CS395T: Structured Models for NLP
Lecture 1: Introduction

Greg Durrett

Administrivia

- Lecture: Tuesdays and Thursdays 9:30am - 10:50am
- Course website: http://www.cs.utexas.edu/~gdurrett/courses/fa2017-cs395t.shtml
- Piazza: https://piazza.com/utexas/fall2017/cs395t/home
- My office hours: Wednesday 10am-noon, GDC 3.420
- TA: Ye Zhang; Office hours:
  - Tuesday 2pm-3pm GDC 1.302 Desk 2
  - Thursday 2pm-3pm, GDC 1.302 Desk 1 (until 2:30), Desk 4 (2:30 onwards)

Course Requirements

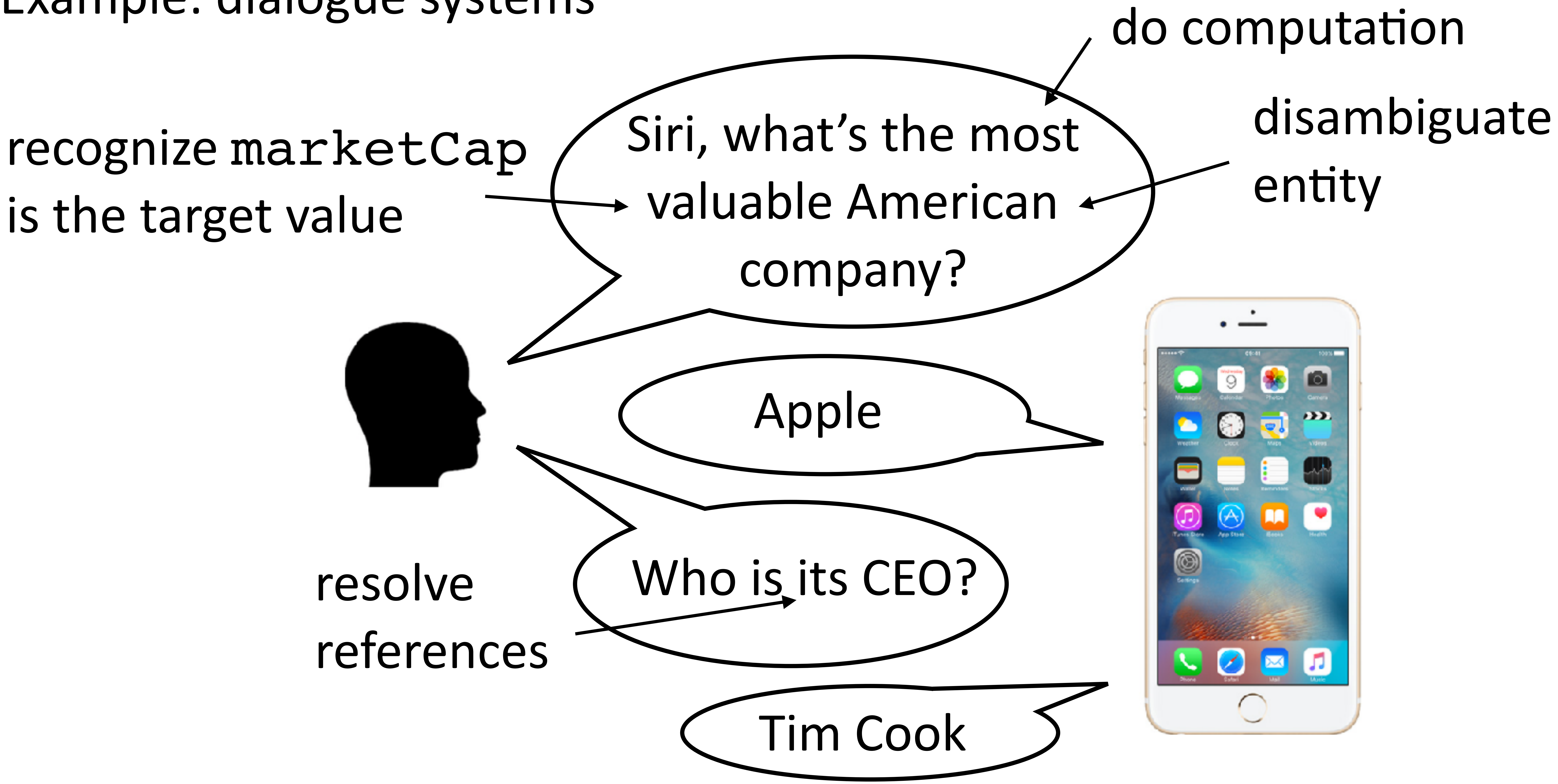
- 391L Machine Learning (or equivalent)
- 311 or 311H Discrete Math for Computer Science (or equivalent)
- Python experience
- Additional prior exposure to probability, linear algebra, optimization, linguistics, and NLP useful but not required

Enrollment

- I want everyone to be able to take this class!
- Priority ordering:
  - CS grad students
  - Other grad students
  - CS undergrads who have satisfied the prerequisites
  - Other undergrads who have satisfied the prerequisites
  - Other undergrads

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

Textual Entailment

Text	Judgments	Hypothesis
A man inspects the uniform of a figure in some East Asian country.	contradiction C C C C C	The man is sleeping
An older and younger man smiling.	neutral N N E N N	Two men are smiling and laughing at the cats playing on the floor.
A black race car starts up in front of a crowd of people.	contradiction C C C C C	A man is driving down a lonely road.

SNLI (Bowman et al., 2015)

- Text is connected to intelligence and knowledge in a fundamental way!
- Goal of NLP (solving problems with text) requires *analyzing* and *understanding* text
- What makes this analysis hard?

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)
- Can try to use the web to learn pragmatics, but that's not giving us a **deep** understanding of text

Language is Ambiguous!

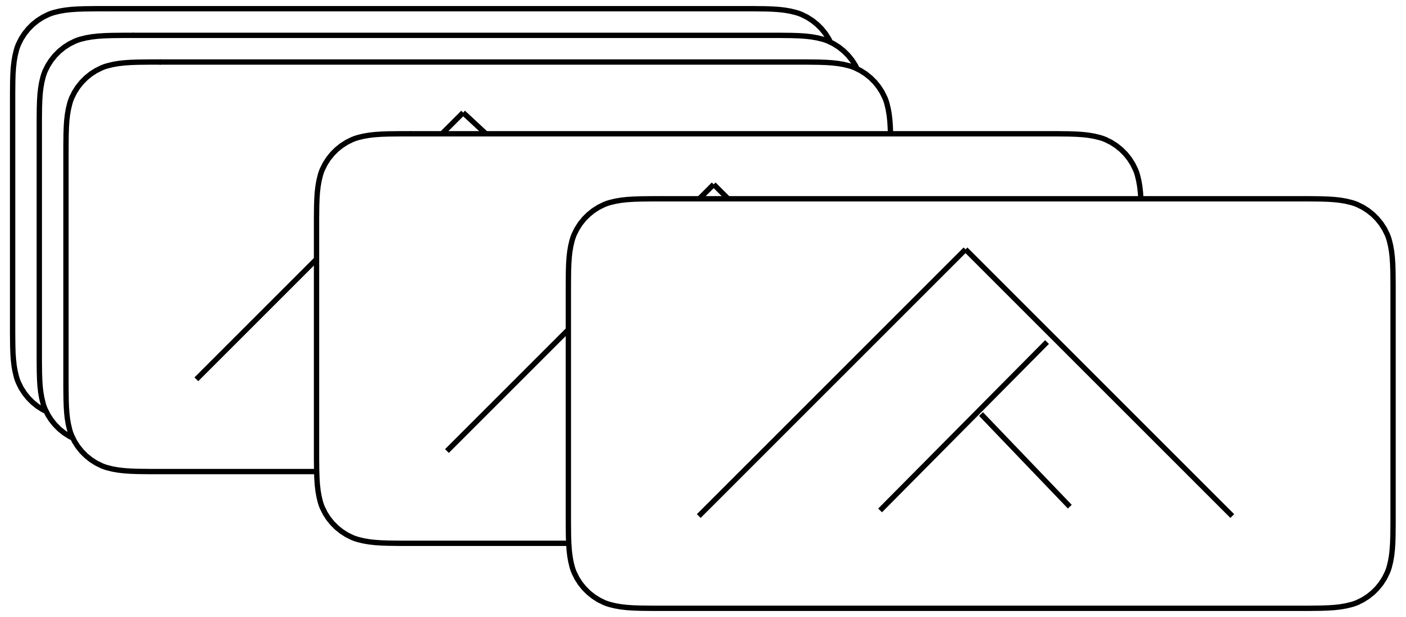
- Headlines
  - Teacher Strikes Idle Kids
  - Hospitals Sued by 7 Foot Doctors
  - Ban on Nude Dancing on Governor's Desk
  - Iraqi Head Seeks Arms
  - Stolen Painting Found by Tree
  - Kids Make Nutritious Snacks
  - Local HS Dropouts Cut in Half
- Why are these funny?
- Pragmatics can resolve this…right?

slide credit: Dan Klein

Language is **Really** Ambiguous!

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

sentence → syntactic parser →



… exponential number

"fish"

→ fish
Phish
ghoti

gh as in tough
o as in women
ti as in motion

- Combinatorially many possibilities, many you won't even register as ambiguities, but systems still have to resolve them!
- So our goal (analyze text) is harder than we thought…how do we do it?

A brief history of (modern) NLP

"AI winter"
rule-based,
expert systems

earliest stat MT
work at IBM

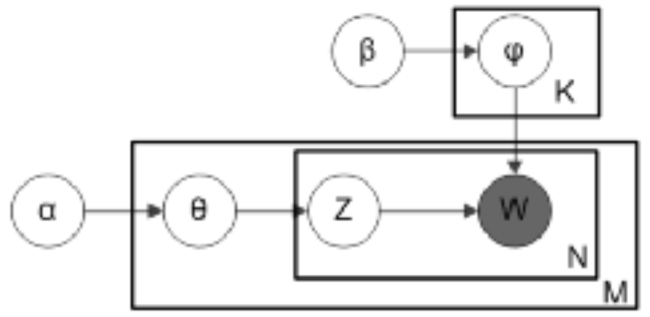
Penn
treebank
S
NP VP

Ratnaparkhi
tagger

NNP VBZ

Collins vs.
Charniak
parsers

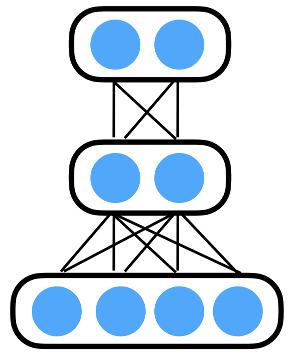
Unsup: topic
models,
grammar induction



Sup: SVMs,
CRFs, NER,
Sentiment

Semi-sup,
structured
prediction

Neural

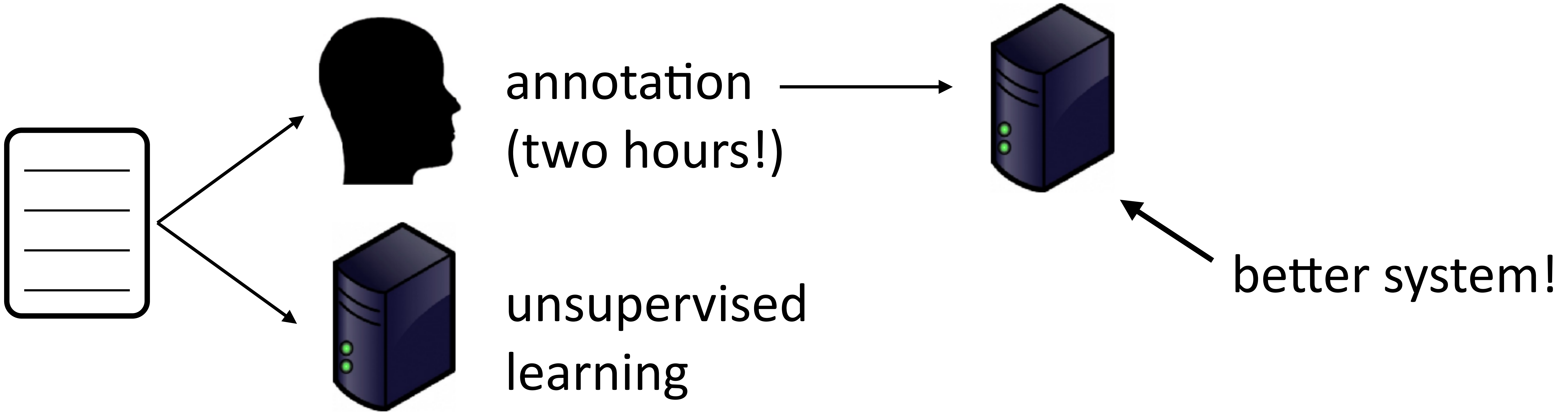


1980 1990 2000 2010 2017

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

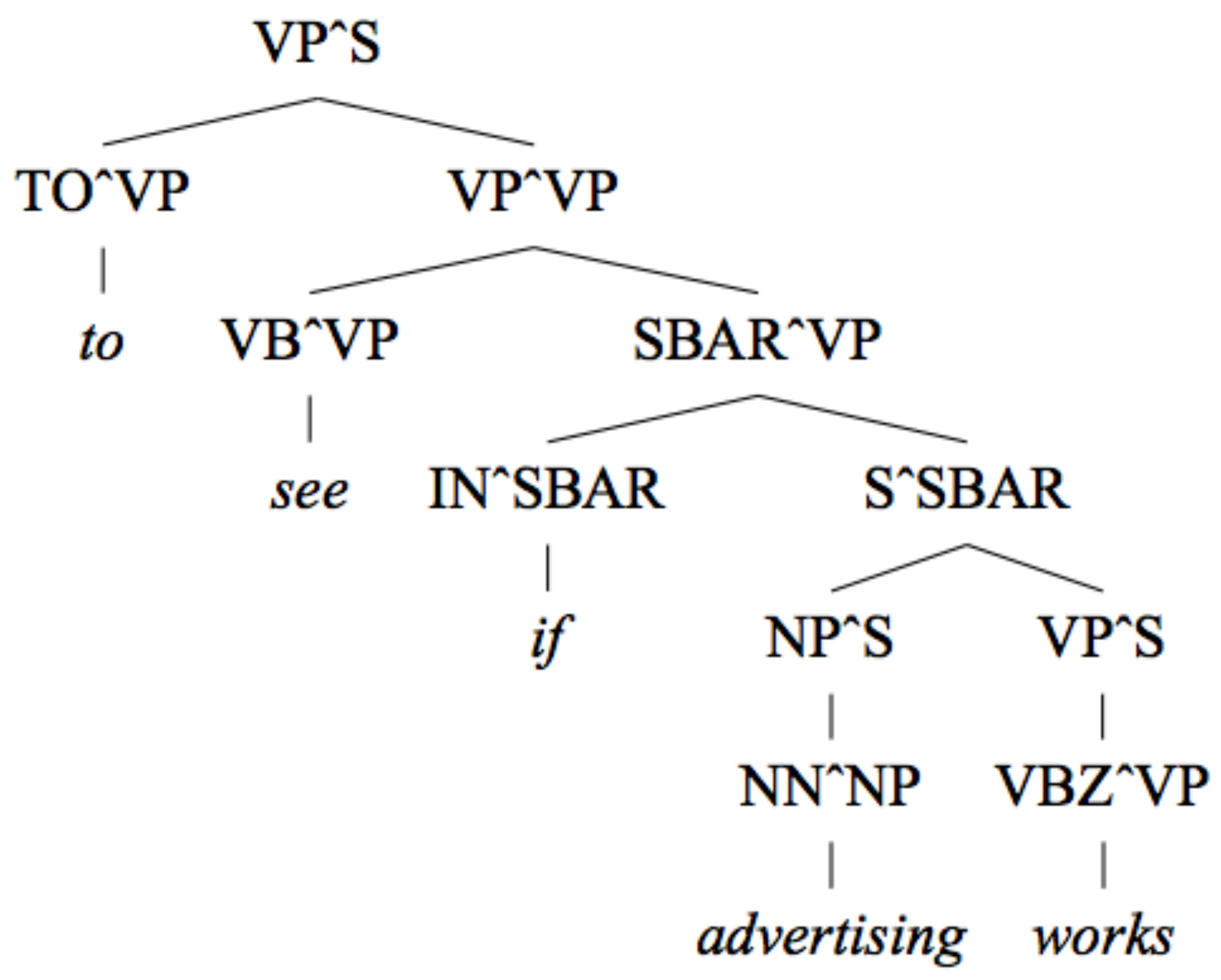


- Even neural nets can do pretty well!

- Balance tradeoff of data/algorithms/compute

Garrette and Baldridge (2013)

Less Manual Structure?

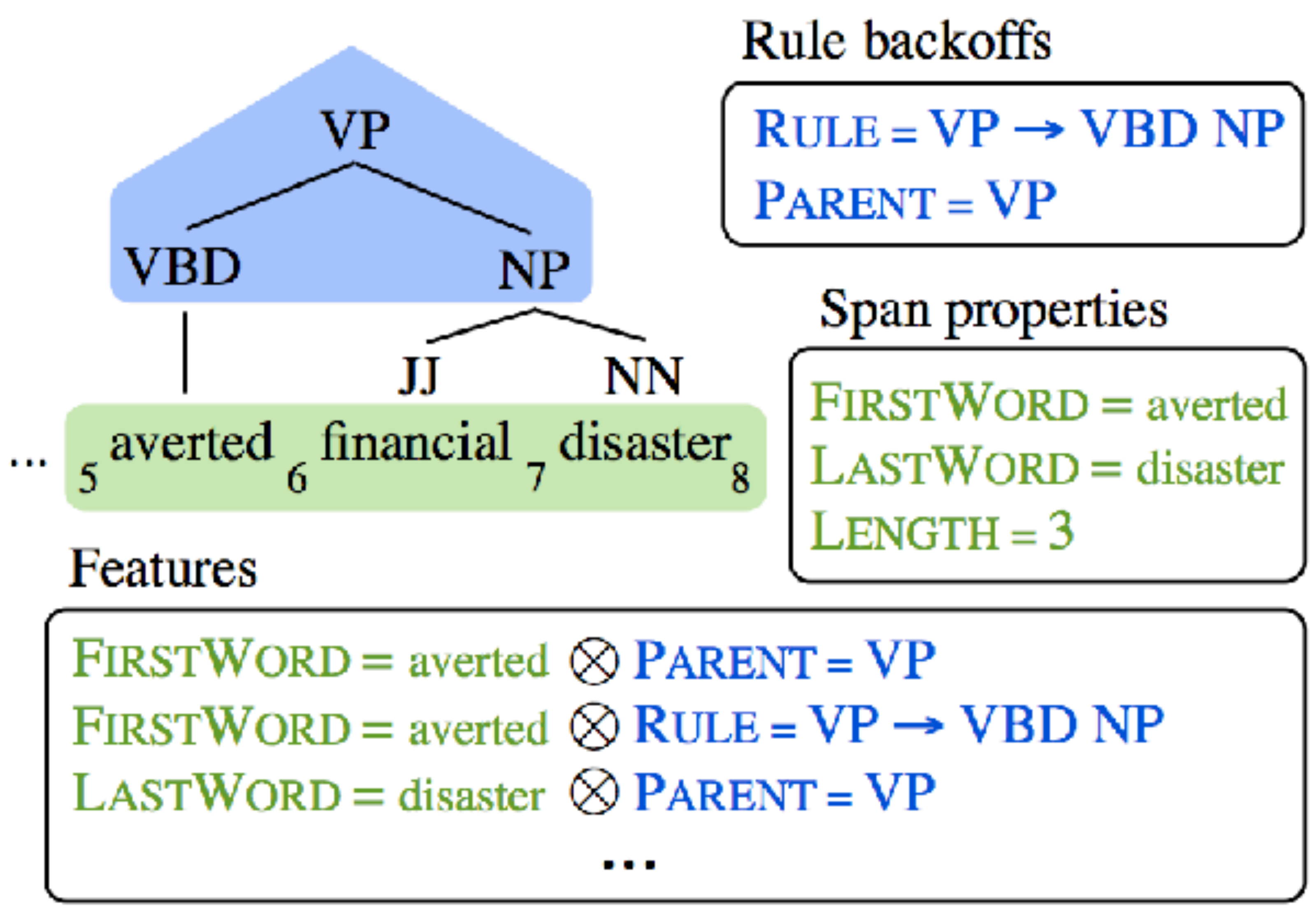


Klein and Manning (2003)

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

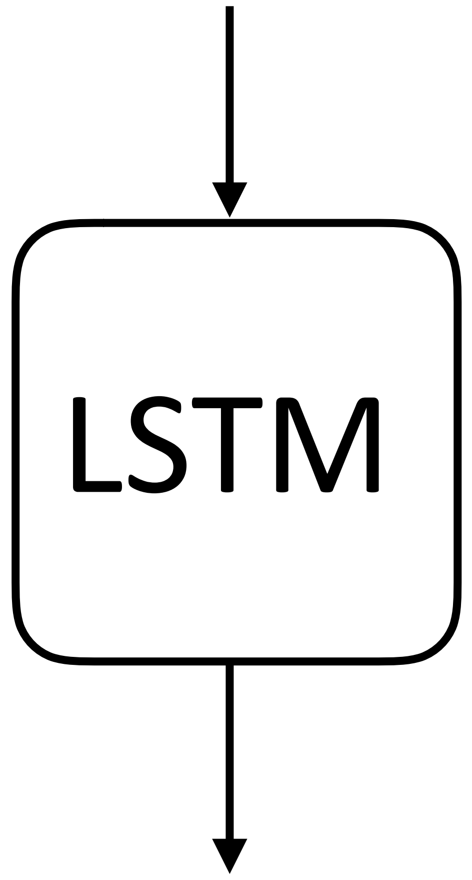
Petrov et al. (2006)



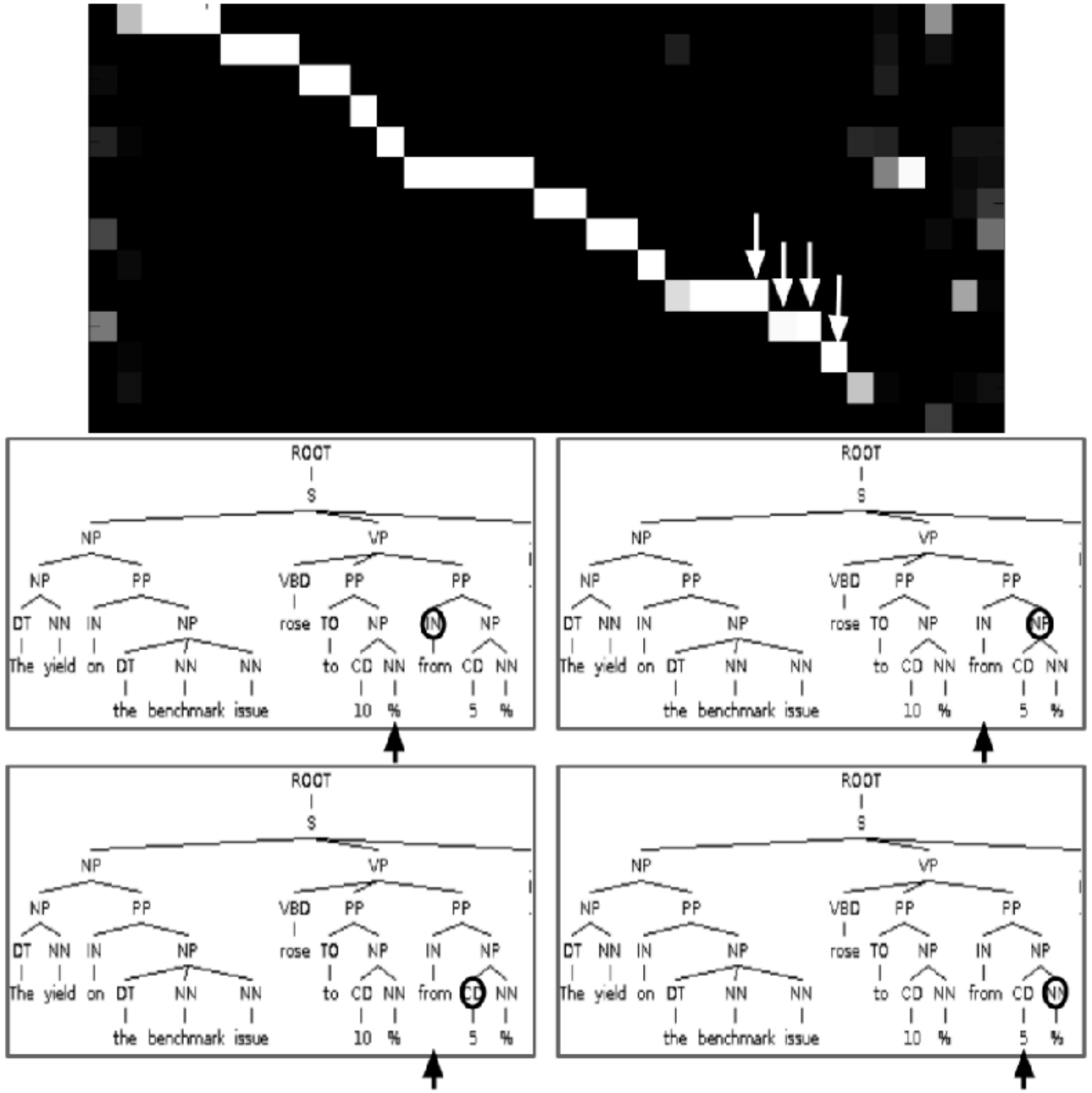
Hall, Durrett, Klein (2014)

Less Manual Structure?

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

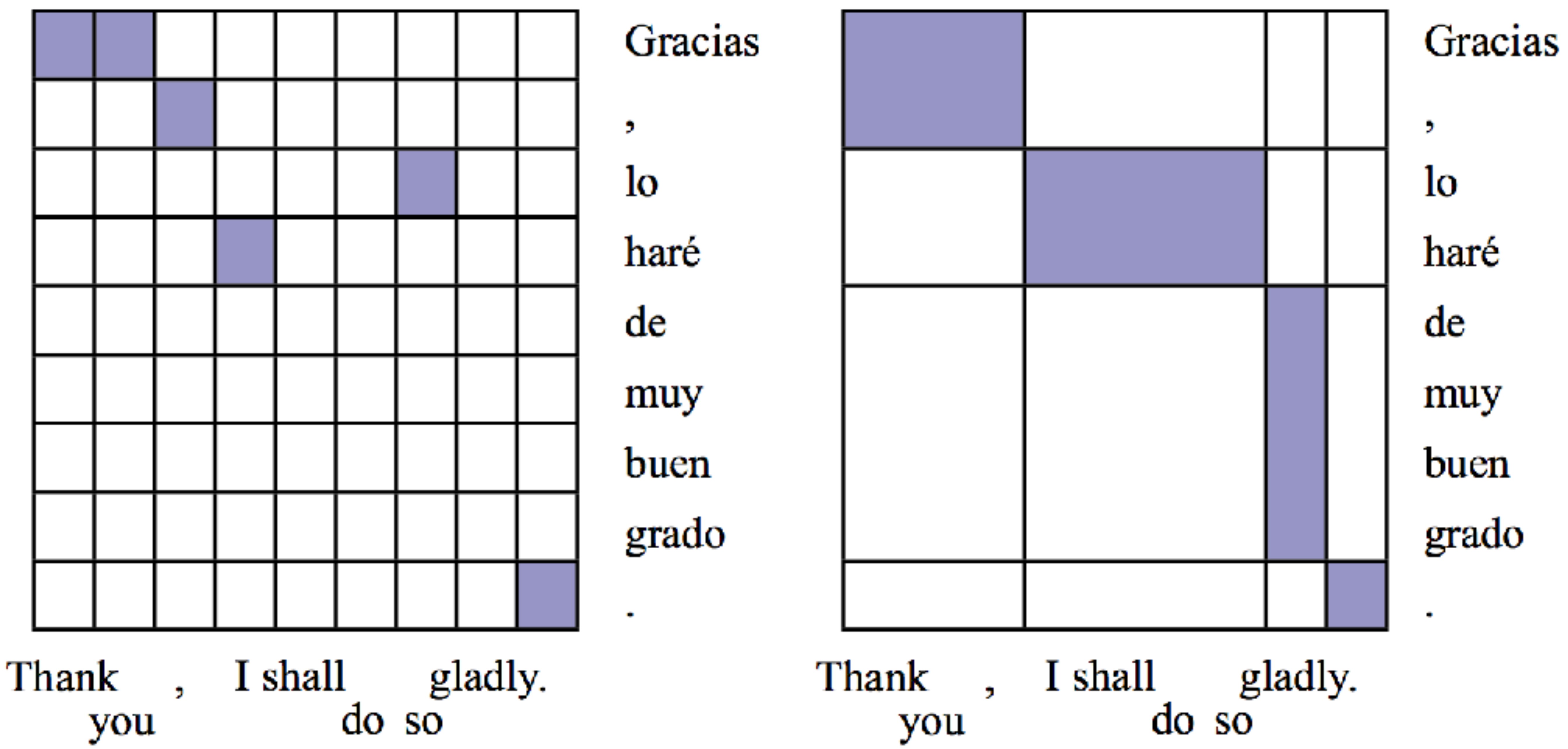


(S (NP (NP (DT The) (NN yield …

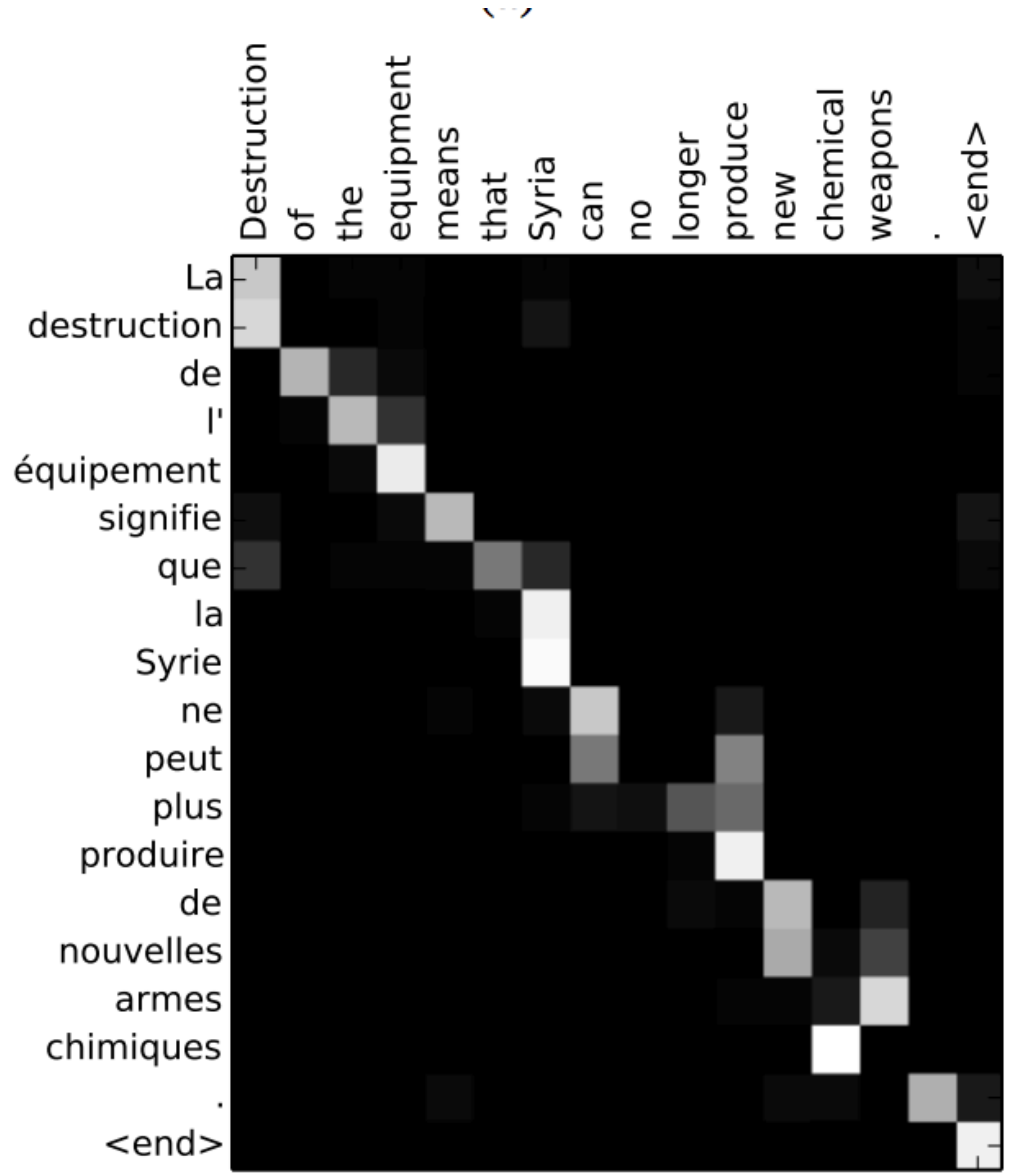


Sutskever et al. (2015), Bahdanau et al. (2014)

Less Manual Structure?



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



DeNero et al. (2008)

Bahdanau et al. (2014)

Less Manual Structure?

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

- Maybe manual structure would help…

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?

- Solving problems with text requires analyzing text
- Many possibilities: rule-based systems, CRFs, neural networks, …
- Knowing which of these to use requires understanding dataset size, problem complexity, and a lot of tricks!
- What do all of these models have in common? What do they need to capture in order to be successful?

Break!

What's important?

- High-capacity models + 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's important?

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

IBM Watson

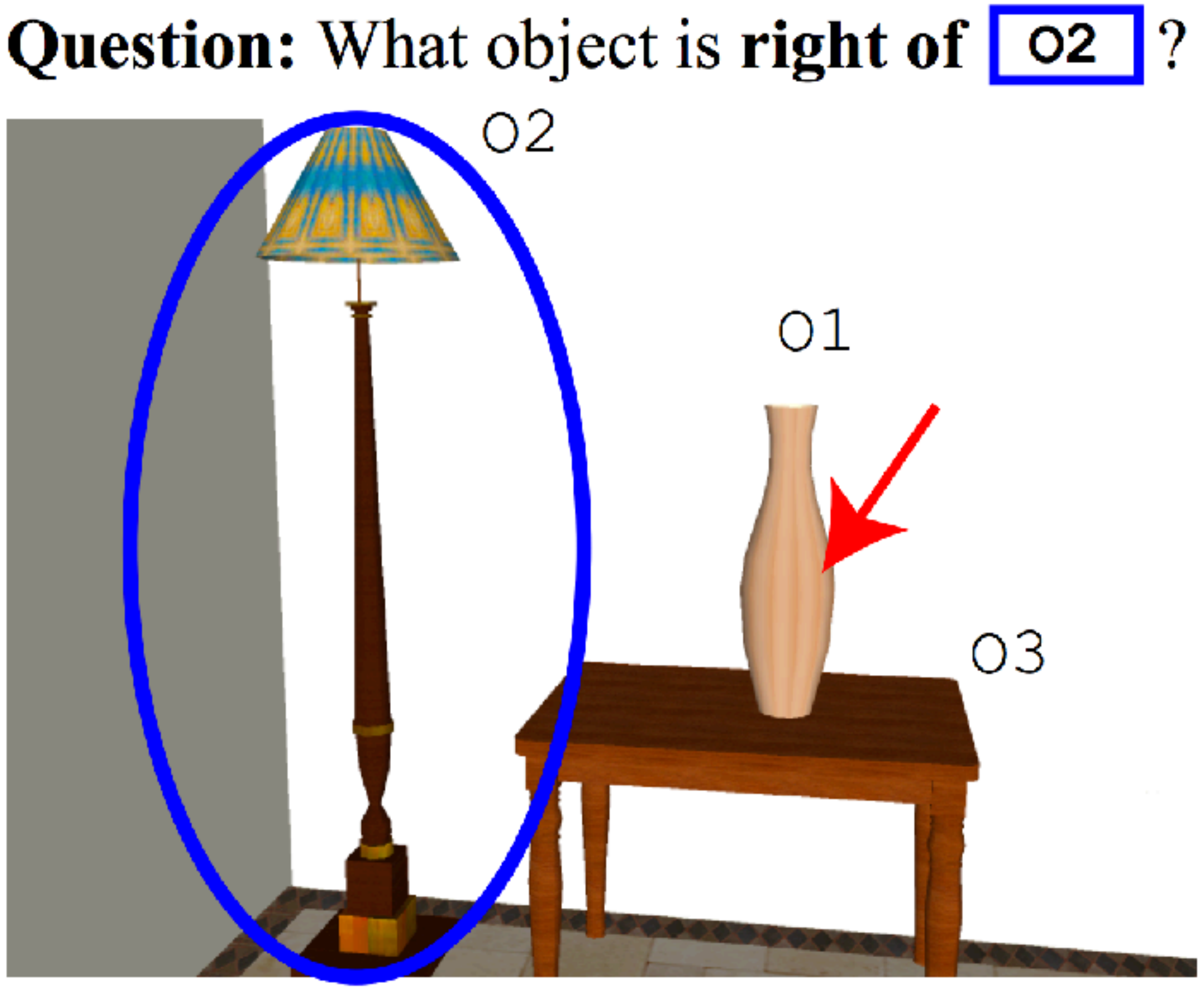
On Sept. 1, 1715 Louis XIV died in this city, site of a fabulous palace he built.

Died	1 September 1715 (aged 76) Palace of Versailles, Versailles, France

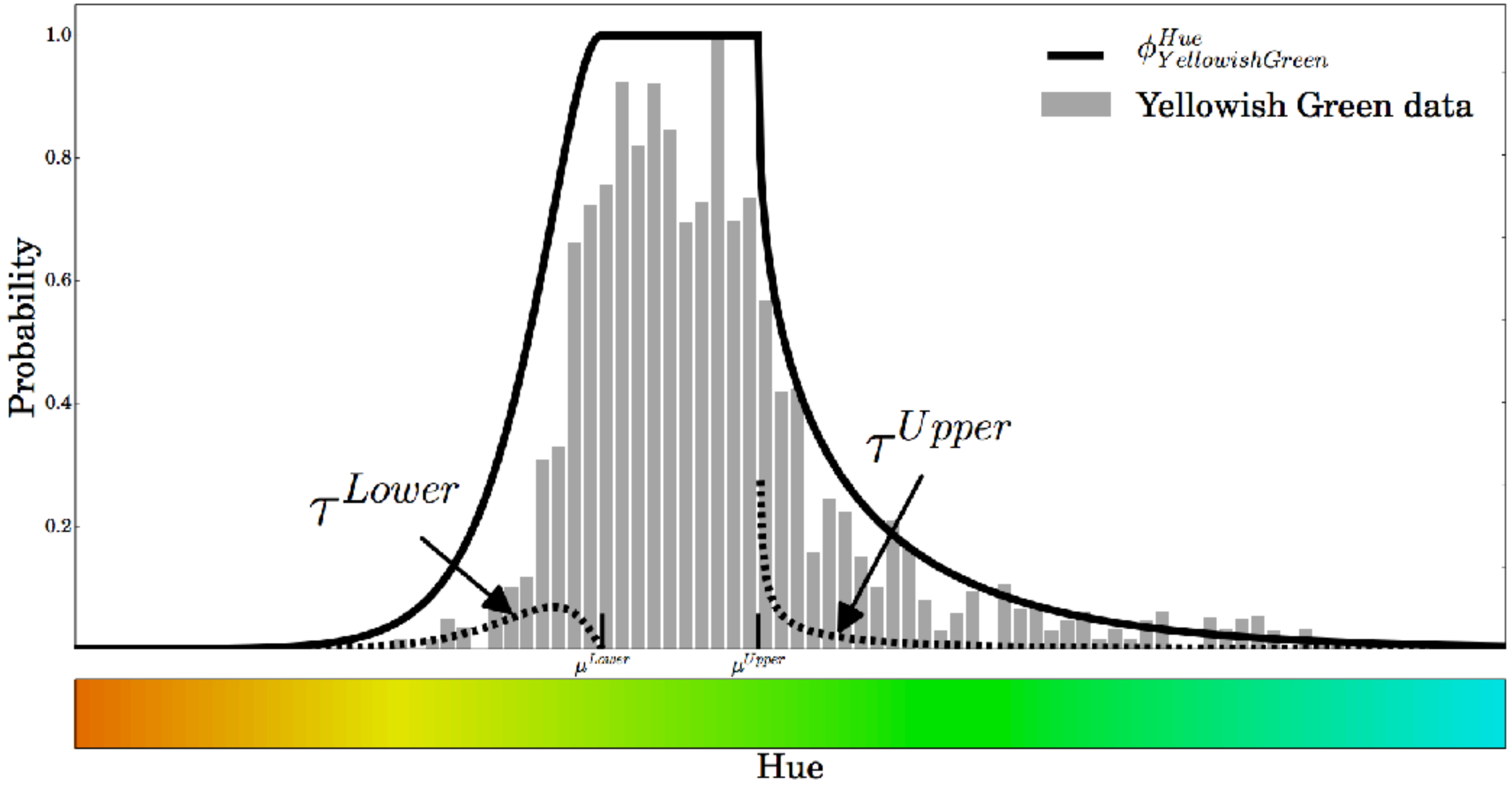
Answer: What is **Versailles**?

What's important?

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



Golland et al. (2010)



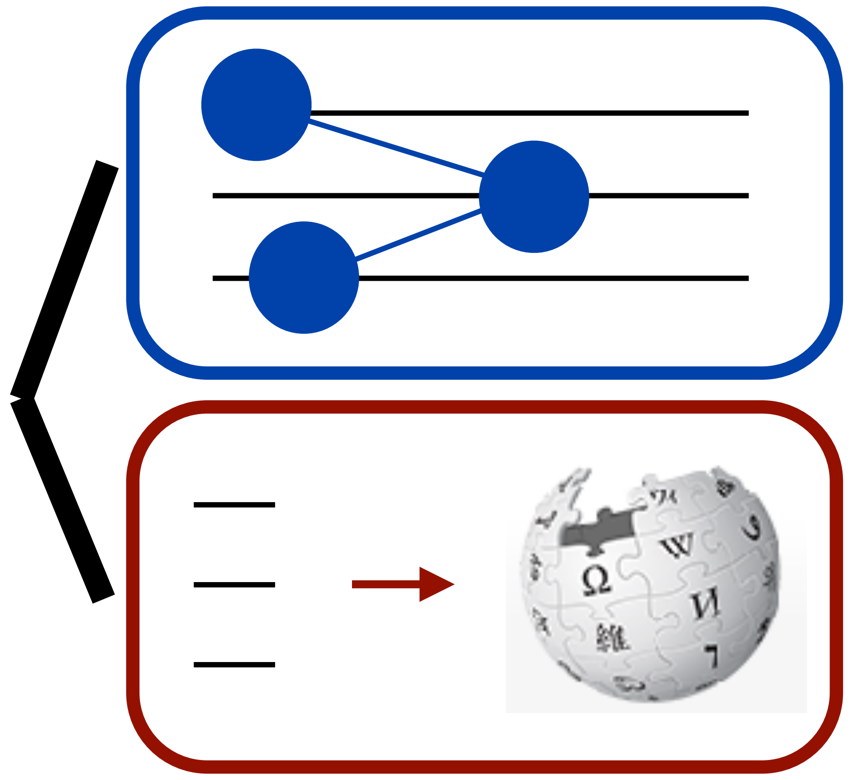
McMahan and Stone (2015)

What’s important?

- Multitask interactions: recognize constraints to be more statistically efficient (and humanlike!) in our reasoning

Dell is headquartered just outside Austin. The company…

coreference resolution



entity disambiguation

Durrett and Klein (2014)

What's important?

- 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)

How do we build systems to do all this?

- Structured statistical models

- **Structured:** lets us incorporate cross-task constraints, inductive biases from linguistics, knowledge, etc.

- **Statistical:** harness the power of data to do really large-scale pattern recognition and learn from labeled + unlabeled data + interaction with the world

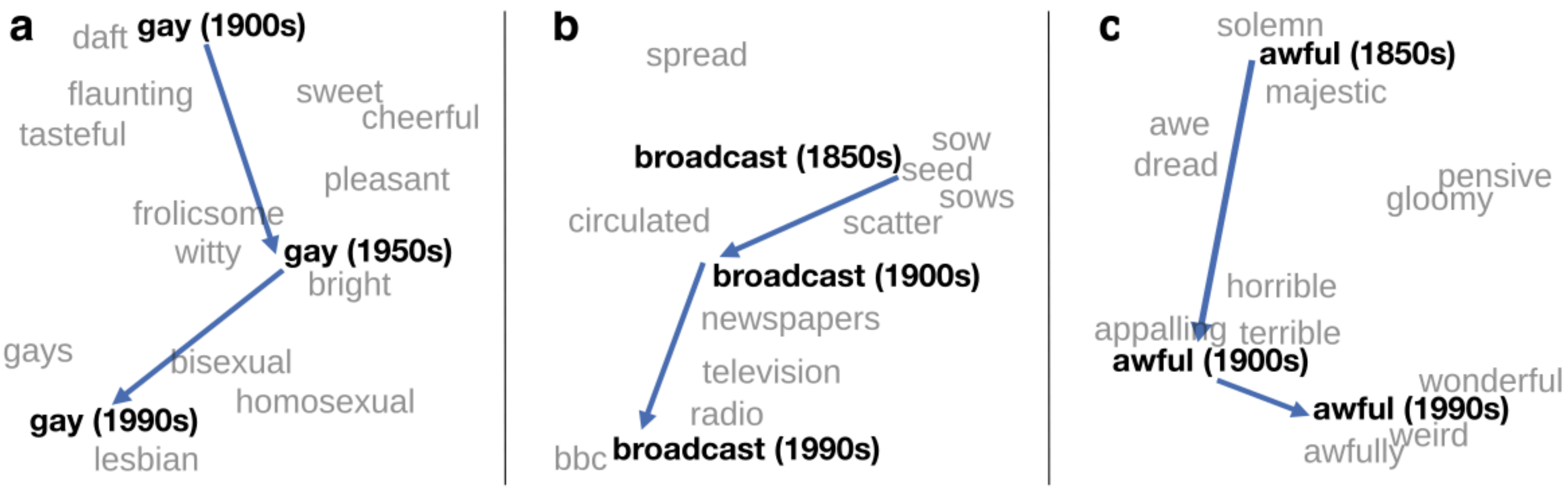
Outline of the Course

- First half: structured prediction
  - Machine learning basics
  - Sequences, trees
  - Inference, learning
- Second half: deep learning
  - RNNs/LSTMs, convolutional networks
  - Word representations
  - Inference, learning

Date	Topics	Readings	Assignments
Aug 31	Introduction		
Sept 5	Machine learning / Classification review	JM 6.1-6.3	
Sept 7	Multiclass classification	JM 7	
Sept 12	Sequence models I: HMMs	JM 9, JM 10.4	P1 out
Sept 14	Sequence models II: CRFs	Sutton CRFs 2.3, Illinois NER	
Sept 19	Sequence models III: Unsupervised	Painless	
Sept 21	Tree-structured models I: Constituency	JM 13.1-13.7, Unlexicalized, Lexicalized, State-split	
Sept 26	Tree-structured models II: Dependency I	JM 14.1-14.4	
Sept 28	Tree-structured models III: Dependency II		P1 due / P2 out
Oct 3	General graphical models I: "loopy" models, etc.	Skip-chain NER, Joint entity	
Oct 5	General graphical models II: ILP models		
Oct 10	Machine Translation		
Oct 12	Neural net basics, word representations	Goldberg 1-6	P2 due
Oct 17	RNNs I: LSTMs, encoder-decoder	Goldberg 10	P3 out
Oct 19	RNNs II: Attention	MT	
Oct 24	RNNs III		
Oct 26	CNNs I	Goldberg 9	
Oct 31	CNNs II / Advanced NNs		
Nov 2	Advanced NNs: Memory networks / pointer networks		P3 due / FP out
Nov 7	Advanced NNs II		
Nov 9	Deep Generative Models: GANs, VAE		Proposals due
Nov 14	Special Topics I		
Nov 16	Special guest lecture: Katrin Erk		
Nov 21	Special Topics II		
Nov 23	NO CLASS (Thanksgiving)		
Nov 28	Special Topics III		
Nov 30	Special Topics IV		
Dec 5	Wrapup		
Dec 7	Project presentations (TBD)		
Dec 15			FP due

NLP vs. Computational Linguistics

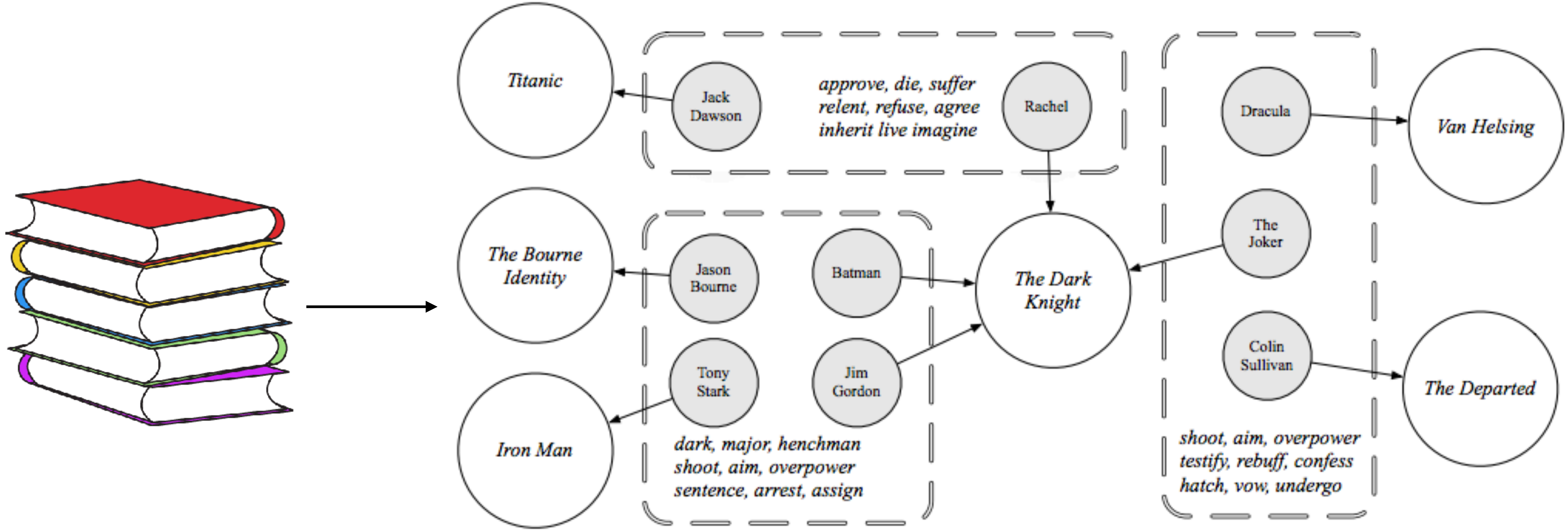
- NLP: build systems that deal with language data
- CL: use computational tools to study language



Hamilton et al. (2016)

NLP vs. Computational Linguistics

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



Bamman, O'Connor, Smith (2013)

Course Goals

- Cover structured machine learning approaches to NLP
  - Show connections between structured algorithms: generative and discriminative, margin and likelihood, neural and linear, etc.: these are all closely related!
  - Dissect the pieces of these structured models: modeling, inference, learning
- Make you a "producer" rather than a "consumer" of NLP tools
- Expose you to classic problems in NLP

Assignments

- Three projects (16.6% each = 50%)
  - Implementation-oriented, open-ended component to each
  - First will be out on 9/12
  - 2-page writeup with statement of what you did
  - ~2 weeks per project, 7 "slip days" for automatic extensions

- Grading: 10-point scale
  - 6 points for minimal code completion
  - 1 point for minimal extension
  - 1 point for minimal 2-page writeup
  - 2 points for better extension, better writeup

8 points ~ A-

Assignments

- Final project (50%)
  - Groups of 1-2
  - (Brief!) proposal to be approved by me
  - Written in the style and tone of an ACL paper
  - Same 10-point grading scheme, 8 points for minimal completion of proposed work

Survey

1. Fill in: I am a [CS / linguistics / other] [grad / 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. HMMs
    4. EM
    5. Part-of-speech tagging
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%