Question Answering as Global Reasoning over Semantic Abstractions

Daniel Khashabi, Dan Roth *(University of Pennsylvania)*
Tushar Khot, Ashish Sabharawal *(Alllen Institute for Artificial Intelligence)*

Feb 6, 2018

ALLEN INSTITUTE
for ARTIFICIAL INTELLIGENCE

Exams

Standardized science exams (Clark et al, 2015):

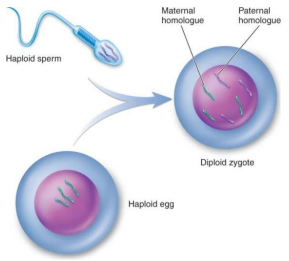
- Broad knowledge (general and scientific)
- Connecting question to

Q: Which physical structure would best help a bear to **survive a winter** in New York State?
A: (A) big ears (B) black nose (C**) thick fur** (D) brown eyes
P: … Polar bears, saved from the bitter cold by their thick fur coats**,** are among the animals in danger …

Biology exams (Berant et al, 2014):

- Technical terms and answer not easy to find.
- Requires understanding complex relations.

Q: What does meiosis directly produce?
(A) Gametes (B) **Haploid cells**
P: … Meiosis produces not gametes but haploid cells that then divide by mitosis and give rise to either unicellular descendants or a haploid multicellular adult organism. Subsequently, the haploid organism carries out further **mitoses, producing the cells** that develop into gametes.



Penn
University of Pennsylvania

Linguistic variability

Which physical structure would best **help a bear to survive a winter**?

(A) big ears (B) black nose (C**)** **thick fur** (D) brown eyes

Thick fur helps a bear survive a winter.

A thick coat of white fur helps bears survive in these cold latitudes.

Polar bears, saved from the bitter cold by their thick fur coats, are among the animals in danger of extinction because of the global warming and human activities.

A given “meaning” can be phrased many surface forms!

QA is a language understanding problem!

verb

Which physical structure would best help a bear to survive a winter?

(A) big ears (B) black nose (C) **thick fur** (D) brown eyes

comma

preposition

Polar bears, saved from the bitter cold by their thick fur coats, are among the animals in danger of extinction because of the global warming and human activities.

QA is fundamentally a NLU problem

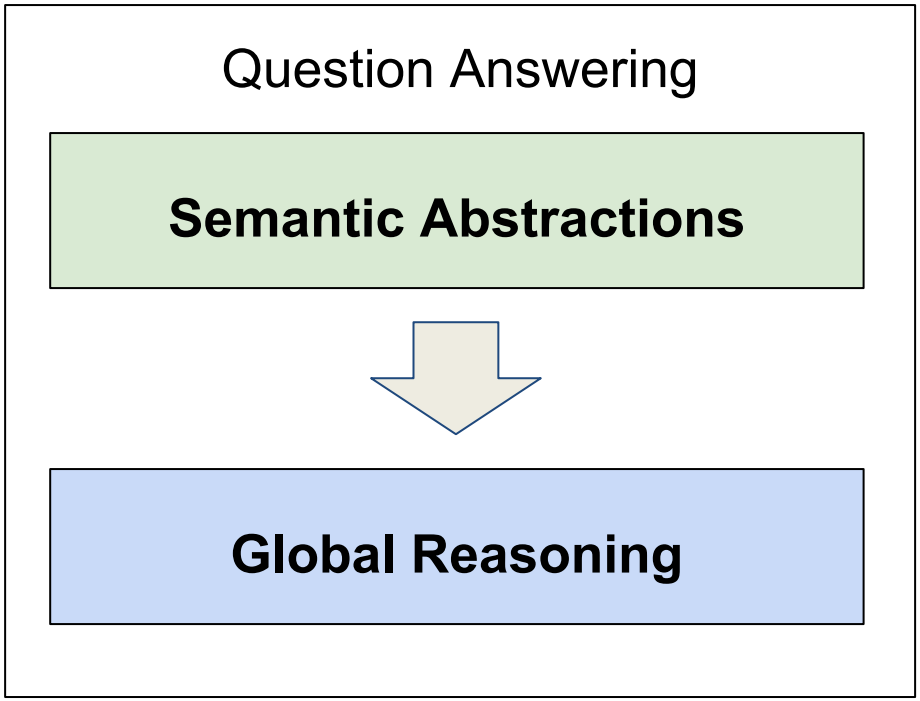
A single abstraction is not enough

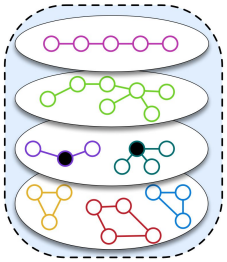
High-level view

Question Answering

as Global Reasoning

over Semantic Abstractions



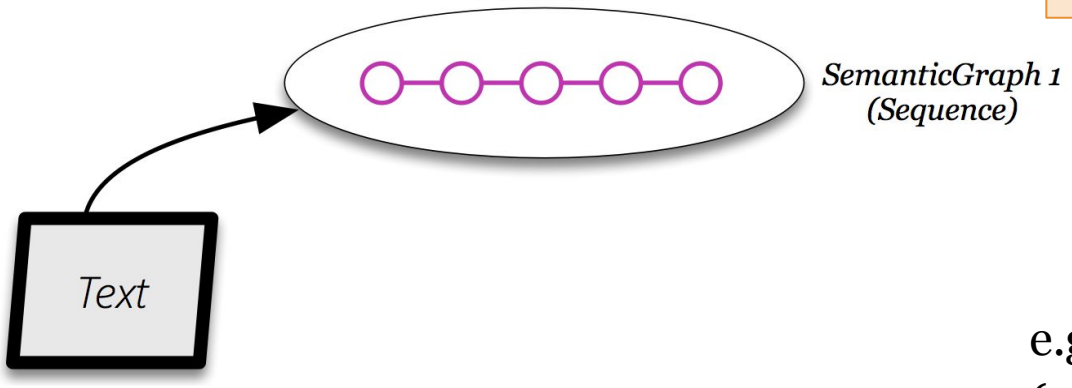


Collections of semantic graphs

Create a unified representation as a family of graphs

- predicate-argument, trees, clusters, sequences

A single representation is not enough to capture complexity of language



e.g named-entities

e.g dependency parse

e.g semantic role labeling (verb, preposition, comma)

e.g co-reference

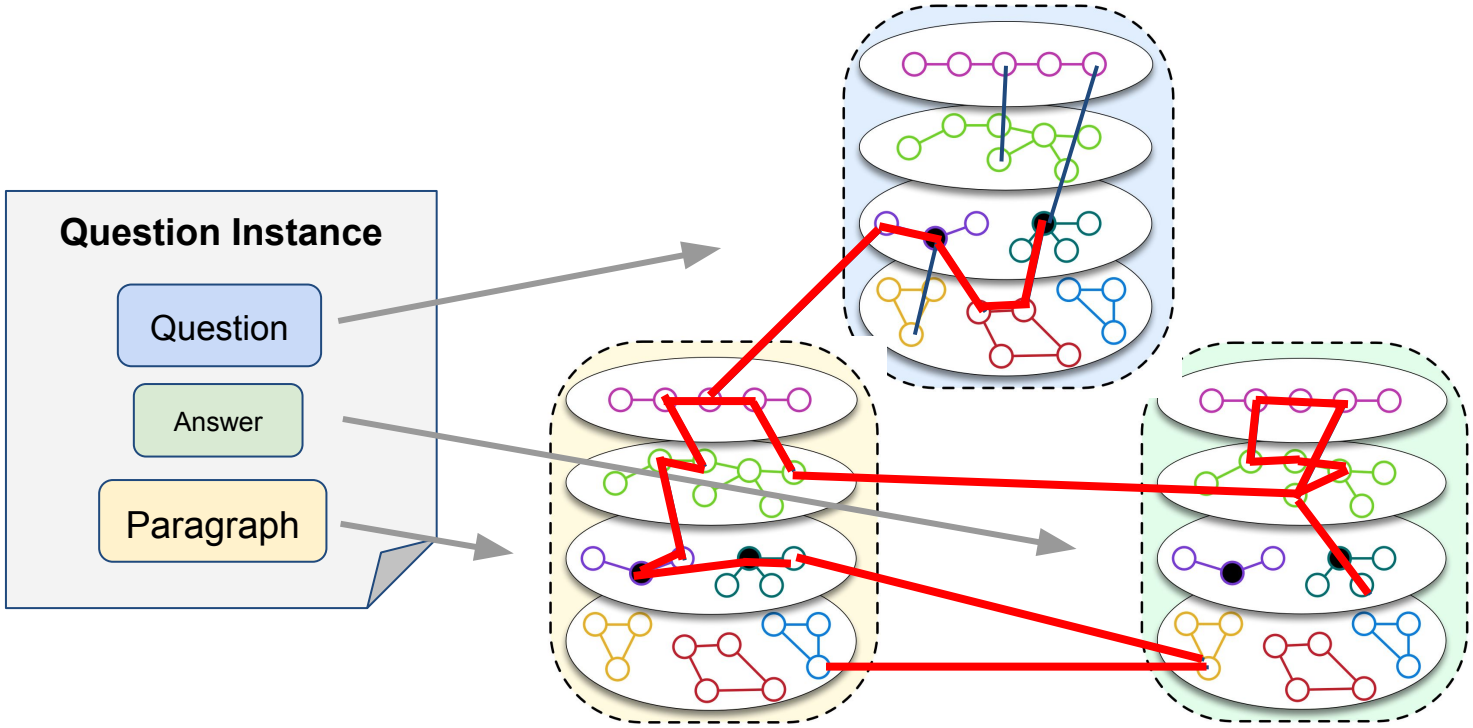
e.g tables

Our representation has nothing to do with the QA task. It reflects our understanding of the language

Reasoning With a Meaning Reprentation

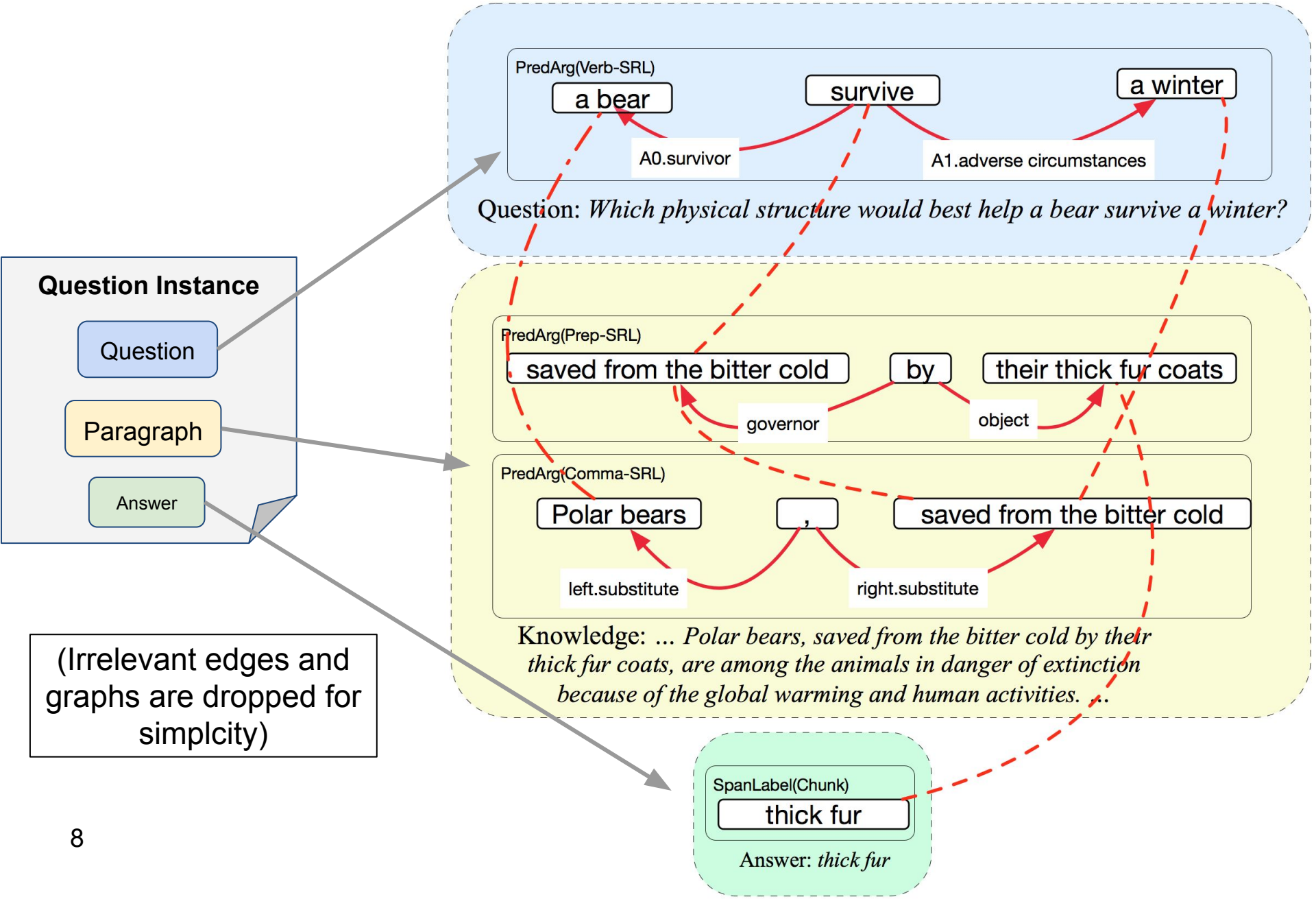
- **Augmented Graph** is the graph which contains potential alignments between elements of any two graphs

Connections via similarity / entailment



Reasoning formulated as best subgraph reasoning

Example subgraph



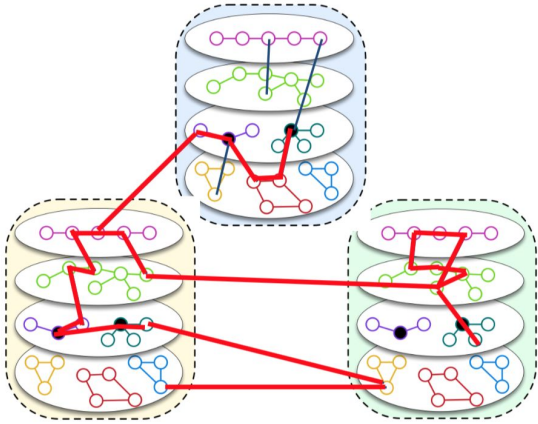
(Irrelevant edges and graphs are dropped for simplcity)

SemanticILP, some details.

Translate QA into a search for an optimal subgraph

Constraint: Incorporate **global** and **local** constraints

- **Global** e.g.
  - Have ends in question and paragraph
  - Connected graph
- **Local** e.g.
  - If using a pred-arg graphs,
    - use at least predicate and argument, or
    - use at least two arguments



Objective: Capture what's a valid reasoning, what's preferred

- **Preferences** e.g.
  - Use sentences nearby
  - If using a pred-arg graph, give priority to the subject

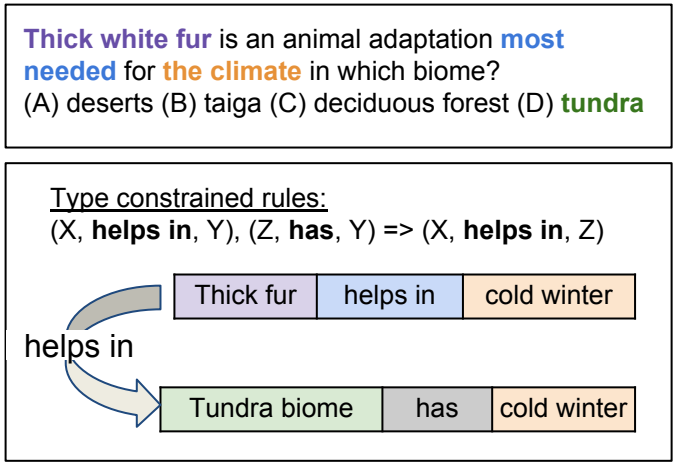
Formulate as Integer Linear Program (ILP) optimization

- Solution points to the best supported answer

Evaluation: notable baselines

- IR **(Clark et al, AAAI'15)**
  - Informatin retrieval baseline (Lucene)
  - Using 280 GB of plain text

- TupleINF **(Khot et al, ACL'17)**
  - Inference over **independent rows**
  - **Auto-generated short triples**
  - And type-constrained rules



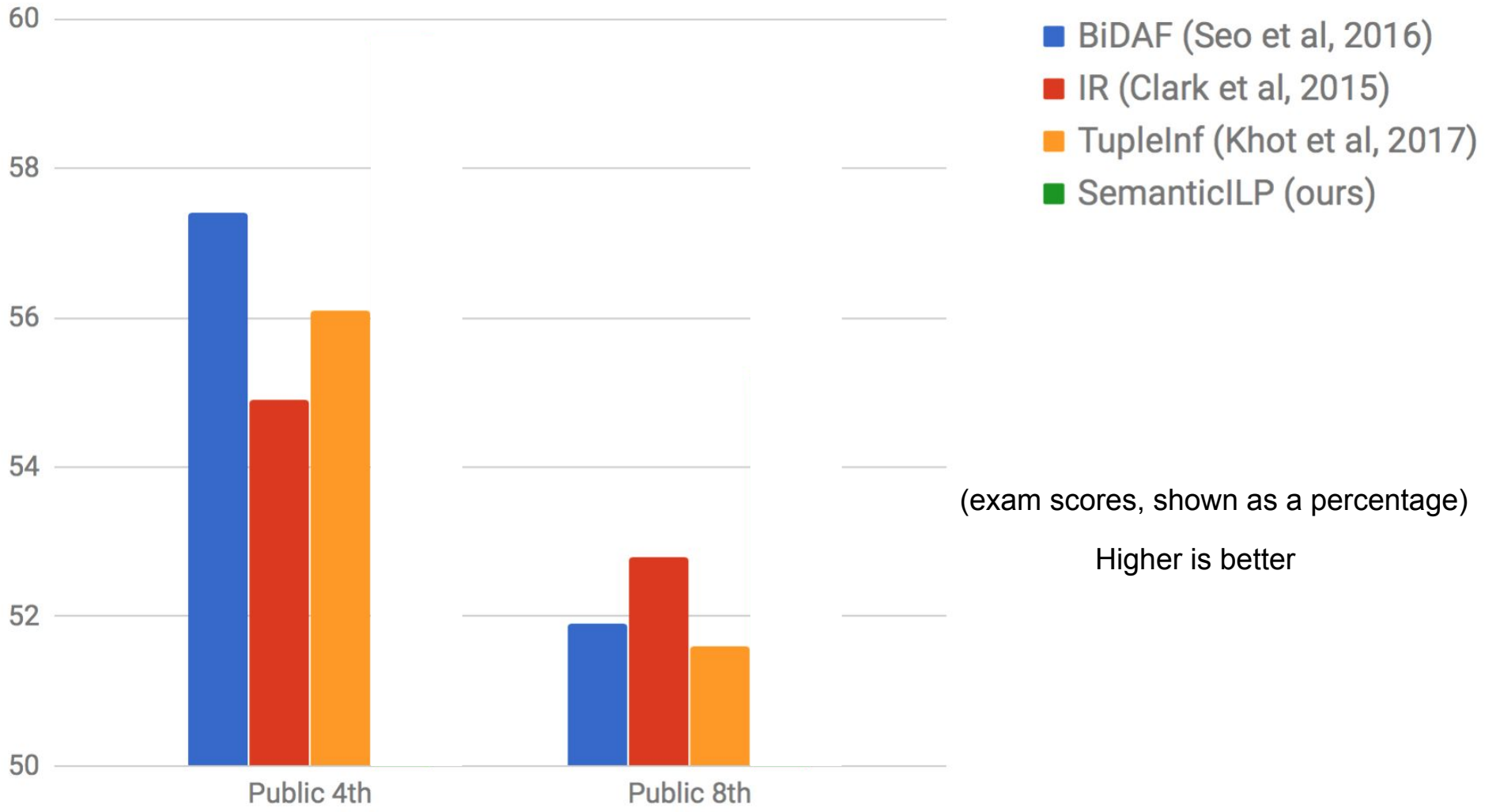
- BiD
  -
  -

We compare with the best baseline on each domain.

However we use one version of our systems across all the datasets.

Penn
University of Pennsylvania

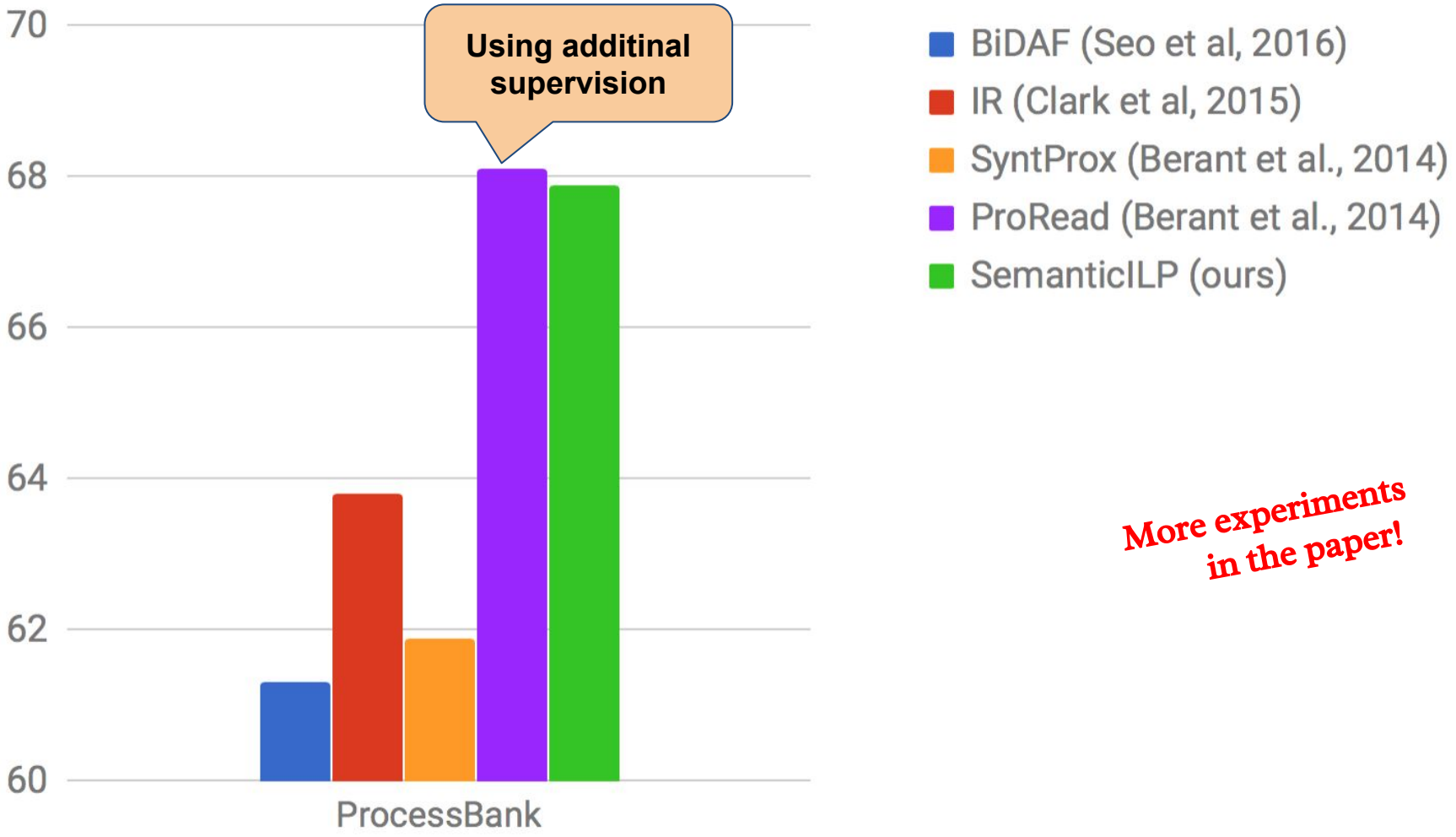
Results #1: Science Questions



(exam scores, shown as a percentage)

Higher is better

Results #2: Biology Questions



More experiments in the paper!

One single system tested on different datasets.

Concluding remarks

- Reasoning over language requires dealing with diverse set of linguistic phenomena.
- Linguistic variability ⇒ collection of semantic abstraction that are linguistically informed
- We decoupled “reasoning for QA” from “abstraction”
- Strong performance on two domains simultaneously

Give it a try:
`https://github.com/allenai/semanticilp`

CogComp-NLP:
`https://github.com/CogComp/cogcomp-nlp`

Extra slides

- Reasoning over language requires dealing with linguistic phenomena.
- Reasons over a wide range of semantic abstractions of the text
- Strong performance on two domains simultaneously

Give it a try:
`https://github.com/allenai/semanticilp`

CogComp-NLP:
`https://github.com/CogComp/cogcomp-nlp`

Concluding

Linguistic variability

Which physical structure would best **help a bear to survive a winter**?

(A) big ears (B) black nose (C) **thick fur** (D) brown eyes

Thick fur helps a bear survive a winter.

A thick coat of white fur helps bears survive in these cold latitudes.

Polar bears, saved from the bitter cold by their thick fur coats, are among the animals in danger of extinction because of the ~~global warming~~ **brown noses** and human activities.

A given "meaning" can be phrased many surface forms!

Semantic understanding can help!

verb

Which physical structure would best help **a bear to survive a winter**?

(A) big ears (B) black nose (C**) thick fur** (D) brown eyes

- **Verb predicate: survive**
  - Survivor: "a bear"
  - Adverse circumstances: "a winter"

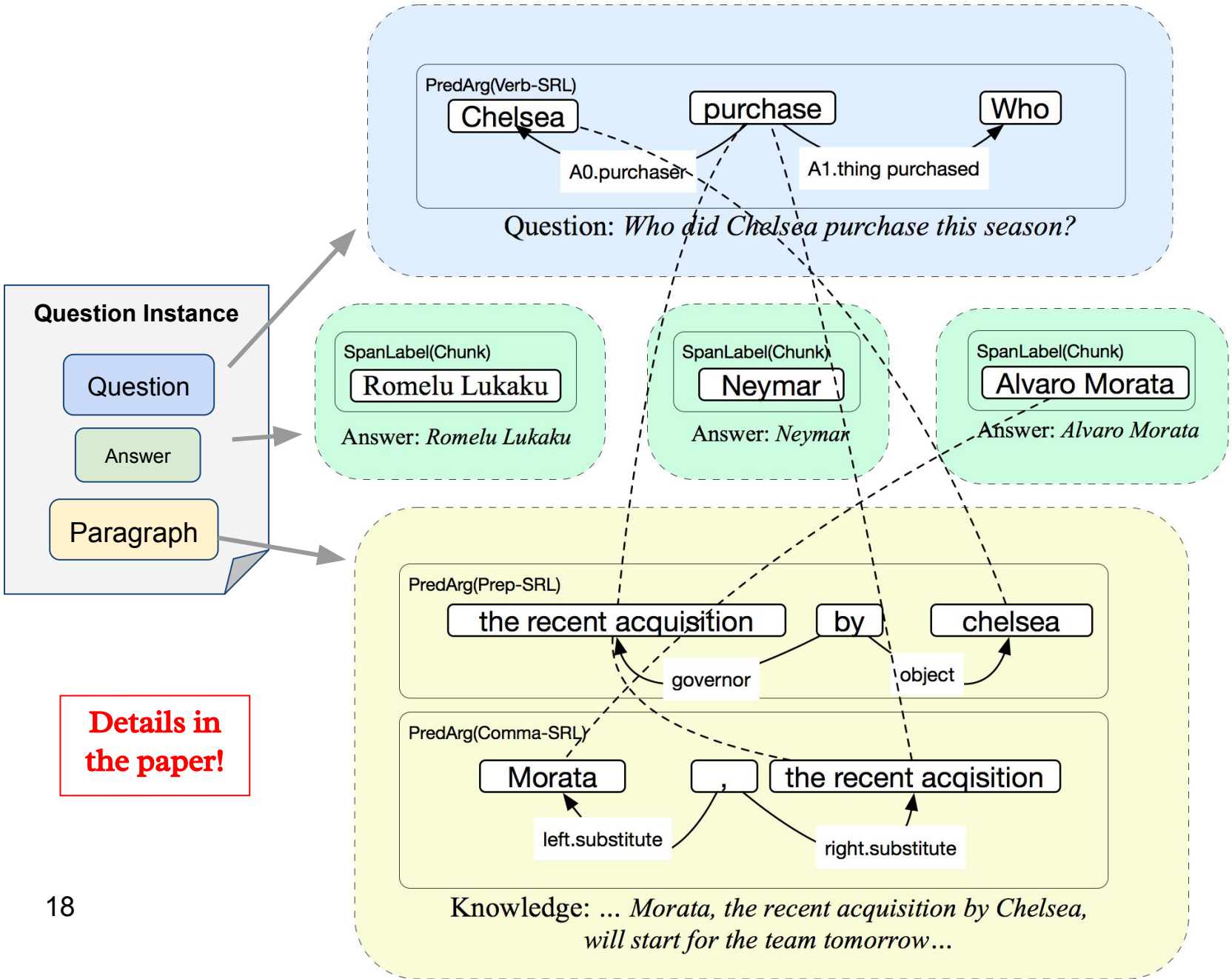
comma

preposition

- A single abstraction is not enough:
  - We use off-the-shelf linguistically-informed annotators
  - Coref, Shallow semantic representations, etc.

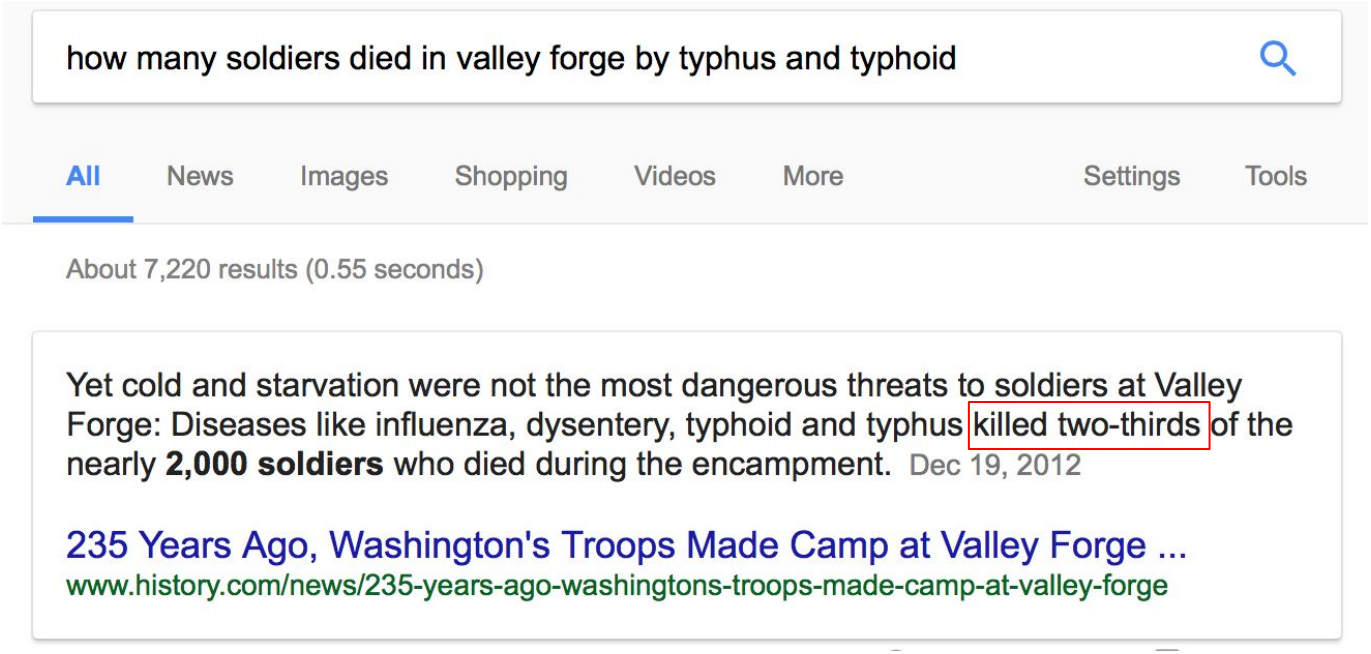
- **Comma predicate: , (sense: substitute)**
  - indicates an apposition structure

SemanticILP: Example

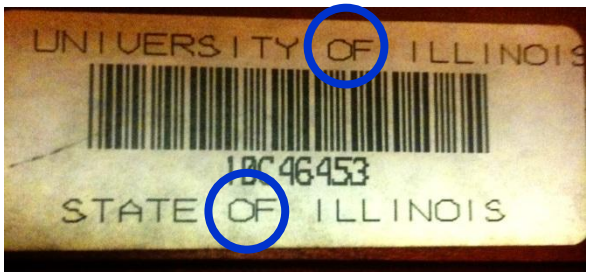


Details in the paper!

Linguistic Complexity



Typhus	
killed	
nearly	
two-thirds	Governor
of	PartWhole (of)
the	
2000	
soldiers	Object
.	

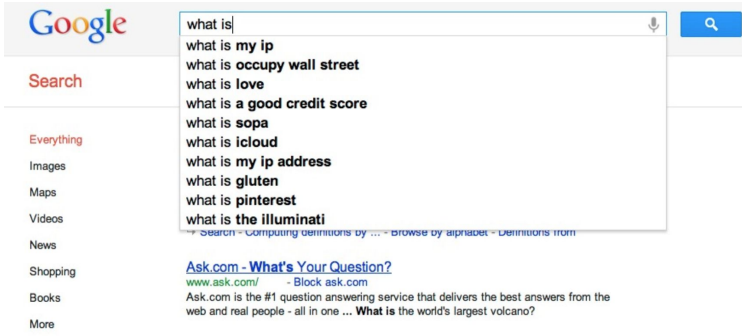


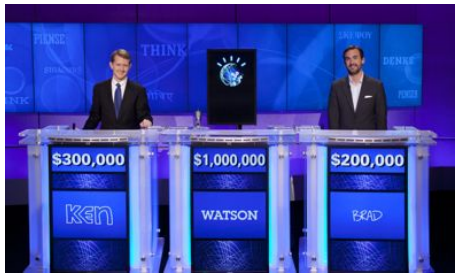
Not all “of”s are the same

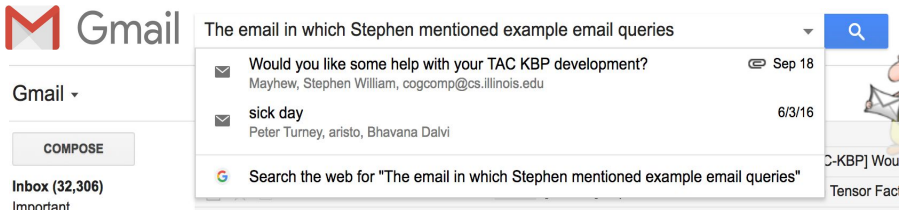
- One argument is a part of another.
  - The governor is a number
  - The object is a group modified by the governor.

QA is everywhere

- One of the oldest problems in AI
- Remarkable features of QA







QA systems are still far from exhibiting human-like intelligence, even in relatively simple ways (vs. human-level)

Results #2: Biology Questions

Dataset	BiDAF	BiDAF tuned	IR	SyntProx Baseline*	ProRead* (structural supervision)	SemanticILP
Process Bank**	58.7	61.3	63.8	61.9	68.1	67.9

SemanticILP does not rely on domain-specific process structure annotation

- Close to the specialized, state-of-the-art ProRead system
- Substantially better than syntax-based and neural baselines

One single system tested on different datasets.

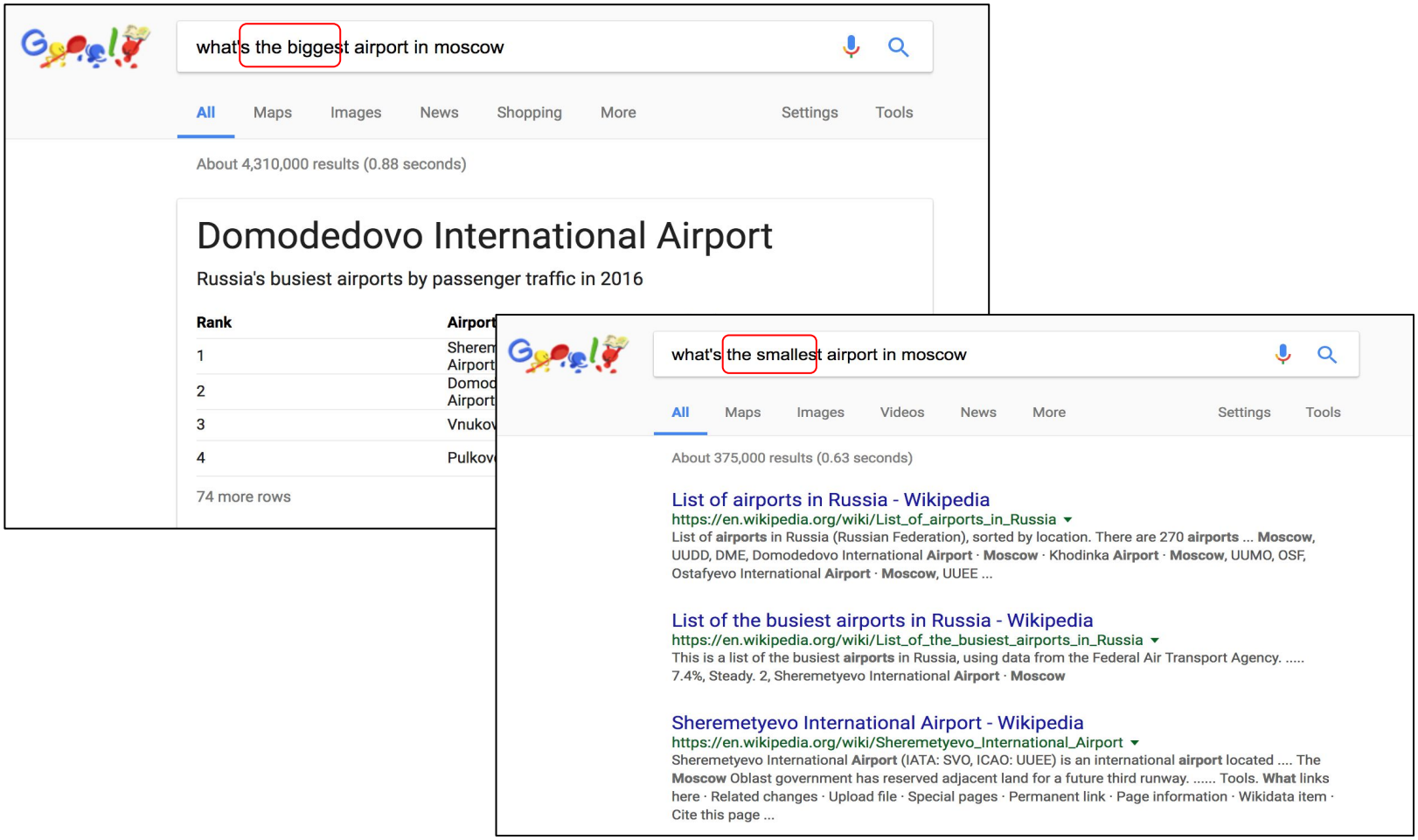
More experiments in the paper!

* Berant et al. (EMNLP, 2014)

** ~70% of the original dataset; true/false and temporal questions currently out of scope

QA solvers; getting lucky

Widely accepted that QA systems get lucky when answering questions, because they fail with small *variations*.



Results #1: Science Questions

- Paragraphs obtained by concatenating top k Lucene-retrieved sentences for various answer options

best overall	best baseline

Dataset
Regents 4th
Public 4th
Regents 8th
Public 8th

(exam scores, shown as a percentage)

Linguistic Nuances

How many ways can we encode the knowledge required for the following question?

verb

Which physical structure would best **help** **a bear to** **survive** **a winter**?

(A) big ears (B) black nose (C**) thick fur** (D) brown eyes

comma preposition

Polar bears**,** saved from the bitter cold **by** their thick fur coats**,** are among the animals in danger of extinction because of the global warming and human activities.

Linguistic understanding can help!

- Verbs, preposition, punctuation
- Domain agnostic => can use pre-trained NLP modules