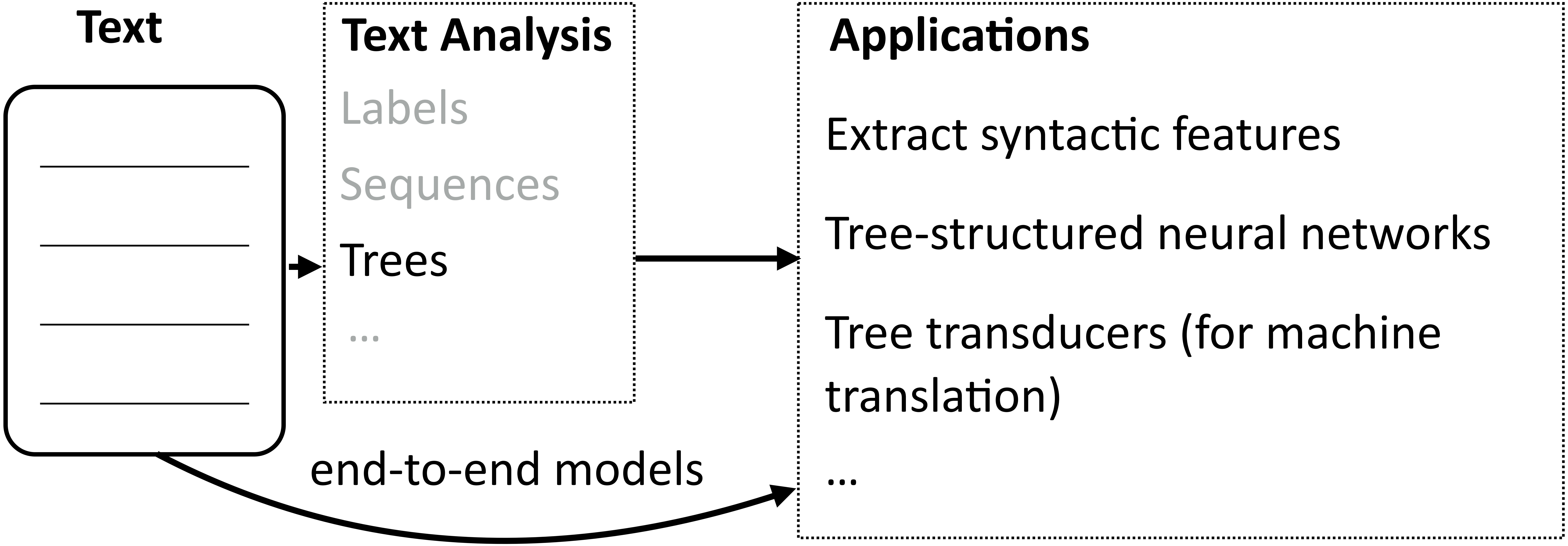
Tree-structured neural networks

Tree transducers (for machine translation)

…

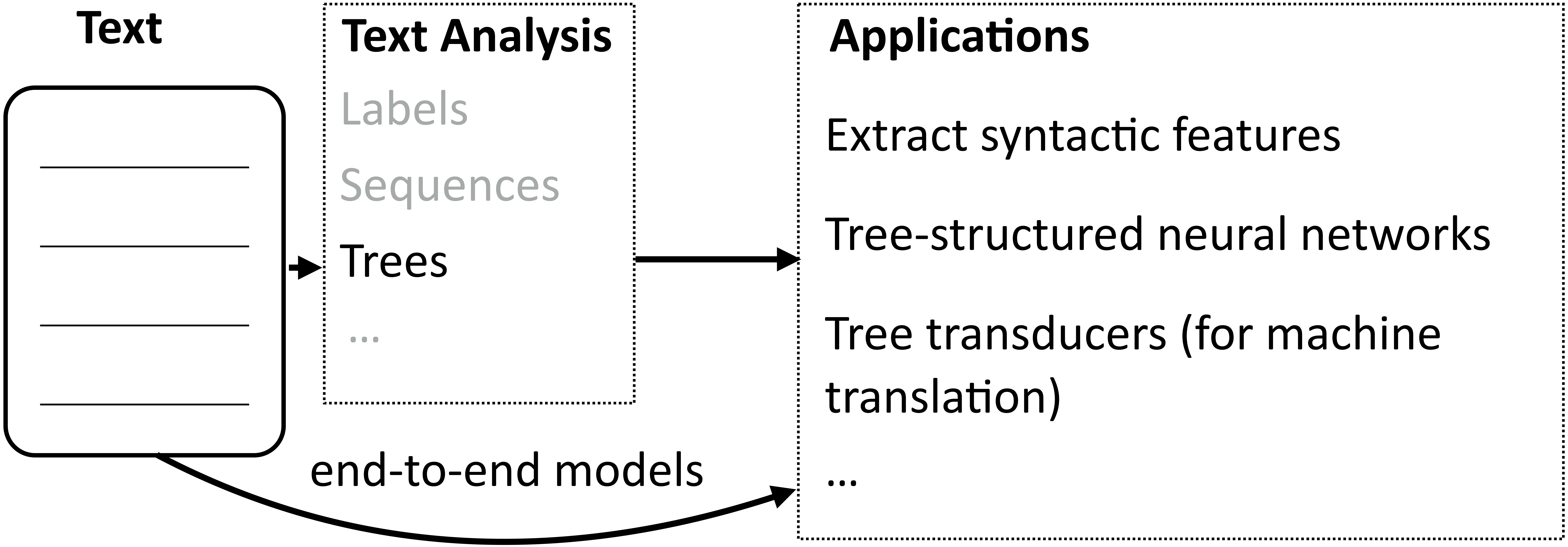
end-to-end models

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)

they advocated

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)

they advocated

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)

they advocated

The city council refused the demonstrators a permit because they _______ violence

they feared

Language is Ambiguous!

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

they advocated

The city council refused the demonstrators a permit because they _______ violence

they feared

Language is Ambiguous!

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

they advocated

The city council refused the demonstrators a permit because they _______ violence

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)

they advocated

The city council refused the demonstrators a permit because they _______ violence

they feared

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

Language is Ambiguous!

slide credit: Dan Klein

Language is Ambiguous!

- Ambiguous News Headlines:

slide credit: Dan Klein

Language is Ambiguous!

- Ambiguous News Headlines:
  - Teacher Strikes Idle Kids

slide credit: Dan Klein

Language is Ambiguous!

- Ambiguous News Headlines:
  - Teacher Strikes Idle Kids
  - Hospitals Sued by 7 Foot Doctors

slide credit: Dan Klein

Language is Ambiguous!

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

slide credit: Dan Klein

Language is Ambiguous!

- Ambiguous News Headlines:
  - Teacher Strikes Idle Kids
  - Hospitals Sued by 7 Foot Doctors
  - Ban on Nude Dancing on Governor’s Desk
  - Iraqi Head Seeks Arms

slide credit: Dan Klein

Language is Ambiguous!

- Ambiguous News 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

slide credit: Dan Klein

Language is Ambiguous!

- Ambiguous News 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

slide credit: Dan Klein

Language is Ambiguous!

- Ambiguous News 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

slide credit: Dan Klein

Language is Ambiguous!

- Ambiguous News 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

slide credit: Dan Klein

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] [.]

slide credit: Dan Klein

What techniques do we use?

(to combine data, knowledge, linguistics, etc.)

A brief history of (modern) NLP

1980 1990 2000 2010 2020

A brief history of (modern) NLP

"AI winter"
rule-based,
expert systems

1980 1990 2000 2010 2020

A brief history of (modern) NLP

“AI winter”
rule-based,
expert systems

earliest stat MT
work at IBM

1980 1990 2000 2010 2020

A brief history of (modern) NLP

"AI winter"
rule-based,
expert systems

Penn
treebank

S
NP VP

earliest stat MT
work at IBM

1980 1990 2000 2010 2020

A brief history of (modern) NLP

"AI winter"
rule-based,
expert systems

earliest stat MT
work at IBM

Penn
treebank

S
NP VP

Ratnaparkhi
tagger

NNP VBZ

1980 1990 2000 2010 2020

A brief history of (modern) NLP

"AI winter"
rule-based,
expert systems

earliest stat MT
work at IBM

Penn
treebank

S
NP VP

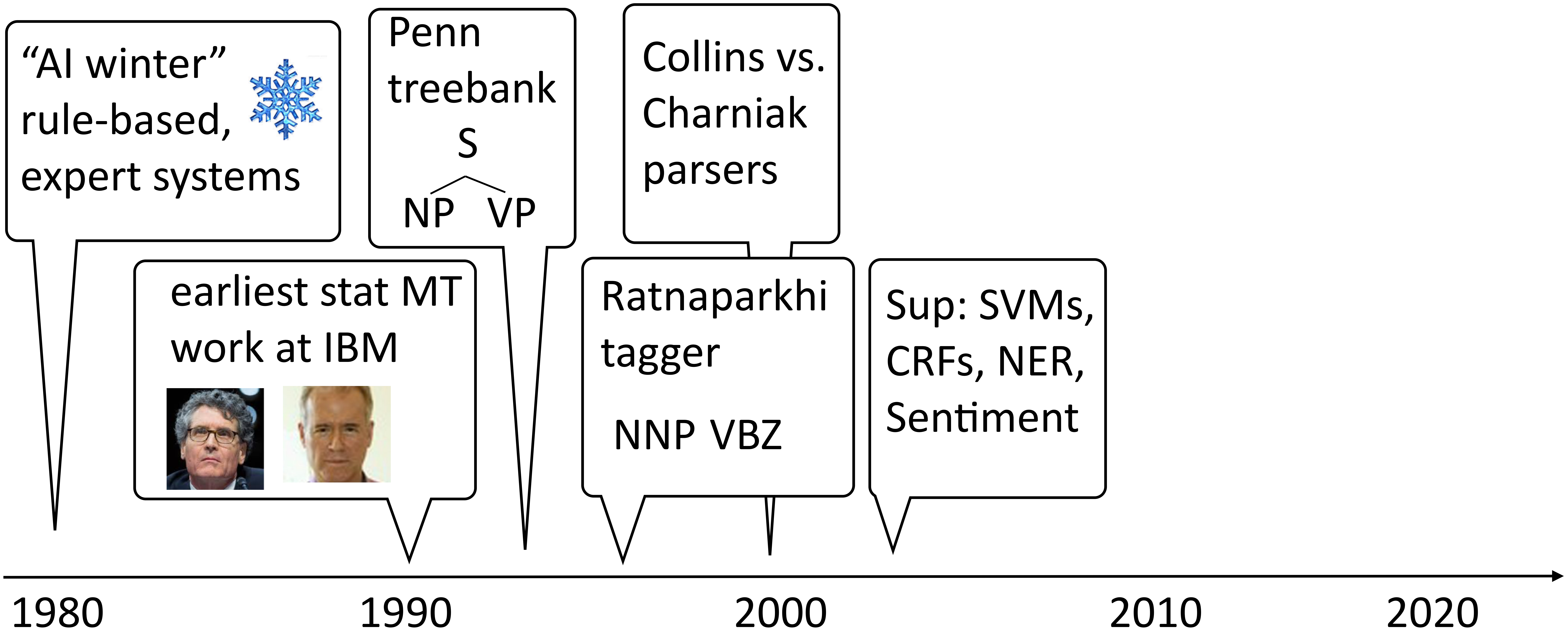
Collins vs.
Charniak
parsers

Ratnaparkhi
tagger

NNP VBZ

1980 1990 2000 2010 2020

A brief history of (modern) NLP



A brief history of (modern) NLP

"AI winter"
rule-based,
expert systems

earliest stat MT
work at IBM

Penn
treebank

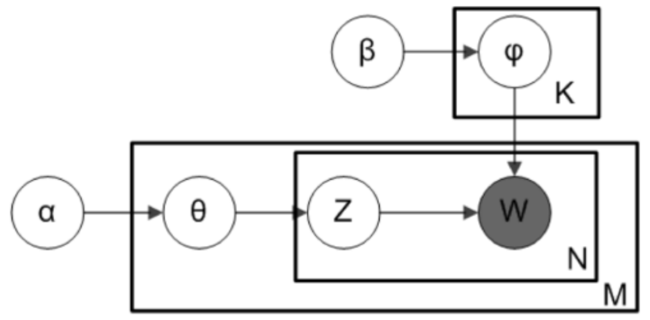
S
NP VP

Ratnaparkhi
tagger

NNP VBZ

Collins vs.
Charniak
parsers

Unsup: topic
models, etc.



Sup: SVMs,
CRFs, NER,
Sentiment

1980 1990 2000 2010 2020

A brief history of (modern) NLP

"AI winter"
rule-based,
expert systems

earliest stat MT
work at IBM

Penn
treebank

S
NP VP

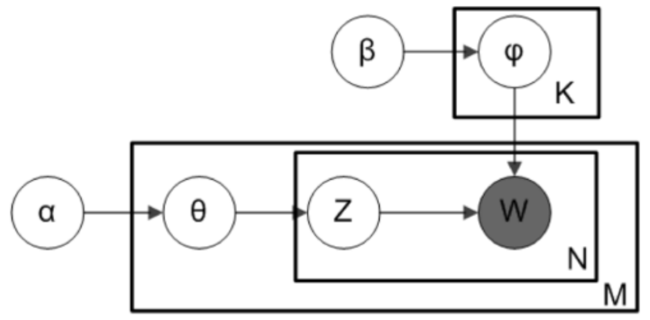
Ratnaparkhi
tagger

NNP VBZ

Collins vs.
Charniak
parsers

Sup: SVMs,
CRFs, NER,
Sentiment

Unsup: topic
models, etc.



Semi-sup,
structured
prediction

1980 1990 2000 2010 2020

A brief history of (modern) NLP

"AI winter"
rule-based,
expert systems

earliest stat MT
work at IBM

Penn
treebank

S
NP VP

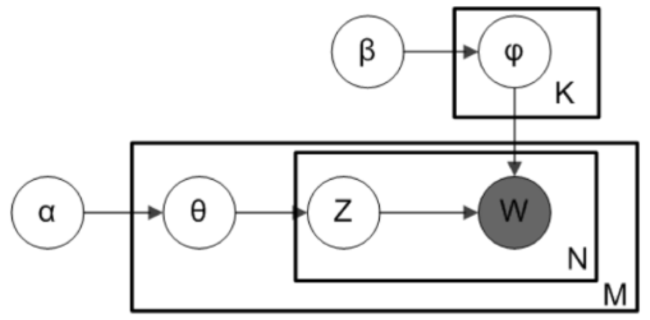
Ratnaparkhi
tagger

NNP VBZ

Collins vs.
Charniak
parsers

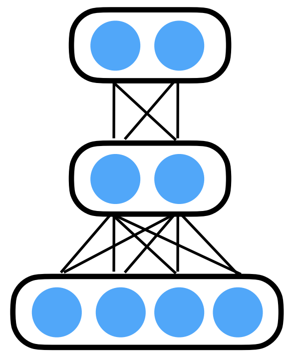
Sup: SVMs,
CRFs, NER,
Sentiment

Unsup: topic
models, etc.



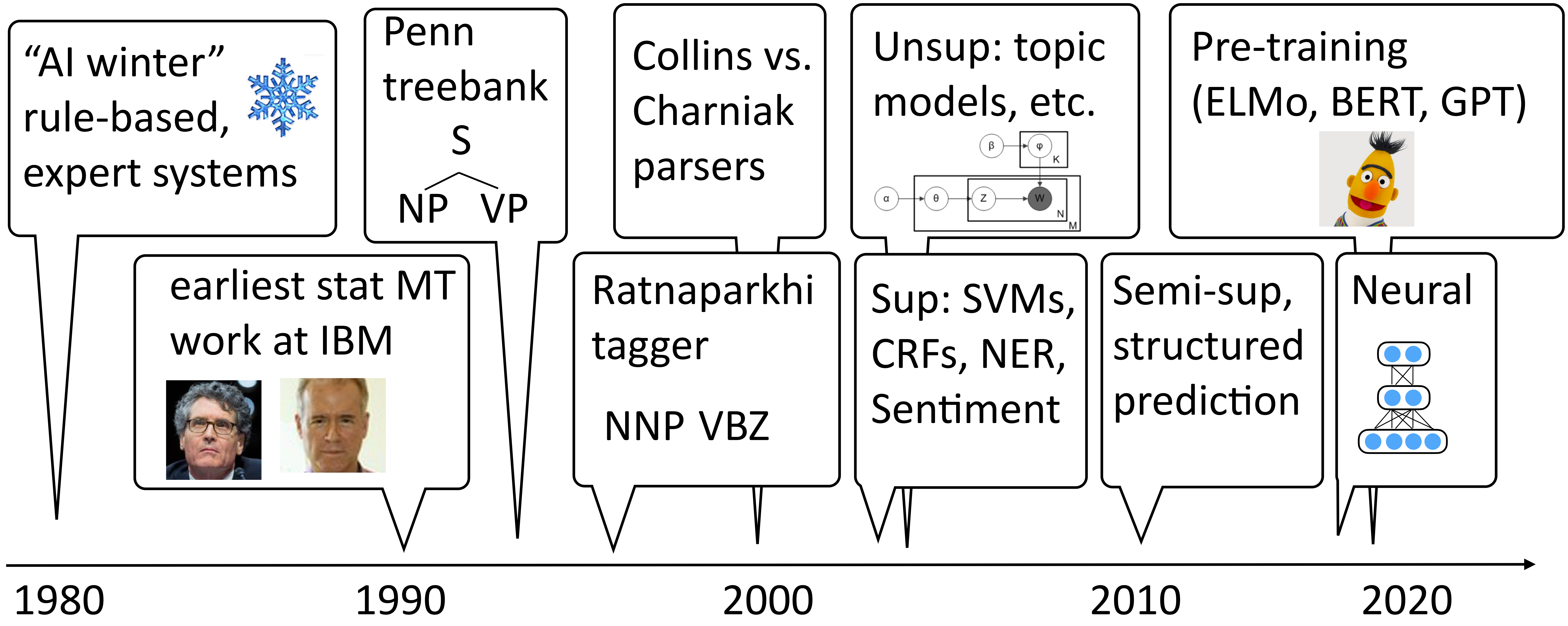
Semi-sup,
structured
prediction

Neural



1980 1990 2000 2010 2020

A brief history of (modern) NLP



How Much Training Data do we Need?

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

How Much Training Data do we Need?

- 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)

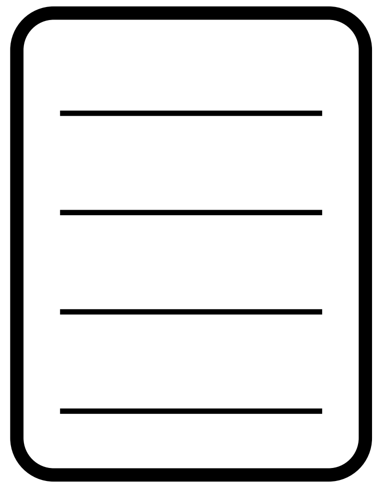
How Much Training Data do we Need?

- 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)

How Much Training Data do we Need?

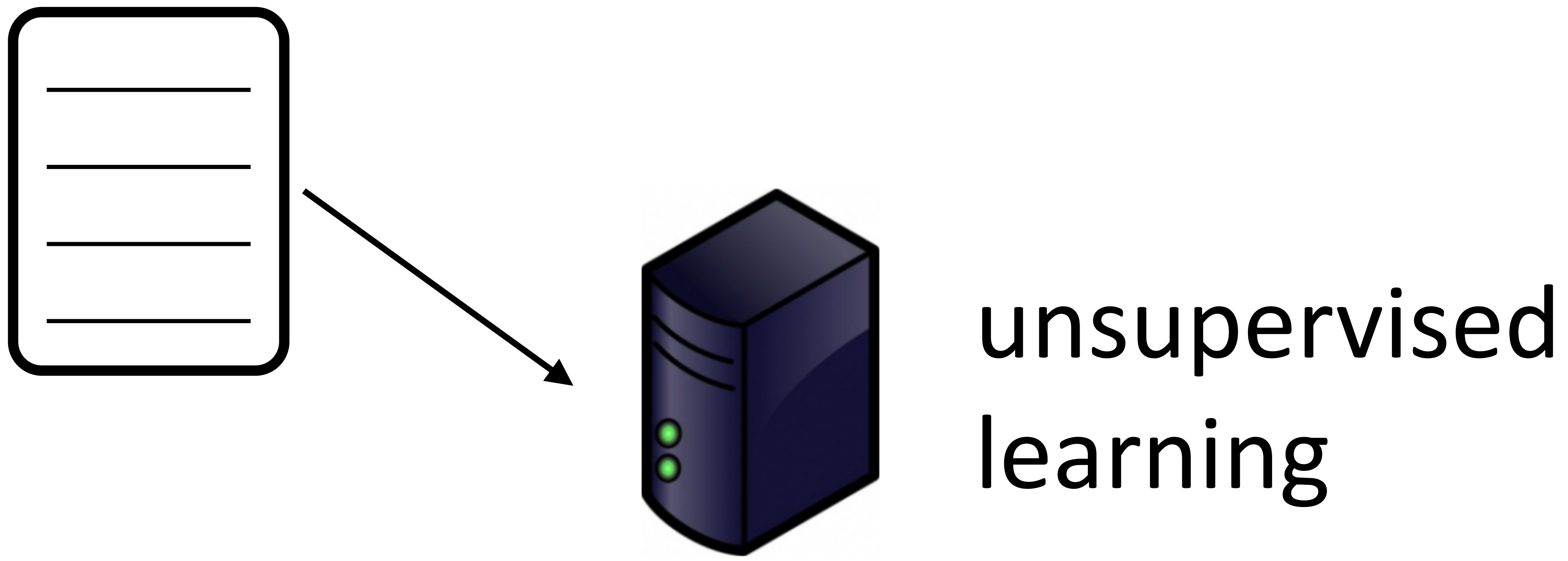
- All of these techniques are data-driven! Some data is naturally occurring, but may need to label
- Supervised techniques work well on very little data



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

How Much Training Data do we Need?

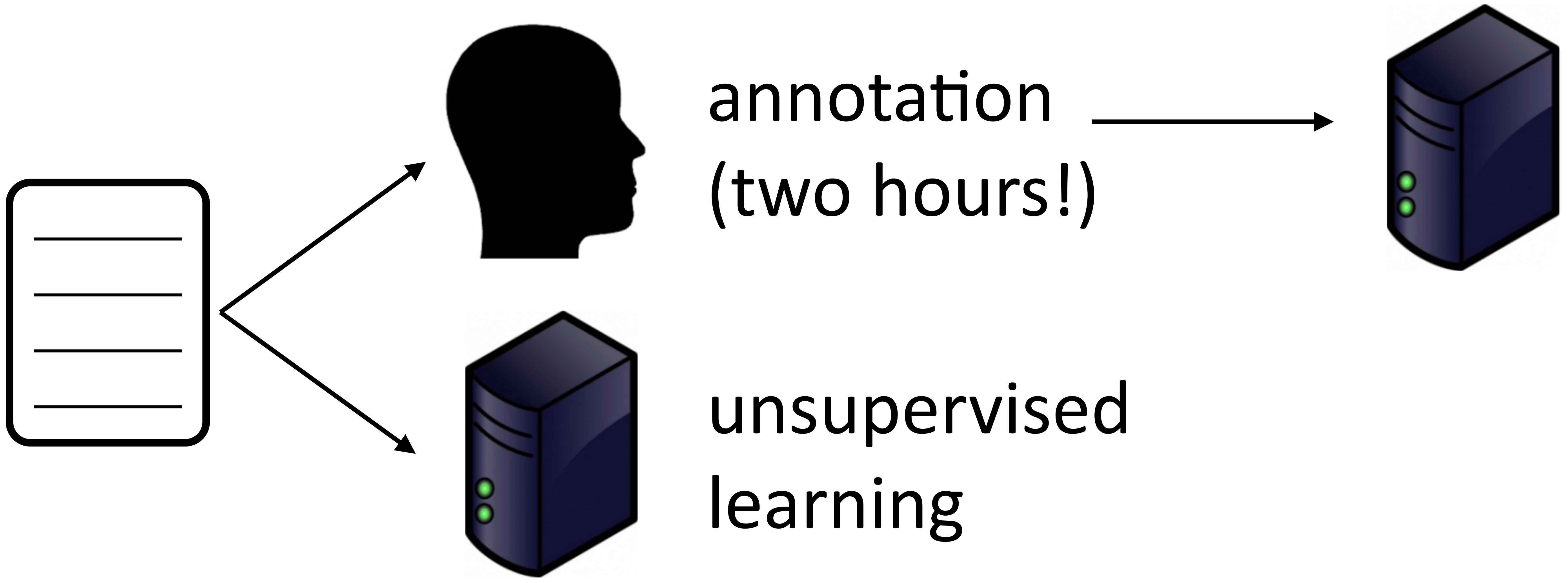
- All of these techniques are data-driven! Some data is naturally occurring, but may need to label
- Supervised techniques work well on very little data



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

How Much Training Data do we Need?

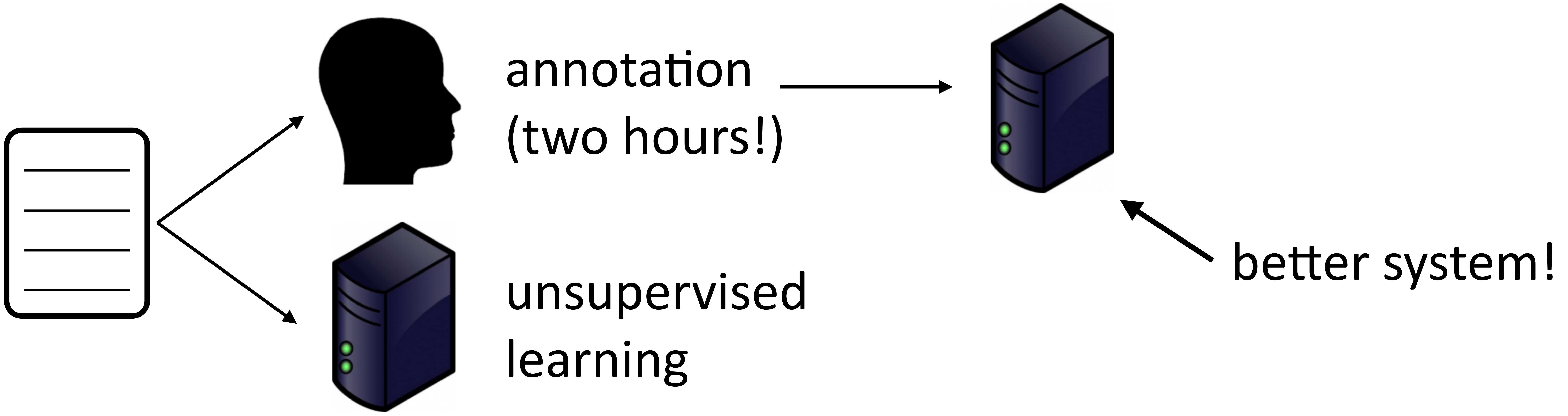
- All of these techniques are data-driven! Some data is naturally occurring, but may need to label
- Supervised techniques work well on very little data



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

How Much Training Data do we Need?

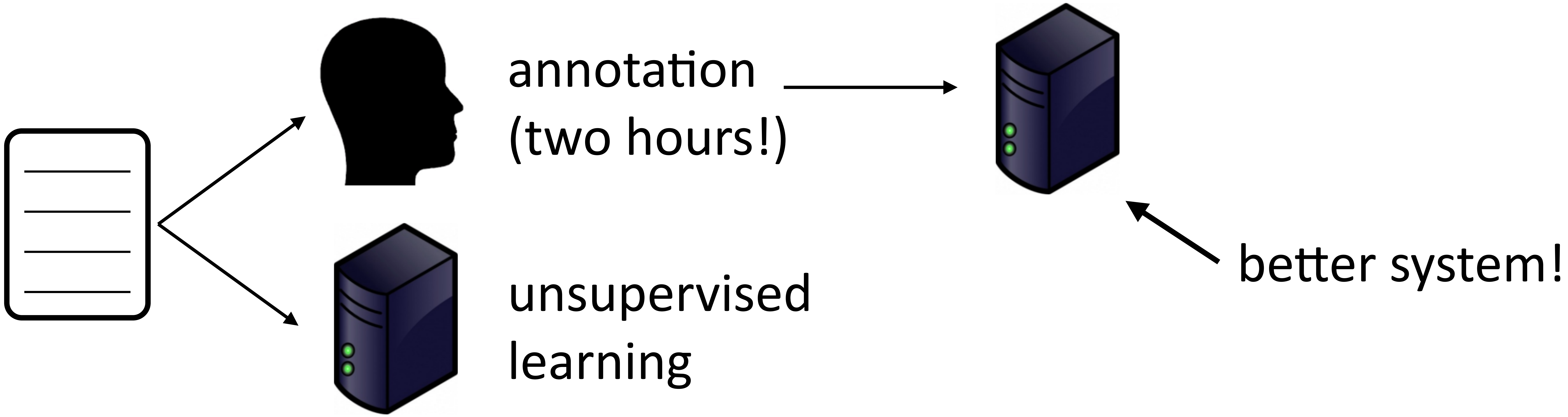
- All of these techniques are data-driven! Some data is naturally occurring, but may need to label
- Supervised techniques work well on very little data



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

How Much Training Data do we Need?

- 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)

Pretraining

- Language modeling: predict the next word in a text $P(w_i | w_1, \ldots, w_{i-1})$

P(*w* | I want to go to) = 0.01 Hawai'i

0.005 LA

0.0001 class

 : use this model for other purposes

P(*w* | the acting was horrible, I think the movie was) = 0.1 bad

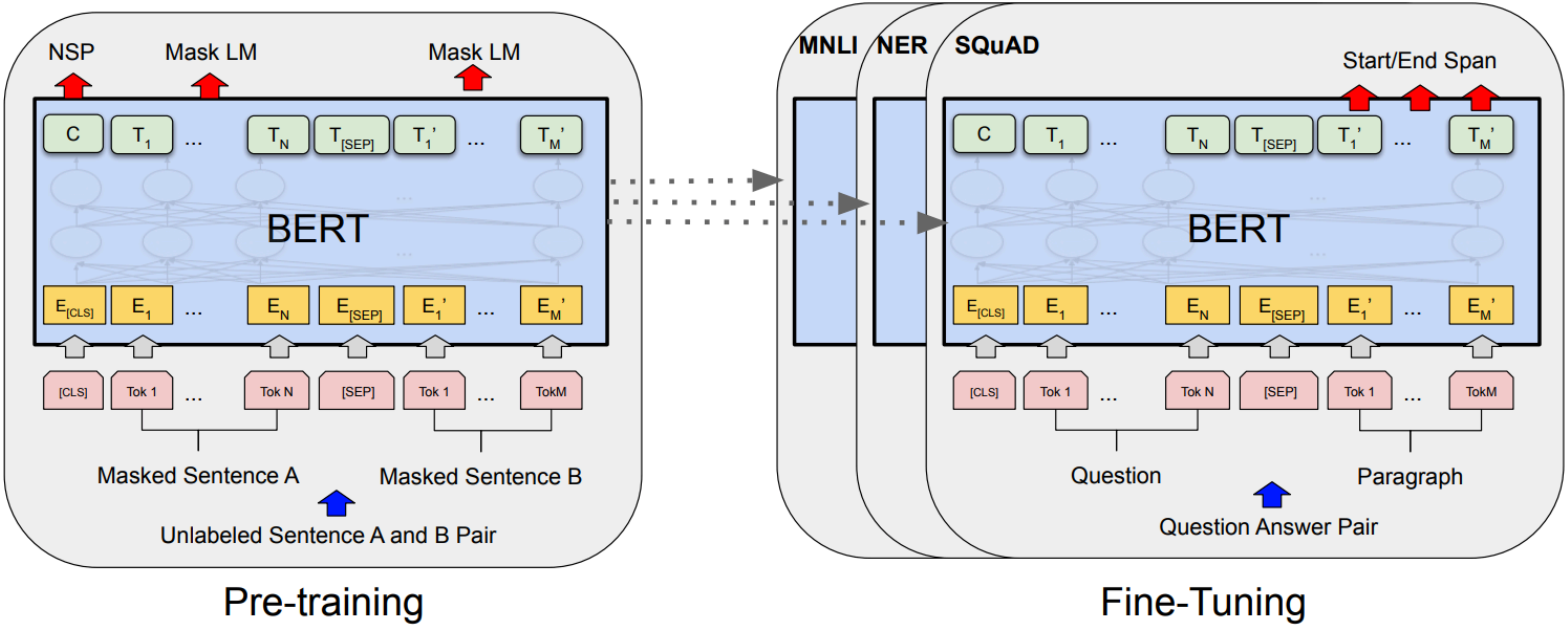
0.001 good

- Model understands some sentiment?

- Train a neural network to do language modeling on massive unlabeled text, fine-tune it to do {tagging, sentiment, question answering, ...}

Peters et al. (2018), Devlin et al. (2019)

BERT



- Key parts which we will study: (1) Transformer architecture; (2) what data is used (both for pre-training and fine-tuning)

Devlin et al. (2019)

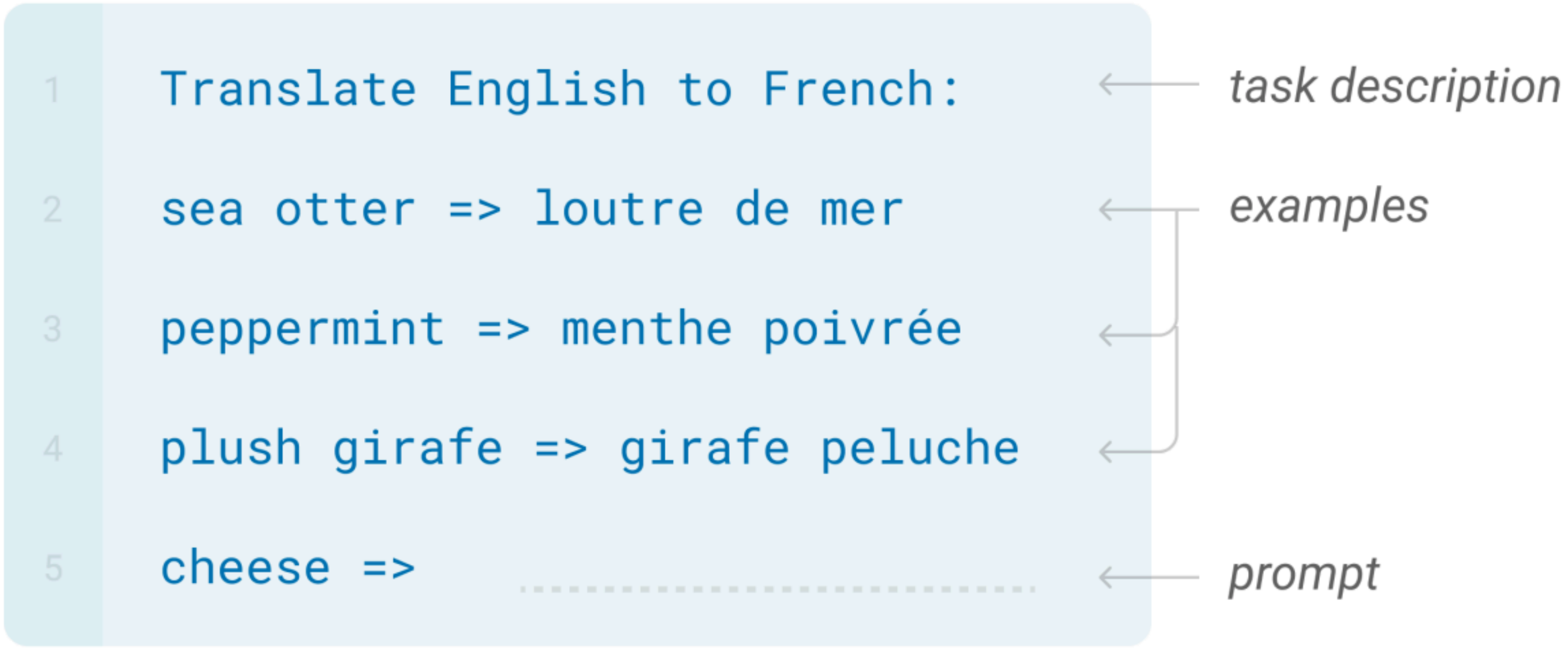
GPT and In-Context Learning

- Even more "extreme" setting: no gradient updates to model, instead large language models "learn" from examples in their context

- Many papers studying why this works. We will read some!

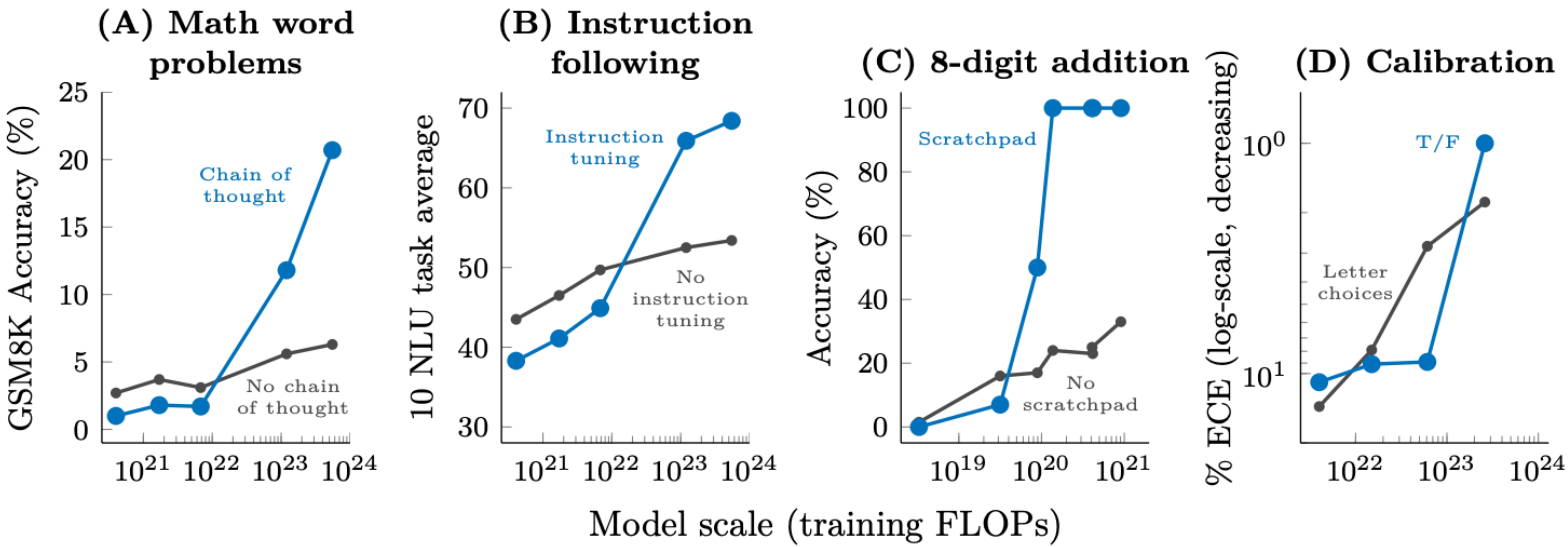
Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.



Brown et al. (2020)

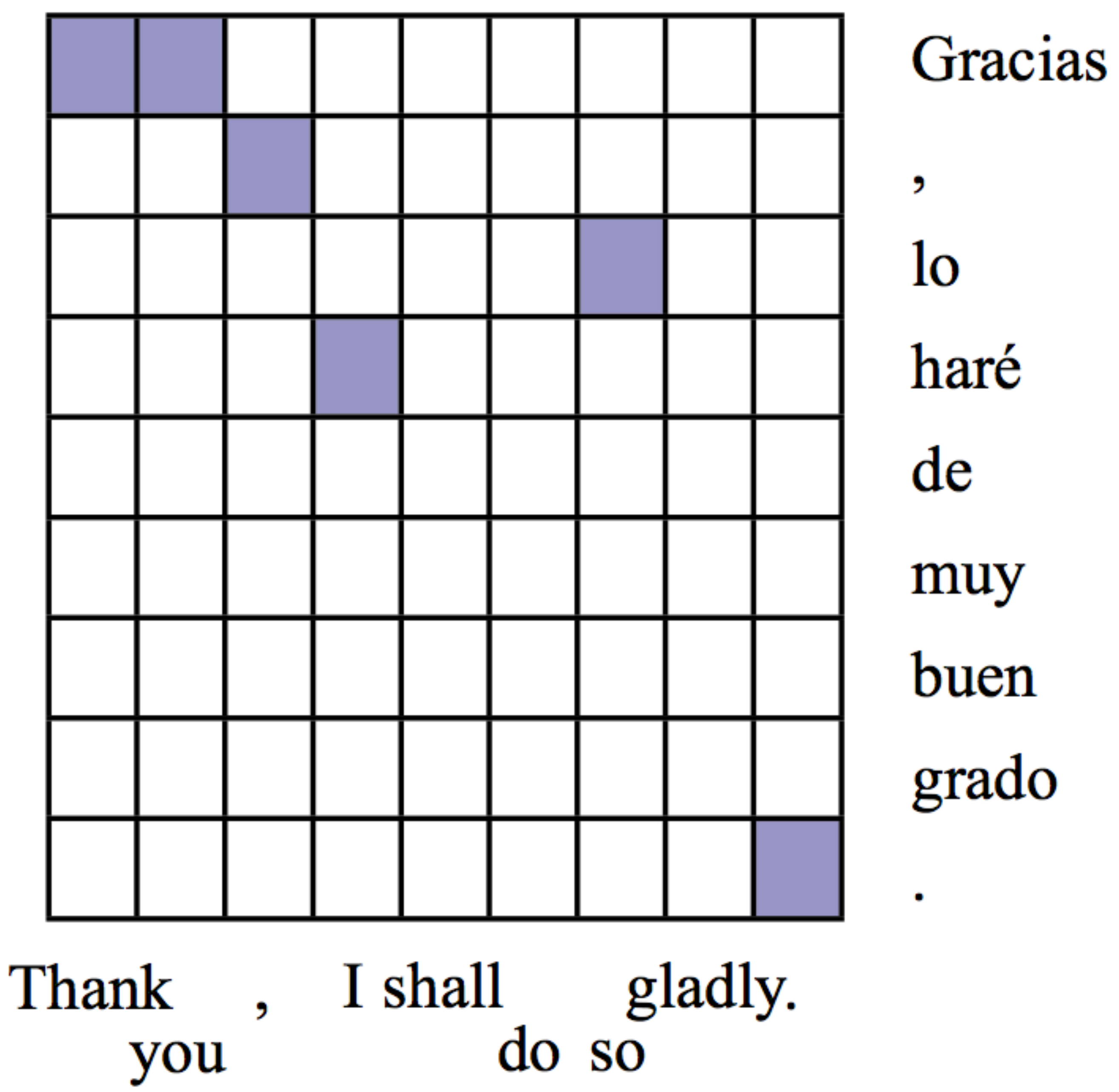
Scaling Laws



- Many of the ideas that are big in 2023 only make sense and only work because the models are so big!

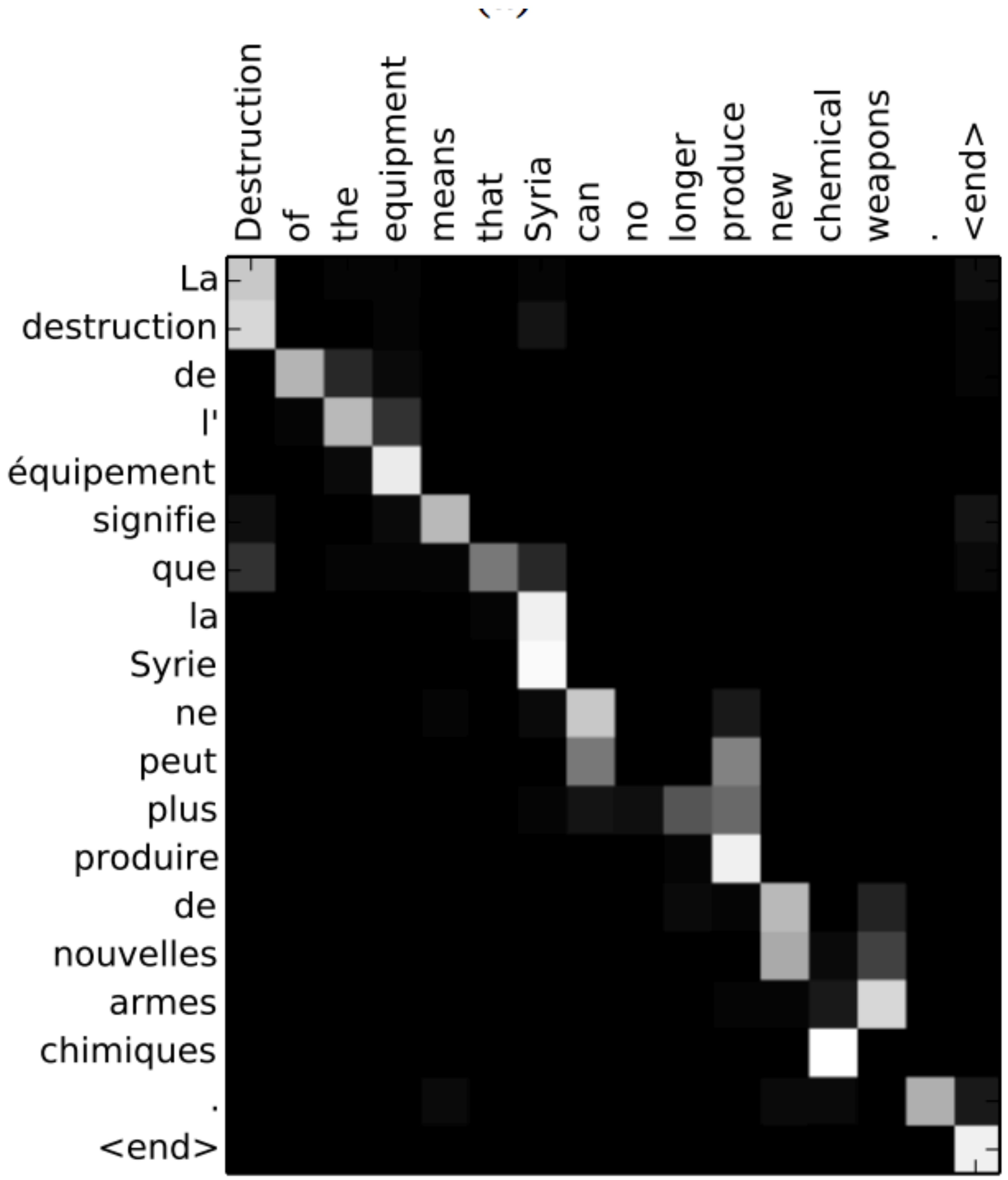
Kaplan et al. (2020), Jason Wei et al. (2022)

Less Manual Structure?



(a) example word alignment (b) example phrase alignment

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_1
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

	CoNLL Avg. F_1
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!

	CoNLL Avg. F_1
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_1
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)

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 to 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 to use requires understanding dataset size, problem complexity, and a lot of tricks!
- NLP encompasses all of these things