Reading Comprehension

THE
UNIVERSITY
OF UTAH

Outline

- The task

- The reading comprehension dataset explosion

- What goes into a reading comprehension dataset?

Outline

- The task

- The reading comprehension dataset explosion

- What goes into a reading comprehension dataset?

Natural language understanding

Suppose someone claims that a program can understand natural language, how can we test for that?

Natural language understanding

Suppose someone claims that a program can understand natural language, how can we test for that?

How might you conduct this test?

Natural language understanding

Suppose someone claims that a program can understand natural language, how can we test for that?

Some ideas:

- Play the imitation game
- Have it answer questions
- See if it makes the same kind of inferences as people

Natural language understanding

Suppose someone claims that a program can understand natural language, how can we test for that?

Some ideas:

- Play the imitation game

- Have it answer questions

- See if it makes the same kind of inferences as people

The tricky part: How do we conduct these tests without having a human in the loop?

One answer: Reading Comprehension

One answer: Reading Comprehension

Water is split, providing a source of electrons and protons (hydrogen ions, H^+) and giving off O_2 as a by-product. Light absorbed by chlorophyll drives a transfer of the electrons and hydrogen ions from water to an acceptor called $NADP^+$.

What can the splitting of water lead to?
A: Light absorption
B: Transfer of ions

One answer: Reading Comprehension

Water is split, providing a source of electrons and protons (hydrogen ions, H^+) and giving off O_2 as a by-product. Light absorbed by chlorophyll drives a transfer of the electrons and hydrogen ions from water to an acceptor called $NADP^+$.

What can the splitting of water lead to?
A: Light absorption
B: Transfer of ions

One answer: Reading Comprehension

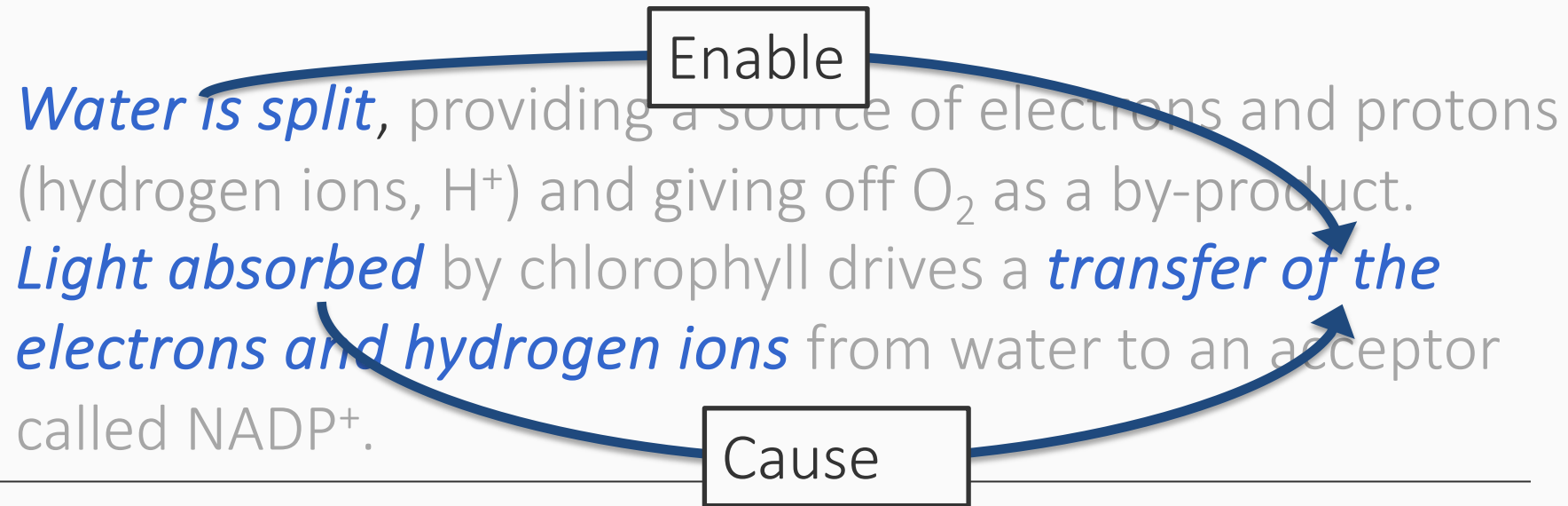*Water is split*, providing a source of electrons and protons (hydrogen ions, H^+) and giving off O_2 as a by-product. *Light absorbed* by chlorophyll drives a *transfer of the electrons and hydrogen ions* from water to an acceptor called $NADP^+$.

What can the splitting of water lead to?
A: Light absorption
B: Transfer of ions

Reading comprehension can be hard!

Enable

Cause

Water is split, providing a source of electrons and protons (hydrogen ions, H^+) and giving off O_2 as a by-product. *Light absorbed* by chlorophyll drives a *transfer of the electrons and hydrogen ions* from water to an acceptor called $NADP^+$.

What can the splitting of water lead to?
A: Light absorption
B: Transfer of ions

Reading comprehension has been a part of NLP for a very long time

A random sampling of NLP work before the year 2000 on this topic:

Eugene Charniak. 1972. *Toward A Model Of Children's Story Comprehension*. PhD Thesis, Massachusetts Institute of Technology.

Jaime G. Carbonell. 1978. POLITICS: Automated ideological reasoning. *Cognitive Science*, 2(1):27–51.

Mooney, Raymond J., and Gerald DeJong. "Learning schemata for natural language processing." In *IJCAI*, pp. 681-687. 1985.

Stefan Wermter and Wendy G. Lehnert. 1990. A survey of question answering in natural language processing. *Poetics*, 19(1):99–120.

Hirschman, Lynette, Marc Light, Eric Breck, and John D. Burger. "Deep read: A reading comprehension system." In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics*, pp. 325-332. 1999.

And many more…

Reading comprehension: The basic structure of the task

- All tests of reading comprehension have the same structure
  - Evidence: Typically one or more passages of text
  - A question that needs to be answered

- The answer is (hopefully) unambiguous

- Evaluation: Does the system get the right answer? Exact match?
  - Maybe give partial credit

Reading comprehension: The basic structure of the task

- All tests of reading comprehension have the same structure
  - Evidence: Typically one or more passages of text
  - A question that needs to be answered

- The answer is (hopefully) unambiguous

- Evaluation: Does the system get the right answer? Exact match?
  - Maybe give partial credit

Let us look at some example datasets

Outline

- The task

- The reading comprehension dataset explosion

- What goes into a reading comprehension dataset?

Reading Comprehension Dataset: SQuAD v1

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

Reading Comprehension Dataset: SQuAD v1

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

Passage

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. "*SQuAD: 100,000+ Questions for Machine Comprehension of Text*". In NAACL 2016

Reading Comprehension Dataset: SQuAD v1

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

Question

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. "*SQuAD: 100,000+ Questions for Machine Comprehension of Text*". In NAACL 2016

Reading Comprehension Dataset: SQuAD v1

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".
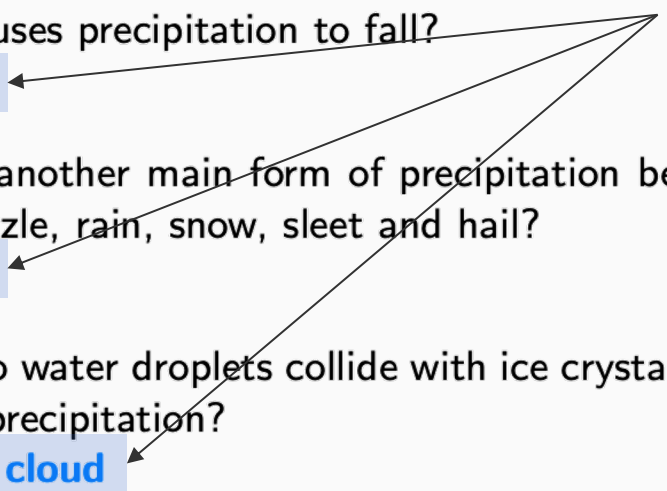
What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

Answers are spans in the passage

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. "*SQuAD: 100,000+ Questions for Machine Comprehension of Text*". In NAACL 2016

Reading Comprehension Dataset: SQuAD v1

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

Spend around 4 minutes on the following paragraph to ask 5 questions! If you can't ask 5 questions, ask 4 or 3 (worse), but do your best to ask 5. Select the answer from the paragraph by clicking on 'Select Answer', and then highlight the smallest segment of the paragraph that answers the question.

Oxygen is a chemical element with symbol O and atomic number 8. It is a member of the chalcogen group on the periodic table and is a highly reactive nonmetal and oxidizing agent that readily forms compounds (notably oxides) with most elements. By mass, oxygen is the third-most abundant element in the universe, after hydrogen and helium. At standard temperature and pressure, two atoms of the element bind to form dioxygen, a colorless and odorless diatomic gas with the formula O

2. Diatomic oxygen gas constitutes 20.8% of the Earth's atmosphere. However, monitoring of atmospheric oxygen levels show a global downward trend, because of fossil-fuel burning. Oxygen is the most abundant element by mass in the Earth's crust as part of oxide compounds such as silicon dioxide, making up almost half of the crust's mass.

When asking questions, **avoid using** the same words/phrases as in the paragraph. Also, you are encouraged to pose **hard questions**.

Ask a question here. Try using your own words

Select Answer

Ask a question here. Try using your own words

Select Answer

Collected by crowdsourcing

Reading Comprehension Dataset: SQuAD v2

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail… Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

What intractability means in practice is open to debate. Saying that a problem is not in P does not imply that all large cases of the problem are hard or even that most of them are. For example, the decision problem in Presburger arithmetic has been shown not to be in P, yet algorithms have been written that solve the problem in reasonable times in most cases. Similarly, algorithms can solve the NP-complete knapsack problem over a wide range of sizes in less than quadratic time and SAT solvers routinely handle large instances of the NP-complete Boolean satisfiability problem.

What has not been done to establish solutions in reasonable period of time?

No Answer

SQuAD v2 adds a set of unanswerable questions

Reading comprehension datasets: A sampler

RACE [Lai et al 2019]: Questions from English exams for Chinese students aged 12-18

Passage:
In a small village in England about 150 years ago, a mail coach was standing on the street. It didn't come to that village often. People had to pay a lot to get a letter. The person who sent the letter didn't have to pay the postage, while the receiver had to. "Here's a letter for Miss Alice Brown," said the mailman.
" I'm Alice Brown," a girl of about 18 said in a low voice.
Alice looked at the envelope for a minute, and then handed it back to the mailman.
"I'm sorry I can't take it, I don't have enough money to pay it", she said.
A gentleman standing around were very sorry for her. Then he came up and paid the postage for her.
When the gentleman gave the letter to her, she said with a smile, " Thank you very much, This letter is from Tom. I'm going to marry him. He went to London to look for work. I've waited a long time for this letter, but now I don't need it, there is nothing in it."
"Really? How do you know that?" the gentleman said in surprise.
"He told me that he would put some signs on the envelope. Look, sir, this cross in the corner means that he is well and this circle means he has found work. That's good news."
The gentleman was Sir Rowland Hill. He didn't forgot Alice and her letter.
"The postage to be paid by the receiver has to be changed," he said to himself and had a good plan.
"The postage has to be much lower, what about a penny? And the person who sends the letter pays the postage. He has to buy a stamp and put it on the envelope." he said . The government accepted his plan. Then the first stamp was put out in 1840. It was called the "Penny Black". It had a picture of the Queen on it.

Questions:

1): The first postage stamp was made _.
A. in England B. in America C. by Alice D. in 1910

2): The girl handed the letter back to the mailman because _
.
A. she didn't know whose letter it was
B. she had no money to pay the postage
C. she received the letter but she didn't want to open it
D. she had already known what was written in the letter

3): We can know from Alice's words that _ .
A. Tom had told her what the signs meant before leaving
B. Alice was clever and could guess the meaning of the signs
C. Alice had put the signs on the envelope herself
D. Tom had put the signs as Alice had told him to

4): The idea of using stamps was thought of by _ .
A. the government
B. Sir Rowland Hill
C. Alice Brown
D. Tom

5): From the passage we know the high postage made _ .
A. people never send each other letters
B. lovers almost lose every touch with each other
C. people try their best to avoid paying it
D. receivers refuse to pay the coming letters

Answer: ADABC

Reading comprehension datasets: A sampler

RACE [Lai et al 2019]: Questions from English exams for Chinese students aged 12-18

HotpotQA [Yang et al 2018]: Multiple documents, requires multiple reasoning hops

Paragraph A, Return to Olympus:
[1] *Return to Olympus is the only album by the alternative rock band Malfunkshun.* [2] *It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990.* [3] Stone Gossard, of Pearl Jam, had compiled the songs and released the album on his label, Loosegroove Records.

Paragraph B, Mother Love Bone:
[4] *Mother Love Bone was an American rock band that formed in Seattle, Washington in 1987.* [5] The band was active from 1987 to 1990. [6] *Frontman Andrew Wood's personality and compositions helped to catapult the group to the top of the burgeoning late 1980s/early 1990s Seattle music scene.* [7] *Wood died only days before the scheduled release of the band's debut album, "Apple", thus ending the group's hopes of success.* [8] The album was finally released a few months later.

Q: What was the former band of the member of Mother Love Bone who died just before the release of "Apple"?
A: Malfunkshun
Supporting facts: 1, 2, 4, 6, 7

Figure 1: An example of the multi-hop questions in HOTPOTQA. We also highlight the supporting facts in *blue italics*, which are also part of the dataset.

Reading comprehension datasets: A sampler

RACE [Lai et al 2019]: Questions from English exams for Chinese students aged 12-18

HotpotQA [Yang et al 2018]: Multiple documents, requires multiple reasoning hops

BoolQ [Clark et al 2019]: Only yes/no questions

Q:	Has the UK been hit by a hurricane?
P:	The Great Storm of 1987 was a violent extratropical cyclone which caused casualties in England, France and the Channel Islands …
A:	Yes. [An example event is given.]
Q:	Does France have a Prime Minister and a President?
P:	… The extent to which those decisions lie with the Prime Minister or President depends upon …
A:	Yes. [Both are mentioned, so it can be inferred both exist.]
Q:	Have the San Jose Sharks won a Stanley Cup?
P:	… The Sharks have advanced to the Stanley Cup finals once, losing to the Pittsburgh Penguins in 2016 …
A:	No. [They were in the finals once, and lost.]

Figure 1: Example yes/no questions from the BoolQ dataset. Each example consists of a question (**Q**), an excerpt from a passage (**P**), and an answer (**A**) with an explanation added for clarity.

Reading comprehension datasets: A sampler

RACE [Lai et al 2019]: Questions from English exams for Chinese students aged 12-18

HotpotQA [Yang et al 2018]: Multiple documents, requires multiple reasoning hops
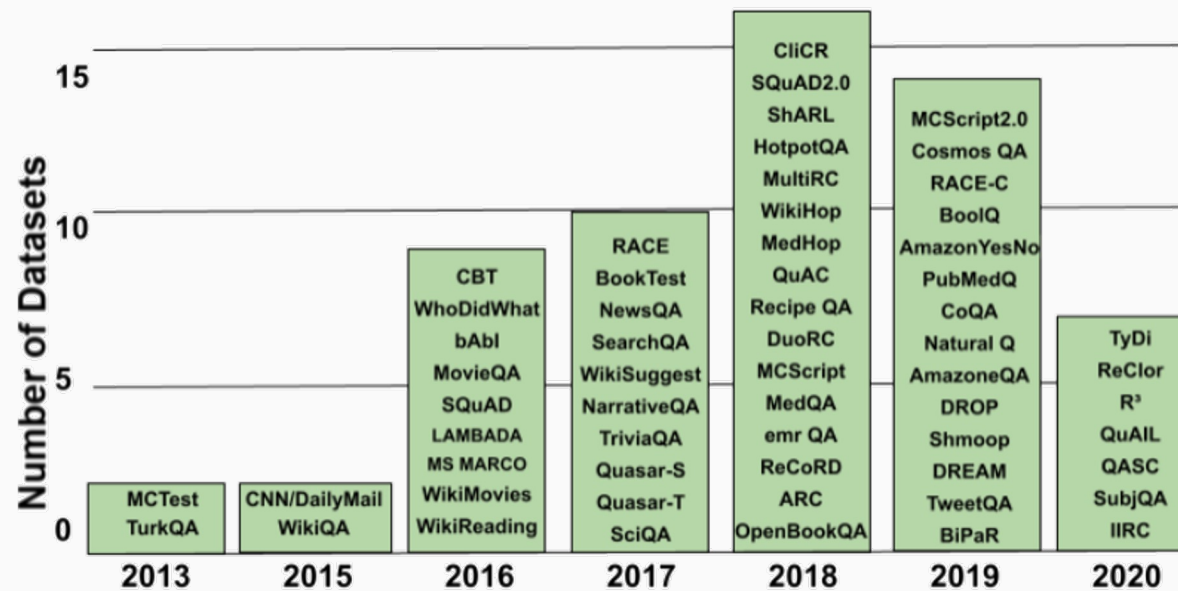
BoolQ [Clark et al 2019]: Only yes/no questions

… and many more

Q:	Has the UK been hit by a hurricane?
P:	The Great Storm of 1987 was a violent extratropical cyclone which caused casualties in England, France and the Channel Islands …
A:	Yes. [An example event is given.]
Q:	Does France have a Prime Minister and a President?
P:	… The extent to which those decisions lie with the Prime Minister or President depends upon …
A:	Yes. [Both are mentioned, so it can be inferred both exist.]
Q:	Have the San Jose Sharks won a Stanley Cup?
P:	… The Sharks have advanced to the Stanley Cup finals once, losing to the Pittsburgh Penguins in 2016 …
A:	No. [They were in the finals once, and lost.]

Figure 1: Example yes/no questions from the BoolQ dataset. Each example consists of a question (**Q**), an excerpt from a passage (**P**), and an answer (**A**) with an explanation added for clarity.

English reading comprehension datasets released per year

Dzendzik, Daria, Jennifer Foster, and Carl Vogel. "English Machine Reading Comprehension Datasets: A Survey." In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 8784-8804. 2021.

Outline

- The task

- The reading comprehension dataset explosion

- What goes into a reading comprehension dataset?

Reading comprehension: The basic structure of the task

- All tests of reading comprehension have the same structure
  - Evidence: Typically one or more passages of text
  - A question that needs to be answered

- The answer is (hopefully) unambiguous

- Evaluation: Does the system get the right answer? Exact match?
  - Maybe give partial credit

Within this structure there are several design choices for constructing datasets
  - Answer type
  - Is the text passage from one document or more?
  - Are there images or other modalities involved?
  - Is there dialogue involved?
  - Where is the data sourced from?

Design choices: Question & Answer types

- Cloze type: Basically fill-in-the-blanks

Text Cloze Style Question	Context Modalities: Images and Descriptions of Steps

Recipe: Last-Minute Lasagna

1. Heat oven to 375 degrees F. Spoon a thin layer of sauce over the bottom of a 9-by-13-inch baking dish.
2. Cover with a single layer of ravioli.
3. Top with half the spinach half the mozzarella and a third of the remaining sauce.
4. Repeat with another layer of ravioli and the remaining spinach mozzarella and half the remaining sauce.
5. Top with another layer of ravioli and the remaining sauce not all the ravioli may be needed. Sprinkle with the Parmesan.
6. Cover with foil and bake for 30 minutes. Uncover and bake until bubbly, 5 to 10 minutes.
7. Let cool 5 minutes before spooning onto individual plates.

Step 1   Step 2   Step 3   Step 4

Step 5   Step 6   Step 7

Question	Choose the best text for the missing blank to correctly complete the recipe Cover. _____. Bake. Cool, serve.
Answer	**A. Top, sprinkle** B. Finishing touches C. Layer it up D. Ravioli bonus round

Semih Yagcioglu, Aykut Erdem, Erkut Erdem, and Nazli Ikizler-Cinbis. *"RecipeQA: A Challenge Dataset for Multimodal Comprehension of Cooking Recipes"*. In EMNLP 2018

Design choices: Question & Answer types

- Cloze type: Basically fill-in-the-blanks

Text Cloze Style Question	Context Modalities: Images and Descriptions of Steps

Recipe: Last-Minute Lasagna

1. Heat oven to 375 degrees F. Spoon a thin layer of sauce over the bottom of a 9-by-13-inch baking dish.
2. Cover with a single layer of ravioli.
3. Top with half the spinach half the mozzarella and a third of the remaining sauce.
4. Repeat with another layer of ravioli and the remaining spinach mozzarella and half the remaining sauce.
5. Top with another layer of ravioli and the remaining sauce not all the ravioli may be needed. Sprinkle with the Parmesan.
6. Cover with foil and bake for 30 minutes. Uncover and bake until bubbly, 5 to 10 minutes.
7. Let cool 5 minutes before spooning onto individual plates.

Question	Choose the best text for the missing blank to correctly complete the recipe Cover. _____. Bake. Cool, serve.
Answer	**A. Top, sprinkle** B. Finishing touches C. Layer it up D. Ravioli bonus round

Semih Yagcioglu, Aykut Erdem, Erkut Erdem, and Nazli Ikizler-Cinbis. "*RecipeQA: A Challenge Dataset for Multimodal Comprehension of Cooking Recipes*". In EMNLP 2018

Design choices: Question & Answer types

- Cloze type: Basically fill-in-the-blanks
- Multiple-choice questions

James the Turtle was always getting in trouble.
Sometimes he'd reach into the freezer and empty out
all the food. Other times he'd sled on the deck and get
a splinter. His aunt Jane tried as hard as she could to
keep him out of trouble, but he was sneaky and got
into lots of trouble behind her back.
One day, James thought he would go into town and
see what kind of trouble he could get into. He went to
the grocery store and pulled all the pudding off the
shelves and ate two jars. Then he walked to the fast
food restaurant and ordered 15 bags of fries. He did-
n't pay, and instead headed home.
His aunt was waiting for him in his room. She told
James that she loved him, but he would have to start
acting like a well-behaved turtle.
After about a month, and after getting into lots of
trouble, James finally made up his mind to be a better
turtle.

1) What is the name of the trouble making turtle?
A) Fries
B) Pudding
C) James
D) Jane

Matthew Richardson, Christopher J.C. Burges, and Erin Renshaw.
"*MCTest: A Challenge Dataset for the Open-Domain Machine
Comprehension of Text*". In EMNLP 2013

Design choices: Question & Answer types

- Cloze type: Basically fill-in-the-blanks

- Multiple-choice questions

- Boolean: Yes/No questions

Question:
Do preoperative statins reduce atrial fibrillation after coronary artery bypass grafting?

Context:
(Objective) Recent studies have demonstrated that statins have pleiotropic effects, including anti-inflammatory effects and atrial fibrillation (AF) preventive effects [...]
(Methods) 221 patients underwent CABG in our hospital from 2004 to 2007. 14 patients with preoperative AF and 4 patients with concomitant valve surgery [...]
(Results) The overall incidence of postoperative AF was 26%. *Postoperative AF was significantly lower in the Statin group compared with the Non-statin group (16% versus 33%, p=0.005).* Multivariate analysis demonstrated that independent predictors of AF [...]

Long Answer:
(Conclusion) Our study indicated that preoperative statin therapy seems to reduce AF development after CABG.

Answer: yes

Figure 1: An instance (Sakamoto et al., 2011) of PubMedQA dataset: Question is the original question title; Context includes the structured abstract except its conclusive part, which serves as the Long Answer; Human experts annotated the Answer yes. Supporting fact for the answer is *highlighted*.

Qiao Jin, Bhuwan Dhingra, Zhengping Liu, William Cohen, and Xinghua Lu. *"PubMedQA: A Dataset for Biomedical Research Question Answering"*. EMNLP 2019

Design choices: Question & Answer types

- Cloze type: Basically fill-in-the-blanks
- Multiple-choice questions
- Boolean: Yes/No questions
- Extractive: The answer is a span in the passage

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. "*SQuAD: 100,000+ Questions for Machine Comprehension of Text*". In NAACL 2016

Design choices: Question & Answer types

- Cloze type: Basically fill-in-the-blanks

- Multiple-choice questions

- Boolean: Yes/No questions

- Extractive: The answer is a span in the passage

- Free-form answers

Title: Ghostbusters II
Question: How is Oscar related to Dana?
Answer: her son
Summary snippet: …Peter's former girlfriend Dana Barrett has had a son, Oscar…
Story snippet:

DANA (setting the wheel brakes on the buggy)
Thank you, Frank. I'll get the hang of this eventually.

She continues digging in her purse while Frank leans over the buggy and makes funny faces at the baby, OSCAR, a very cute nine-month old boy.

FRANK (to the baby)
Hiya, Oscar. What do you say, slugger?

FRANK (to Dana)
That's a good-looking kid you got there, Ms. Barrett.

Tomáš Kočiský, Jonathan Schwarz, Phil Blunsom, Chris Dyer, Karl Moritz Hermann, Gábor Melis, Edward Grefenstette. "*The NarrativeQA Reading Comprehension Challenge*". In TACL 2018

Design choices: The evidence

The evidence can take different forms

- Usually unstructured text
- Sometimes semi-structured text, e.g. WikiTableQuestions, TableQA
- Sometimes structured knowledge
- Images (i.e., visual QA), e.g. ChartQA
- Audio, e.g. DAQA, Spoken-SQuAD
- Movies, e.g. MovieQA
- Mixed modalities

Rogers, Anna, Matt Gardner, and Isabelle Augenstein. "*QA dataset explosion: A taxonomy of NLP resources for question answering and reading comprehension.*" *ACM Computing Surveys* 55, no. 10 (2023): 1-45.

Design choices: The evidence
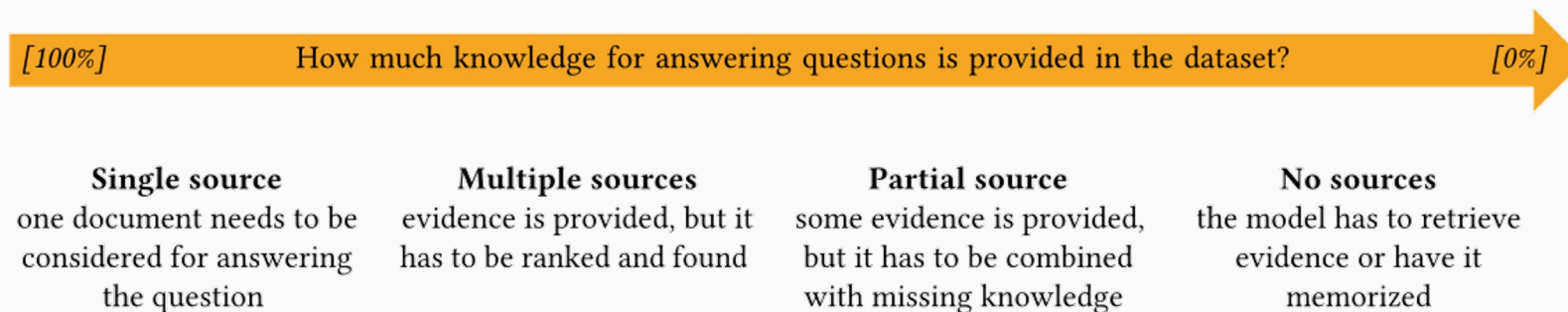
The evidence can take different forms

- Usually unstructured text
- Sometimes semi-structured text, e.g. WikiTableQuestions, TableQA
- Sometimes structured knowledge
- Images (i.e., visual QA), e.g. ChartQA
- Audio, e.g. DAQA, Spoken-SQuAD
- Movies, e.g. MovieQA
- Mixed modalities

Varying amounts of evidence can be presented

[100%] How much knowledge for answering questions is provided in the dataset? [0%]

Single source
one document needs to be considered for answering the question

Multiple sources
evidence is provided, but it has to be ranked and found

Partial source
some evidence is provided, but it has to be combined with missing knowledge

No sources
the model has to retrieve evidence or have it memorized

Rogers, Anna, Matt Gardner, and Isabelle Augenstein. "*QA dataset explosion: A taxonomy of NLP resources for question answering and reading comprehension.*" *ACM Computing Surveys* 55, no. 10 (2023): 1-45.

Data sources

Different datasets draw from different sources

- **News**: NewsQA, MultiRC, ReCoRD, QuAIL,CNN/DailyMail, WhoDidWhat, CoQA

- **Books**: ChildrenBookTest, BookTest, LAMBADA, BiPaR, partly CoQA, Shmoop, SciQ

- **Movie scripts**: MovieQA, WikiMovies, DuoRC, MultiRC and NarrativeQA

- **Clinical reports**, **medical abstracts**: CliCR, PubMedQuestions, MedQA, emrQA, QAngaroo MedHop

- **Exams**:
  - English as a Foreign Language examinations: RACE, RACE-C, and DREAM
  - GMAT and LSAT: ReClor
  - Medical: MedQA
  - Science exams: SciQ, ARC, OpenBookQA, QASC

- **Online reviews**: AmazonQA, AmazonYesNo, SubjQA

Dzendzik, Daria, Jennifer Foster, and Carl Vogel. "English Machine Reading Comprehension Datasets: A Survey." In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 8784-8804. 2021.

Data sources

Different datasets draw from different sources

- **News**: NewsQA, MultiRC, ReCoRD, QuAIL,CNN/DailyMail, WhoDidWhat, CoQA

- **Books**: ChildrenBookTest, BookTest, LAMBADA, BiPaR, partly CoQA, Shmoop, SciQ

- **Movie scripts**: MovieQA, WikiMovies, DuoRC, MultiRC and NarrativeQA

- **Clinical reports**, **medical abstracts**: CliCR, PubMedQuestions, MedQA, emrQA, QAngaroo MedHop

- **Exams**:
  - English as a Foreign Language examinations: RACE, RACE-C, and DREAM
  - GMAT and LSAT: ReClor
  - Medical: MedQA
  - Science exams: SciQ, ARC, OpenBookQA, QASC

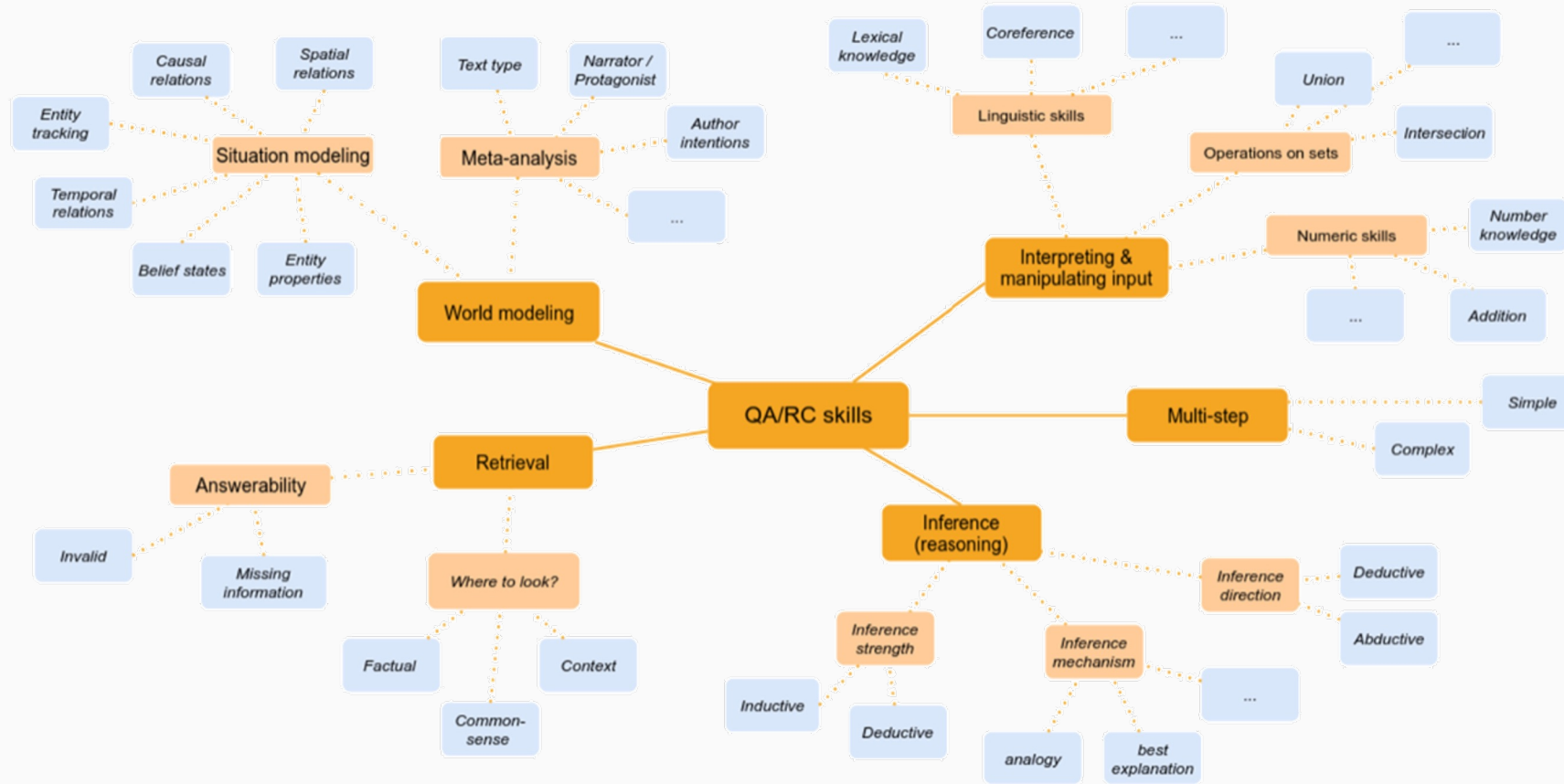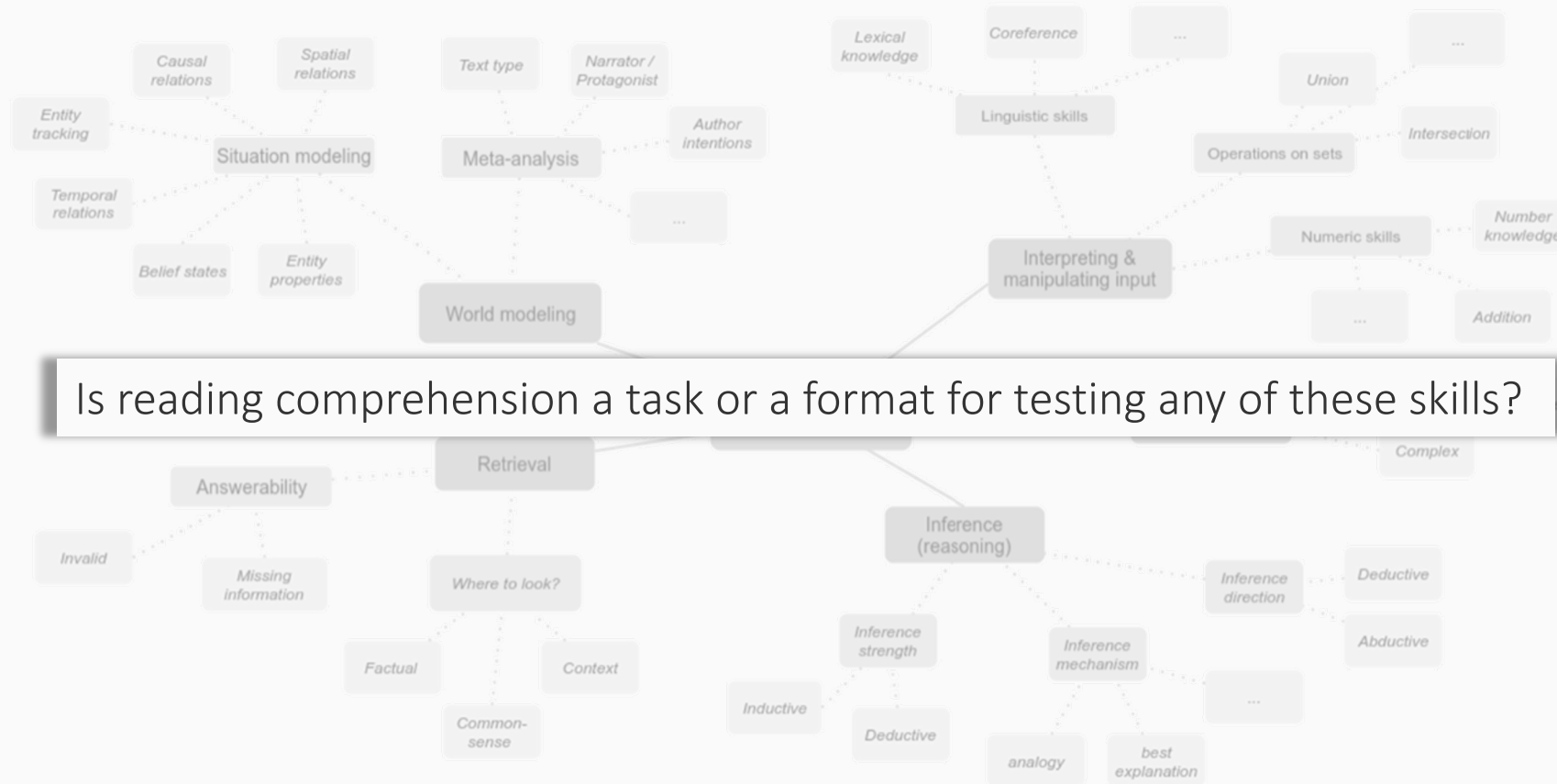- **Online reviews**: AmazonQA, AmazonYesNo, SubjQA

And many more!

Dzendzik, Daria, Jennifer Foster, and Carl Vogel. "English Machine Reading Comprehension Datasets: A Survey." In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 8784-8804. 2021.

What skills are needed to perform reading comprehension?

Rogers, Anna, Matt Gardner, and Isabelle Augenstein. "*QA dataset explosion: A taxonomy of NLP resources for question answering and reading comprehension.*" *ACM Computing Surveys* 55, no. 10 (2023): 1-45.

What skills are needed to perform reading comprehension?

Is reading comprehension a task or a format for testing any of these skills?

Gardner, Matt, Jonathan Berant, Hannaneh Hajishirzi, Alon Talmor, and Sewon Min. "*Question Answering is a Format; When is it Useful?*" arXiv preprint arXiv:1909.11291 (2019).

Summary

Reading comprehension as a test for text understanding

Many datasets out there that differ in their sources of data and their assumptions about reading comprehension
- Mostly English and Chinese, some other languages like Russian, Korean, etc.

A few standard approaches for modeling
- Train an encoder model to extract the answer or to predict an answer for multiple choice
- Use a decoder or an encoder decoder model to predict the text of the answer
- …

What are some problems with the definition of the task? Do we have a full fledged test of reasoning here?