AmbigQA: Answering Ambiguous Open-domain Questions

Sewon Min, Julian Michael, Hannaneh Hajishirzi, Luke Zettlemoyer




UWNLP facebook AI2

EMNLP 2020

Motivation

What season does Meredith and Derek get married in Grey's Anatomy?

Who was England's prime minister during WWI?

Questions from NQ-open (Kwiakowski et al 2020)

Motivation

What season does Meredith and Derek get married in Grey's Anatomy?

Now or Never (Grey's Anatomy)

... Season 5 ... Meredith and Derek have decided not to wait any longer to get married and just go to City Hall that evening ... writes their vows down on a post-it note.

Grey's Anatomy (Season 7)

... She and Derek decide to adopt Zola, an orphaned baby, and make their marriage legal.

Who was England's prime minister during WWI?

H. H. Asquith

... served as Prime Minister of the United Kingdom from 1908 to 1916. ... In August 1914, Asquith took Great Britain and the British Empire into the First World War.

David Lloyd George

... served as Prime Minister of the United Kingdom from 1916 to 1922. ... victory during the First World War.

Motivation

What season does Meredith and Derek get married in Grey's Anatomy?

Now or Never (Grey's Anatomy)

... Season 5 ... Meredith and Derek have decided not to wait any longer to get married and just go to City Hall that evening ... writes their vows down on a post-it note.

Grey's Anatomy (S

...dopt Zola, an orphaned baby, ...their marriage legal.

Over 50% of questions from NQ are ambiguous

Who was England's prime minister during WWI?

H. H. Asquith

... served as Prime Minister of the United Kingdom from 1908 to 1916. ... In August 1914, Asquith took Great Britain and the British Empire into the First World War.

David Lloyd George

... served as Prime Minister of the United Kingdom from 1916 to 1922. ... victory during the First World War.

Motivation

What season does Meredith and Derek get married in Grey's Anatomy?

Now or Never (Grey's Anatomy)

... Season 5 ... Meredith and Derek have decided not to wait any longer to get married and just go to City Hall that evening ... writes their vows down on a post-it note.

Open-domain questions are inherently ambiguous

When people ask questions in new / unfamiliar topics,
questions cannot be guaranteed to have a single clear answer

ed baby,

Who was England's prime minister during WWI?

H. H. Asquith

... served as Prime Minister of the United Kingdom from 1908 to 1916. ... In August 1914, Asquith took Great Britain and the British Empire into the First World War.

David Lloyd George

... served as Prime Minister of the United Kingdom from 1916 to 1922. ... victory during the First World War.

Task– options

What season does Meredith and Derek get married in Grey's Anatomy?

Season 5

Season 7

Task– options

What season does Meredith and Derek get married in Grey's Anatomy?

Season 5

Less helpful for users due to lack of context

Task– options

What season does Meredith and Derek get married in Grey's Anatomy?

Season 5

Less helpful for users due to lack of context

What season does Meredith and Derek get married in Grey's Anatomy?

Meredith and Derek got informally married with a post-it note in Season 5 and then they later made it legal when they adopt a baby in Season 7...

Task– options

What season does Meredith and Derek get married in Grey's Anatomy?

Season 5

Less helpful for users due to lack of context

What season does Meredith and Derek get married in Grey's Anatomy?

Meredith and Derek got inform… …ried with a post-it note in Se… …ter made it l… …eason 7…

Less well-defined, Non-trivial evaluation,
Cannot separate answer prediction vs. disambiguation

AmbigQA task

What season does Meredith and Derek get married in Grey's Anatomy?

Q: What season does Meredith and Derek get informally married in Grey's Anatomy?
A: Season 5

Q: What season does Meredith and Derek get legally married in Grey's Anatomy?
A: Season 7

Who was England's prime minister during WWI?

Q: Who was England's prime minister in the beginning of WWI?
A: H. H. Asquith

Q: Who was England's prime minister in the end of WWI?
A: David Lloyd George

AmbigQA task

What season does Meredith and Derek get married in Grey's Anatomy?

Q: What season does Meredith and Derek get informally married in Grey's Anatomy?
A: Season 5

Q: What season does Meredith and Derek get legally married in Grey's Anatomy?
A: Season 7

Who was England's prime minister during WWI?

Q: Who was England's prime minister in the beginning of WWI?
A: H. H. Asquith

Q: Who was England's prime minister in the end of WWI?
A: David Lloyd George

Explicit answers to the original question
+ disambiguation in a more well-defined way

Contribution

- **New task (AmbigQA)** that answers to the questions by identifying all plausible answers along with their disambiguations
- **Dataset** with 14,042 annotations, with frequent, diverse & subtle ambiguity
- **First baseline models**, with experiments showing their effectiveness in learning from our data while highlighting avenues for future work

Content

Introduction

Related Work

Task & Data

Baselines & Experiments

Related work (1/2)

Open-domain question answering

- Questions vary from information-seeking (Berant et al 2013, Kwiatkowski et al 2019, Clark et al 2019) to more specialized trivia/quiz (Joshi et al 2017, Dunn et al 2017)

- Assume each question has a single clear answer

- Nonetheless, the answers in the previous data are often debatable; an average agreement of the answers in Natural Questions annotations is 49.2%

Related work (1/2)

Open-domain question answering

- Questions vary from information-seeking (Berant et al 2013, Kwiatkowski et al 2019, Clark et al 2019) to more specialized trivia/quiz (Joshi et al 2017, Dunn et al 2017)
- Assume each question has a single clear answer
- Nonetheless, the answers in the previous data are often debatable; an average agreement of the answers in Natural Questions annotations is 49.2%

We **embrace ambiguity** as **inherent** to information-seeking questions

Related work (2/2)

Asking clarification questions

- Questions that are annotated by crowdworkers (Xu et al 2019)
- Simple, vague keywords, e.g. "*dinasour*" (Zhai et al 2003, Aliannejadi et al 2019)

Related work (2/2)

Asking clarification questions

- Questions that are annotated by crowdworkers (Xu et al 2019)
- Simple, vague keywords, e.g. *"dinasour"* (Zhai et al 2003, Aliannejadi et al 2019)

We study **unintentional** & **subtle** ambiguity in natural questions

We provide **complete** & **immediate** solution

Content

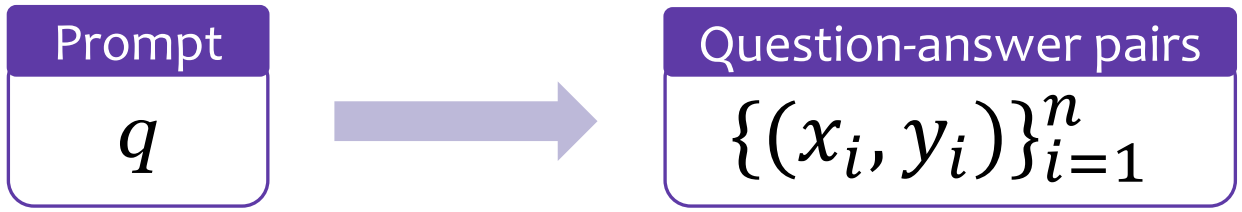
Introduction

Related Work

Task & Data

Baselines & Experiments

Task Definition

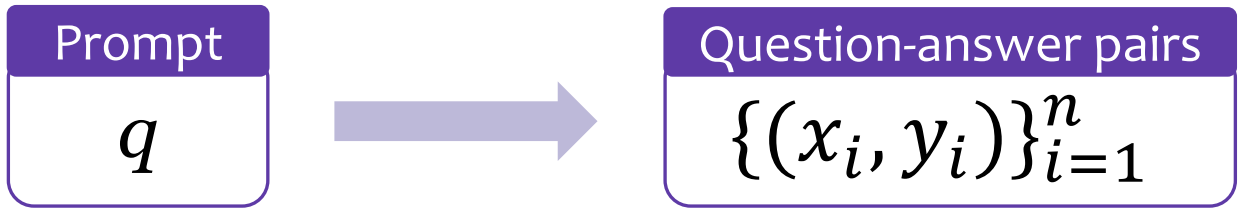


What season does … get married in Grey's Anatomy?

x_1: What season … get informally married in Grey's Anatomy?
y_1: Season 5

x_2: What season … get legally married in Grey's Anatomy?
y_2: Season 7

Task Definition



What season does … get married in Grey's Anatomy?

x_1: What season … get informally married in Grey's Anatomy?

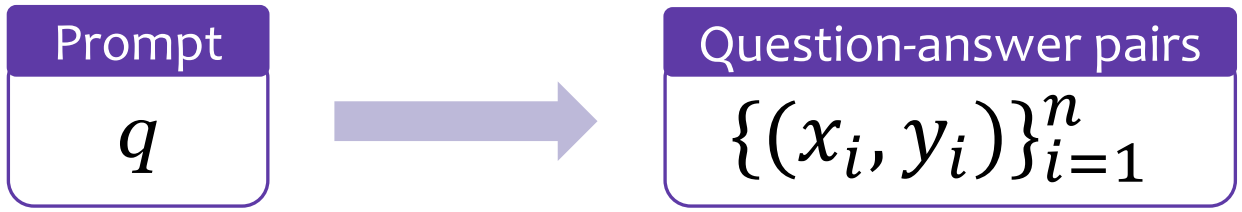
y_1: Season 5

x_2: What season … get legally married in Grey's Anatomy?

y_2: Season 7

y_i: an equally plausible answer

Task Definition



What season does … get married in Grey's Anatomy?

x_1: What season … get informally married in Grey's Anatomy?

y_1: Season 5

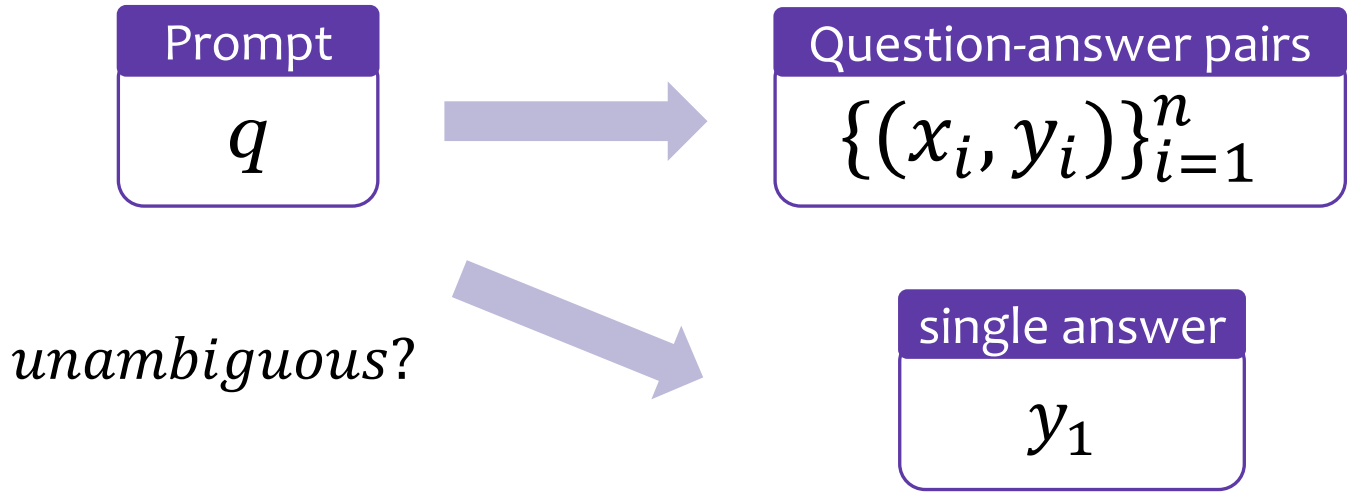
x_2: What season … get legally married in Grey's Anatomy?

y_2: Season 7

y_i: an equally plausible answer

x_i: a *minimal* modification of q whose answer is unambiguously y_i

Task Definition



Evaluation metrics

- Multiple Answer Prediction (F_{answer})
- Full task (F_{BLEU}, $F_{\text{EDIT-F1}}$)
  - Consider a similarity score between reference and generated question

Please see the paper for the exact formulas

Evaluation metrics

- Multiple Answer Prediction (F_{answer})
- Full task (F_{BLEU}, $F_{\text{EDIT-F1}}$)
  - Consider a similarity score between reference and generated question

BLEU **EDIT-F1** NEW

Evaluation metrics

- Multiple Answer Prediction (F_{answer})
- Full task (F_{BLEU}, $F_{\text{EDIT-F1}}$)
  - Consider a similarity score between reference and generated question

BLEU **EDIT-F1** NEW

Prompt	Who made the play the crucible?
Reference	Q: Who wrote the play the crucible? (-made, +wrote)
Prediction	Q: Who made the play the crucible in 2012? (+in, +2012)

Edit-F1 = 0.0

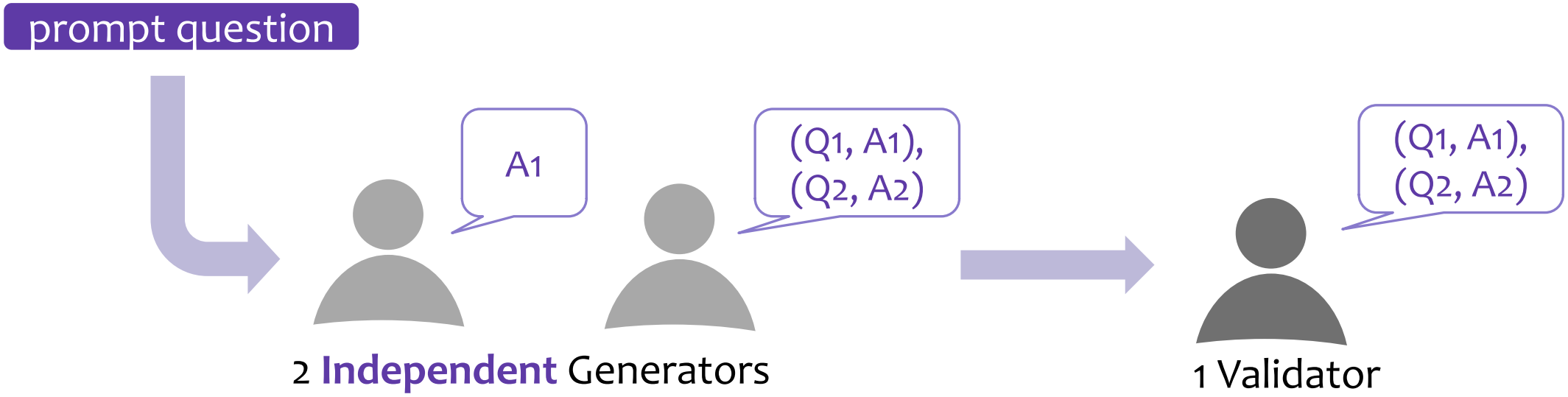
Data Collection

- Maximizing recall is difficult even for humans
- We were able to collect high quality data with high levels of ambiguity using **careful worker selection** and **an annotation pipeline: generation + validation**

Data Collection

- Maximizing recall is difficult even for humans
- We were able to collect high quality data with high levels of ambiguity using **careful worker selection** and **an annotation pipeline: generation + validation**

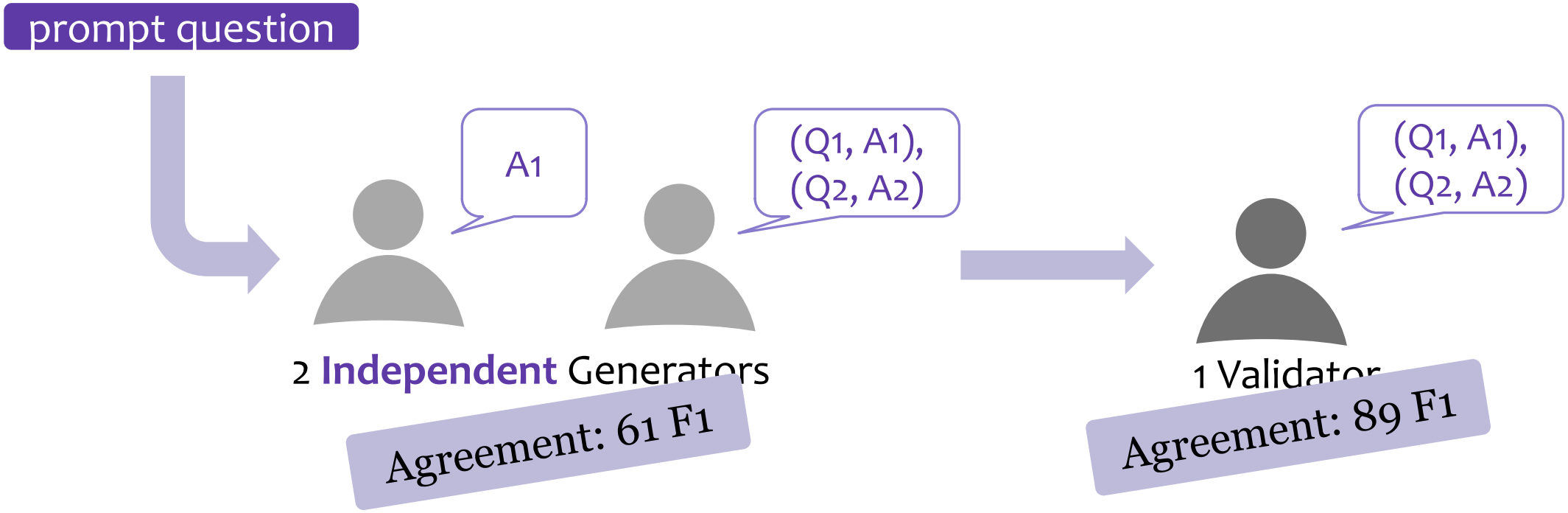
Information-seeking questions from NQ (Kwiatkowski et al 2019)



Data Collection

- Maximizing recall is difficult even for humans
- We were able to collect high quality data with high levels of ambiguity using **careful worker selection** and **an annotation pipeline: generation + validation**

Information-seeking questions from NQ (Kwiatkowski et al 2019)



Data Analysis

- 14,042 questions
- Over 50% of questions are ambiguous
- Diverse types of ambiguity

Type of Ambiguity	%
Event references	39
Properties	27
Entity references	23
Answer types	16
Time-dependency	13
Multiple sub-questions	3

Data Analysis

- 14,042 questions
- Over 50% of questions are ambiguous
- Diverse types of ambiguity

Type of Ambiguity	%
	27
Entity references	23
Answer types	16
Time-dependency	13
Multiple sub-questions	3

More analyses in the paper

Content

Introduction

Related Work

Task & Data

Baselines & Experiments

Please see the paper for details!

Baselines

Step 1: Multi-answer prediction

Step 2: Question disambiguation

Baselines

Step 1: Multi-answer prediction

DPR (Karpukhin et al 2020)

Thresholding over likelihood

SpanSeqGen

Generate a sequence of answers, separated by [SEP]

Step 2: Question disambiguation

Baselines

Step 1: Multi-answer prediction

DPR (Karpukhin et al 2020)

Thresholding over likelihood

SpanSeqGen

Generate a sequence of answers, separated by [SEP]

Step 2: Question disambiguation

Prompt question
Targeted answer
Untargeted answers
Passages

BART
(Lewis et al 2020)

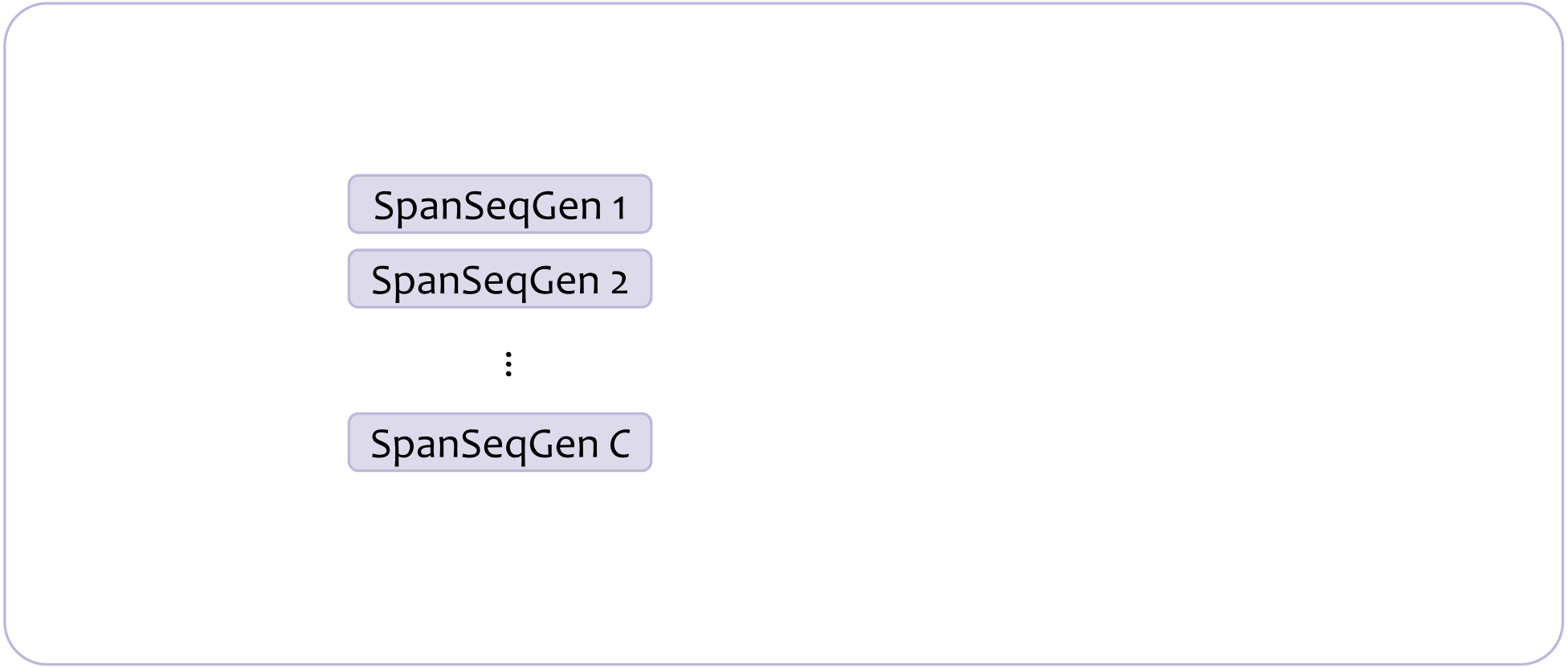
Edited question

Modified Democratic Co-training (Zhou & Goldman, 2004)

- Prevalence of unlabeled ambiguity in NQ-open → Let's use ***a single known answer*** from NQ-open as ***weak supervision***

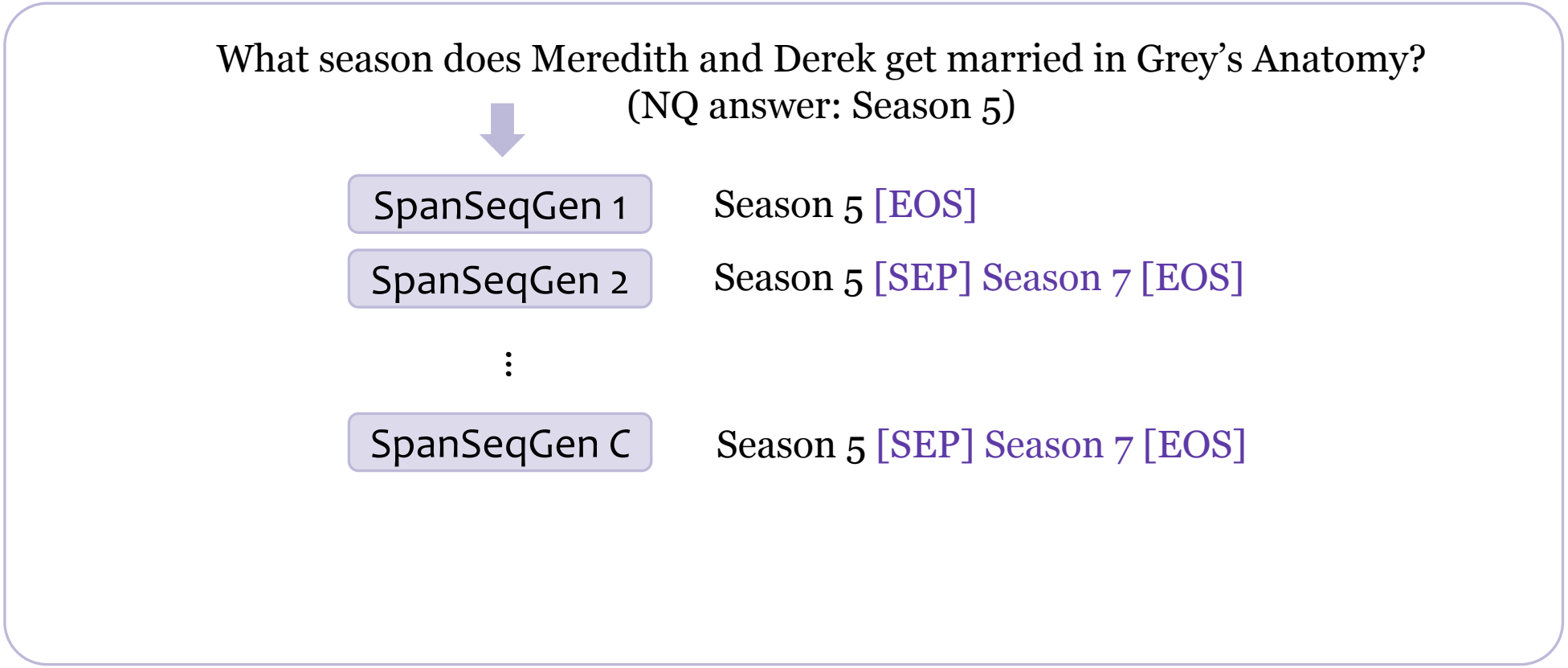
Modified Democratic Co-training (Zhou & Goldman, 2004)

- Prevalence of unlabeled ambiguity in NQ-open → Let's use ***a single known answer*** from NQ-open as ***weak supervision***



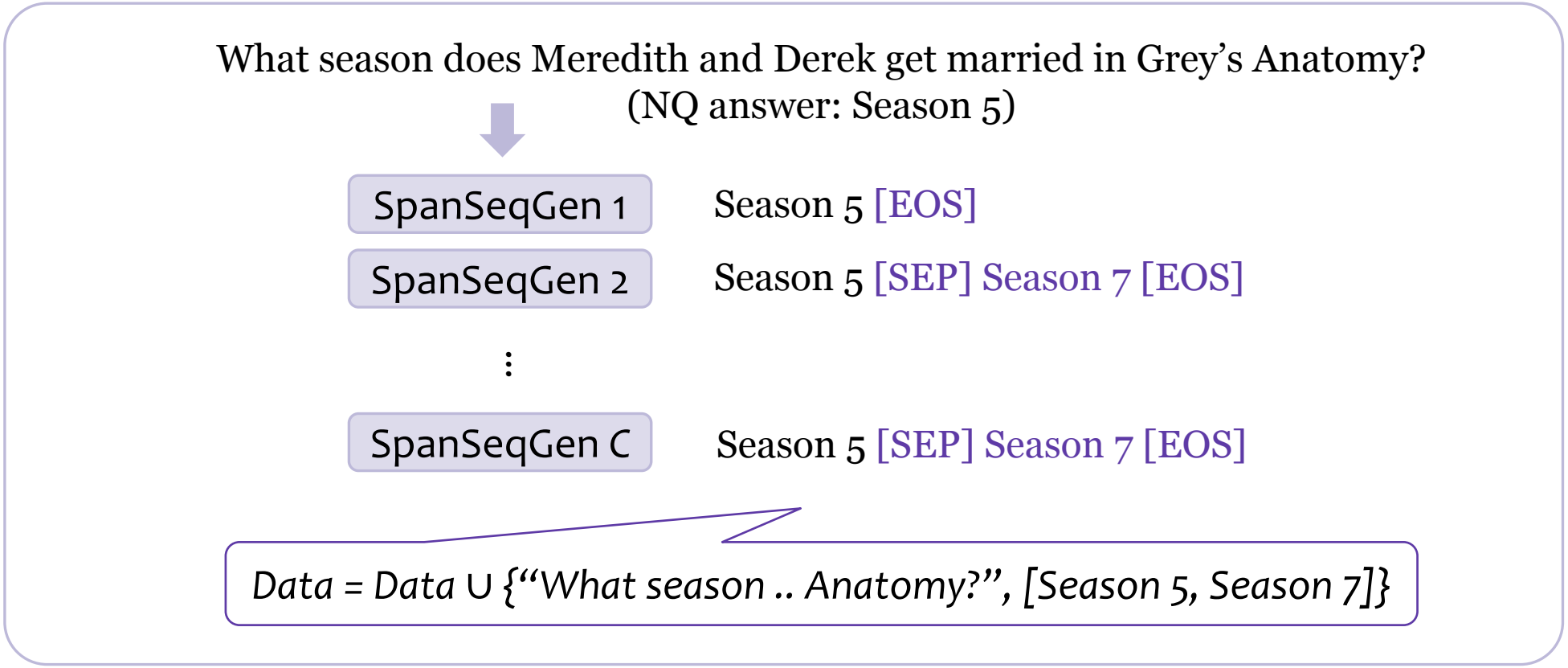
Modified Democratic Co-training (Zhou & Goldman, 2004)

- Prevalence of unlabeled ambiguity in NQ-open → Let's use ***a single known answer*** from NQ-open as ***weak supervision***



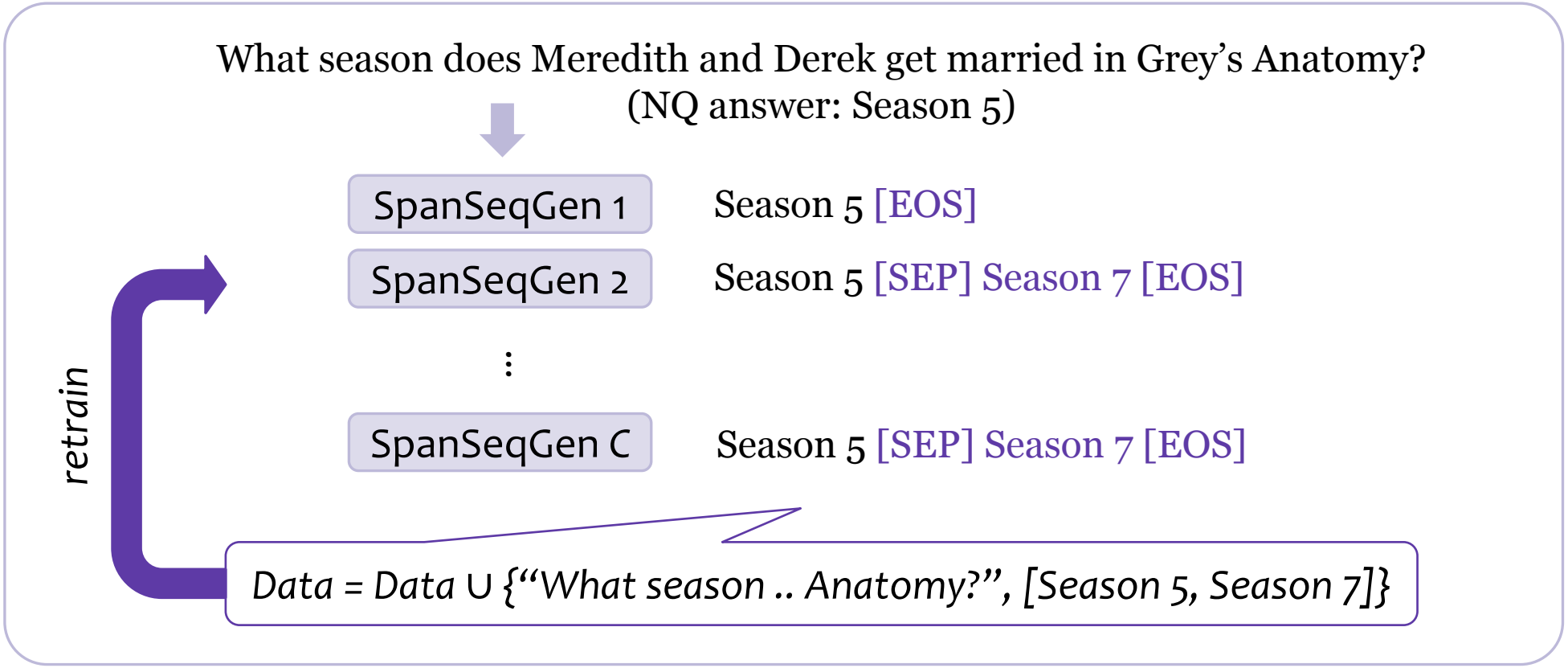
Modified Democratic Co-training (Zhou & Goldman, 2004)

- Prevalence of unlabeled ambiguity in NQ-open → Let's use ***a single known answer*** from NQ-open as ***weak supervision***

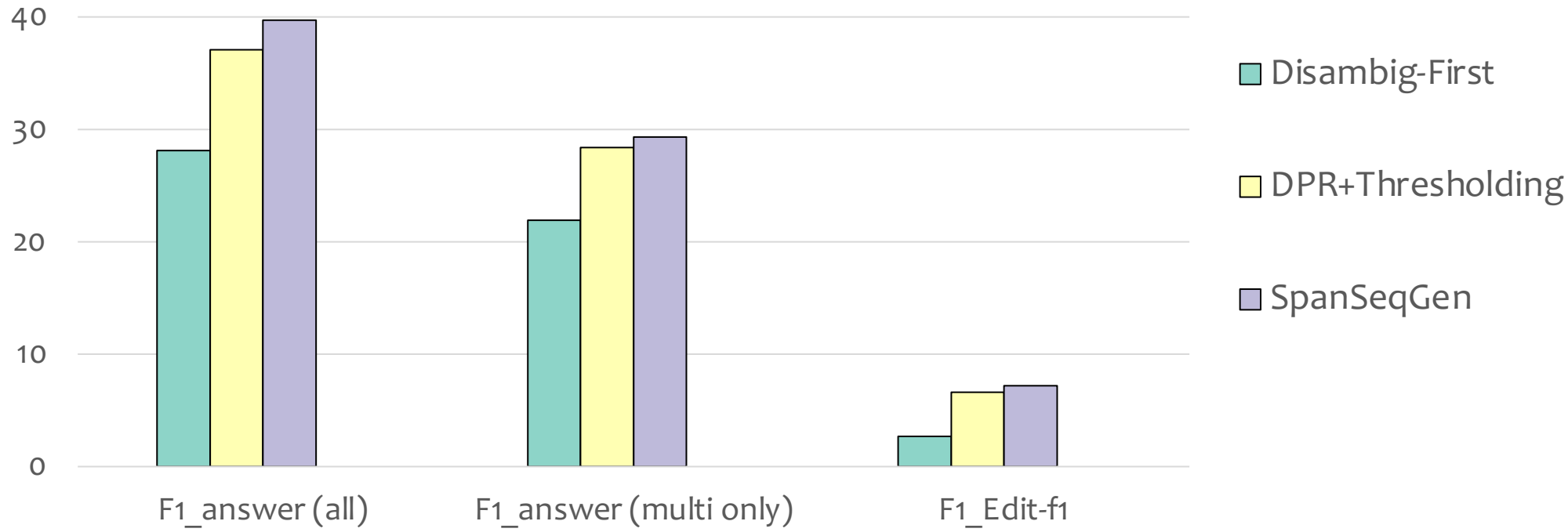


Modified Democratic Co-training (Zhou & Goldman, 2004)

- Prevalence of unlabeled ambiguity in NQ-open → Let's use ***a single known answer*** from NQ-open as ***weak supervision***



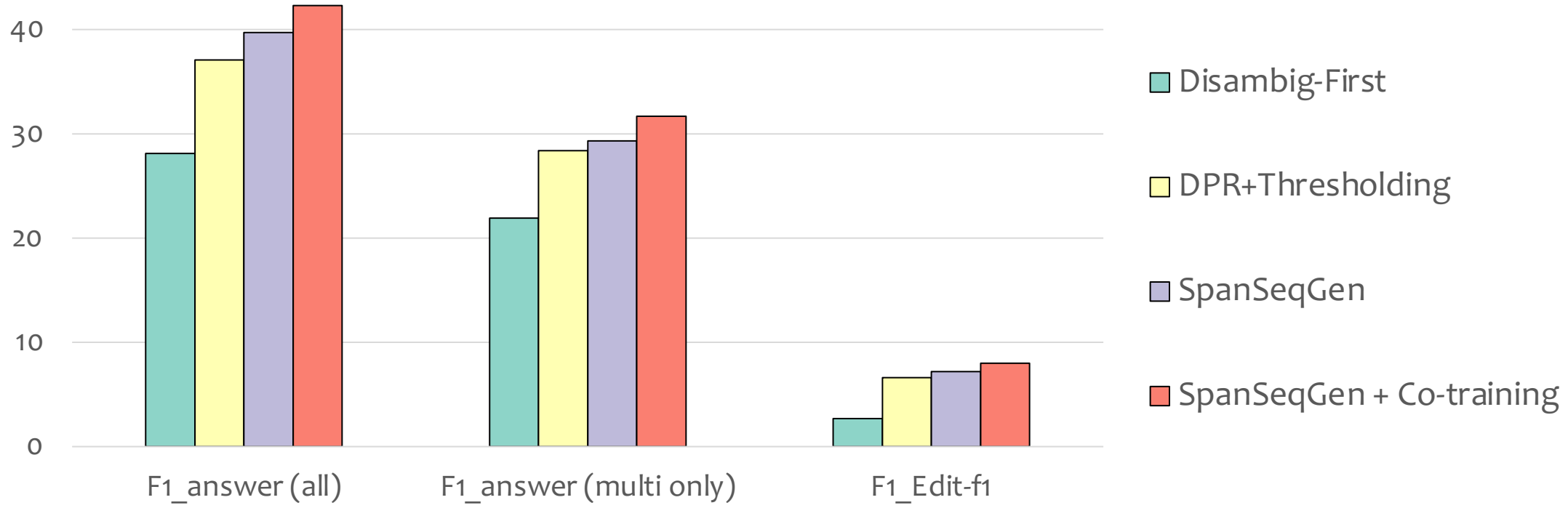
Results



SpanSeqGen outperforms other baselines

(Esp. Disambig-First, which does disambiguation before reading passages)

Results

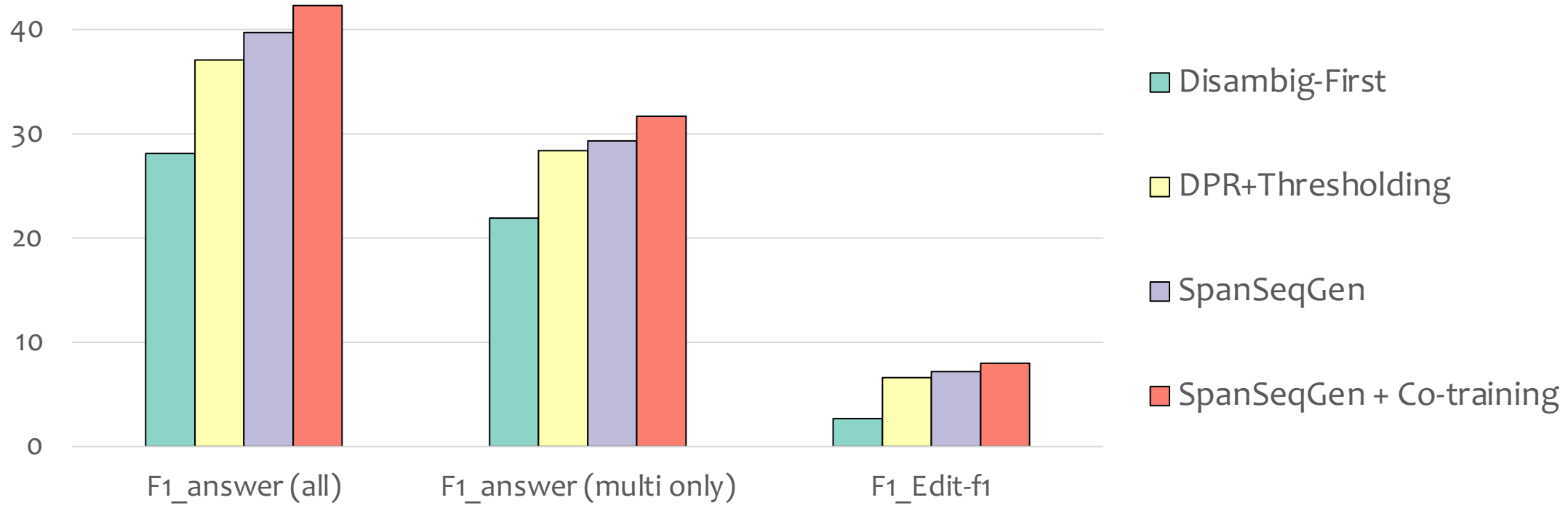


SpanSeqGen outperforms other baselines

(Esp. Disambig-First, which does disambiguation before reading passages)

SpanSeqGen + Co-training further boosts the performance

Results



SpanSeqGen outperforms other baselines

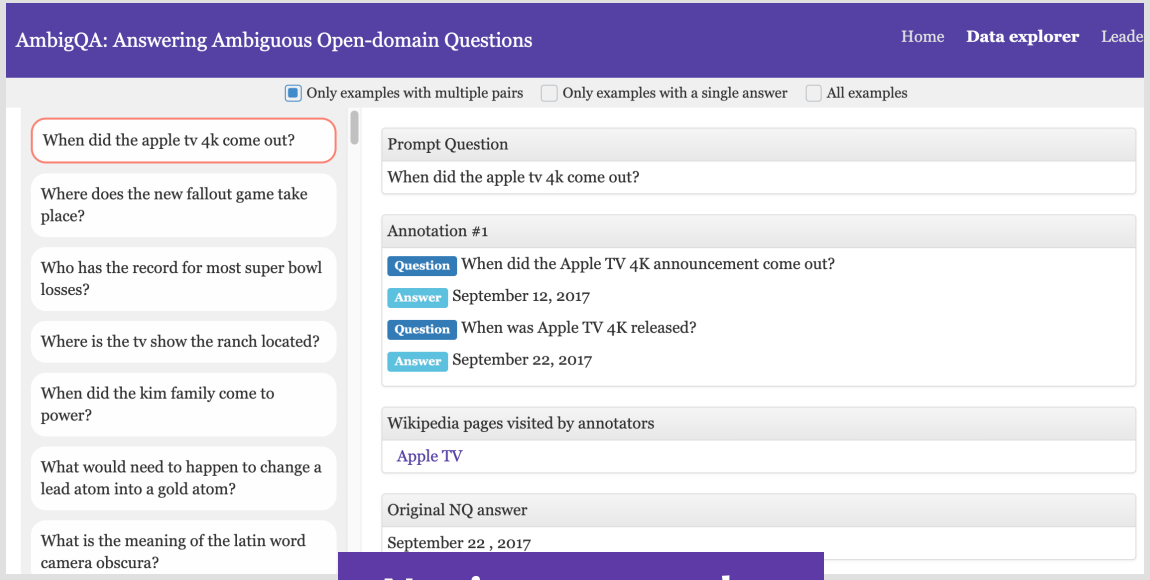
(Esp. Disambig-First, which does disambiguation before reading passages)

SpanSeqGen + Co-training further boosts the performance

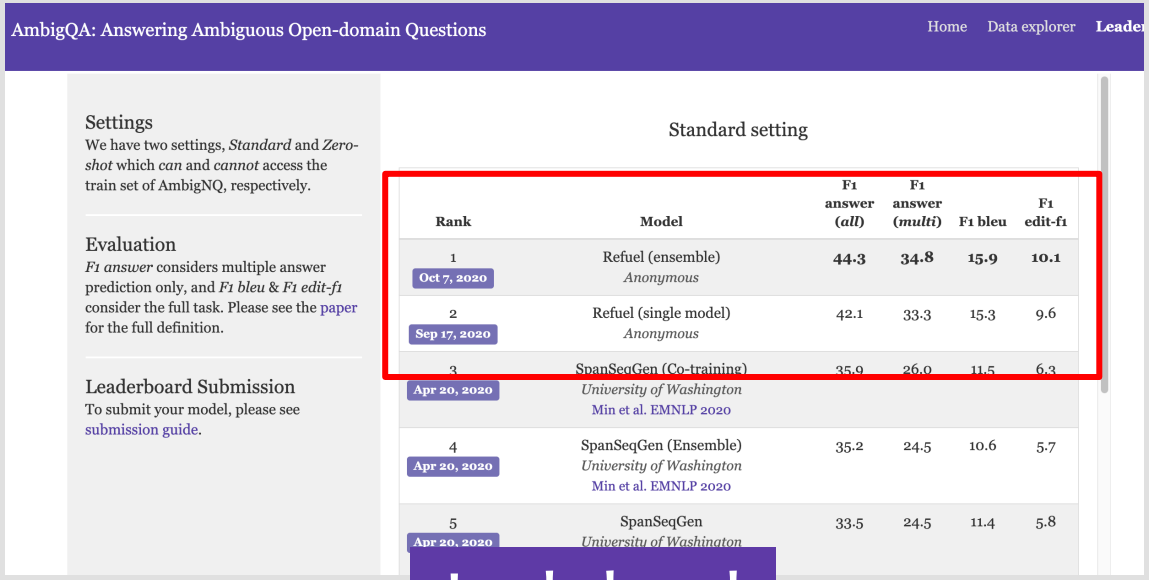
Still huge room for improvements

Thank you!

https://nlp.cs.washington.edu/ambigqa/



Navigate samples



Leaderboard

Already progress from the community!

Join us in the QnA session!