



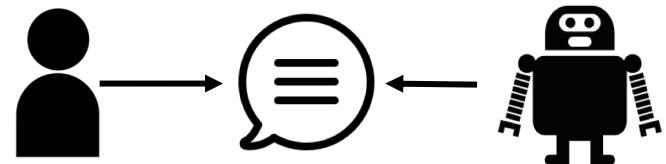
Penn
UNIVERSITY *of* PENNSYLVANIA

Reasoning-Driven Question Answering for Natural Language Understanding

Daniel Khashabi

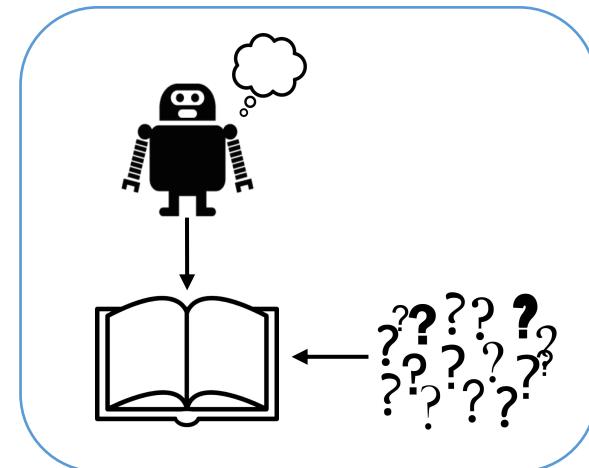
Natural Language Understanding

- Interpret a given text similar to humans.



- Measuring the progress by answering questions.

- A system that is better in understanding language, should have a higher chance of answering these questions.
 - This has been used in the field for many years
[Winograd, 1972; Lehnert, 1977b; others]
 - Question Answering (QA),
 - Reading Comprehension (RC),
 - Textual Entailment (TE), etc.



NLU Challenges: Ambiguity

- Making sense of strings.



“A 61-year old furniture salesman was pushed down the shaft **of** a freight elevator yesterday.”

A 61-year old furniture salesman

An antique furniture salesman

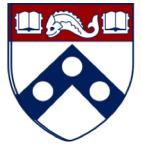
~~shaft of a freight elevator~~

height of a freight elevator

Oxford English Dictionary lists
10 primary meanings for “of”.

part-whole

attribute



NLU Challenges: Variability

- A single meaning mentioned in many different ways.



“The buffer springs at the bottom of the shaft prevented the car from crushing the salesman, John J. Hug, after he was pushed from the first floor to the basement. The car stopped about 12 inches above him as he flattened himself at the bottom of the pit. Mr. Hug was pinned in the shaft for about half an hour until his cries attracted the attention of a porter.”

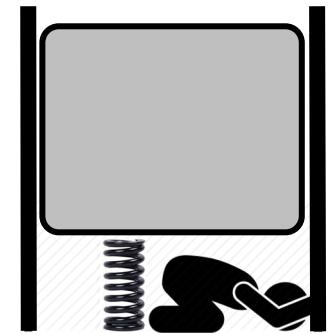
- Even more variability in bigger units (phrases, sentences, paragraphs, etc.)



NLU Challenges: Reading Between the Lines

- Lots of understanding is only implied from text.

“The car stopped about 12 inches above him as he flattened himself at the bottom of the pit.”



- *the car is significantly heavier than the man;*
- *he had nowhere to go at the bottom of the pit;*
- *if he didn't flatten himself, he would have died;*
- ...

Common-sense understanding

[Aristotle; Avicenna; Descartes; others]



NLU Challenges: Small bits, Big conclusions

- “Reasoning” as the process of combining facts and beliefs, to make decisions.

[Johnson-Laird, 1980]

- Why was he crying?

- Maybe he was scared.
 - Maybe he was injured.

Abductive reasoning

[Peirce, 1883]

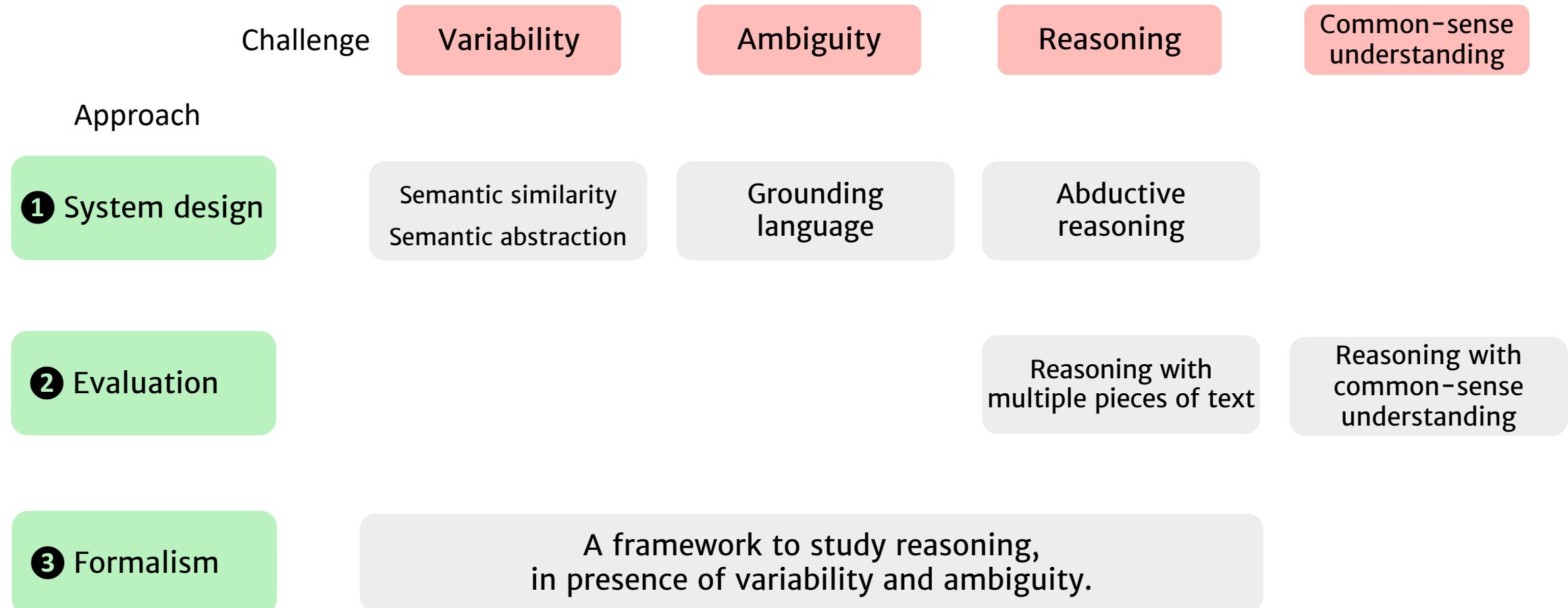
“... his weeps attracted
the attention of a porter”

- Many other forms of “reasoning”:

- Inductive, deductive, analogy, quantitative, etc.



Thesis Structure and Challenges Addressed



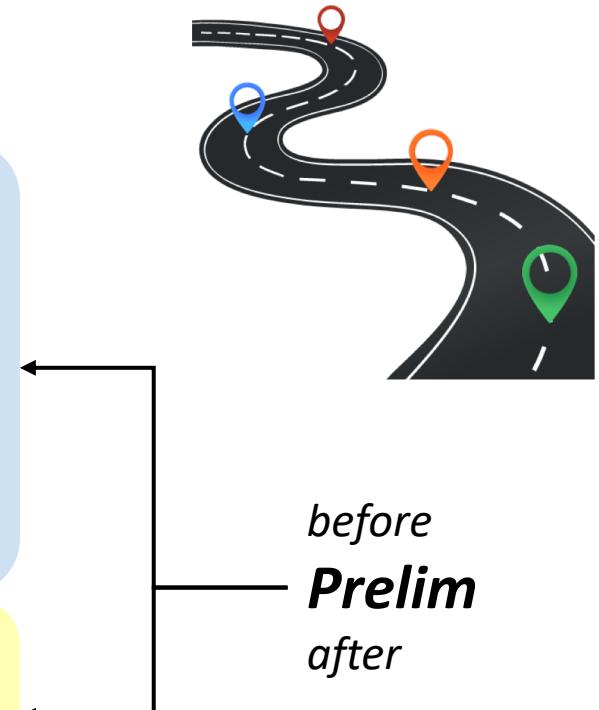
Thesis Statement

- Progress in automated question answering could be facilitated by incorporating the ability to reason over natural language abstractions and world knowledge.
- More challenging, yet realistic QA datasets pose problems to current technologies; hence, more opportunities for improvement.



Road map

- Introduction and motivation
- Part 1: Reasoning-Driven System Design
 - QA as Subgraph Optimization on Tabular Knowledge [[IJCAI'16](#)]
 - QA with Semantic Abstractions of Raw Text [[AAAI'18](#)]
 - Learning to Pay Attention to Essential Terms in Questions [[CoNLL'17](#)]
- Part 2: Moving the Peaks Higher: More Challenging Datasets
 - A QA Benchmark for Reasoning on Multiple Sentences [[NAACL'18](#)]
 - A QA Benchmark for Temporal Common-sense [[Submitted](#)]
- Part 3: Formal Study of Reasoning in Natural Language
 - Capabilities and Limitations of Reasoning in Natural Language [[In submission](#)]
- Conclusion



Road Map

- **Part 1: Reasoning-Driven System Design**

- QA as Subgraph Optimization on Tabular Knowledge [\[IJCAI'16\]](#)

- Motivation
 - Knowledge as Tables
 - Reasoning on Knowledge
 - Experimental results

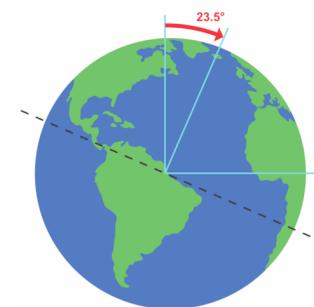


QA as Subgraph Optimization on Tabular Internal Knowledge: Overview

- Standardized science exams. [Clark et al, 2015]
- Simple language; machines require the ability use the knowledge and abstract over it.
- The “knowledge” encoded within the solver.

Question: In New York State, the longest period of daylight occurs during which month?

Candidates: (A) June (B) March (C) December (D) September



New Zealand

shortest

night

In ~~New York State~~, the ~~longest~~ period of ~~daylight~~ occurs during which month?

- (A) June
- (B) March
- (C) December
- (D) September

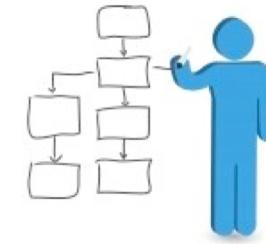
Premise: *a system that “understands” this phenomenon can correctly answer many variations!*



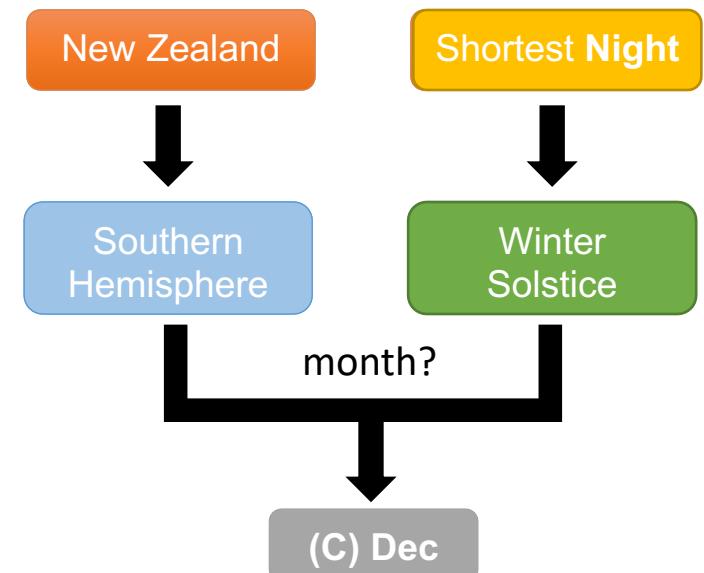
Semi-Structured Inference

Question: In New York State, the longest period of daylight occurs during which month?

- (A) June (B) March (C) December (D) September



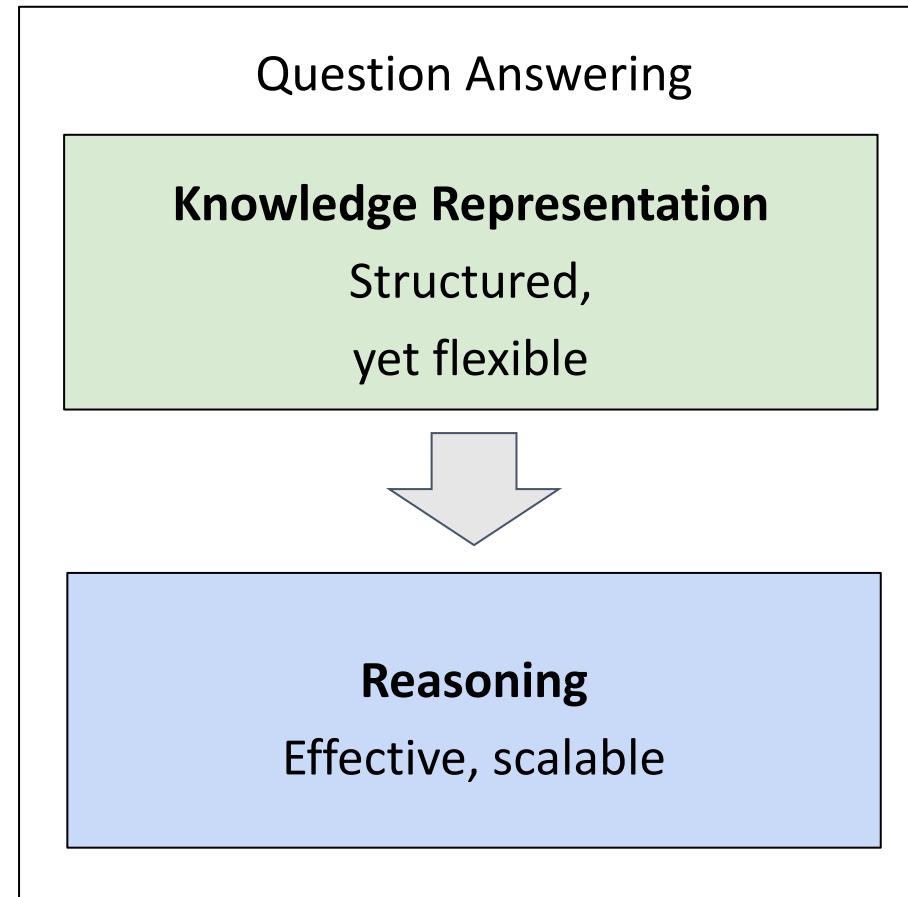
- Structured, Multi-Step Reasoning
 - Science knowledge in small, manageable, swappable pieces: *regions, hemispheres, solstice,*
 - Reasoning: putting together pieces of knowledge in a principled way.
 - Goal: **overcome brittleness**
- ✓ principled approach, explainable answers
✓ robust to variations



How can we achieve this?

Semi-Structured Inference: High-level View

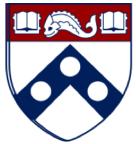
Question Answering
as **Global Reasoning**
over **Semi-Structured Knowledge**



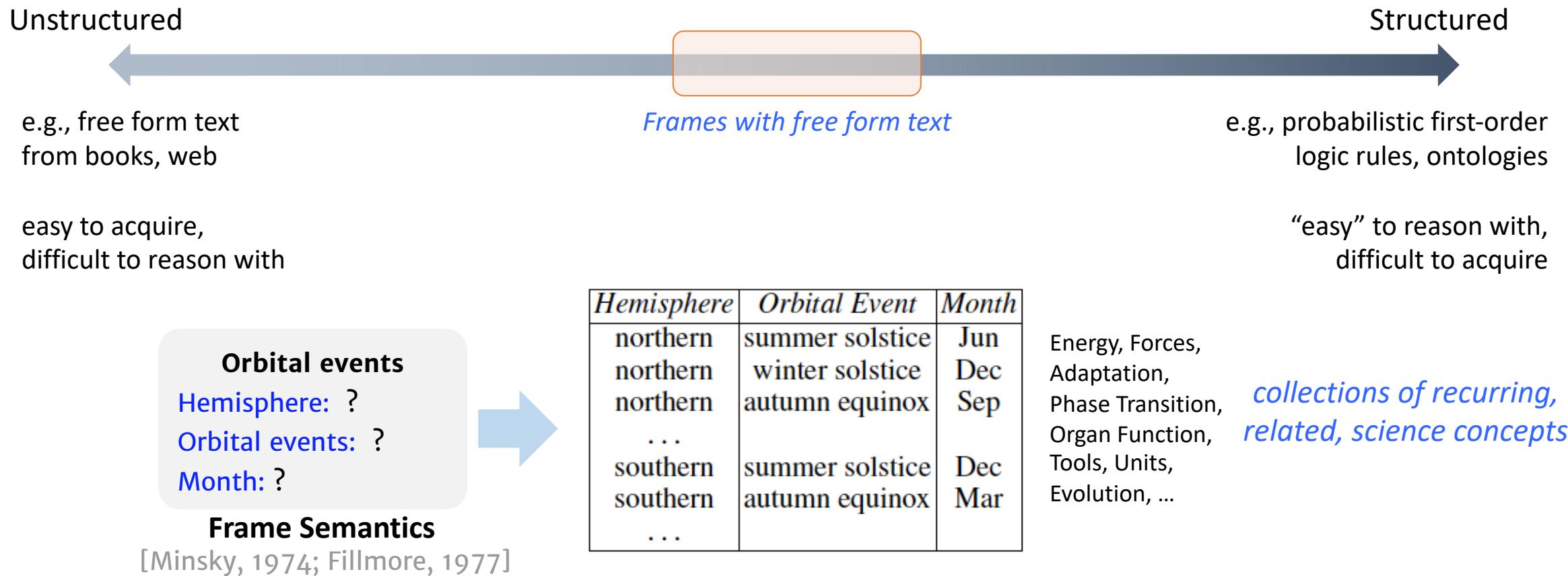
The Necessary Knowledge

The Knowledge Atlas: 12 key sections

Celestial Phenomena sun moon stars day/night, rotation revolution	The Earth air water land weather precipitation erosion	Matter solid/liquid/gas properties conductivity texture temperature measuring tools	Energy forms energy transfer heat electricity chemical energy energy conversion
Forces gravity magnetism force friction pull/pushing attraction	Living things living nonliving characteristics animals plants fish	Inheritance inherited traits resemblance acquired traits learned traits body features skills	The Environment and Adaptation senses habitats behavior camouflage survival
Continuity of Life life cycle life span offspring reproduction coloration mating	Life Functions breathing growing eating food air water	Interdependence food web producers consumers decomposers predators prey	Human Impact human activities environment ecosystem pollution conservation deforestation



Knowledge as Frames



Simple structure, flexible content

- Can acquire knowledge in automated and semi-automated ways [Dalvi et al, 2016]

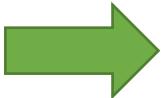


Road Map

- **Part 1: Reasoning-Driven System Design**

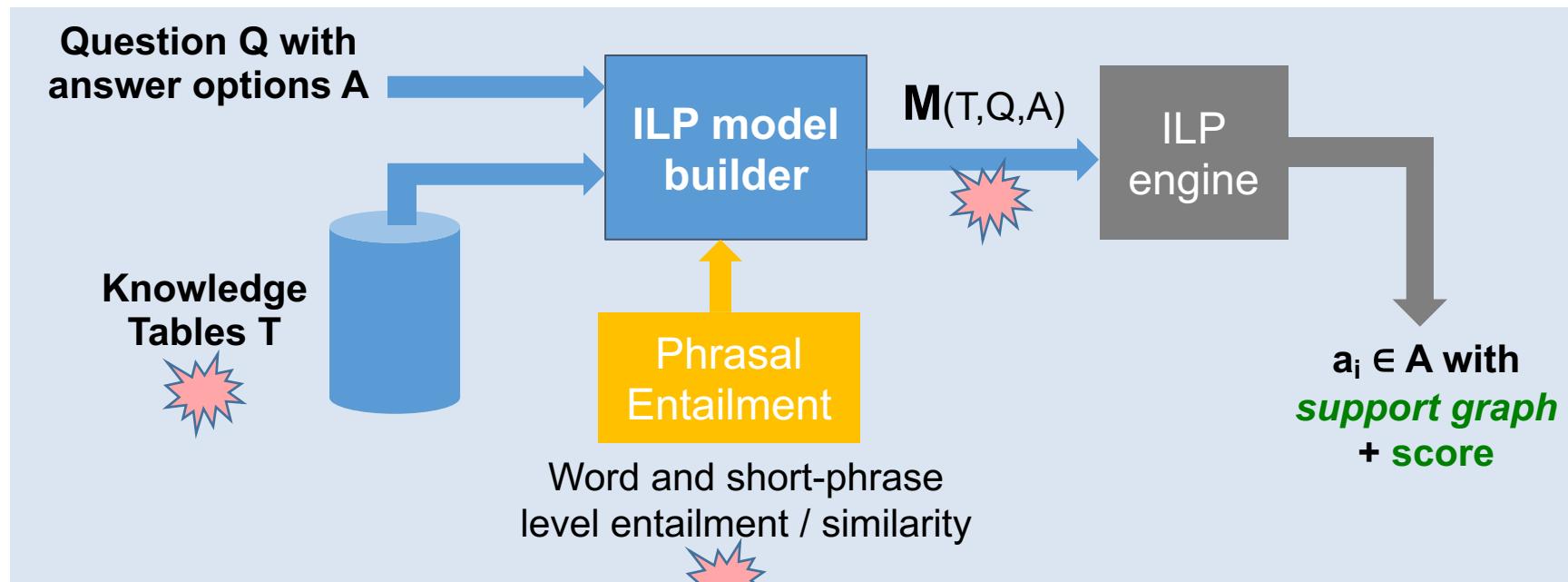
- QA as Subgraph Optimization on Tabular Knowledge *[IJCAI'16]*

- Motivation
 - Knowledge as Frames
 - Reasoning on Knowledge
 - Experimental results



TableILP Solver: An Overview

- A discrete **optimization** approach to QA for multiple-choice questions



$$M(T,Q,A) \rightarrow$$

$$\max \sum_i c_i x_i$$
$$\forall x_i \in \mathbb{N} \cup \{0\}$$
$$\begin{cases} \sum_i a_{1i} x_i \leq b_1 \\ \dots \\ \sum_i a_{ki} x_i \leq b_k \end{cases}$$

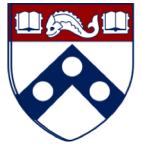
Optimization using Integer Linear Program (**ILP**) formalism



TableILP: Main Idea

Q: In New York State, the longest period of daylight occurs during which month?

Cities, States, Countries		(A) December (B) June (C) March (D) September
<i>Potential Link:</i> Regions and Hemispheres	Orbital Events: Geographical properties & Timing	



Search for the best **Support Graph** connecting the Question to an Answer through Tables.

TableILP: Main Idea

Q: In New York State, the longest period of daylight occurs during which month?

Subdivision	Country
New York State	USA
California	USA
Rio de Janeiro	Brazil
...	...

Orbital Event	Day Duration	Night Duration
Summer Solstice	Long	Short
Winter Solstice	Short	Long
....

Country	Hemisphere
United States	Northern
Canada	Northern
Brazil	Southern
....	...

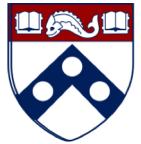
Hemisphere	Orbital Event	Month
North	Summer Solstice	June
North	Winter Solstice	December
South	Summer Solstice	December
South	Winter Solstice	June

Semi-structured Knowledge

- (A) December
- (B) June
- (C) March
- (D) September

Abductive reasoning

[Peirce, 1883]



Search for the best **Support Graph** connecting the Question to an Answer through Tables.

TableILP: Main Idea

An ideal
Support Graph

Q: In **New York State**, the **longest period of daylight** occurs during which **month**?

Subdivision	Country
New York State	USA
California	USA
Rio de Janeiro	Brazil
...	...

Country	Hemisphere
United States	Northern
Canada	Northern
Brazil	Southern
....	...

Orbital Event	Day Duration	Night Duration
Summer Solstice	Long	Short
Winter Solstice	Short	Long
....

Hemisphere	Orbital Event	Month
North	Summer Solstice	June
North	Winter Solstice	December
South	Summer Solstice	December
South	Winter Solstice	June

Semi-structured Knowledge

- (A) December
- (B) June
- (C) March
- (D) September

Link this information
to identify the best
supported answer!

Search for the best Support Graph connecting
the Question to an Answer through Tables.



Approach: ILP model

Goal: Design ILP constraints C and objective function F , s.t.
maximizing F subject to C yields a “desirable” support graph

- Many possible “proof structures”
 - single/multi-table, single/multi-row, answer in table header, answer spanning multiple cells
- Must balance reward for connections with penalty for spurious links
- Imperfect lexical “similarity” blackbox
- Partial or missing knowledge in tables
- Question logic (negation, conjunction, comparison)
- Scalability of ILP solvers

Not so straightforward!

$$\begin{aligned} \max & \sum_i c_i x_i \\ \forall x_i & \in \mathbb{N} \cup \{0\} \end{aligned} \quad \left\{ \begin{array}{l} \sum_i a_{1i} x_i \leq b_1 \\ \dots \\ \sum_i a_{ki} x_i \leq b_k \end{array} \right.$$



ILP model

Operates on lexical units of alignment

- cells + headers of tables T
- question chunks Q
- answer options A

~50 high level constraints + preferences

Variables define the space of “support graphs” connecting Q, A, T

- Which nodes + edges between lexical units are active?

Objective Function: “better” support graphs = higher objective value

- Reward active units, high lexical match links, column header match, ...
- WH-term boost (“which *form of energy...*”), science-term boost (“*evaporation*”)
- Penalize spurious overuse of frequently occurring terms

$$\begin{aligned} \max & \sum_i c_i x_i \\ \forall x_i & \in \mathbb{N} \cup \{0\} \end{aligned} \quad \left\{ \begin{array}{l} \sum_i a_{1i} x_i \leq b_1 \\ \dots \\ \sum_i a_{ki} x_i \leq b_k \end{array} \right.$$



ILP Model: Constraints

Dual goal: scalability, consider only meaningful support graphs

- **Structural Constraints**

- Meaningful proof structures
 - connectedness, question coverage, appropriate table use
 - single/multi-table, single/multi-row, etc
- Simplicity appropriate for 4th / 8th grade

- **Semantic Constraints**

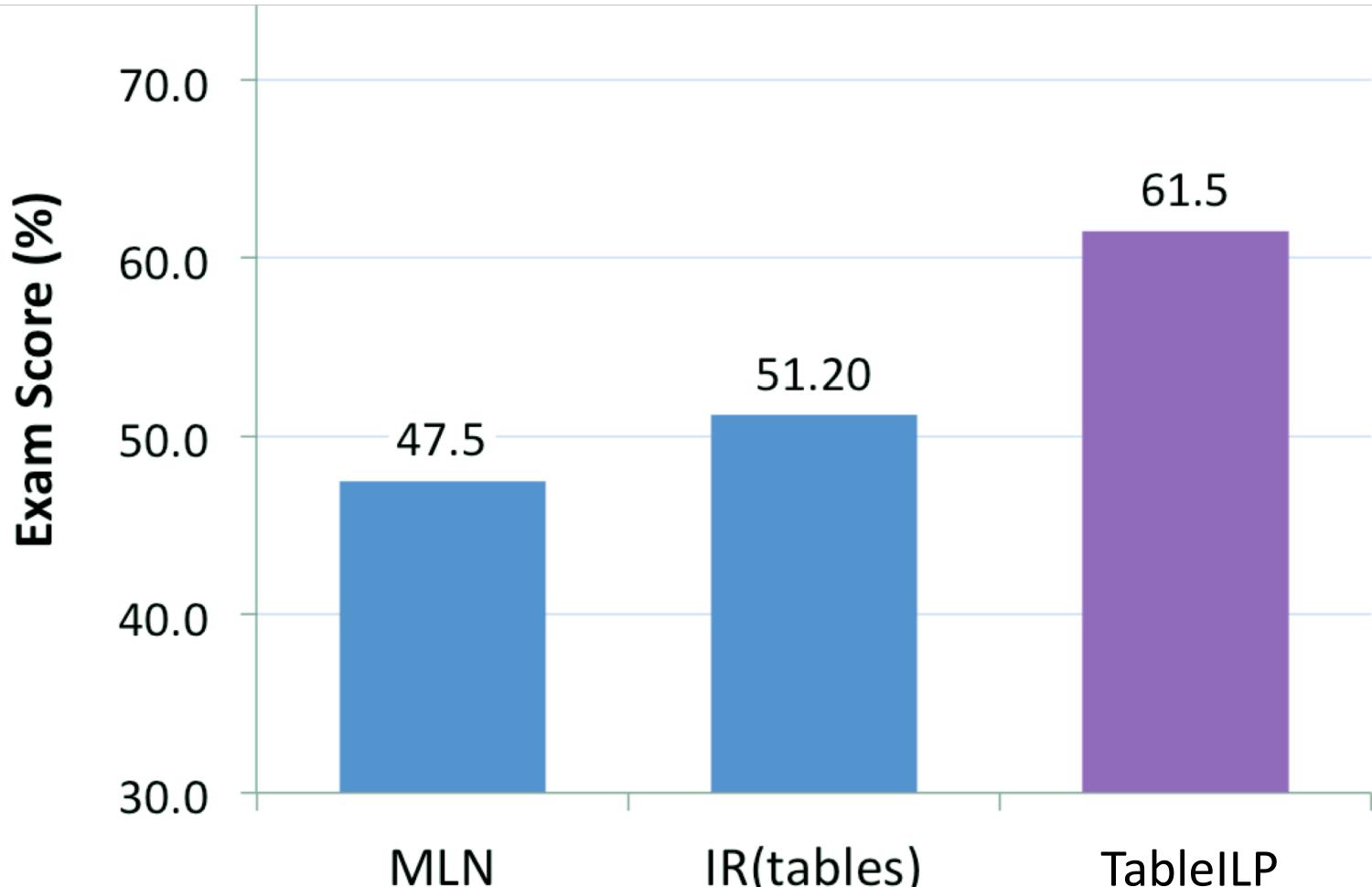
- Chaining => table joins between semantically similar column pairs
- Relation matching (ruler measures length, change from water to liquid)



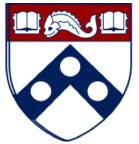
Experimental Results [KKS'16]



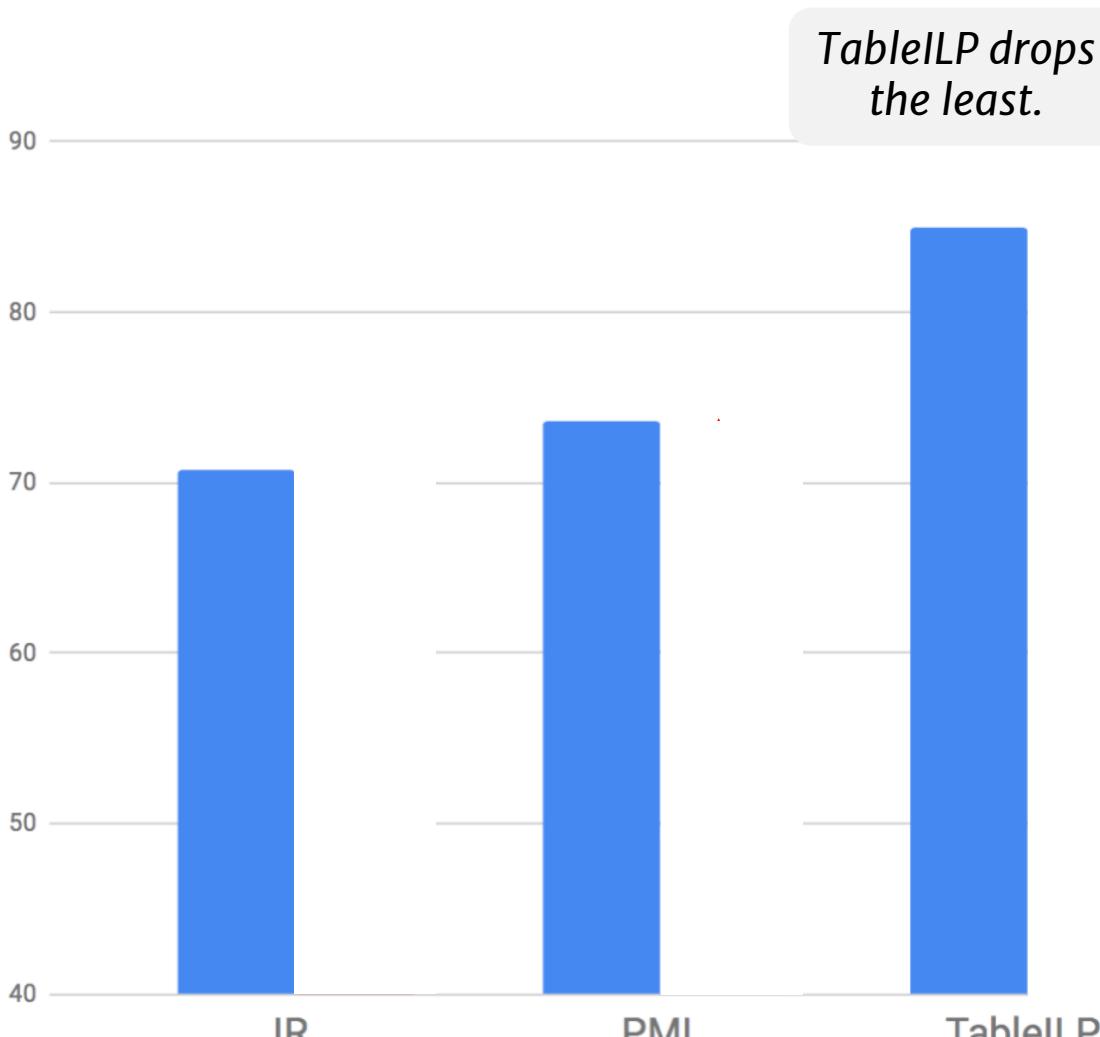
Experimental Results [KKS16]



TableILP is substantially better than IR & MLN, when given knowledge derived from the same, domain-targeted sources.



Assessing Brittleness: Question Perturbation



How robust are approaches to simple question perturbations *that would typically make the question easier for a human?*

- E.g., Replace incorrect answers with arbitrary co-occurring terms

In New York State, the longest period of daylight occurs during which month?
(A) *eastern* (B) June (C) *history* (D) *years*

- Original Questions
- with Adversarial Candidates



Motivating Example: Circling Back!

- Towards “real understanding” of the phenomenon tested in a question.

In New York State, the longest period of daylight occurs during which month?

In New Zealand, the longest period of daylight occurs during which month?

In New Zealand, the shortest period of daylight occurs during which month?

In New Zealand, the shortest period of night occurs during which month?



Motivating Example: Circling Back!

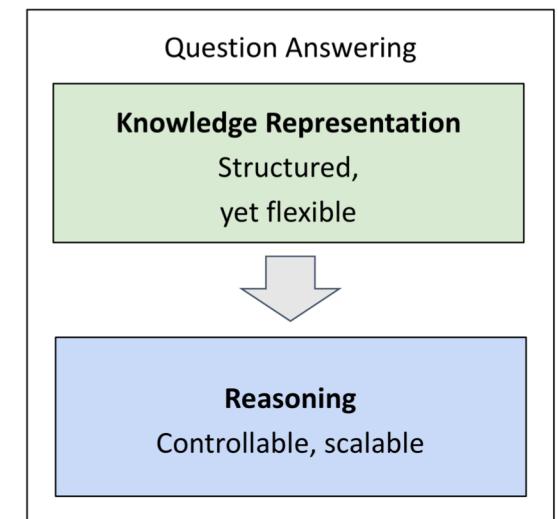
- Towards “real understanding” of the phenomenon tested in a question.

	IR	PMI	TableILP
<i>In New York State, the longest period of daylight occurs during which month?</i>	June	June	June
<i>In New Zealand, the longest period of daylight occurs during which month?</i>	March	Dec	Dec
<i>In New Zealand, the shortest period of daylight occurs during which month?</i>	March	Dec	June
<i>In New Zealand, the shortest period of night occurs during which month?</i>	Dec	Dec	Dec



Summary, So Far

- Elementary school science tests as a challenge for NLU.
- Knowledge as semi-automatically extracted knowledge.
- Abductive reasoning, to provide the best explanations.
- Showed effective and complementary performances.
- Impacts, since publications:
 - Strong performance on new datasets [Clark et al, Arxiv'2018]
 - Inspired other works [Khot et al, ACL'2017]



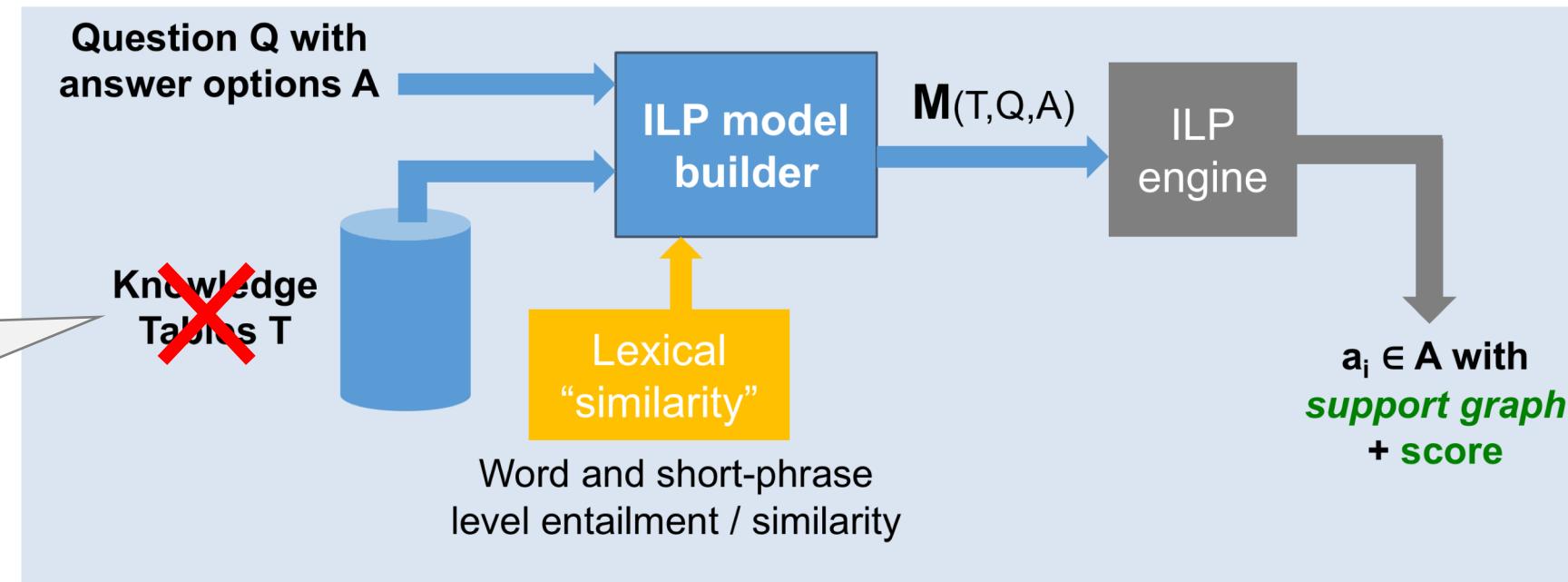
Road Map

- **Part 1: Reasoning-Driven System Design**

- QA as Subgraph Optimization on Tabular Knowledge [IJCAI'16]
- QA with Semantic Abstractions of Raw Text [AAAI'18]
- Learning to Pay Attention to Essential Terms in Questions [CoNLL'17]



Reasoning With Semantic Abstractions [KKS'18]

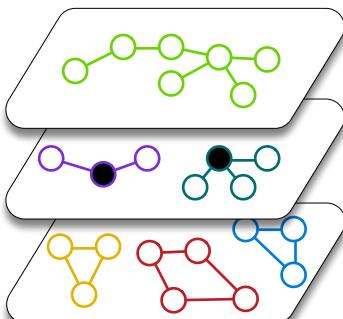


Knowledge provided as raw text.

~~Knowledge Tables T~~

Word and short-phrase level entailment / similarity

$a_i \in A$ with support graph + score

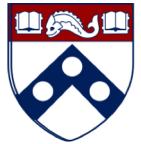


Representing text, as layers of semantic abstractions.

- *Verb-Semantic Roles* [Punyakanok et al, 2008]
- *Preposition-Semantic Roles* [Srikumar & Roth, 2013]
- *Comma-semantic Roles* [Arivazhagan et al, 2016]
- *Coreference* [Chang et al, 2012]

Easier domain transfer!

Improvements in multiple domains.

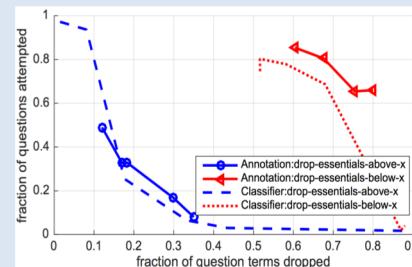
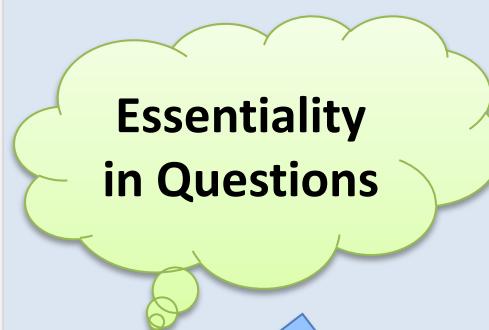


Learning Essential Terms in Questions [KKS'17]

Challenge for QA systems: Is a word in a question ***important, redundant, or distracting?***

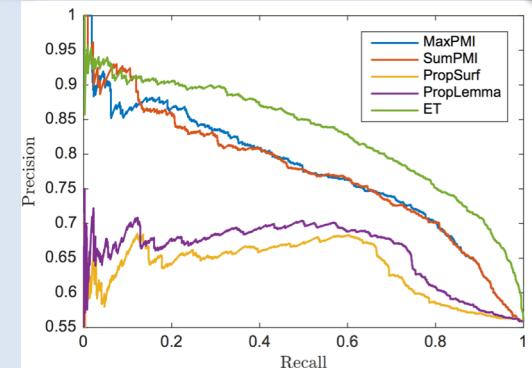


Some **animals** grows **thicker hair** as a season changes. This **adaptation** helps to _____.
 (A) find food (B) **keep warmer** (C) grow stronger (D) escape from predators



Important for humans!

State-of-the-art
Essentiality classifier:
F1 = 0.8, MAP = 0.9



Dataset	Baseline	With ET
Regents	59.11	60.85
AI2Public	57.90	59.10
RegtsPertd	61.84	66.84

Up to 5% increase
in end-to-end QA
performance

Road Map



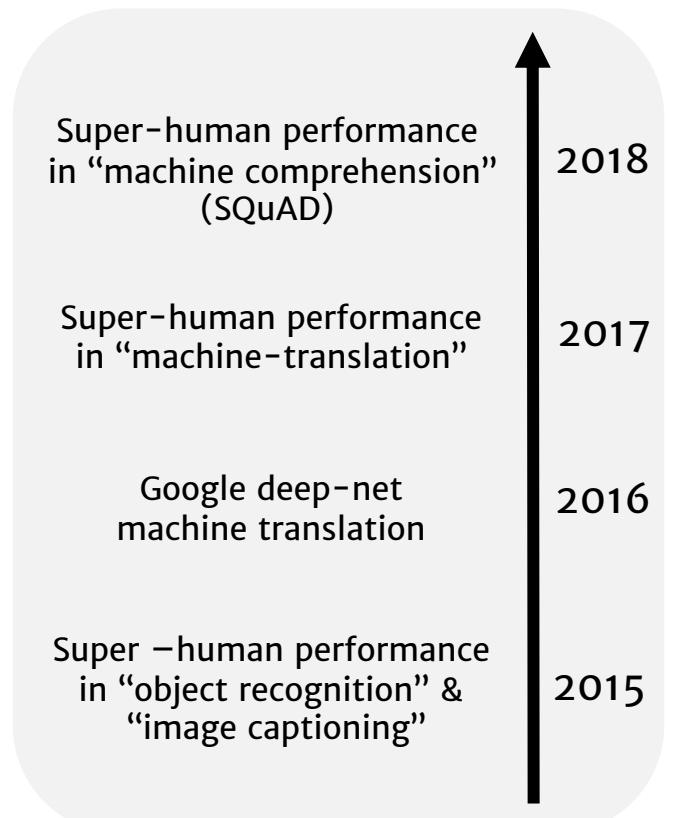
- **Part 2: Moving the Peaks Higher: More Challenging Datasets**

- A QA Benchmark for Temporal Common Sense [*Submitted*]
- A QA Benchmark for Reasoning on Multiple Sentences [*NAACL'18*]



The Recent State of the Field: Good News

- Many large QA datasets [Rajpurkar et al, 2016; others]
- Successes, with neural nets
 - Faster computers (e.g., GPUs)
 - New computational modules (e.g., Attentions)
 - More data



The Recent State of the Field: Bad News

- The systems easily break-down. [K at al, 2016; Jia et al, 2017; Belinkov et al, 2018; others]
- Many problems with no significant success: Math word problems; dialogue; many others.
 - Some are not even defined yet.
- Discoveries are more about tasks (datasets). [Darwiche, 2017]
- The urge to scale up datasets has biased in certain angles and limited their diversity.



Moving the Peaks Higher

Goal here:

Define and create challenges not addressed by the community.

Challenges that require external knowledge, common-sense, complex reasoning, etc.

Including
our systems

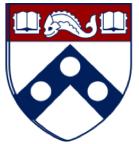


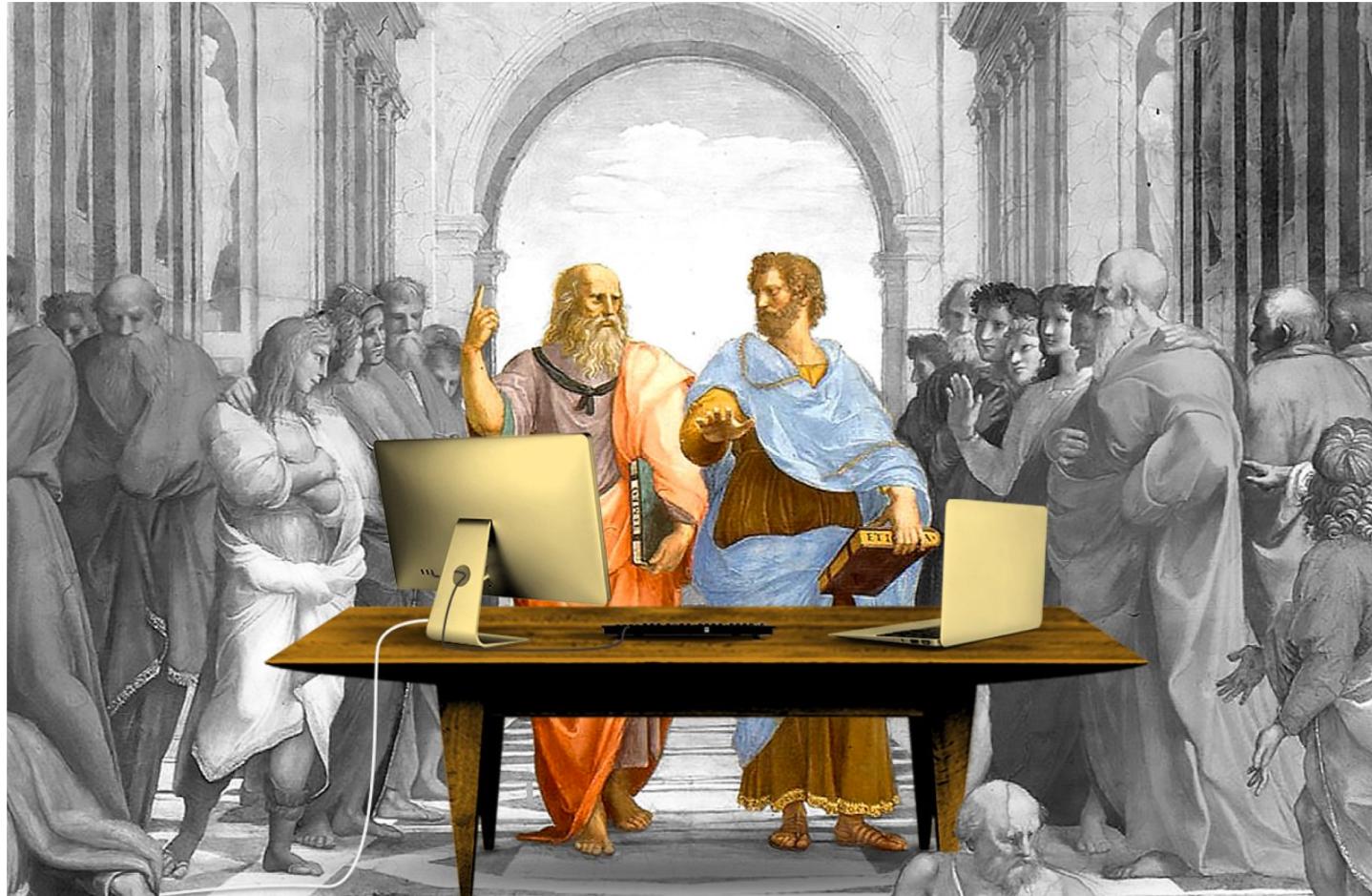
Road Map



- **Part 2: Moving the Peaks Higher: More Challenging Datasets**

- A QA Benchmark for Temporal Common Sense [*Submitted*]
- A QA Benchmark for Reasoning on Multiple Sentences [*NAACL'18*]





Did Aristotle have a laptop?

[Valiant, ?]



Did Aristotle have a laptop?



All

News

Shopping

Videos

Images

More

Settings

Tools

About 1,100,000 results (0.41 seconds)

Aristotle's Laptop | Series on Machine Consciousness - World Scientific

<https://www.worldscientific.com/worldscibooks/10.1142/8113>

This book is about a scientific ingredient that **was** not available to **Aristotle**: the science of information. Would the course of the philosophy of the mind **have** been ...

Aristotle's Laptop: The discovery of our informational mind | Request PDF

https://www.researchgate.net/.../259464884_Aristotle's_Laptop_The_discovery_of_out...

Aristotle's convincing philosophy is likely to **have** shaped (even indirectly) many ... mind **have** been different had **Aristotle** pronounced that the matter of mind **was** ...

Aristotle's Laptop eBook by Igor Aleksander - 9789814425629 ...

https://www.kobo.com/.../Advanced_Computing/Artificial_Intelligence ▾

Read "Aristotle's Laptop The Discovery of our Informational Mind" by Igor ... **Aristotle's** convincing philosophy is likely to **have** shaped (even indirectly) ... **have** been different had **Aristotle** pronounced that the matter of mind **was** information?

Types of “Knowledge”

- [Loosely-defined] types of knowledge

- Encyclopedic knowledge

- *The capital city of Venezuela is Caracas.*
 - *Obama was born in Honolulu, HI.*

- Commonsense knowledge



Focus of
this work

- *If you lived thousands of years ago, you're unlikely to be alive now.*
 - *Laptops didn't exist, before they were invented.*



Common Sense: A Short History

- Many early works

- Since the early days of AI [McCarthy, 68; Charniak, 77; others]
 - Ambitious projects [Lenat, 85; others]

- Recent years:

- Winograd Schema Challenge [Levesque, 2014; Peng, K, Roth, 2015]

“Jack pulled up a picture of Aristotle on his laptop”

“Jack pulled up a picture of Aristotle from his biography”

Incentivizing commonsense understanding as a high-level and well-defined task.



Temporal Common Sense

- Understanding “time” is a key ability in many NLU tasks.

“Going to barbershop” takes a couple of **hours**

Event duration

“Going to college” takes a couple of **years**

“Take a trip to Africa” happens once in a few **years**

Event frequency

“Take a trip to parent’s” happens every few **weeks or months**

“Going to work” usually happens around **morning** time

Typical time

“Going to a bar” usually happens around **evening/night** time

Goal of this section:

QA dataset that requires temporal commonsense.



Temporal Common Sense as QA

- A dataset of natural language questions
 - Questions about a “temporal” understanding
 - About 1k questions and 8k candidate answers
 - 5 temporal phenomena:
 - Event duration, event frequency, absolute time point, event ordering, stationary vs transient

Scenario: I clapped her shoulder to show I was not laughing at her.

Question: How long did they laugh?

Candidates:

- (A) for a few days (B) 20 minutes (C) for a few minutes



Temporal Common Sense as QA

- A dataset of natural language questions
 - Questions about a “temporal” understanding
 - About 1k questions and 8k candidate answers
 - 5 temporal phenomena:
 - Event duration, event frequency, absolute time point, event ordering, stationary vs transient

Scenario: *With the amount of money Diggler was making he was able to support both his and Rothchild's addictions.*

Question: *What time did Diggler go to work?*

Candidates:

- (A) at eight in the late night (B) he leaves around 3 am (C) he leaves around 8 am



Temporal Common Sense as QA

- A dataset of natural language questions

- Questions about a “temporal” understanding

- About 1k questions and 8k candidate answers
 - 5 temporal phenomena:
 - Event duration, event frequency, absolute time point, event ordering, stationary vs transient

Scenario: She checked the kitchen, but didn't find anything missing there except for a clock.

Question: Why was she checking for missing items?

Candidates:

(A) her window had been broken

(B) someone tied her hands



Temporal Common Sense as QA

- A dataset of natural language questions
 - Questions about a “temporal” understanding
 - About 1k questions and 8k candidate answers
 - 5 temporal phenomena:
 - Event duration, event frequency, absolute time point, event ordering, stationary vs transient

Scenario: They then took a boat to Africa and Asia, where they went on a trip through the mountains.

Question: How often do they go on trips?

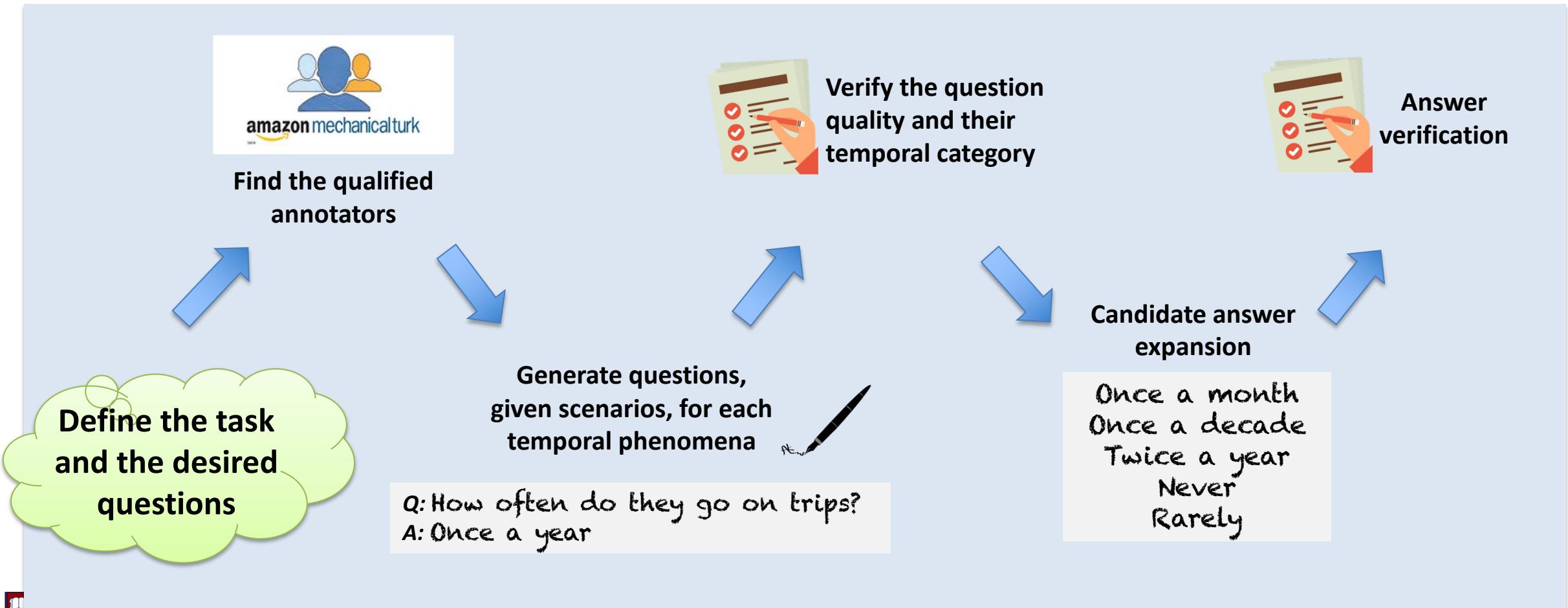
Candidates:

- (A) every night
- (C) twice a year

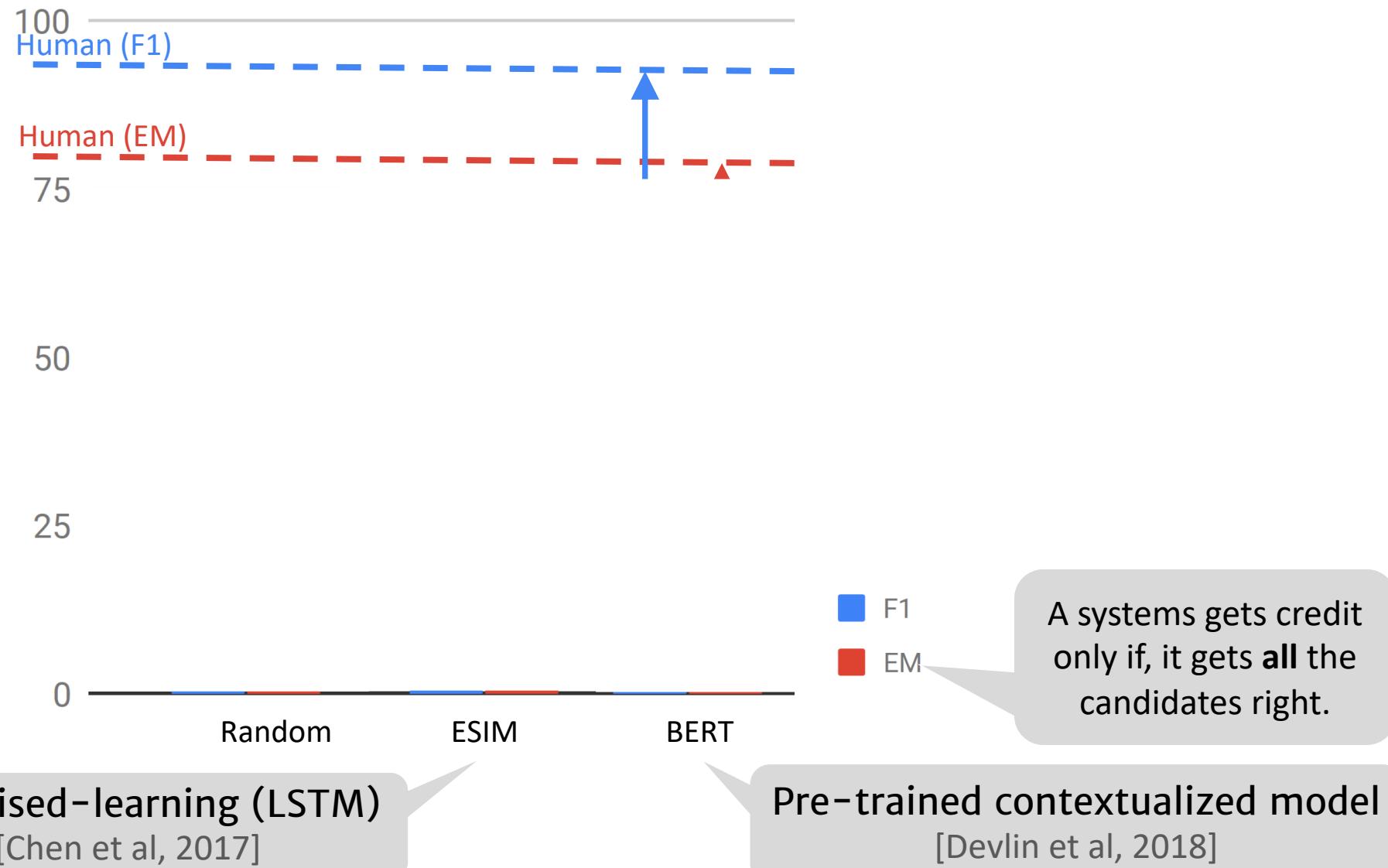
- (B) once a year
- (D) once a week



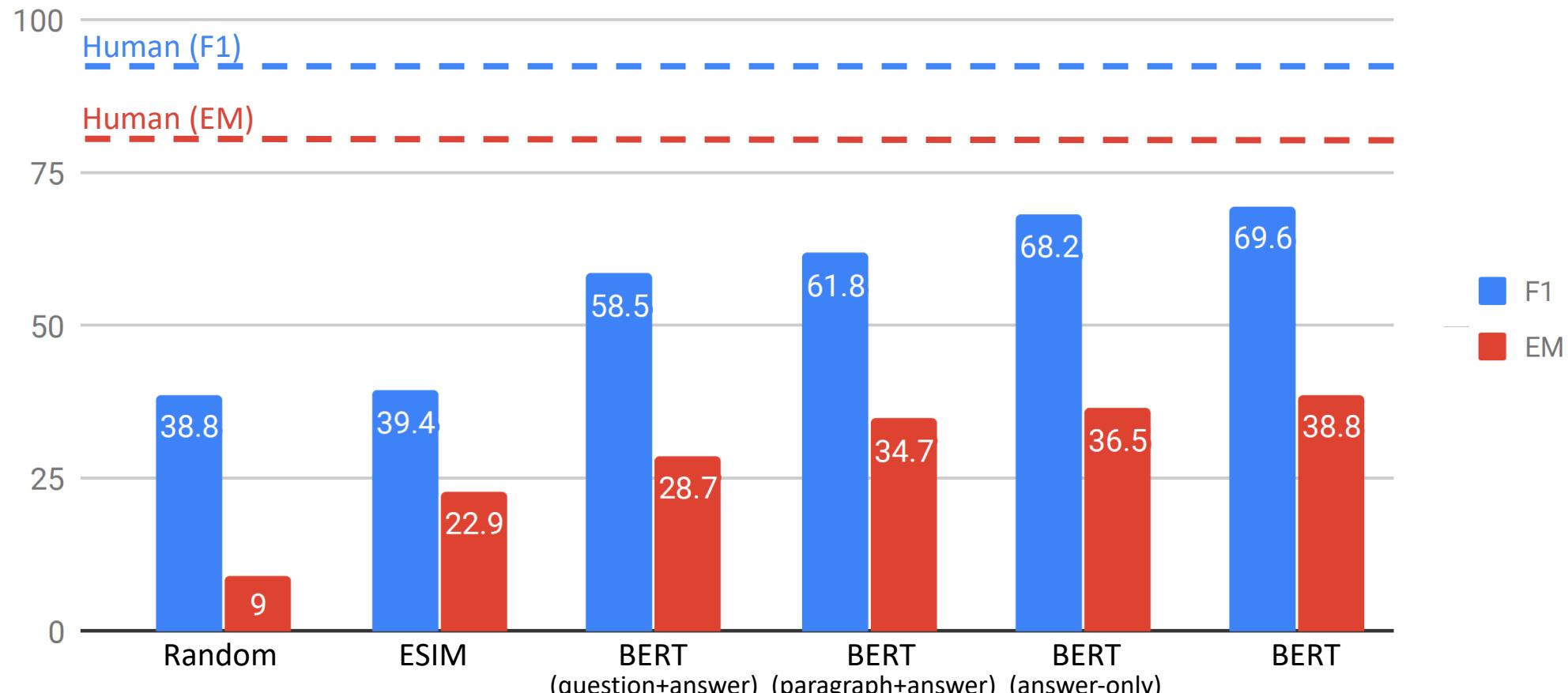
TacoQA: Construction Overview



Experimental Results [ZKNR, under review]

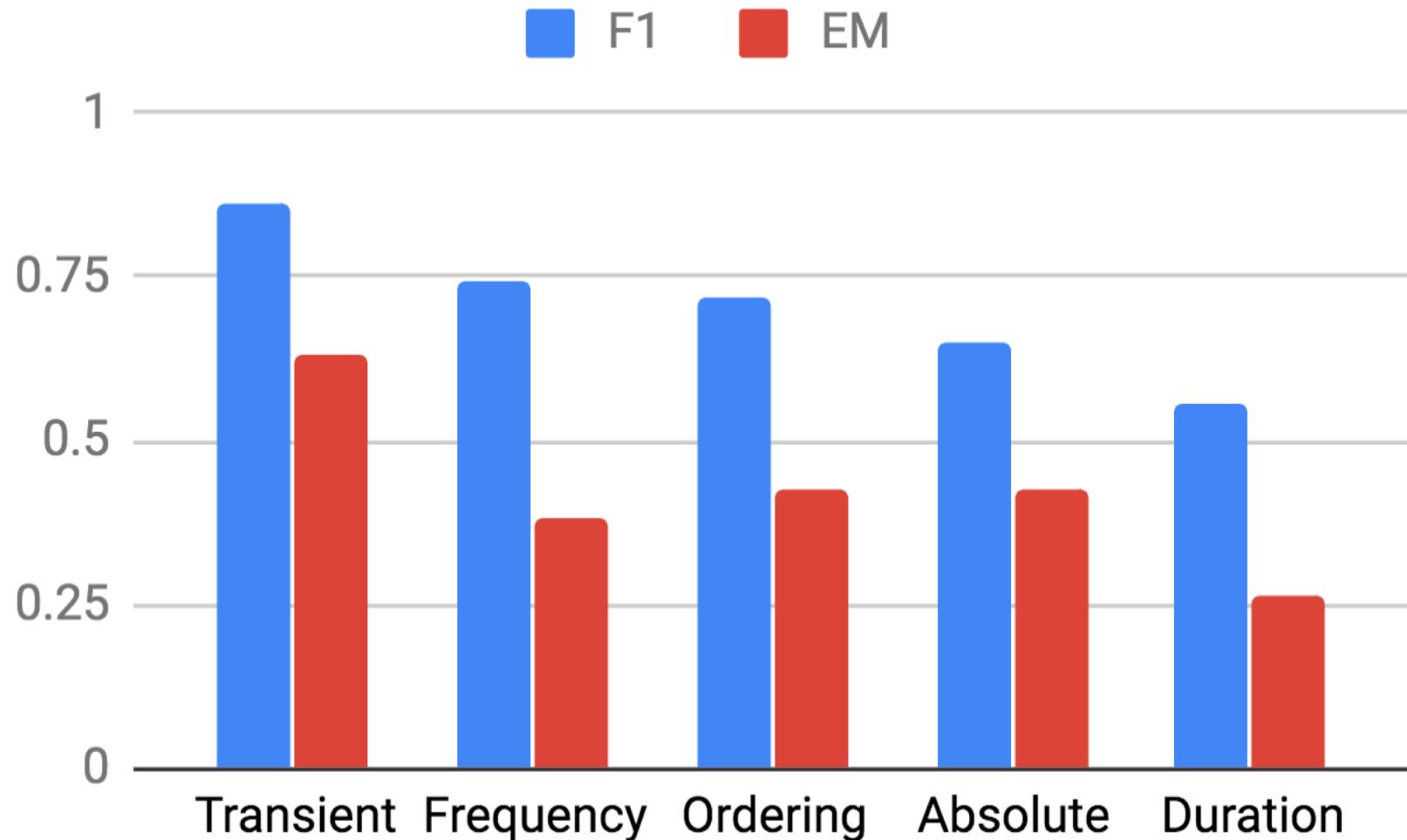


Experimental results [zKNR, under review]



[Devlin et al, 2018]

Experimental results [zKNR, under review]



The State of Current Systems

- Explicit knowledge bases; e.g. ConceptNet [Liu & Singh, 2001]
 - Okay precision but low recall (coverage)
 - Suffer from brittleness
- Language models, soft representations [Devlin et al, 2018; Peters et al, 2018]
 - Can deal with certain associations
 - Suffer from precision issues

Scenario: He laid down on the chair and pawed at her as she ran in a circle under it.

Question: How long did she run in a circle?



BERT selected as answer	Candidate answer
X	weeks
X	days
✓	minutes



Summary of This Section

- Understanding time is crucial aspect of NLU.
- A QA dataset of temporal commonsense questions.
- Evaluated systems and showing few angles they are missing.



Road Map



- **Part 2: Moving the Peaks Higher: More Challenging Datasets**

- A QA Benchmark for Temporal Common-sense [*Submitted*]
- A QA Benchmark for Reasoning on Multiple Sentences [[NAACL'18](#)]



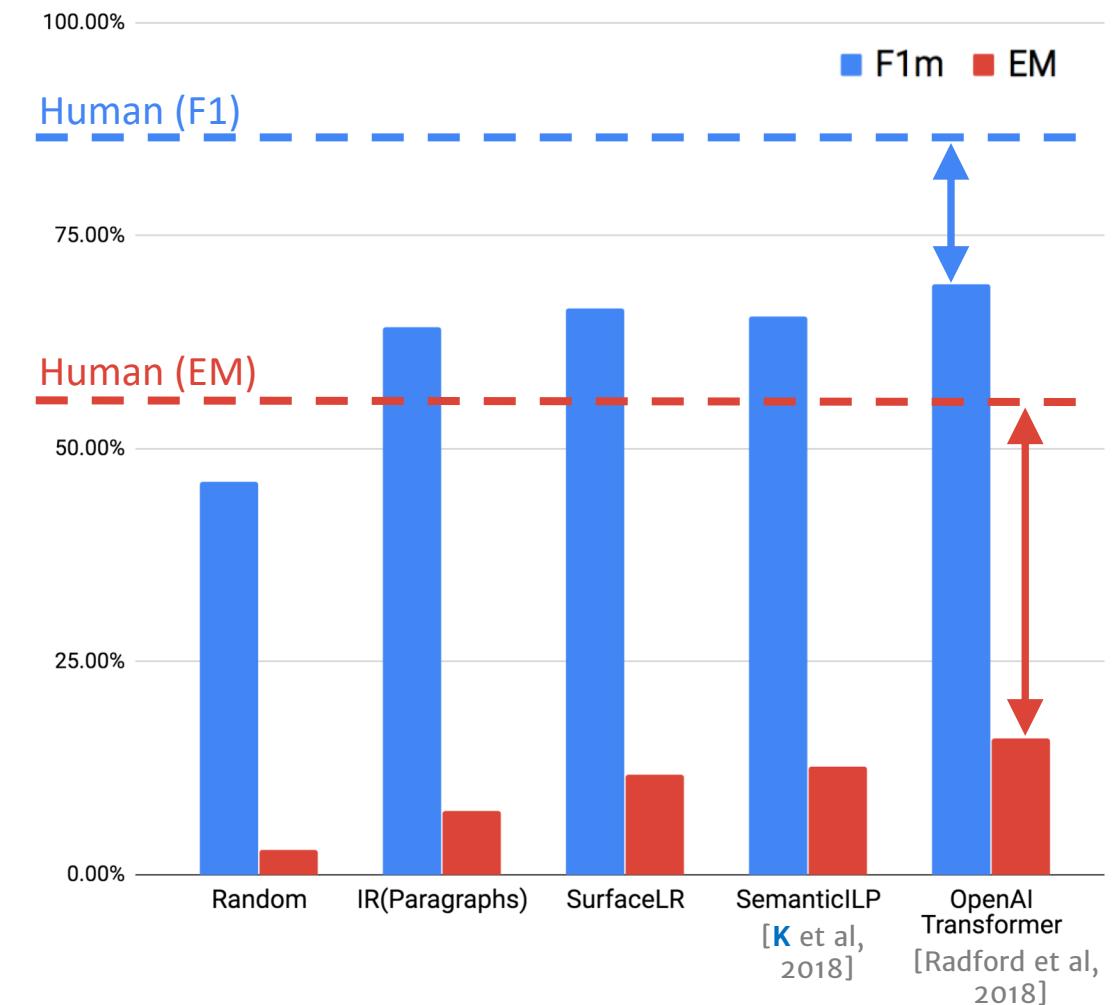
A Benchmark for Reasoning over Multiple Sentences [KCRUR'18]

- The need for creating “reasoning-forcing” challenges

“Multi-sentence” hypothesis:
Questions that require multiple sentences tend to be “hard”.

- 4-step crowdsourcing
- From 8 domains (fiction, news, science, etc)
 - +10k questions
 - 50k candidate answers
 - +700 paragraphs

<https://cogcomp.org/multirc>



Road Map

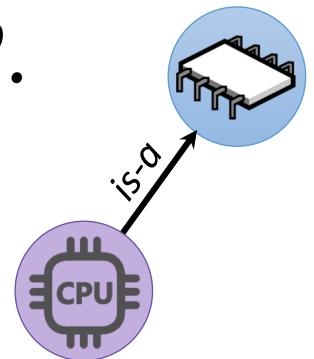
- **Part 3:** Formal Study of Reasoning in Natural Language

- Capabilities and Limitations of Reasoning in Natural Language *[In submission]*



A Formal Study of NL Reasoning: Overview

- We provide a formalized study of reasoning.
- Requires assumptions about “knowledge” and “reasoning”.
 - Information represented as **graphs** (nodes and semantic relations).
 - Any other structure can be thought of an explicit or implicit graph.
 - Incorporate properties like *variability*, *ambiguity*, etc.
 - *Reasoning*: the operation that combines chunks of information to make a conclusion.
- Distinguish **successful** and **failed** reasoning.



What We Do Not Say

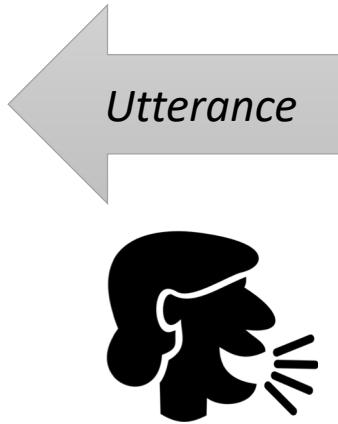
- **Not** making claims about:
 - How “reasoning” should be defined.
 - How “knowledge” should be represented.
 - How systems should be designed.
- Theoretical results based on assumptions (“no free lunch”).
 - which may or may not stand the test of time.



Tale of Two Spaces

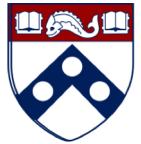
Symbol Graph

- Physical world
- Ambiguity
- Variability
- Missing relations

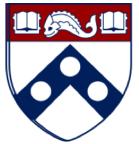
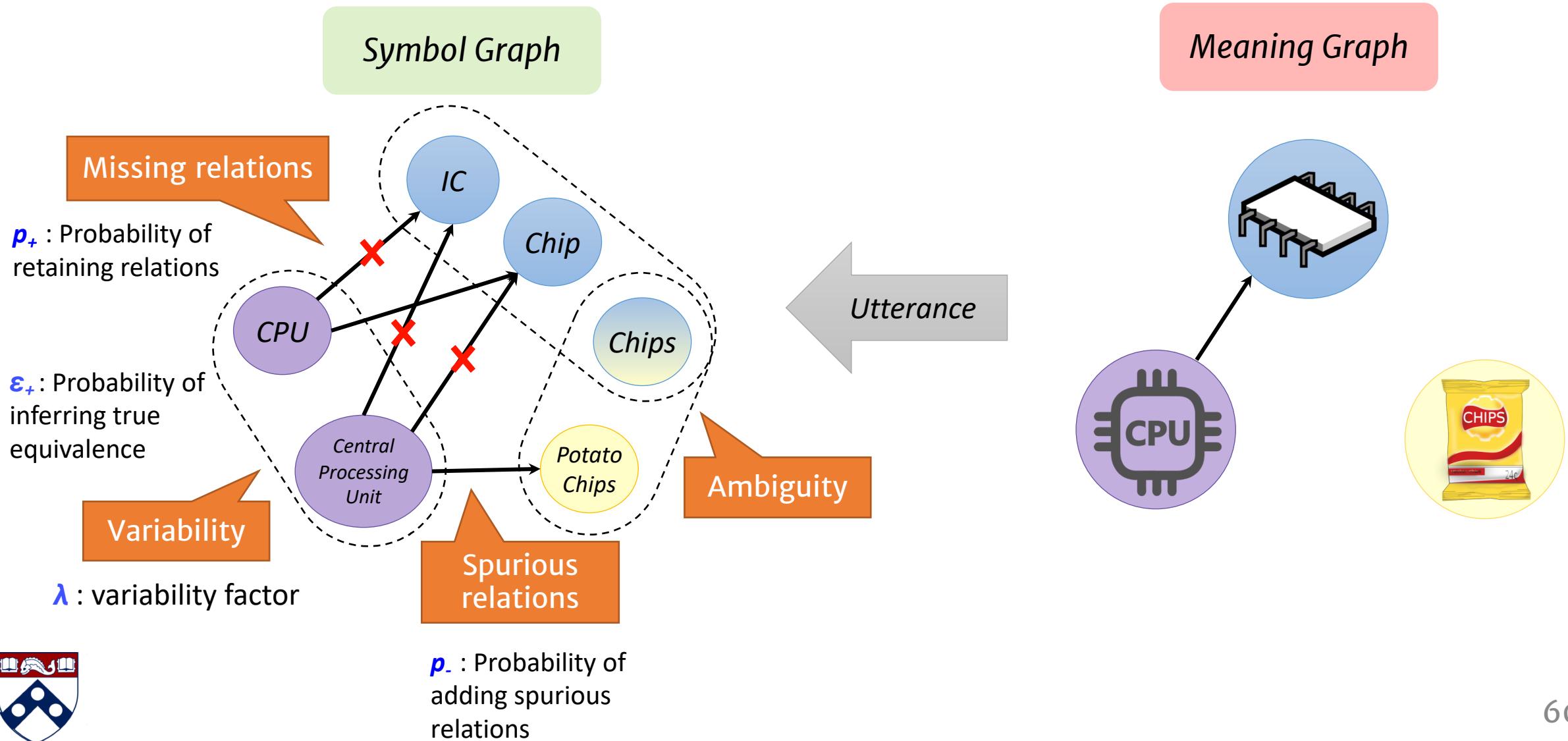


Meaning Graph

- Conceptualization
- No ambiguity
- No variability
- No missing relations

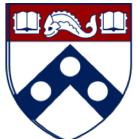
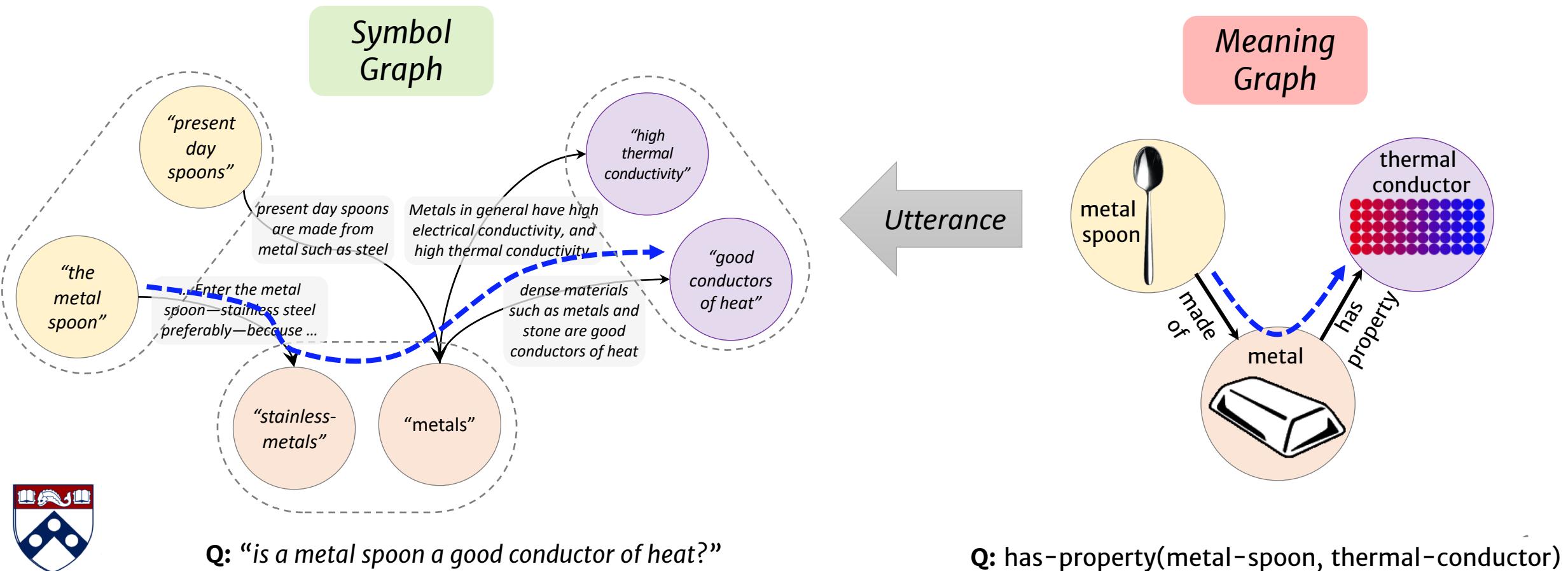


Formalizing Symbol and Meaning Space



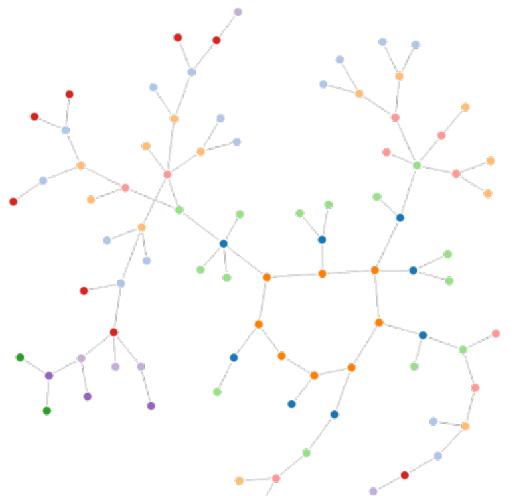
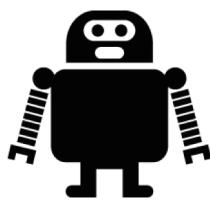
Reasoning by Combining Local Information

- Reasoning itself is hard to define.
- Class of reasoning which functions by combining local information (“multi-hop”)

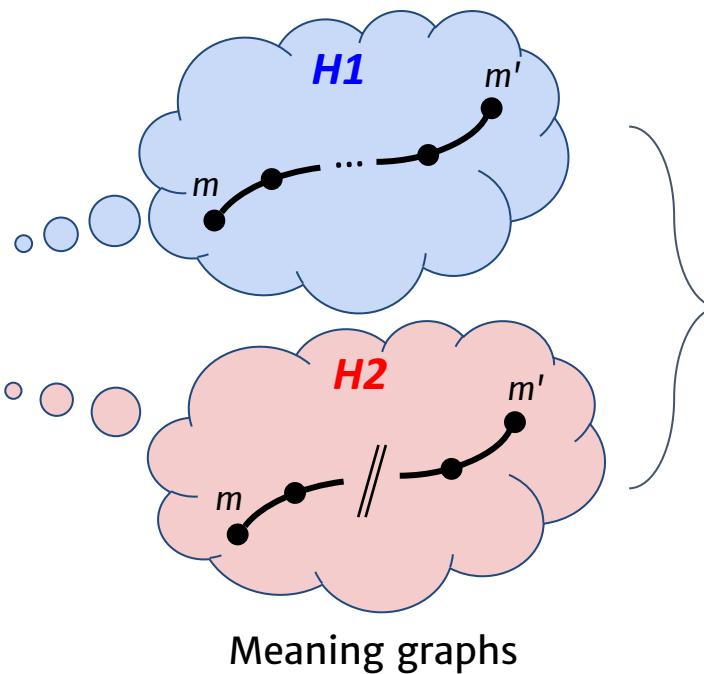
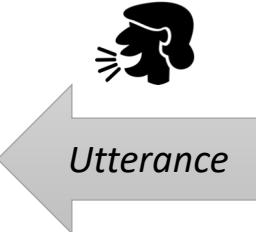


The Inference Problem

- “Inferring” connectivity in the (hidden) meaning graph
 - Given observations (a symbol graph)



Symbol graph

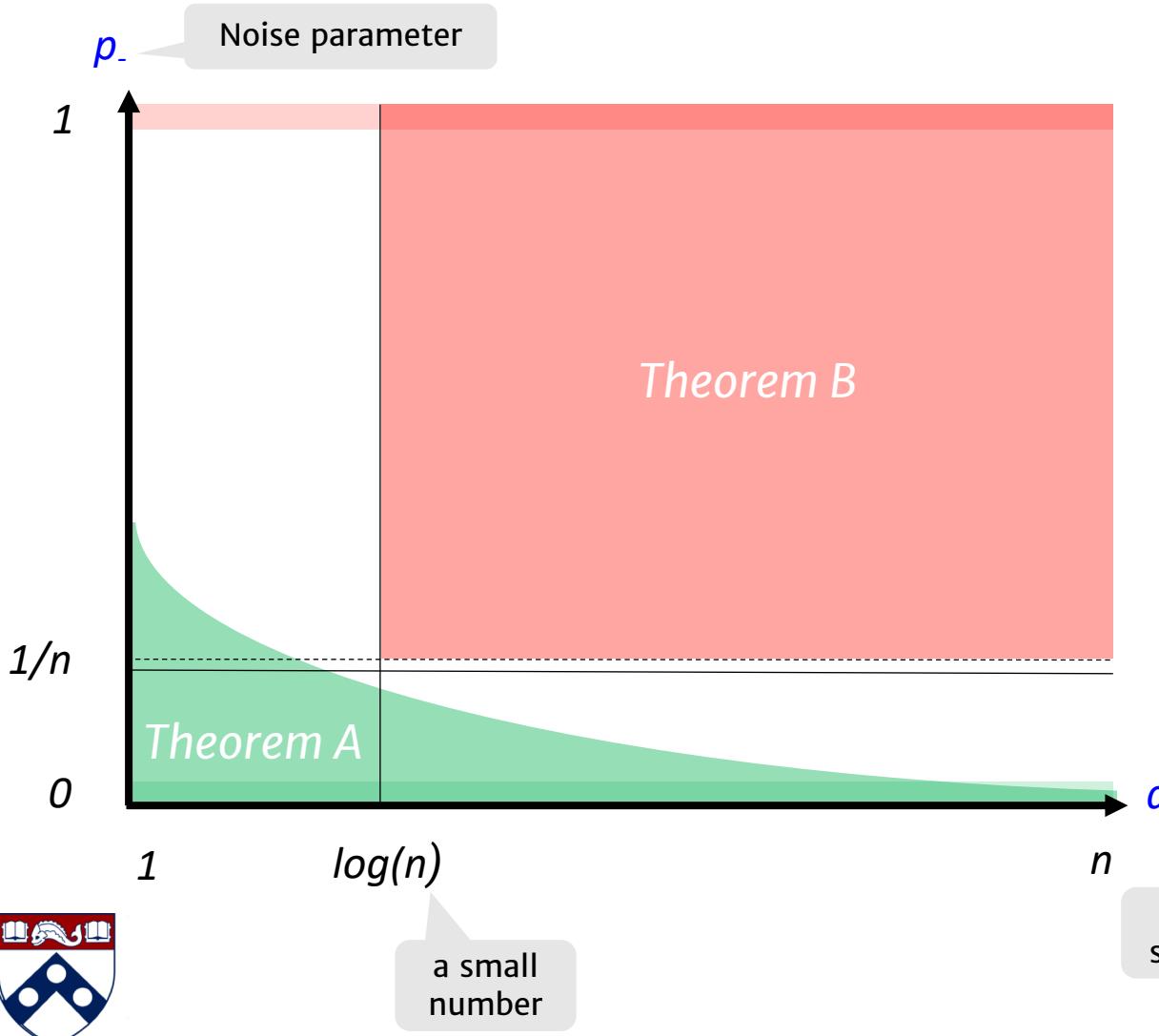


one of the
hypotheses require
d-step connectivity

Goal: Infer the connectivity of two given nodes (in the *unseen* meaning graph),
given observations in the symbol graph.



Results: Big picture [KSKSR, in submission]

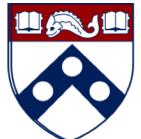


Theorem A (informal)

For any d there is a choice of parameters such that an algorithm can confidently distinguish H_1 and H_2 .

Theorem B (informal)

If $d \in \Omega(\log n)$ and “sufficient” noise, no algorithm can confidently distinguish H_1 and H_2 .



Practical lessons

- Pursuing “very long” multi-hop reasoning is unlikely to result in general results.
- Corollary: one has to focus on richer representations (i.e., dealing with **ambiguity** and **variability**) such that it leads to few number of hops needed.



Summary of this section

- A framework for studying “reasoning”, in the context of language problems.
- Multi-hop reasoning:
 - + There are non-trivial problems where successful reasoning is reliable.
 - - Reasoning with “large”-many hops likely to fail, even with small amount of noise.
- Implications for practice
 - **Hypothesis:** invest in representations that lead to few hops reasonings.



Summary of the talk

- NLU; potentials for significant impacts in the coming years.
- Answering questions: a natural evaluation protocol.
 - Many challenges along the way to this goal: ambiguity, variability, etc.
- Approaches:
 - System design: systems that abstracting over text and reasoning with it.
 - Evaluation: effective benchmarks to measure and incentivize the community.
 - Formalism: to study a class of reasoning algorithms in the context of language.



Thesis Publication

- **KSKSR.** *On the Capabilities and Limitations of Reasoning for Natural Language Understanding*, in submission.
- **ZKNR.** *A Question Answering Benchmark for Temporal Common-sense*, under review.
- **KCRUR.** *Looking Beyond the Surface: A Challenge Set for Reading Comprehension over Multiple Sentences*, NAACL, 2018.
- **KKSR.** *Question Answering as Global Reasoning over Semantic Abstractions*, AAAI, 2018.
- **KKSR.** *Learning What is Essential in Questions*, CoNLL, 2017.
- **KKSR.** *Question Answering via Integer Programming over Semi-Structured Knowledge*, IJCAI, 2016.



Other Publications

- NLP:

- CKWCR. *Seeing Things from a Different Angle: Discovering Diverse Perspectives about Claims*, under review.
- ZKCR. *Zero-Shot Open Entity Typing as Type-Compatible Grounding*, EMNLP, 2018.
- CEKSTT^K. *Combining Retrieval, Statistics, and Inference to Answer Elementary Science Questions*, AAAI, 2016.
- FKPWR. *Illinois-Profiler: Knowledge Schemas at Scale*, Cognitum, 2015.
- PKR. *Solving Hard Co-reference Problems*, NAACL, 2015.

- NLP software/tools:

- K et al. *CogCompNLP: Your Swiss Army Knife for NLP*, LREC, 2018.
- SCCKSVBWR. *EDISON: Feature Extraction for NLP, Simplified*, LREC, 2016.

- ML/Optimization/etc:

- KSCKCSR. *Relational Learning and Feature Extraction by Querying over Heterogeneous Information Networks*, StartAI, 2018.
- KKCMSP. *Better call Saul: Flexible Programming for Learning and Inference in NLP*, COLING, 2016.
- QK. *Online Learning with Adversarial Delays*, NourIPS, 2015.
- KNJF. *Joint Demosaicing and Denoising via Learned Non-parametric Random Fields*, TIP, 2014.





Snigdha Chaturvedi
(UCSC)



Tushar Khot
(AI2)



Dan Roth
(UPenn)



Ashish Sabharwal
(AI2)



Erfan Sadeqi Azer
(Indiana U)



Oren Etzioni
(AI2)



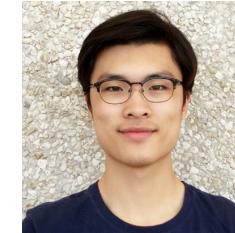
Peter Clark
(AI2)



Michael Roth
(Saarland U)

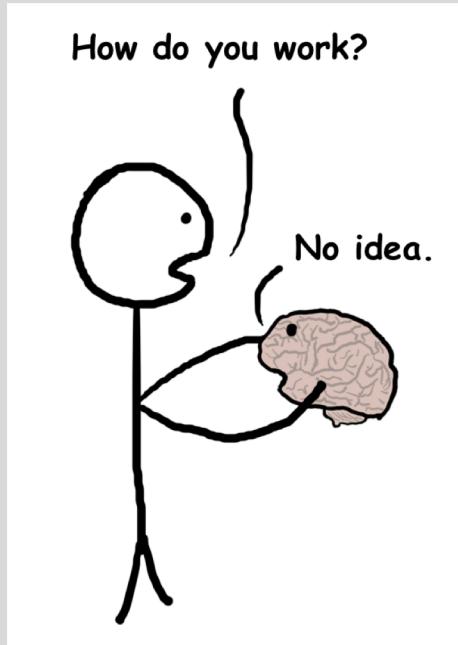


Shyam Upadhyay
(Upenn → Google)



Ben Zhou
(UIUC → UPenn)

- That's it folks



Questions?