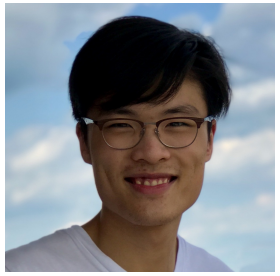
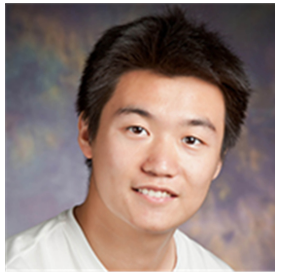
"Going on a vacation" takes longer than *"Going for a walk"*: A Study of Temporal Commonsense Understanding



Ben Zhou

Daniel Khashabi*



Qiang Ning*

Dan Roth

*Currently affiliated with AI2

Temporal Common Sense

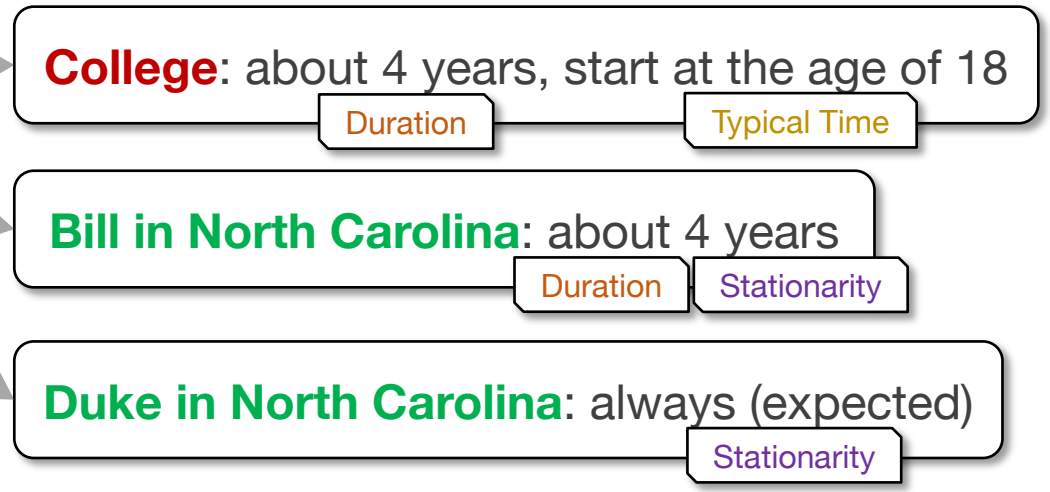
- Humans assume information when reading
  - Not explicitly mentioned
  - Related to time
- Happens all the time
  - To better understand the storyline and beyond

Temporal Common Sense

My friend Bill went to Duke University in North Carolina. With a degree in CS, he joined Google MTV as a software engineer. As a huge basketball fan, he has attended all 3 NBA finals since then. He also plans to visit Duke regularly as an alumnus to attend their home games.

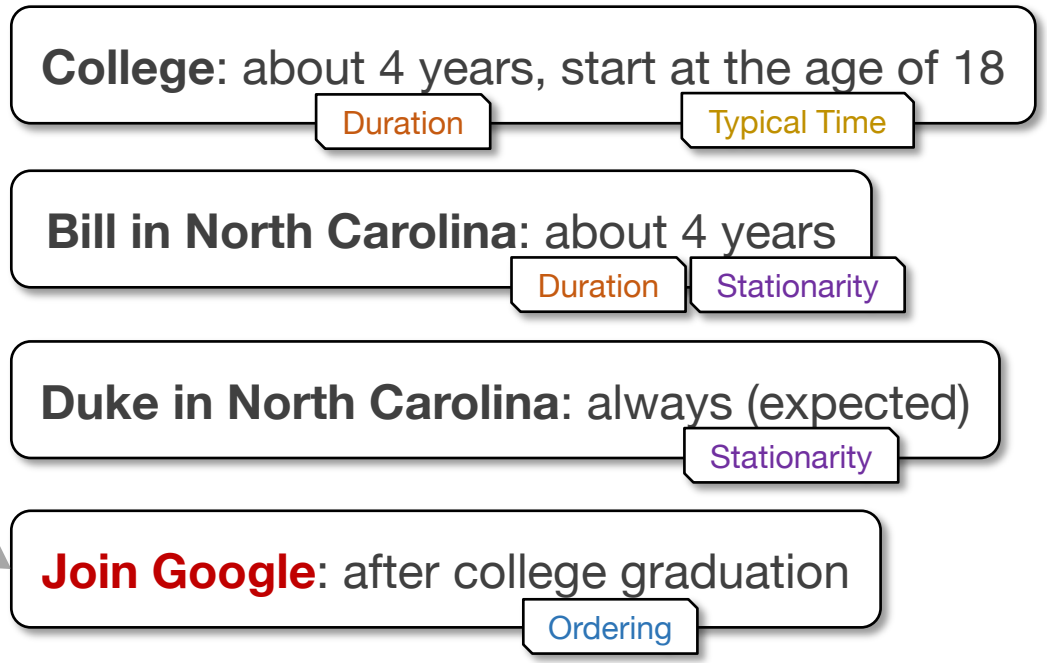
Temporal Common Sense

My friend Bill went to Duke University in North Carolina. With a degree in CS, he joined Google MTV as a software engineer. As a huge basketball fan, he has attended all 3 NBA finals since then. He also plans to visit Duke regularly as an alumnus to attend their home games.



Temporal Common Sense

My friend Bill went to Duke University in North Carolina. With a degree in CS, he joined Google MTV as a software engineer. As a huge basketball fan, he has attended all 3 NBA finals since then. He also plans to visit Duke regularly as an alumnus to attend their home games.



Temporal Common Sense

My friend Bill went to Duke University in North Carolina. With a degree in CS, he joined Google MTV as a software engineer. As a huge basketball fan, he has attended all 3 NBA finals since then. He also plans to visit Duke regularly as an alumnus to attend their home games.

Temporal Common Sense

My friend Bill went to Duke University in North Carolina. With a degree in CS, he joined Google MTV as a software engineer. As a huge basketball fan, he has attended all 3 NBA finals since then. He also plans to visit Duke regularly as an alumnus to attend their home games.

College: about 4 years, start at the age of 18
- Duration
- Typical Time

Bill in North Carolina: about 4 years
- Duration
- Stationarity

Duke in North Carolina: always (expected)
- Stationarity

Join Google: after college graduation
- Ordering

NBA Finals: every year
- Frequency

Visit Alma Mater: 0-2 times per year, 0-2 days each time
- Frequency
- Duration

Attend basketball games: a few hours
- Duration

Temporal Commonsense

My friend Bill went to Duke University in North Carolina. With a degree in CS, he joined Google MTV as a software engineer. As a huge basketball fan, he has attended all 3 NBA finals since then. He also plans to visit Duke regularly as an alumnus to attend their home games.

** Human infer temporal common sense that helps them to better understand the story.*

- **Q: How old is Bill?**
- A: Around 25.
- R: 3 + 4 + 18

- **Q: How long will take Bill to fly to Duke?**
- A: A few (1-5) hours.
- R: Duke is always in NC, Bill is now in CA

- **Q: How often would he visit Duke in the future?**
- A: A few (<5) times a year.

- Q: **Which one happened first, went or joined?**
- A: Went.

Our Contribution

- MC-TACO 🌮 (**m**ultiple **c**hoice **t**empor**a**l **co**mmon-sense) **:**
  - A dataset that focuses on temporal commonsense
  - Input:

			Gold	Prediction	
He went to Duke University.	How long did it take him to graduate?	4 years			✓
He went to Duke University.	How long did it take him to graduate?	10 days			✓
		3.5 years			X
		16 hours			✓
					✓

  - Task: Decide whether each answer is plausible.
  - Metrics:
    - Exact Ma
    - F1: The F
  - Statistics:
    - 1,893 questions
    - 13,225 question-answer pairs
  - Conclusion: current systems are not enough to solve this.

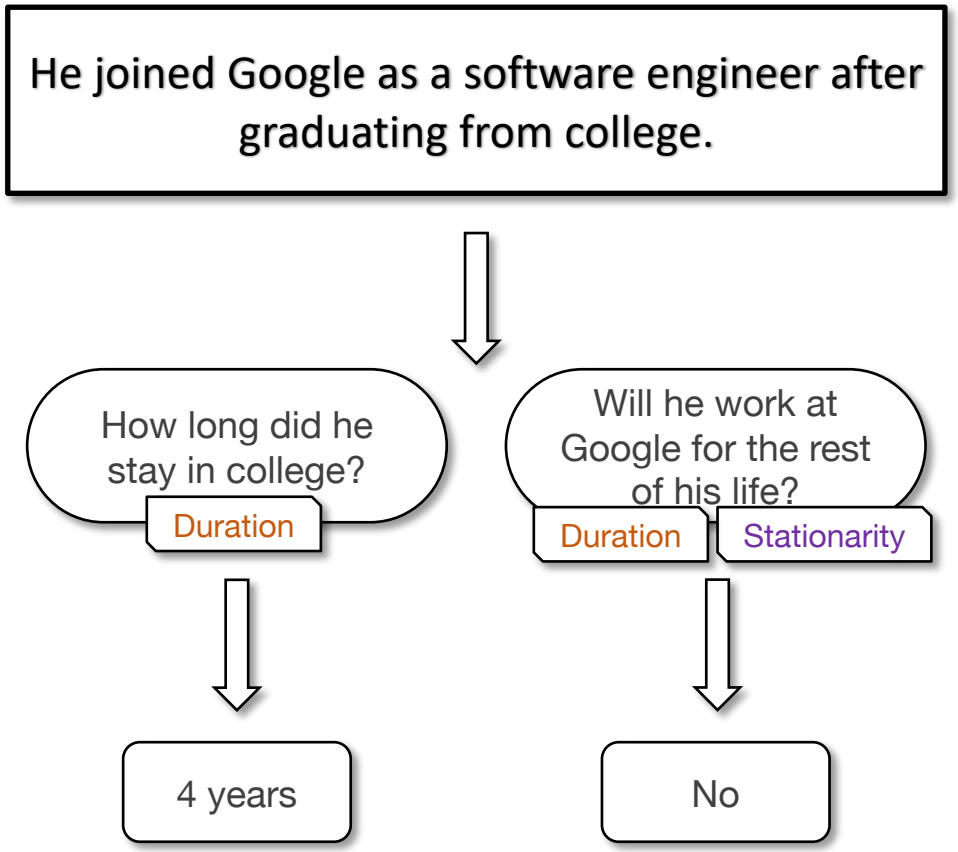
Reading Comprehension: able to answer any questions regarding a piece of text

Exact Match: able to label all candidate answers of a question

.7
Exact Match: 0.0

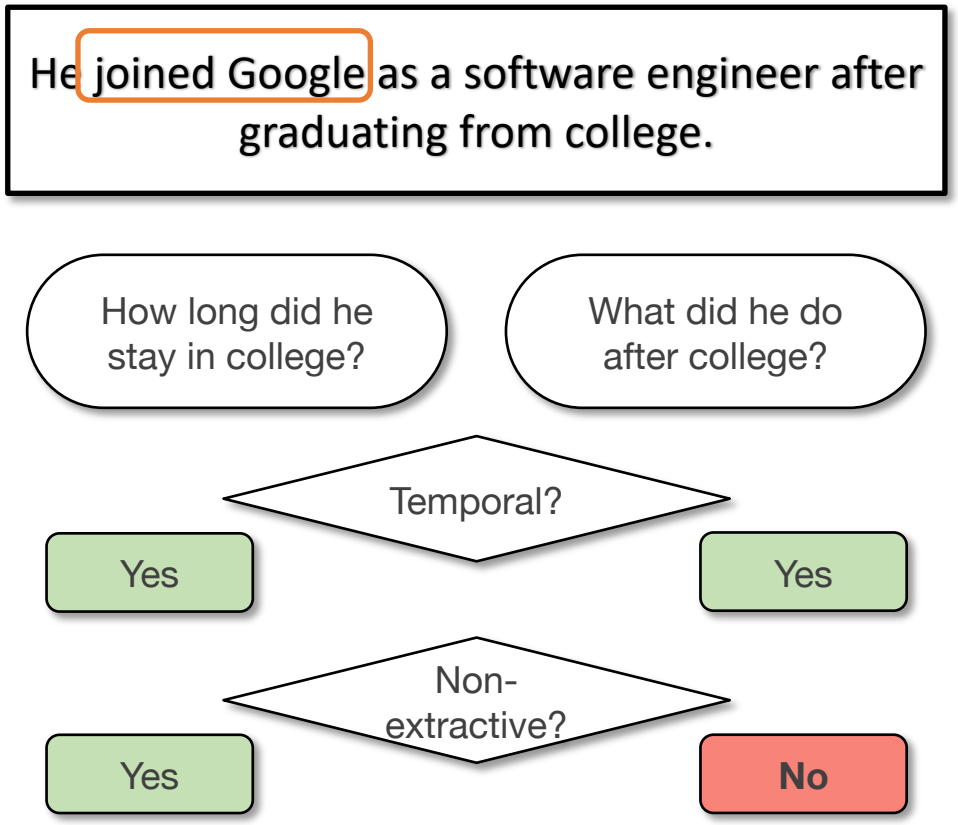
MC-TACO: Construction

- Step 0: Source Sentence Generation
  - Randomly samples sentences
- Step 1: Question Generation
  - Ask people to write questions
    - A) **temporal**
    - B) **non-extractive**
      - To require commonsense
  - Ask for one “plausible” answer



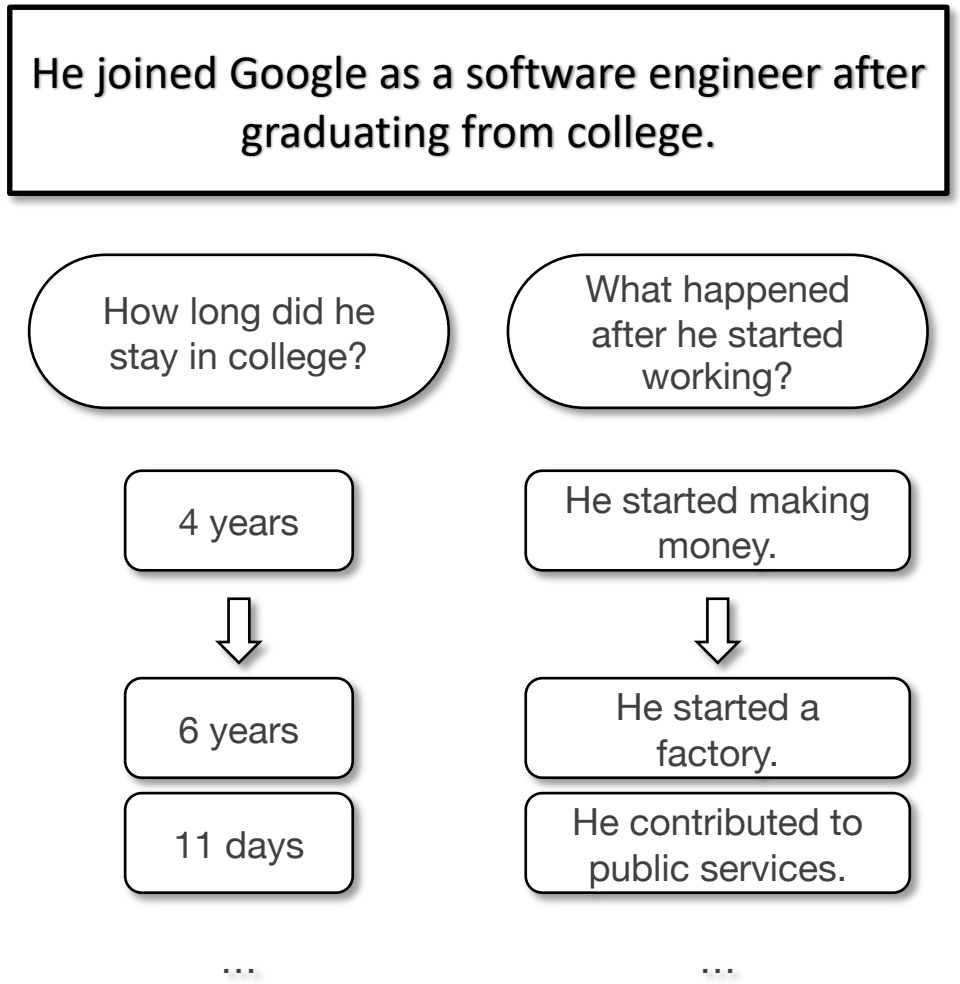
MC-TACO: Construction

- Step 2: Question Verification
  - 2 additional verifications on each question
  - 100% agreement
  - We also ask for
    - 1 “plausible” answer
    - 1 “implausible” answer



MC-TACO: Construction

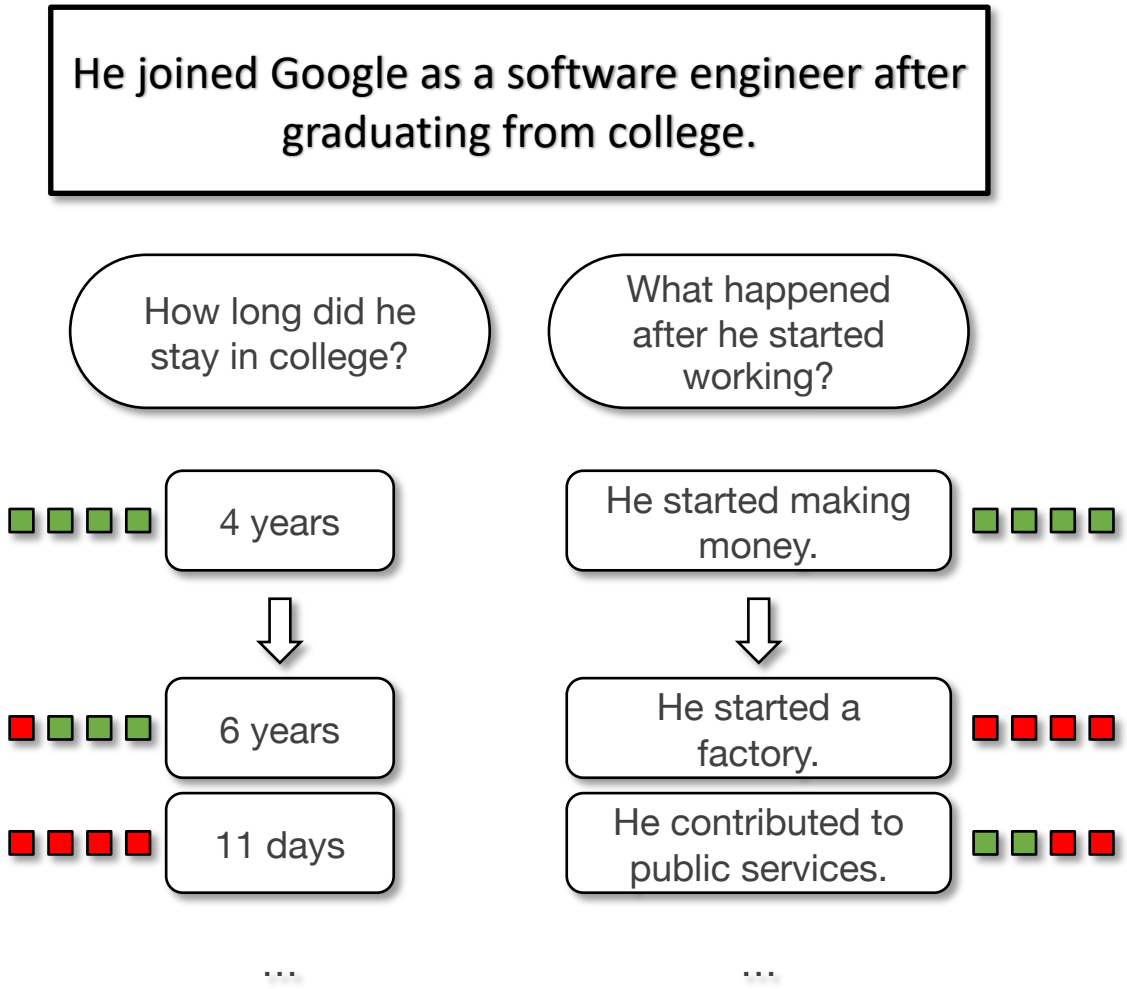
- Step 3: Candidate Answer Expansion
  - ☐ Seed answers from step 1+2
  - ☐ Expand candidates automatically
    - Perturbations
    - Information Retrieval



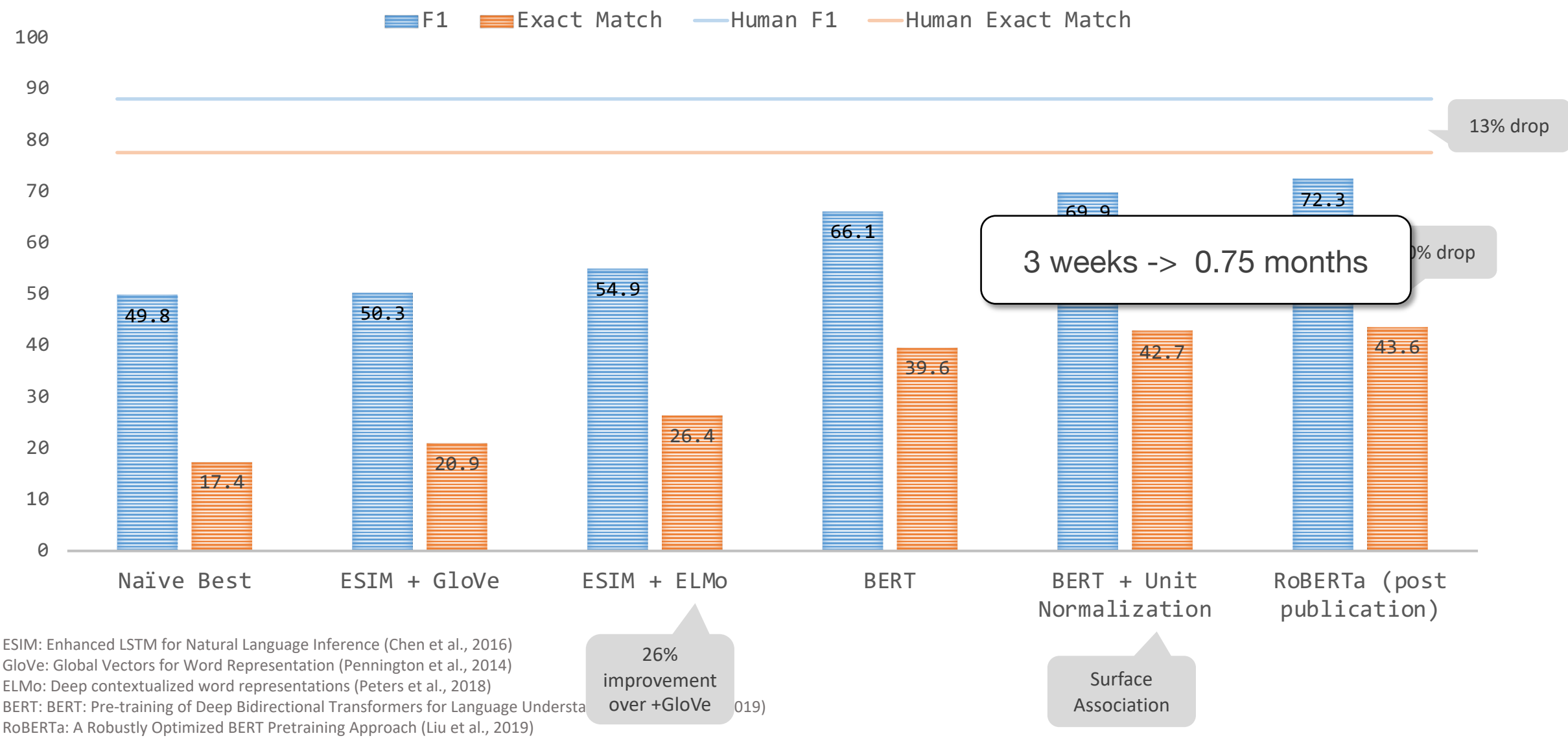
MC-TACO: Construction

- Step 4: Answer Labeling
  - Each answer is labeled by **4** different annotators
  - Either "likely" ■ or "unlikely" ■
  - Enforce 100% agreement

    ■■■■ or ■■■■
    - Eliminate marginal answers with "intermediate" probability



Results



ESIM: Enhanced LSTM for Natural Language Inference (Chen et al., 2016)
GloVe: Global Vectors for Word Representation (Pennington et al., 2014)
ELMo: Deep contextualized word representations (Peters et al., 2018)
BERT: BERT: Pre-training of Deep Bidirectional Transformers for Language Understa 019)
RoBERTa: A Robustly Optimized BERT Pretraining Approach (Liu et al., 2019)

Summary

- Define 5 temporal commonsense phenomena
- Present MC-TACO, a QA dataset focused on temporal commonsense
- Show that existing systems are not enough to solve it
- Encourage further research
- Thanks!

Leaderboard
https://leaderboard.allenai.org/mctaco/

GitHub (data, baseline, evaluator)
https://github.com/CogComp/MCTACO