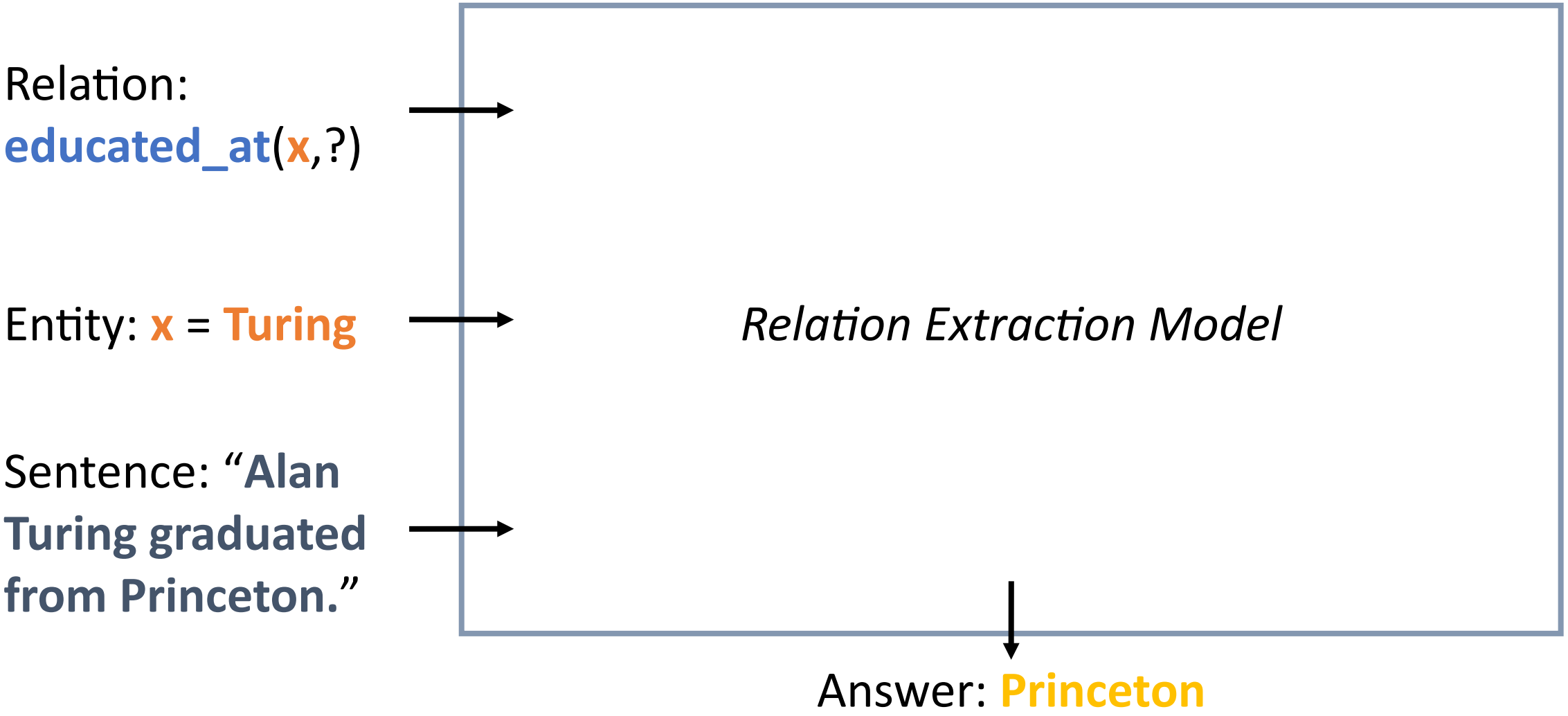
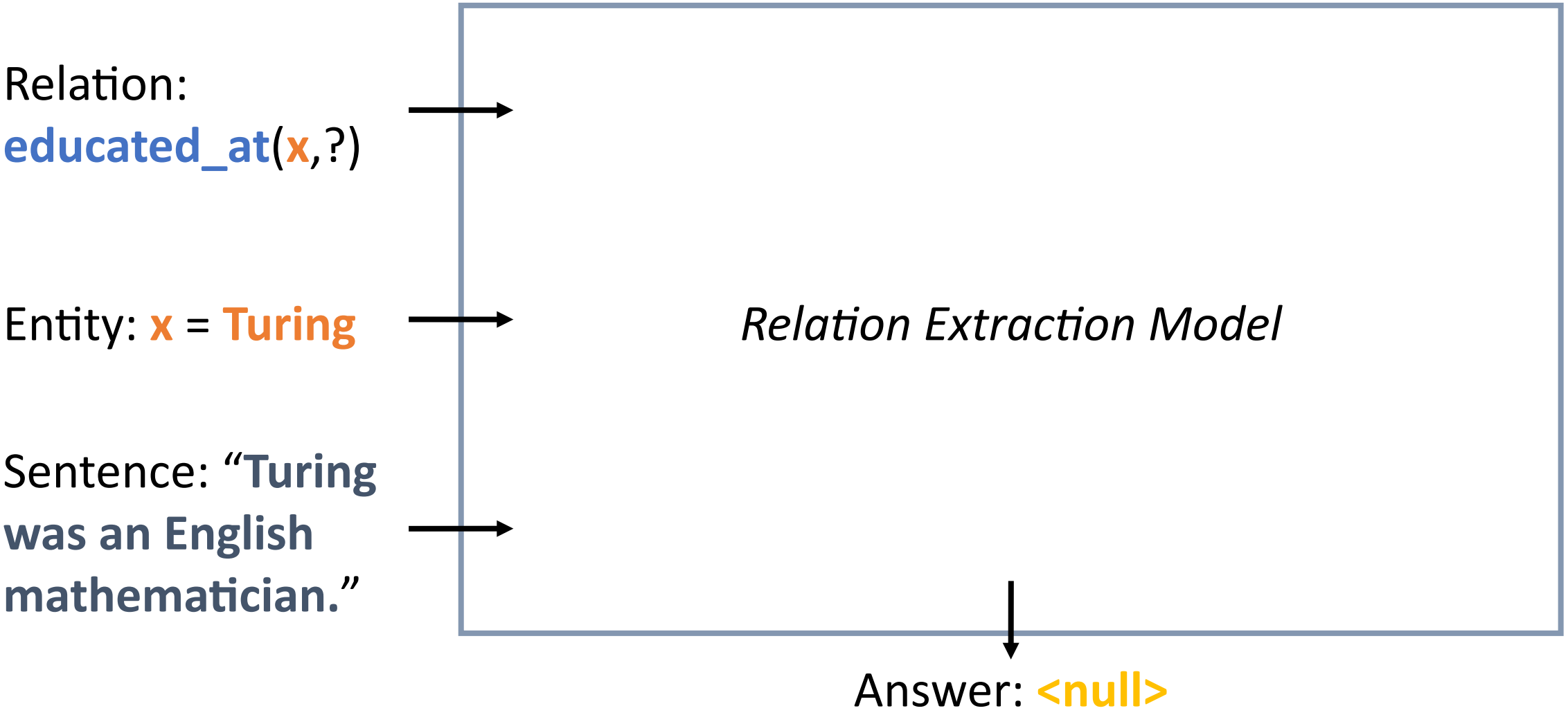
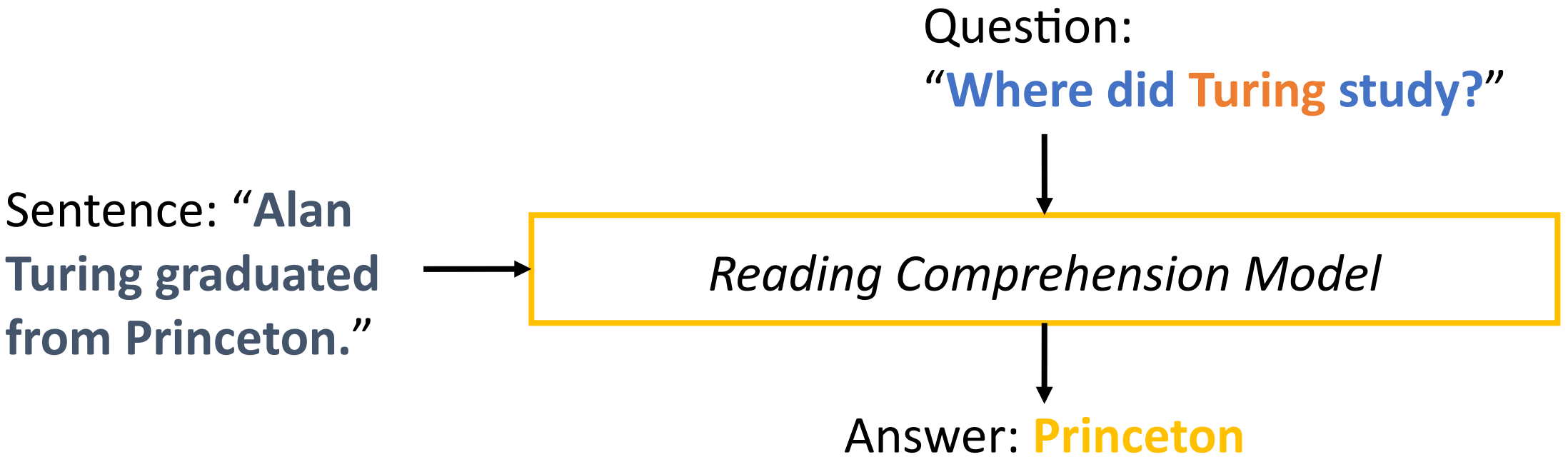
Relation Extraction (Slot Filling)



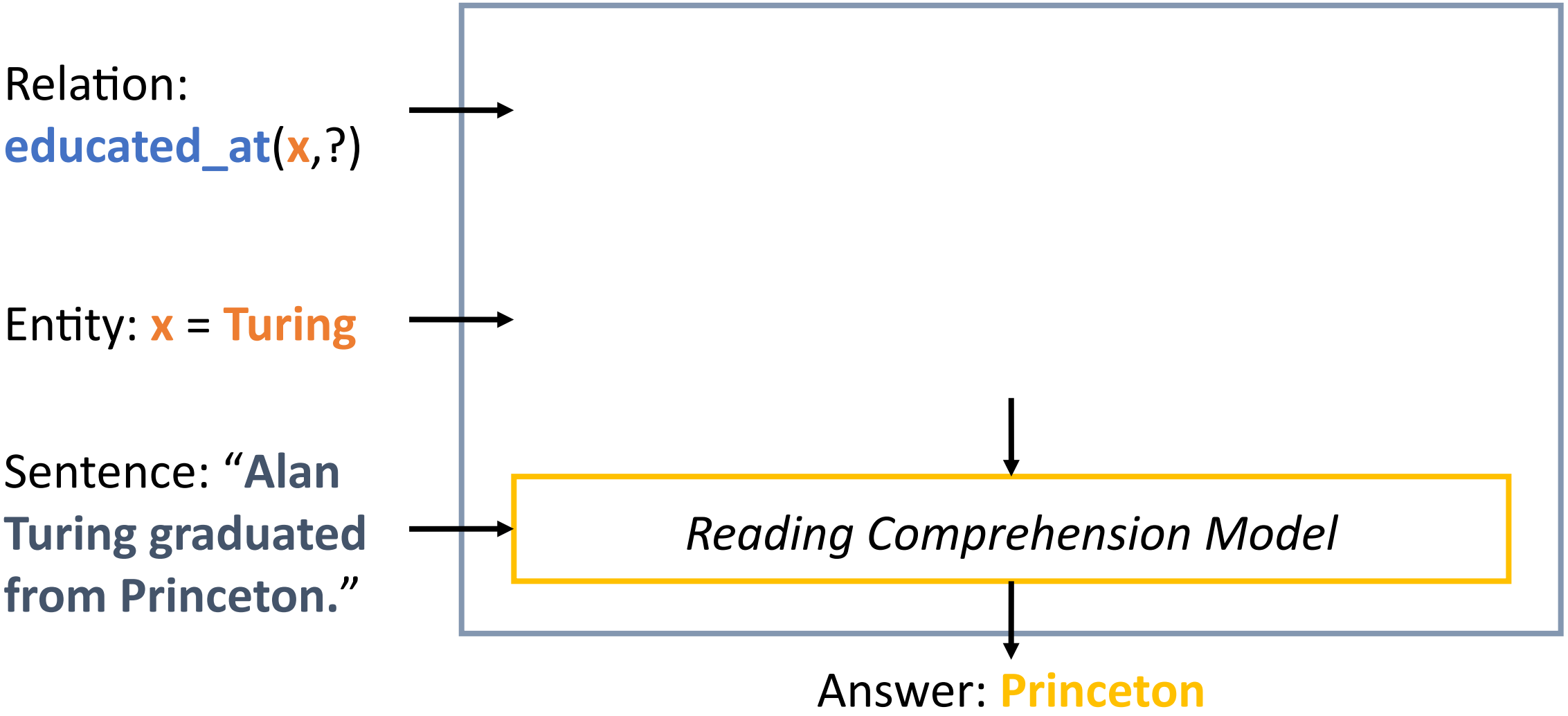
Relation Extraction (Slot Filling)



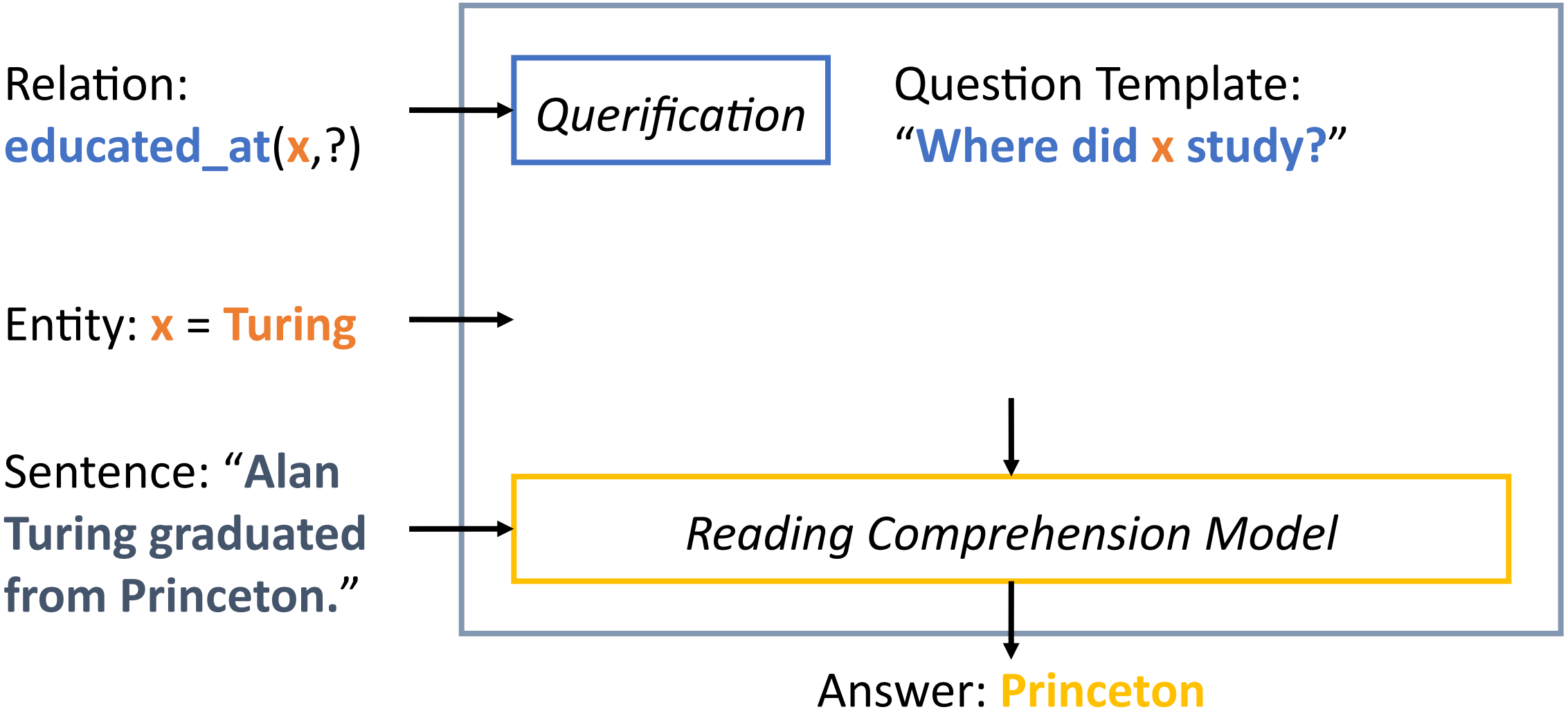
Reading Comprehension



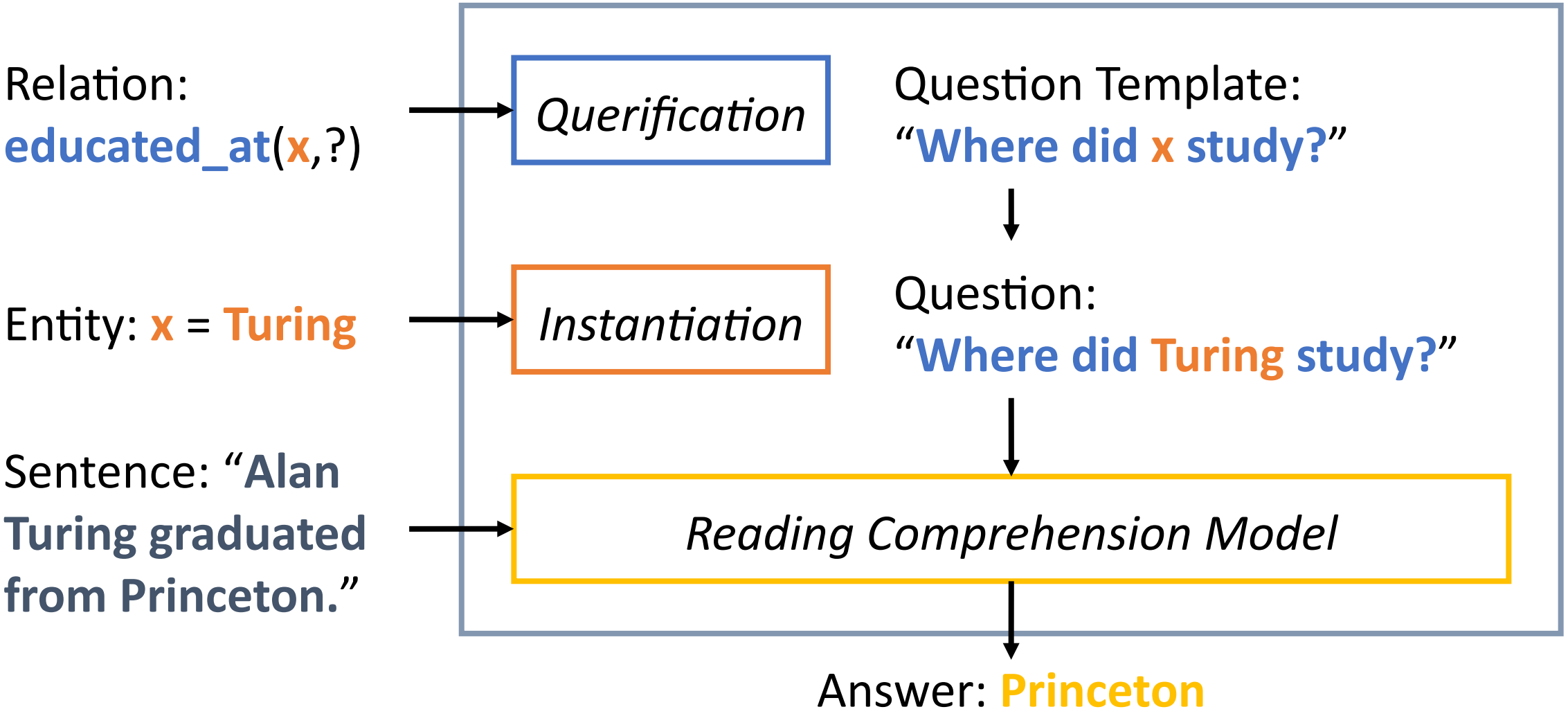
Relation Extraction via Reading Comprehension



Relation Extraction via Reading Comprehension



Relation Extraction via Reading Comprehension



Advantages

Advantage: Generalize to Unseen Questions

- Provides a **natural-language API** for defining and querying relations

educated_at(**Turing**, ?)

“Where did Turing study?”

“Which university did Turing go to?”

Advantage: Generalize to Unseen Relations

- Enables **zero-shot** relation extraction

  Train: **educated_at**, **occupation**, **spouse**, …

  Test: **country**

- Impossible for many relation-extraction systems

Challenges

- Translating relations into question templates
  - Schema Querification
  - Generated over 30,000,000 examples

- Modeling reading comprehension
  - Plenty of research on SQuAD (Rajpurkar et al, EMNLP 2016)
  - Model based on BiDAF (Seo et al, ICLR 2017)

- Predicting negative instances
  - Modified BiDAF can indicate no answer

Challenges

Instance Querification

educated_at(**Turing**, **Princeton**)

“Where did Turing study?”

“Where did Turing graduate from?”

“Which university did Turing go to?”

Problem: scaling to millions of examples

Large-Scale Simple Question Answering with Memory Networks (Bordes et al, 2015)

Schema Querification: The Challenge

educated_at(**x**,?)

"Where did x study?"

"Where did x graduate from?"

"Which university did x go to?"

Problem: not enough information

Schema Querification: Crowdsourcing Solution

Ask a single question about x whose answer is, for each sentence, the underlined spans.

1) The wine is produced in the **x** region of **France**.
2) **x**, the capital of **Mexico**, is the most populous city in North America.
3) **x** is an unincorporated and organized territory of **the United States**.
4) The **x** mountain range stretched across **the United States** and **Canada**.

"In which country is x located?"

Dataset

- Annotated **120 relations** from WikiReading (Hewlett et al, ACL 2016)
- Collected **10 templates per relation** with **high agreement**
- Generated over **30,000,000 reading comprehension examples**
- Generated **negative examples** by mixing questions about same entity

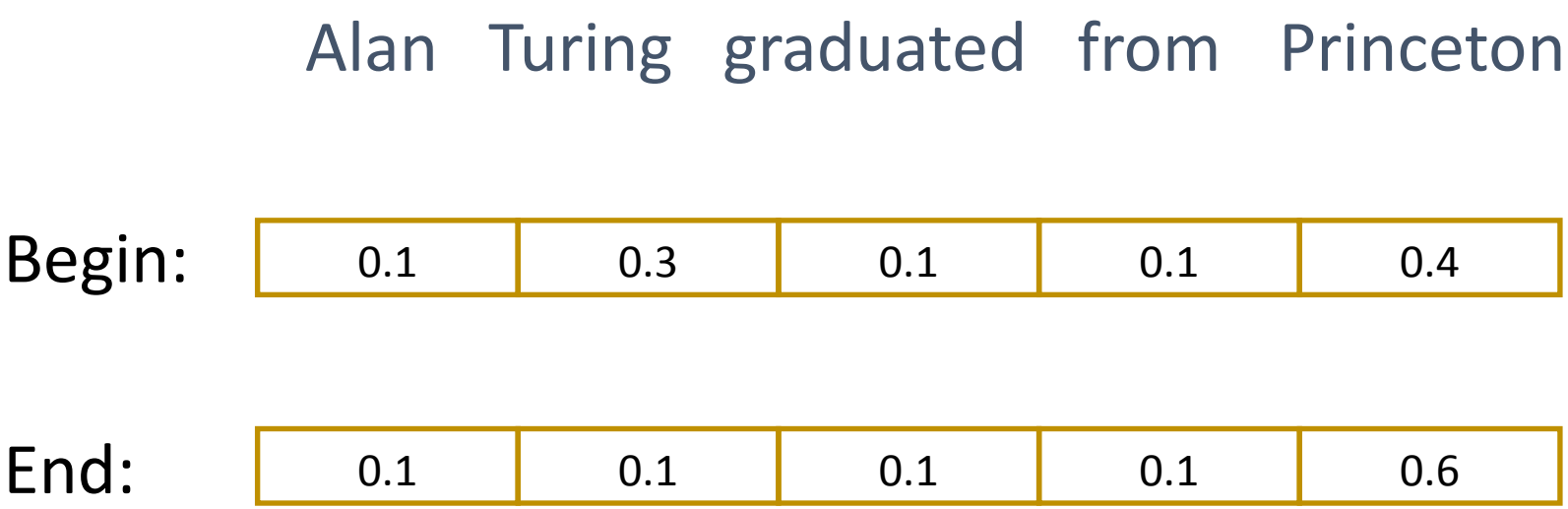
Reading Comprehension Model: BiDAF

- Pre-trained word embeddings
- Character embeddings
- Bi-directional LSTMs for contextualization
- Special attention mechanism:
  - Attends on both question and sentence
  - Computed independently for each token in the sentence

Bi-Directional Attention Flow for Machine Comprehension
(Seo et al, ICLR 2017)

Reading Comprehension Model: BiDAF

- Output Layer:

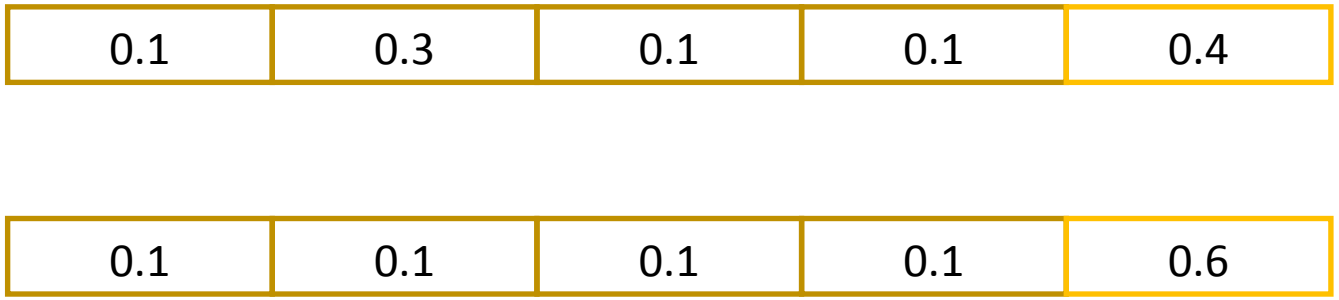


Bi-Directional Attention Flow for Machine Comprehension
(Seo et al, ICLR 2017)

Reading Comprehension Model: BiDAF

- Output Layer:

Alan Turing graduated from [Princeton]



Bi-Directional Attention Flow for Machine Comprehension
(Seo et al, ICLR 2017)

Predicting Negative Instances

- Output Layer:

	Alan	Turing	graduated	from	Princeton	<null>
Begin:	0.01	0.03	0.01	0.01	0.04	0.9
End:	0.01	0.01	0.01	0.01	0.06	0.9

Add <null> token to the sentence

Predicting Negative Instances

- Output Layer:

	Alan	Turing	graduated	from	Princeton	[<null>]
Begin:	0.01	0.03	0.01	0.01	0.04	0.9
End:	0.01	0.01	0.01	0.01	0.06	0.9

if argmax = <null>, predict **no answer**

Experiments

Generalizing to Unseen Questions

- Model is trained on several question templates per relation
  - “**Where did Alan Turing study?**”
  - “**Where did Claude Shannon graduate from?**”
  - “**Which university did Edsger Dijkstra go to?**”

- User asks about the relation using a different form
  - “**Which university awarded Noam Chomsky a PhD?**”

Generalizing to Unseen Questions

- **Experiment:** split the data by **question templates**

- Performance on **seen** question templates: 86.6% F1
- Performance on **unseen** question templates: 83.1% F1

- Our method is robust to new descriptions of existing relations

Generalizing to Unseen Relations

- Model is trained on several relations
  - “**Where did Alan Turing study?**” (**educated_at**)
  - “**What is Ivanka Trump’s job?**” (**occupation**)
  - “**Who is Justin Trudeau married to?**” (**spouse**)

- User asks about a new, unseen relation
  - “**In which country is Seattle located?**” (**country**)

Generalizing to Unseen Relations

- **Experiment:** split the data by **relations**

Results

- Random named-entity baseline: 12.2% F1
- Off-the-shelf RE system: *impossible*
- BiDAF w/ relation name as query: 33.4% F1
- BiDAF w/ querified relation as query: 39.6% F1

  \+ multiple questions at test: 41.1% F1

Why does a reading comprehension model enable zero-shot relation extraction?

- It can learn **answer types** that are used across relations

  Q: **When** was the Snow Hawk released?

  S: The Snow Hawk is a **1925** film…

- It can detect **paraphrases of relations**

  Q: Who **started** the Furstenberg China Factory?

  S: The Furstenberg China Factory **was founded by** Johann Georg…