CS388: Natural Language Processing
Lecture 1: Introduction

Greg Durrett

Administrivia

- Lecture: Tuesdays and Thursdays 9:30am - 10:50am
- Course website: http://www.cs.utexas.edu/~gdurrett/courses/fa2018/cs388.shtml
- Piazza: link on the course website
- My office hours: Wednesday 10am-noon, GDC 3.420
- TA: Jifan Chen; Office hours:
  - Monday + Tuesday, 1pm-2pm GDC 1.302

Course Requirements

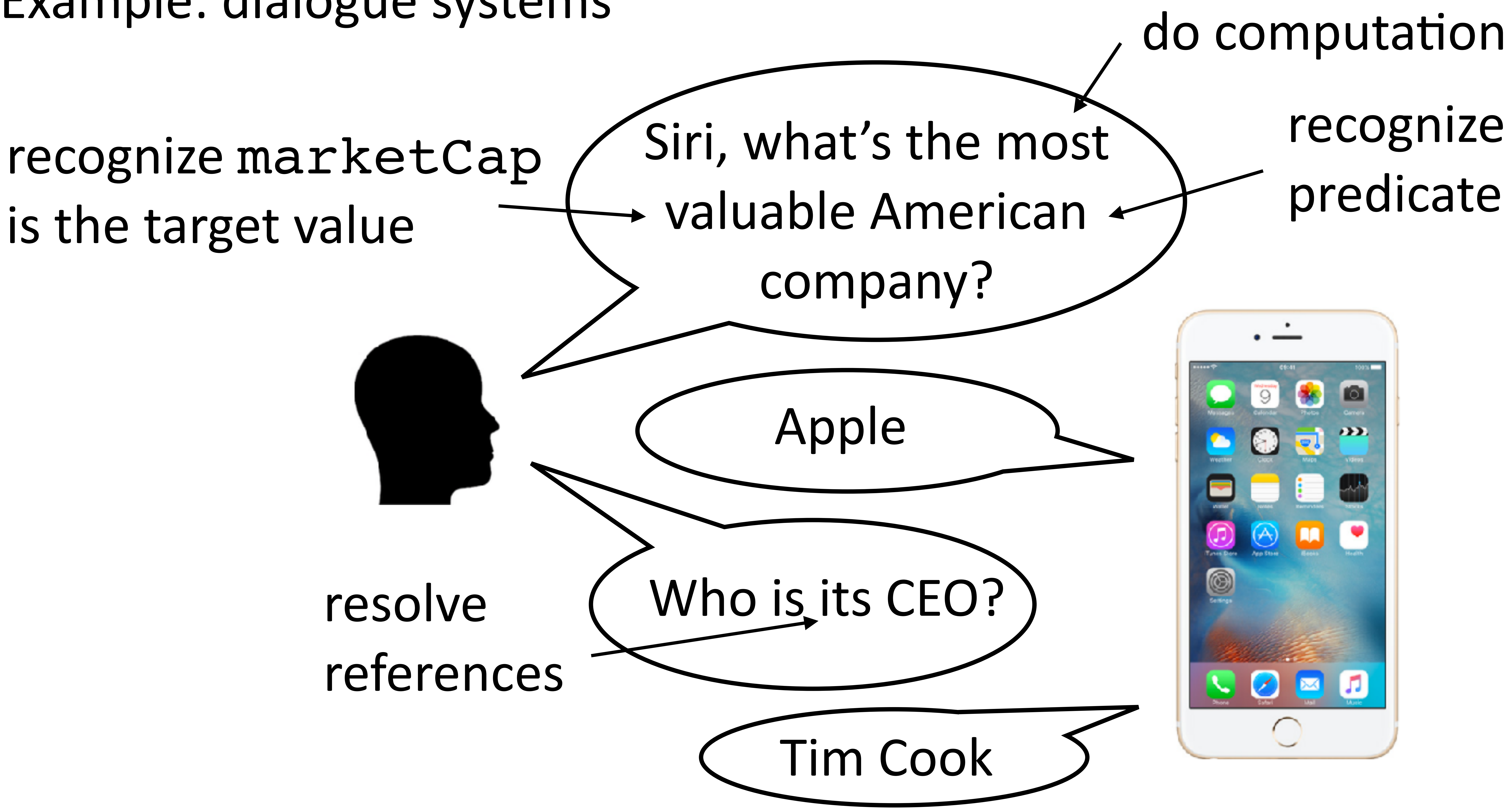
- 391L Machine Learning (or equivalent)
- 311 or 311H Discrete Math for Computer Science (or equivalent)
- Python experience
- Additional prior exposure to probability, linear algebra, optimization, linguistics, and NLP useful but not required

Enrollment

- I want everyone to be able to take this class!
- Mini1 is out now (due September 11):
  - 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 projects, which are smaller-scale than the final project)

What’s the goal of NLP?

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



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

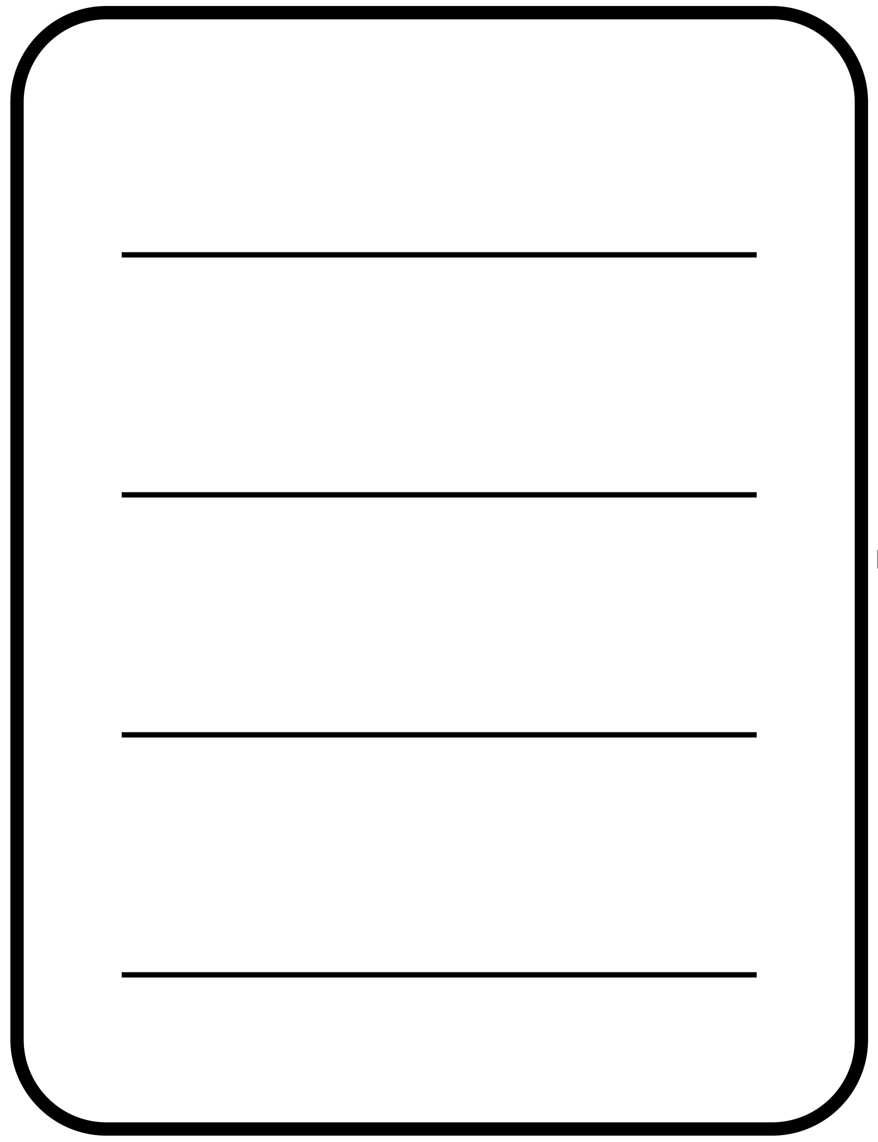
Machine Translation

People’s Daily, August 30, 2017

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

NLP Analysis Pipeline

Text



Text Analysis

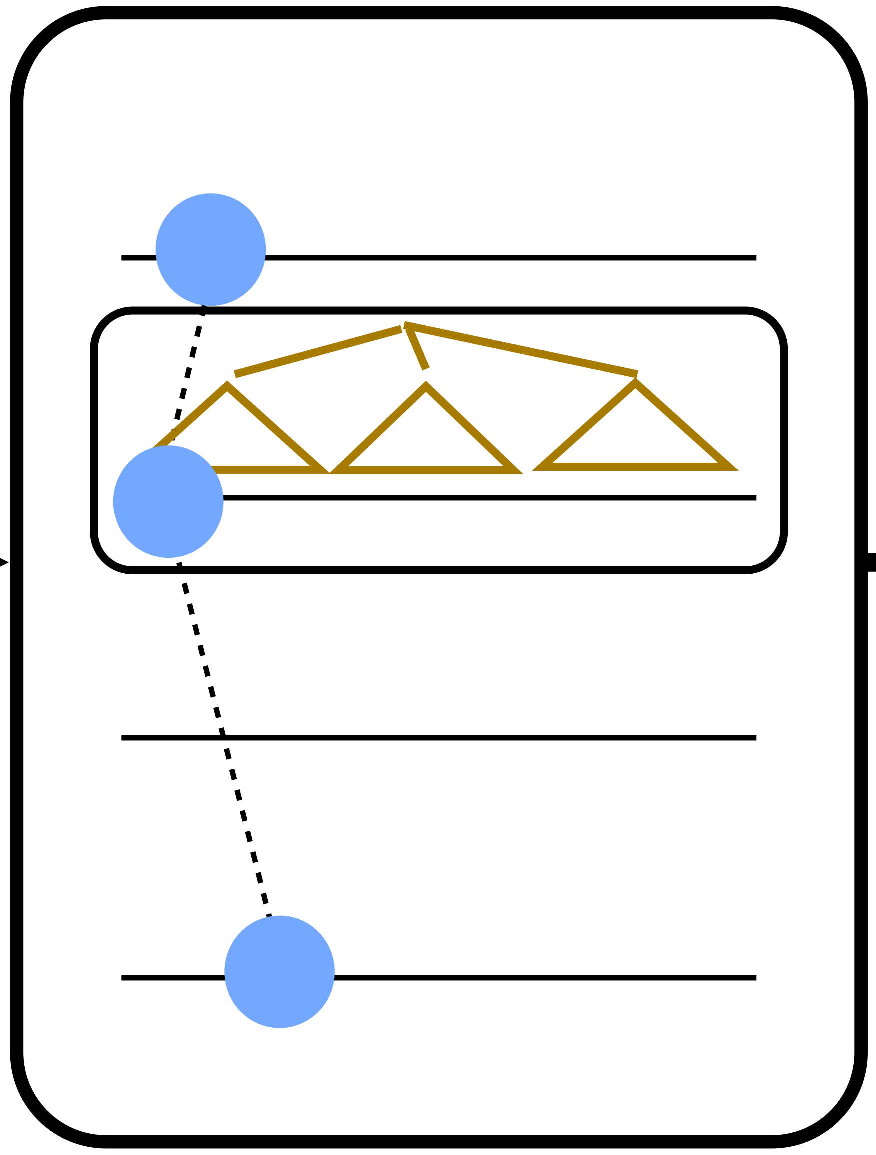
Syntactic parses

Coreference resolution

Entity disambiguation

Discourse analysis

Annotations



Applications

Summarize

Extract information

Answer questions

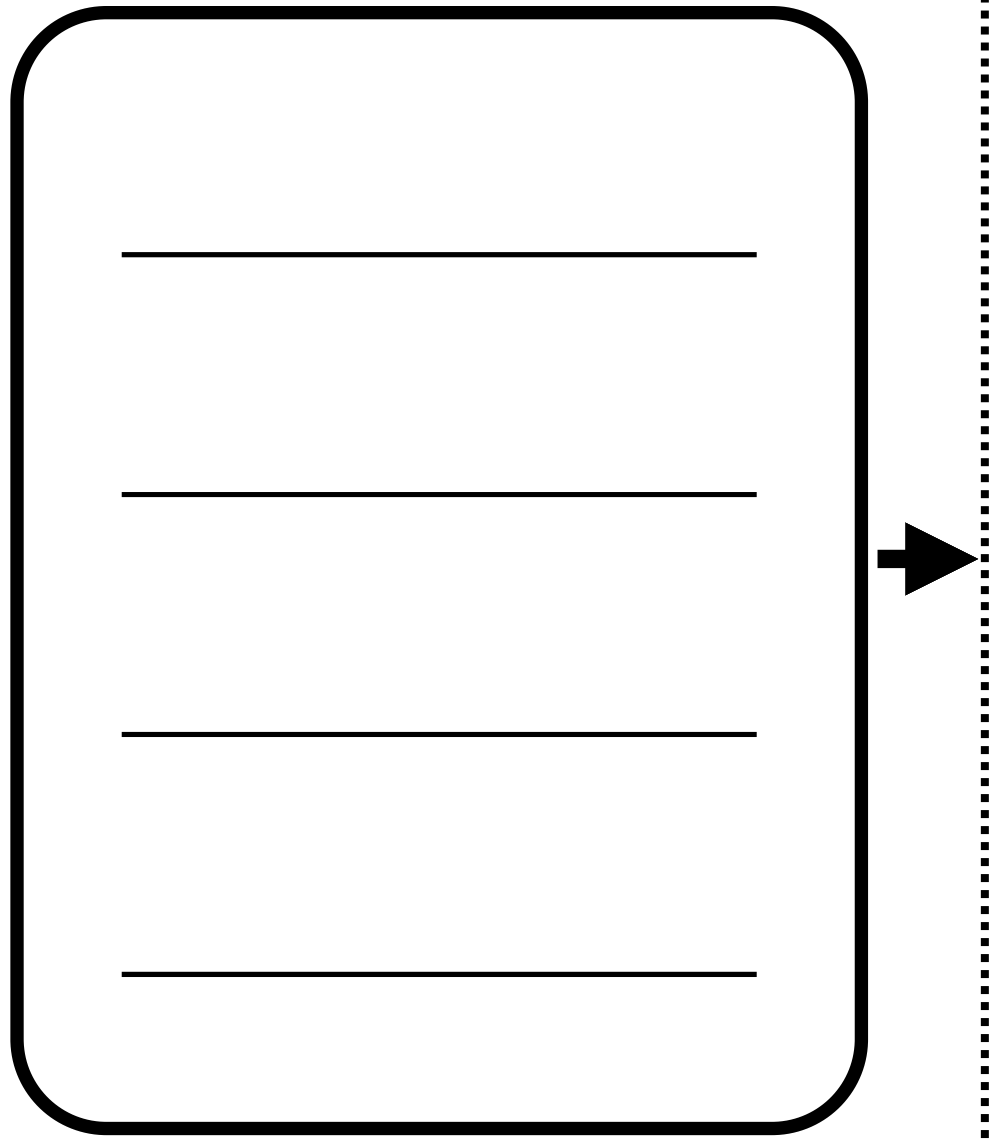
Identify sentiment

Translate

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

How do we represent text?

Text



Labels

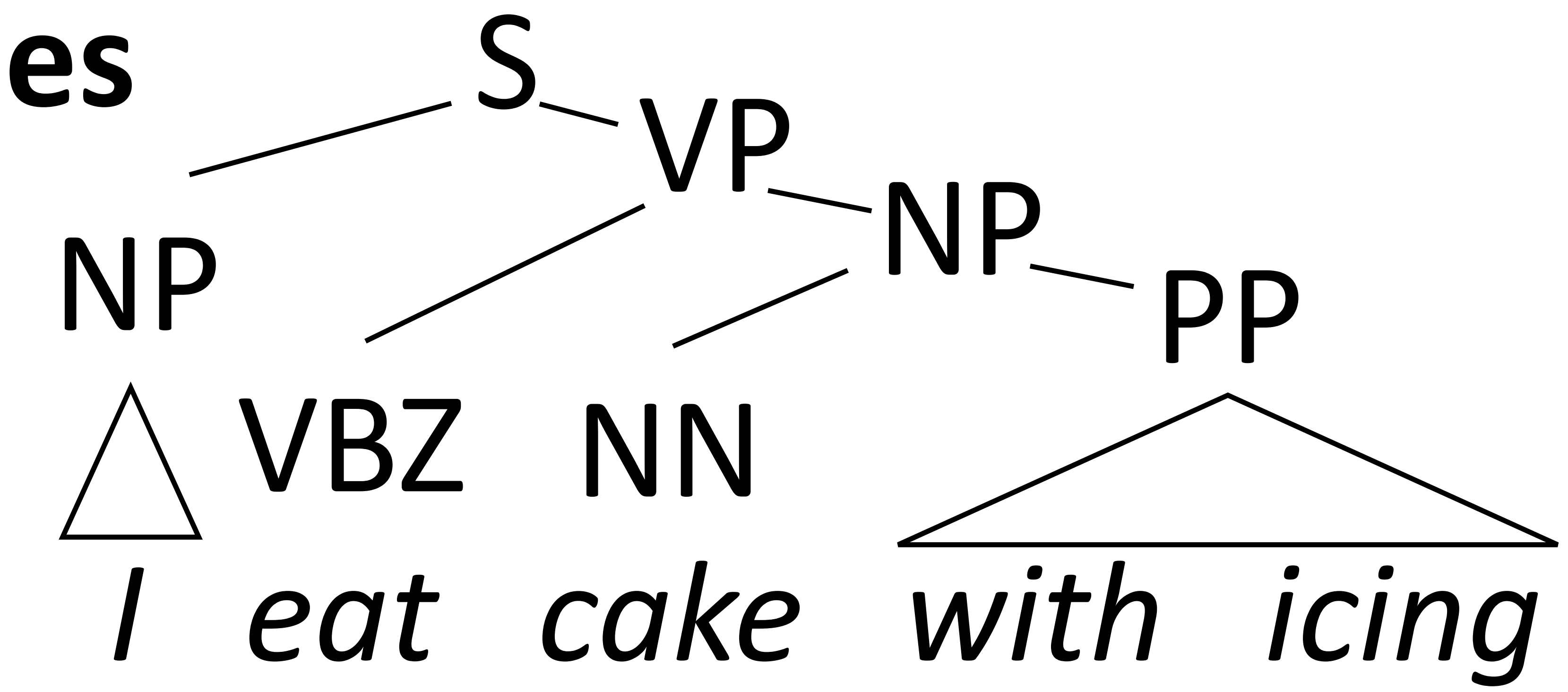
the movie was good +

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*

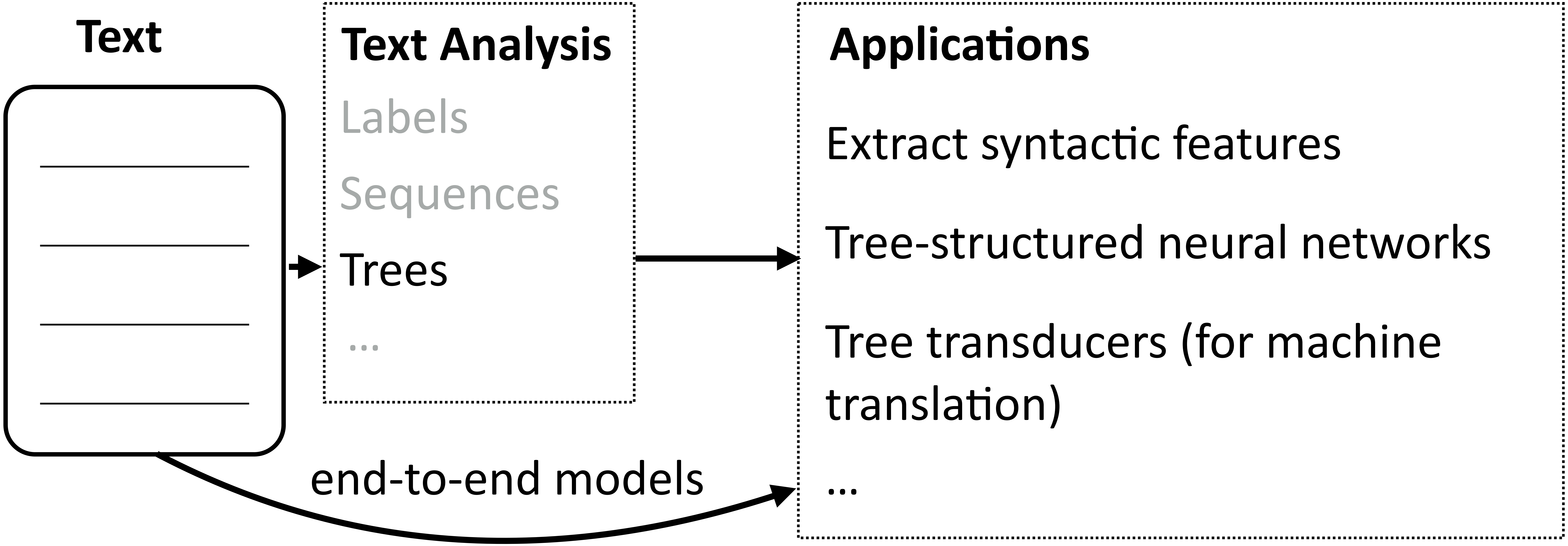
Trees



$\lambda x.\ flight(x) \wedge dest(x)=Miami$

flights to Miami

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)

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!

- 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

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 do we need to understand language?

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

DOJ greenlights Disney - Fox merger

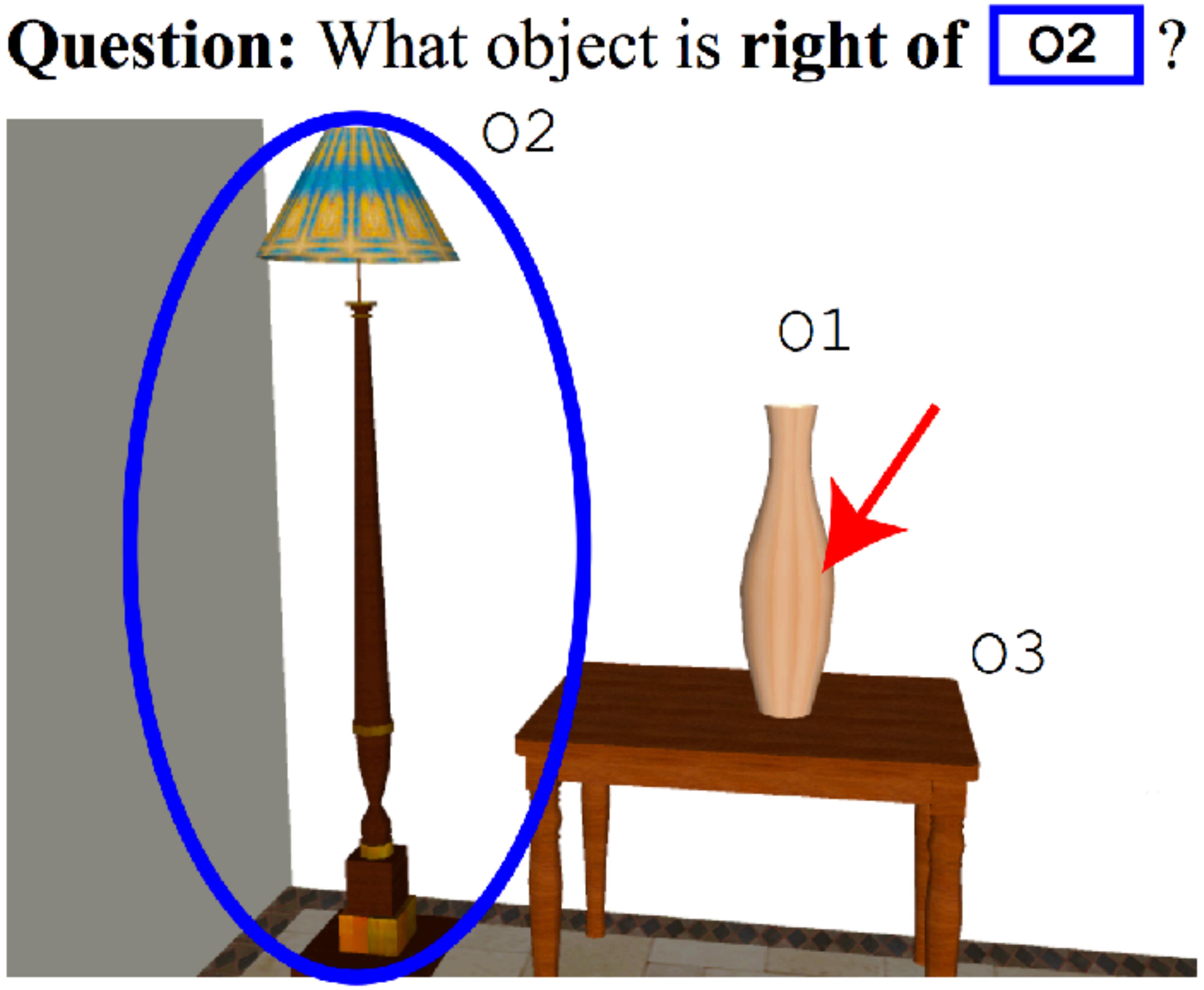
Department of Justice

metaphor;
“approves”

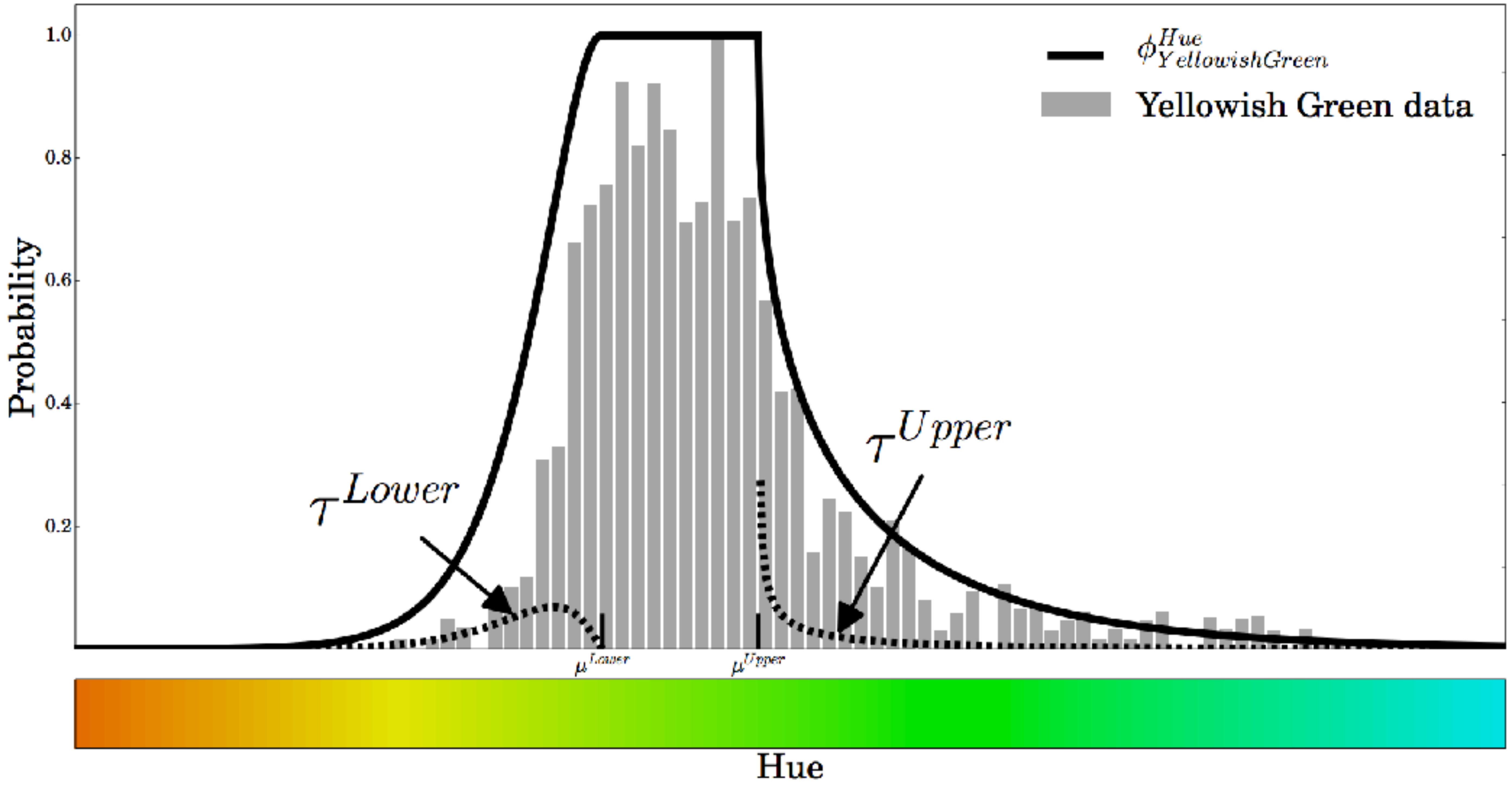
- What is a green light? How do we understand what “green lighting” does?

What do we need to understand language?

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



Golland et al. (2010)



McMahan and Stone (2015)

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 = John; C_f = {John}

c. He called up Mike yesterday to work out a plan. (he = John)
C_b = John; C_f = {John, Mike} (CONTINUE)

d. Mike has annoyed him a lot recently.
C_b = John; C_f = {Mike, John} (RETAIN)

e. He called John at 5 AM on Friday last week. (he = Mike)
C_b = Mike; C_f = {Mike, John} (SHIFT)

Centering Theory
Grosz et al. (1995)

What techniques do we use?

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

A brief history of (modern) NLP

"AI winter"
rule-based,
expert systems

earliest stat MT
work at IBM

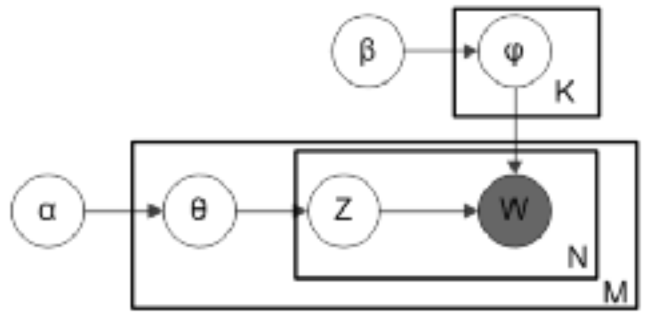
Penn
treebank
S
NP VP

Ratnaparkhi
tagger
NNP VBZ

Collins vs.
Charniak
parsers

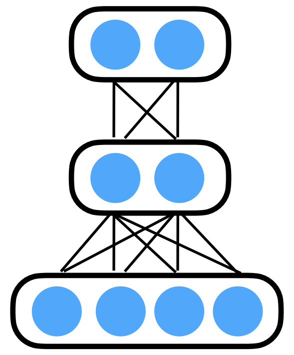
Sup: SVMs,
CRFs, NER,
Sentiment

Unsup: topic
models,
grammar induction



Semi-sup,
structured
prediction

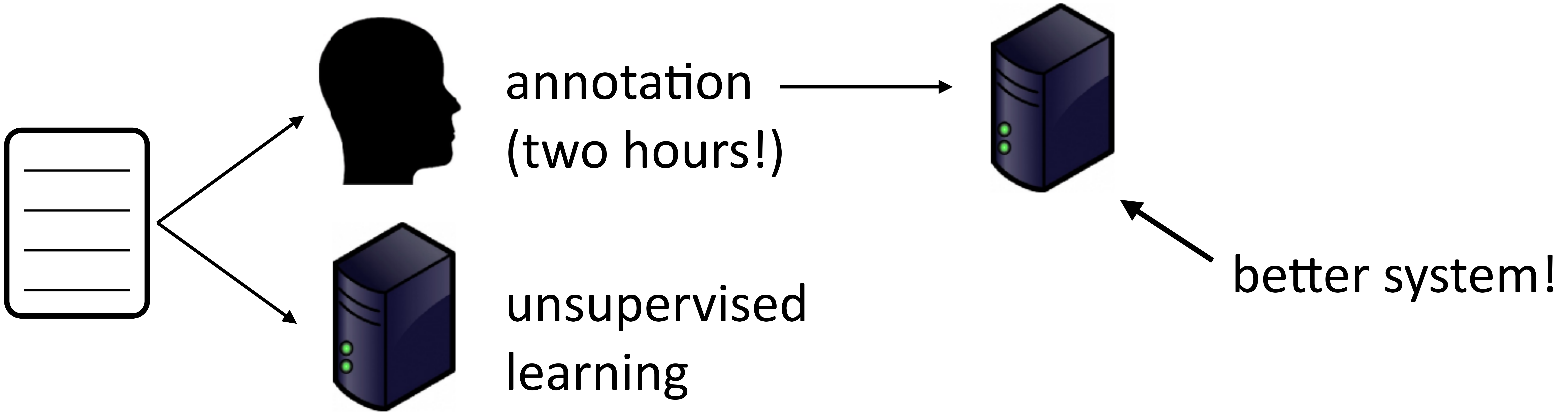
Neural



1980 1990 2000 2010 2018

Structured Prediction

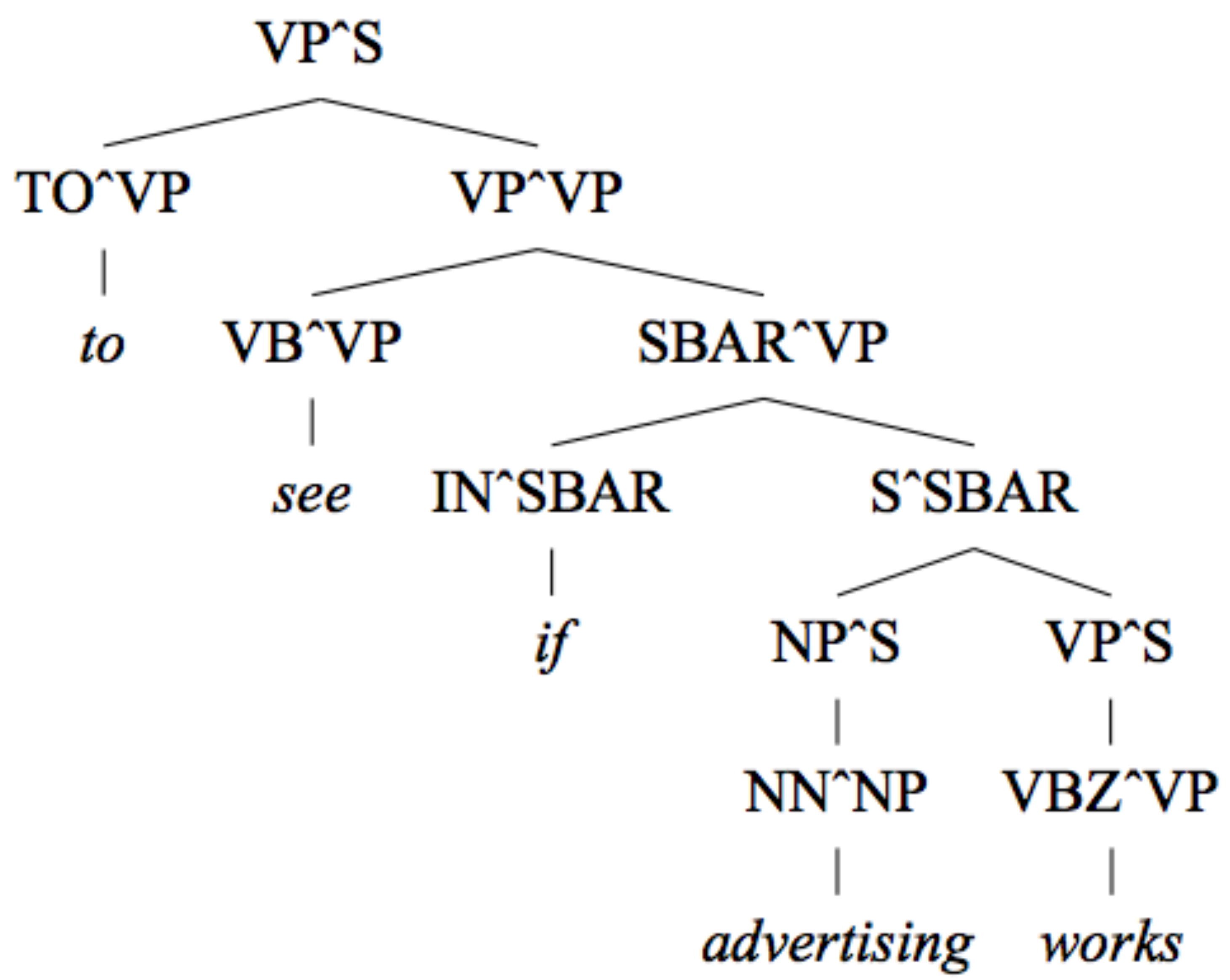
- All of these techniques are data-driven! Some data is naturally occurring, but may need to label
- Supervised techniques work well on very little data



- 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?

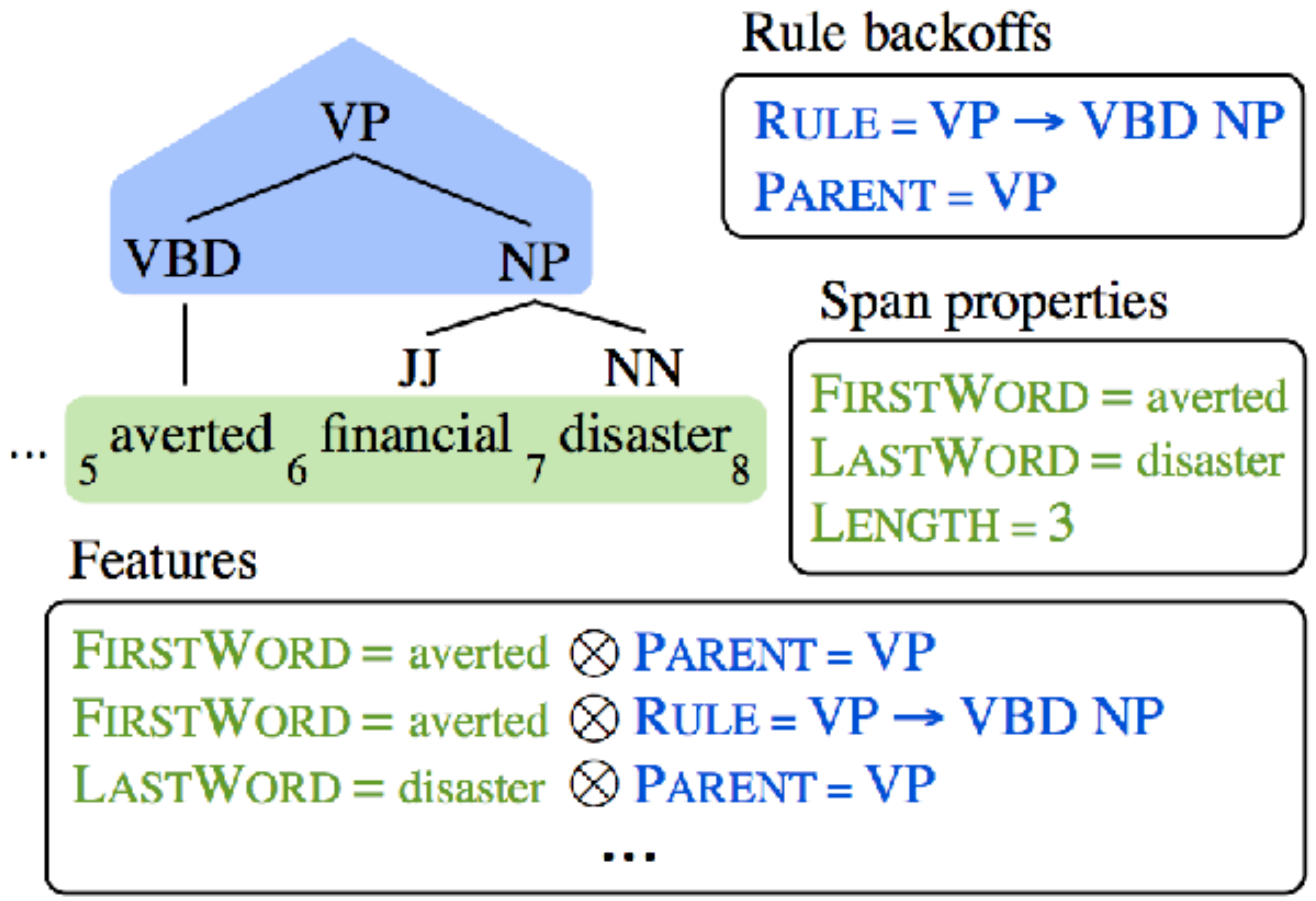


Klein and Manning (2003)

VBZ

VBZ-0	gives	sells	takes
VBZ-1	comes	goes	works
VBZ-2	includes	owns	is
VBZ-3	puts	provides	takes
VBZ-4	says	adds	Says
VBZ-5	believes	means	thinks
VBZ-6	expects	makes	calls
VBZ-7	plans	expects	wants
VBZ-8	is	's	gets
VBZ-9	's	is	remains
VBZ-10	has	's	is
VBZ-11	does	Is	Does

Petrov et al. (2006)

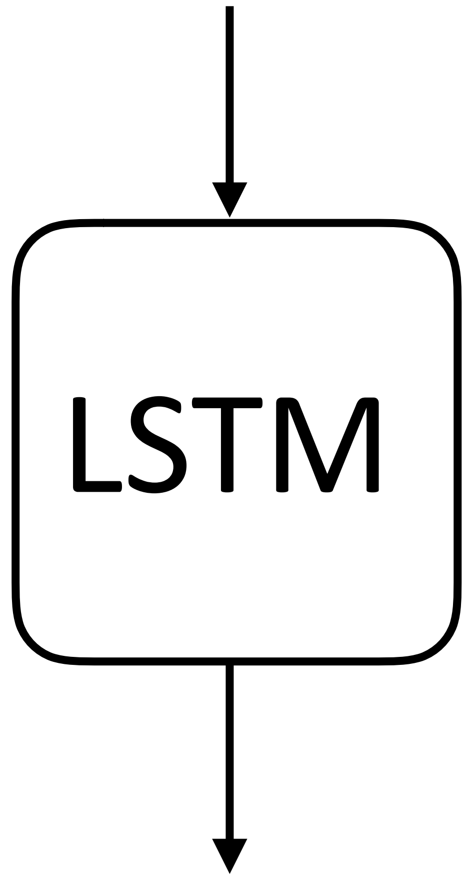


Hall, Durrett, Klein (2014)

- Manually constructed grammars -> EM-induced grammars -> basic grammars + features -> ...

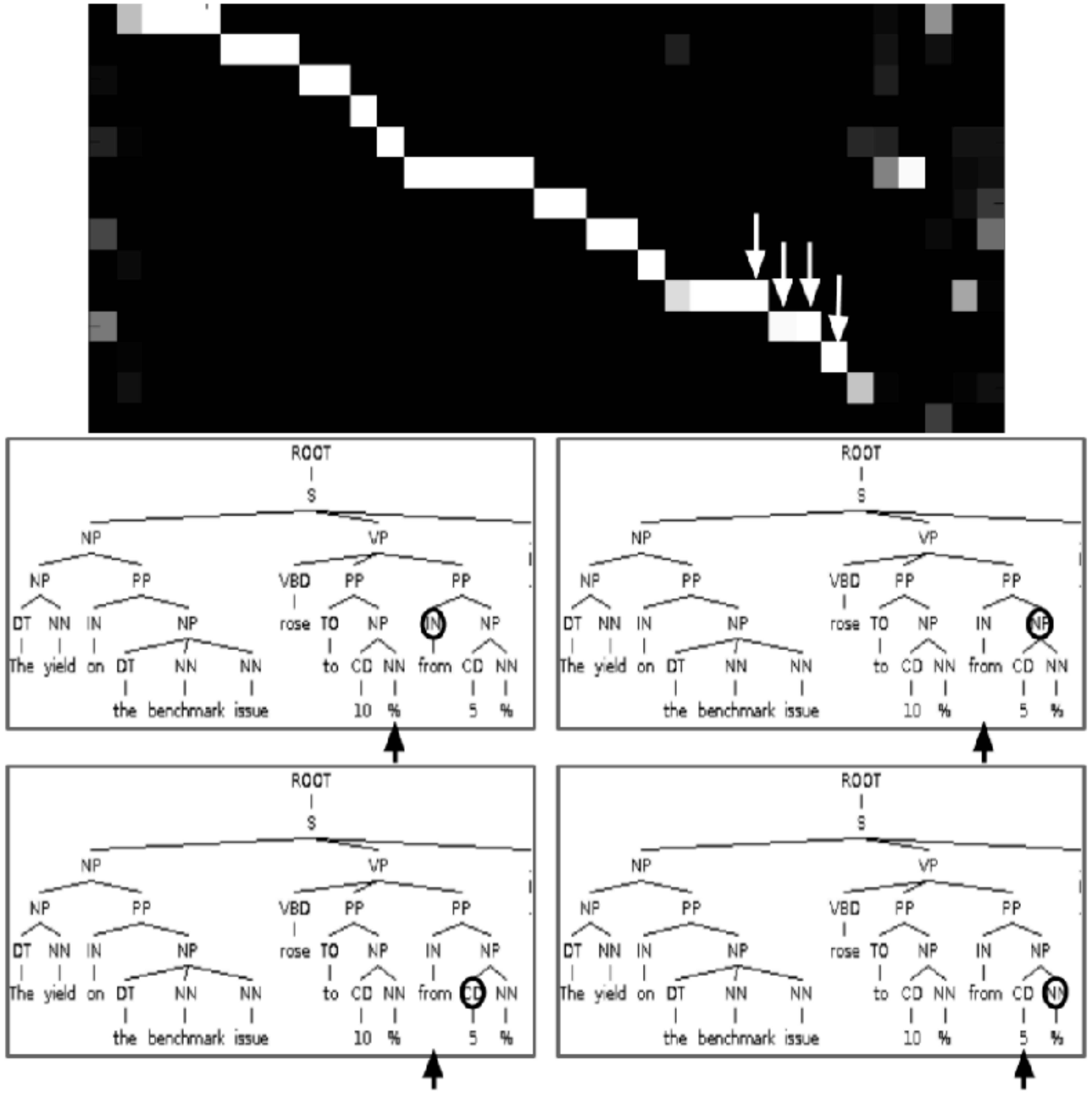
Less Manual Structure?

The yield on the benchmark issue rose to 10% from 5%



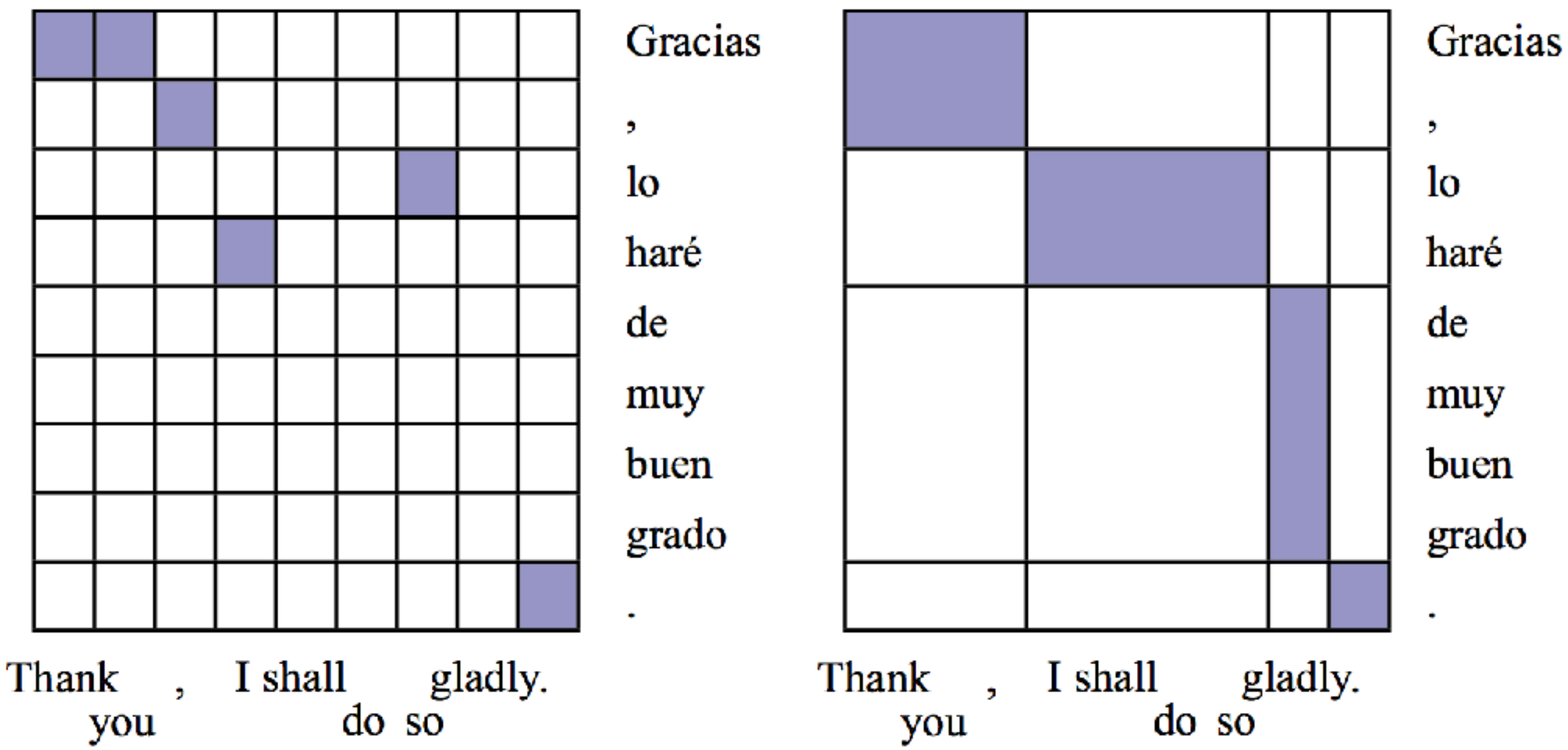
(S (NP (NP (DT The) (NN yield …

- No grammars!

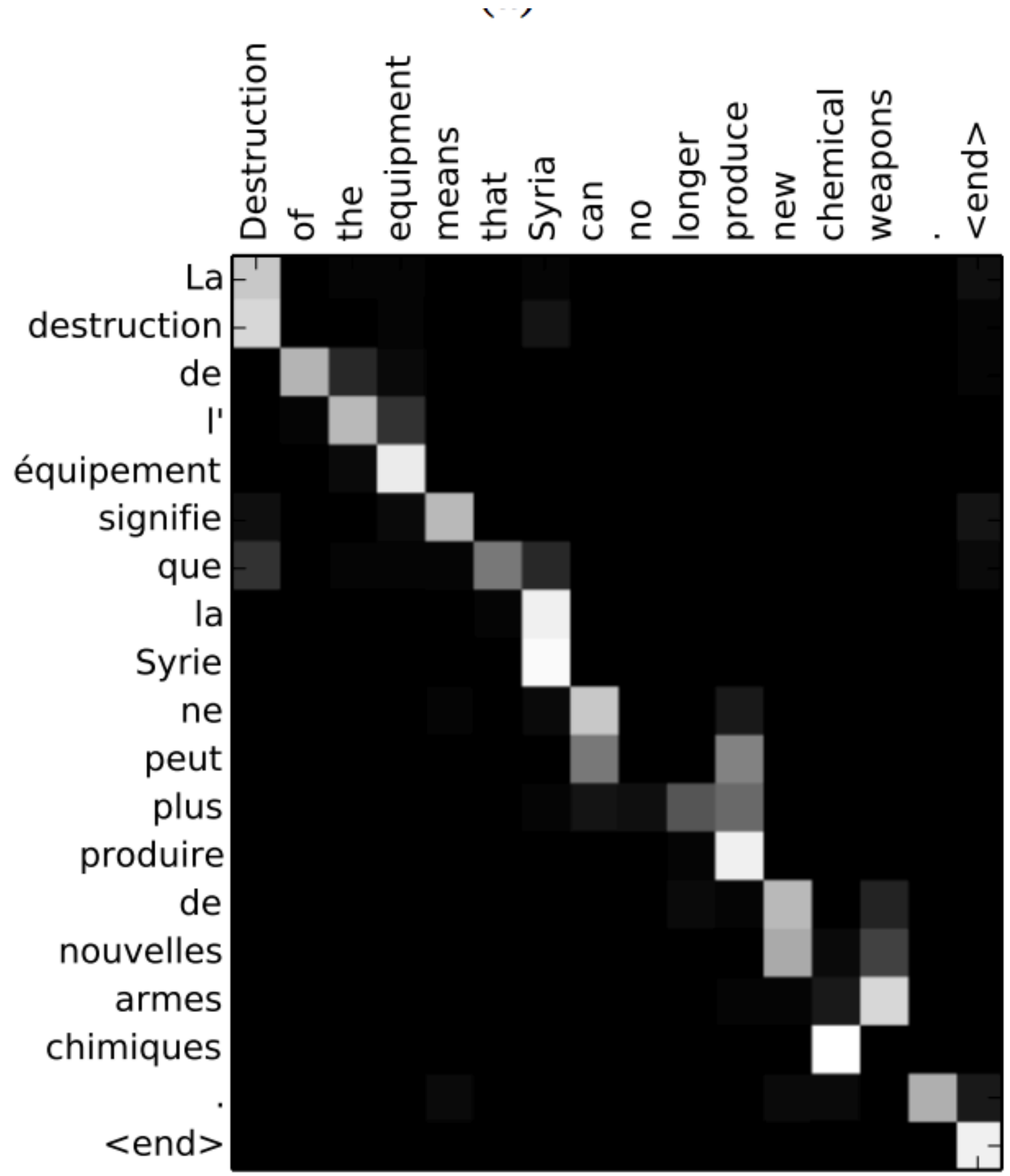


Sutskever et al. (2015), Bahdanau et al. (2014)

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?

- 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)

Does manual structure have a place?

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

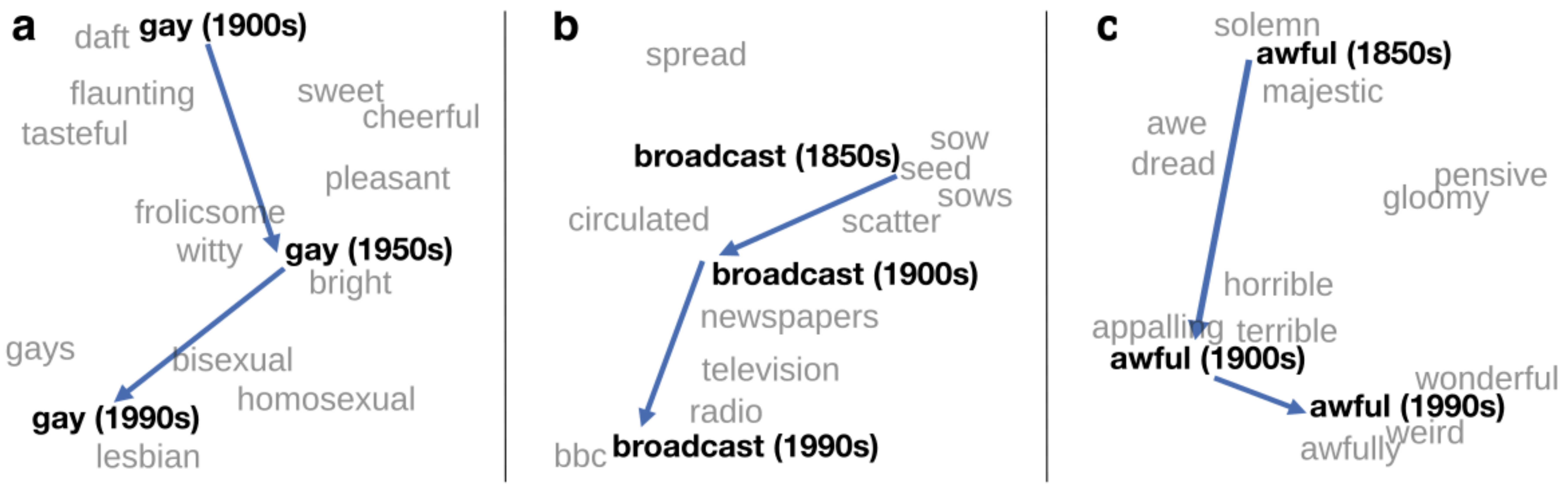
- Maybe manual structure would help...

Where are we?

- 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

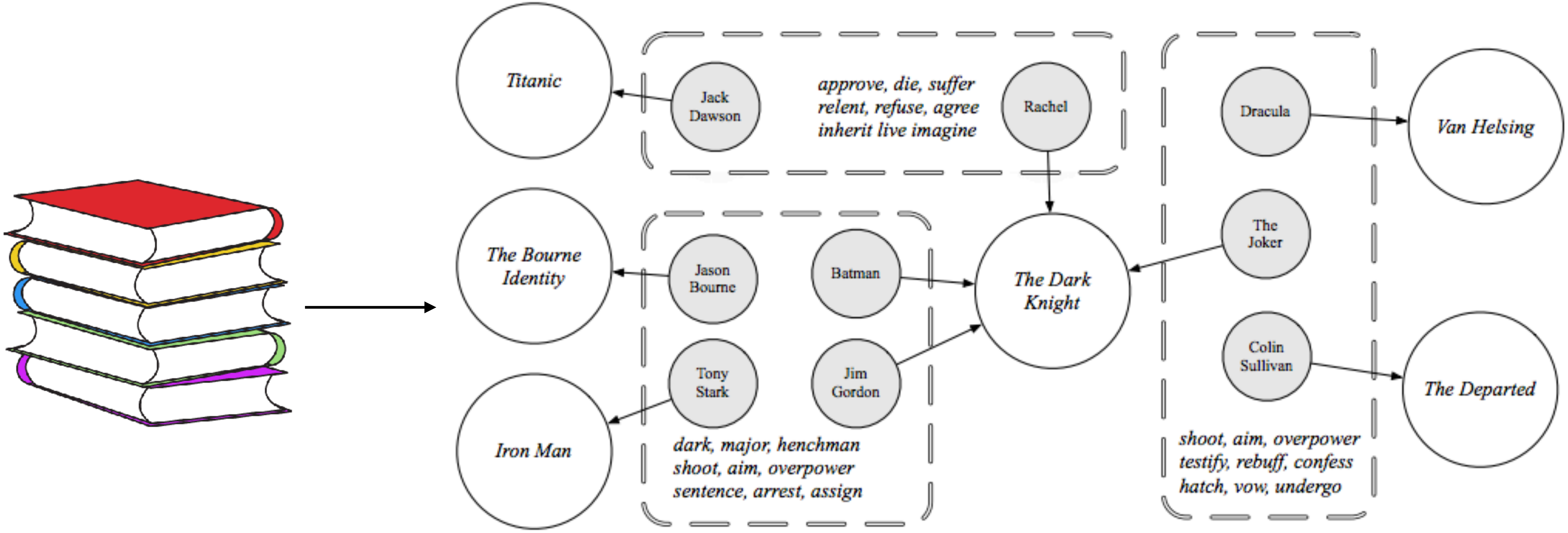
- NLP: build systems that deal with language data
- CL: use computational tools to study language



Hamilton et al. (2016)

NLP vs. Computational Linguistics

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



Bamman, O'Connor, Smith (2013)

Outline of the Course

ML and structured prediction for NLP

Neural nets

Syntax/ semantics

Applications: MT, IE, summarization, dialogue, etc.

Date	Topics	Readings	Assignments
Aug 30	Introduction [4pp]		Mini1 out
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	Mini1 due / Proj1 out
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	Proj1 due / Mini2 out
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		Mini2 due
Oct 16	Semantics II / Seq2seq I		
Oct 18	Seq2seq II: Beam search, attention	Seq2seq, Attention, Luong Attention	Proj2 out
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		Proj2 due
Nov 6	Applications I: Reading comprehension / MemNets	E2E Memory Networks, CBT, SQuAD, BiDAF	
Nov 8	Applications II: Language grounding		FP Proposals due
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			FP due

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 (e.g.: state-of-the-art NER system = project 1 + mini 2, basic MT system = project 2)

Assignments

- Two minis (10% each), two projects (20% each)
  - Implementation-oriented, with an open-ended component to each
  - Mini 1 (classification) is out NOW
  - ~2 weeks per assignment, 5 “slip days” for automatic extensions
- Grading:
  - Minis: 80% for reaching the performance threshold, 20% writeup
  - Projects: 60% for reaching the performance threshold, 20% writeup, 20% extension

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!**

Assignments

- Final project (40%)
  - Groups of 2 preferred, 1 is possible
  - (Brief!) proposal to be approved by me
  - Written in the style and tone of an ACL paper

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. Fill in: Assuming I can enroll, my probability of taking this class is X%
5. One interesting fact about yourself, or what you like to do in your spare time