



TORQUE: A Reading Comprehension Dataset of Temporal ORdering QUEstions

Qiang Ning, Hao Wu, Rujun Han
Nanyun Peng, Matt Gardner, Dan Roth



Time is important for story understanding



People were angry



Police used tear gas



People **were angry** first, and then the police **used tear gas**.

Same events, different temporal ordering



Police used tear gas



People were angry



Police **used tear gas** first, and then people **were angry**.

Same events, different temporal ordering



Police used tear gas

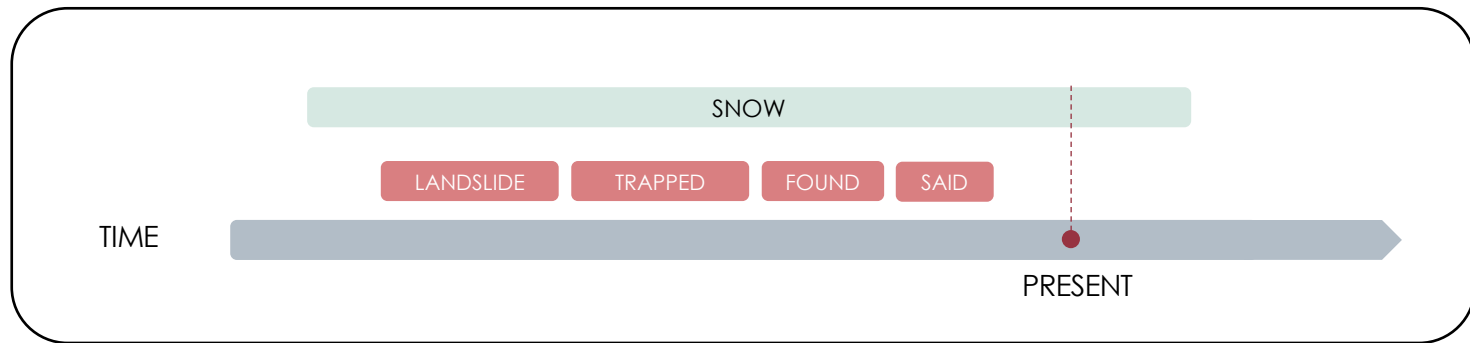


People were angry



This paper studies **temporal relationship** as a **machine reading comprehension** (MRC) problem.

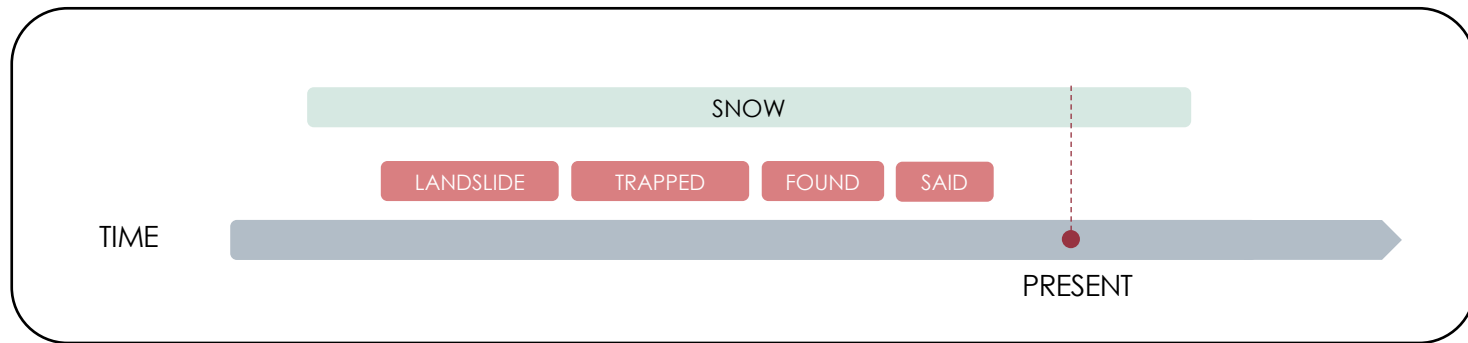
Text: Heavy **snow** is causing disruption to transport across the UK, with heavy rainfall bringing flooding to the south-west of England. Rescuers searching for a woman **trapped** in a **landslide** at her home **said** they had **found** a body.



Existing systems don't really use "time" to answer these questions.

Question	ELMo-BiDAF (SQuAD)	BERT (SQuAD)	NAQANet (DROP)	BERT (SQuAD 2.0)
What happened <u>before</u> a woman was trapped?	they had found a body	a landslide	in a landslide	a landslide
What happened <u>after</u> a woman was trapped?	they had found a body	they had found a body	in a landslide	a landslide
What happened <u>while</u> a woman was trapped?	they had found a body	a landslide	in a landslide	a landslide
What happened <u>before</u> the snow started?	disruption to ... UK	landslide	woman trapped in a landslide	heavy rainfall ... landslide
What happened <u>after</u> the snow started?	disruption to ... UK	flooding to ... England	woman trapped in a landslide	heavy rainfall ... England
What happened <u>during</u> the snow?	disruption to ... UK	a landslide	woman trapped	landslide
What happened <u>before</u> the rescuers found a body?	Rescuers searching ... landslide	a landslide	woman trapped	a landslide
What happened <u>after</u> the rescuers found a body?	Rescuers searching ... landslide	Rescuers searching ... Cornwall	woman trapped in a landslide	landslide
What happened <u>during</u> the rescue?	Rescuers searching ... landslide	a landslide	woman trapped in a landslide	they had found a body

Text: Heavy **snow** is causing disruption to transport across the UK, with heavy rainfall bringing flooding to the south-west of England. Rescuers searching for a woman **trapped** in a **landslide** at her home **said** they had **found** a body.



Existing systems don't really use "time" to answer these questions.

Question	ELMo-BiDAF (SQuAD)	BERT (SQuAD)	NAQANet (DROP)	BERT (SQuAD 2.0)
What happened <u>before</u> a woman was trapped?	they had found a body	a landslide	in a landslide	a landslide
What happened <u>after</u> a woman was trapped?	they had found a body	they had found a body	in a landslide	a landslide
What happened <u>while</u> a woman was trapped?	they had found a body	a landslide	in a landslide	a landslide
What happened <u>before</u> the snow started?	disruption to ... UK	landslide	woman trapped in a landslide	heavy rainfall ... landslide
What happened <u>after</u> the snow started?	disruption to ... UK	flooding to ... England	woman trapped in a landslide	heavy rainfall ... England
What happened <u>during</u> the snow?	disruption to ... UK	a landslide	woman trapped	landslide
What happened <u>before</u> the rescuers found a body?	Rescuers searching ... landslide	a landslide	woman trapped	a landslide
What happened <u>after</u> the rescuers found a body?	Rescuers searching ... landslide	Rescuers searching ... Cornwall	woman trapped in a landslide	landslide
What happened <u>during</u> the rescue?	Rescuers searching ... landslide	a landslide	woman trapped in a landslide	they had found a body

We aim to fix this problem using "TORQUE"

Challenges

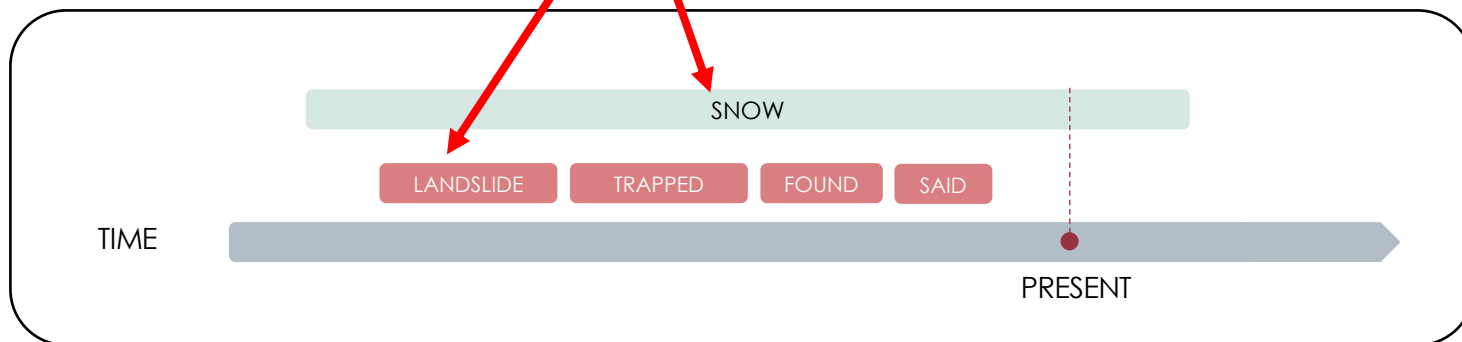
1. Requires event understanding

- ❑ *Text: “Rescuers searching for a woman trapped in a landslide at her home said they had found a body.”*
- ❑ **Q1: What was a woman trapped in?** A typical SQuAD question that only requires predicate-argument understanding.
- ❑ **Q2: What started before a woman was trapped?** This requires event understanding: e.g., “landslide” is an event while “body” is not.

Challenges

2. There are many events in a single passage, and they form a complex temporal structure, such that

- ❑ One question can have multiple events as answers
 - ❑ Or, it can also have no answers at all
- What happened after the landslide? – multiple answers
 - What happened before the snow started? – no answer



Challenges

3. Temporal relation questions are often sensitive to a few key words such as “before,” “after,” and “start.”

- ❑ What happened **before** the rescuers said something?
- ❑ What happened **after** the rescuers said something?
- ❑ What event **started** before the rescuers said something?
- ❑ ...

Models that don't really understand time will perform poorly on these questions

TORQUE: Data collection process

Given a passage, an annotator needs to

1. Label all the events For challenge1
2. Repeat the following steps
 - ▣ Ask a question querying temporal relations For challenge2
 - ▣ Select all the correct answers from the event list
 - ▣ Slightly perturb the question and answer it For challenge3

TORQUE

Heavy snow is causing disruption to transport across the UK, with heavy rainfall bringing flooding to the south-west of England. Rescuers searching for a woman trapped in a landslide at her home said they had found a body.

The loan may be extended by the McAlpine group for an additional year with an increase in the conversion price to \$2.50 a share. The sale of shares to the McAlpine family along with the recent sale of 750,000 shares of Meridian stock to Haden MacLellan Holding PLC of Surrey, England and a recent public offering have increased Meridian's net worth to \$8.5 million, said William Feniger, chief executive officer of Toledo, Ohio-based Meridian.

This time, however, some analysts think he could face a real battle. "Without some unexpected coup de theatre, I don't see what will block the Paribas bid," said Philippe de Cholet, analyst at the brokerage Cholet-Dupont amp Cie. Mr. de Cholet said Mr. Fournier's biggest hope was to somehow persuade regulatory authorities to block the bid.



3.2k passages from the 2.8k
articles in TempEval3

Heavy snow is causing disruption to transport across the UK, with heavy rainfall bringing flooding to the south-west of England. Rescuers searching for a woman trapped in a landslide at her home said they had found a body.

The loan may be extended by the McAlpine group for an additional year with an increase in the conversion price to \$2.50 a share. The sale of shares to the McAlpine family along with the recent sale of 750,000 shares of Meridian stock to Haden MacLellan Holding PLC of Surrey, England and a recent public offering have increased Meridian's net worth to \$8.5 million, said William Feniger, chief executive officer of Toledo, Ohio-based Meridian.

This time, however, some analysts think he could face a real battle. "Without some unexpected coup de theatre, I don't see what will block the Paribas bid," said Philippe de Cholet, analyst at the brokerage Cholet-Dupont amp Cie. Mr. de Cholet said Mr. Fournier's biggest hope was to somehow persuade regulatory authorities to block the bid.

3.2k passages from the 2.8k articles in TempEval3

- Two sentences per passage
- 50 tokens per passage

Heavy snow is causing disruption to transport across the UK, with heavy rainfall bringing flooding to the south-west of England. Rescuers searching for a woman trapped in a landslide at her home said they had found a body.

The loan may be extended by the McAlpine group for an additional year with an increase in the conversion price to \$2.50 a share. The sale of shares to the McAlpine family along with the recent sale of 750,000 shares of Meridian stock to Haden MacLellan Holding PLC of Surrey, England and a recent public offering have increased Meridian's net worth to \$8.5 million, said William Feniger, chief executive officer of Toledo, Ohio-based Meridian.

This time, however, some analysts think he could face a real battle. "Without some unexpected coup de theatre, I don't see what will block the Paribas bid," said Philippe de Cholet, analyst at the brokerage Cholet-Dupont amp Cie. Mr. de Cholet said Mr. Fournier's biggest hope was to somehow persuade regulatory authorities to block the bid.

3.2k passages from the 2.8k articles in TempEval3

- Two sentences per passage
- 50 tokens per passage

24.9k events

Heavy snow is causing disruption to transport across the UK, with heavy rainfall bringing flooding to the south-west of England. Rescuers searching for a woman trapped in a landslide at her home said they had found a body.

The loan may be extended by the McAlpine group for an additional year with an increase in the conversion price to \$2.50 a share. The sale of shares to the McAlpine family along with the recent sale of 750,000 shares of Meridian stock to Haden MacLellan Holding PLC of Surrey, England and a recent public offering have increased Meridian's net worth to \$8.5 million, said William Feniger, chief executive officer of Toledo, Ohio-based Meridian.

This time, however, some analysts think he could face a real battle. "Without some unexpected coup de theatre, I don't see what will block the Paribas bid," said Philippe de Cholet, analyst at the brokerage Cholet-Dupont amp Cie. Mr. de Cholet said Mr. Fournier's biggest hope was to somehow persuade regulatory authorities to block the bid.



TORQUE

3.2k passages from the 2.8k articles in TempEval3

- Two sentences per passage
- 50 tokens per passage

24.9k events

For each passage, we have

- 3 hard-coded questions
- on average, 7 user-generated questions

Heavy snow is causing disruption to transport across the UK, with heavy rainfall bringing flooding to the south-west of England. Rescuers searching for a woman trapped in a landslide at her home said they had found a body.

Q1: What event has already finished?

Q2: What event has begun but has not finished?

Q3: What will happen in the future?

Hard-coded questions

Q4: What happened before a woman was trapped?

Q5: What had started before a woman was trapped?

Q6: What happened while a woman was trapped?

Q7: What happened after a woman was trapped?

Q8: What happened at about the same time as the snow?

Q9: What happened after the snow started?

Q10: What happened before the snow started?

User-provided

TORQUE

3.2k passages from the 2.8k articles in TempEval3

- Two sentences per passage
- 50 tokens per passage

24.9k events

For each passage, we have

- 3 hard-coded questions
- on average, 7 user-generated questions
- Grouped as contrast sets

Heavy snow is causing disruption to transport across the UK, with heavy rainfall bringing flooding to the south-west of England. Rescuers searching for a woman trapped in a landslide at her home said they had found a body.

Q1: What event has already finished?

Q2: What event has begun but has not finished?

Q3: What will happen in the future?

Hard-coded questions

Q4: What happened before a woman was trapped?

Q5: What had started before a woman was trapped?

Q6: What happened while a woman was trapped?

Q7: What happened after a woman was trapped?

Group of contrast questions

Q8: What happened at about the same time as the snow?

Q9: What happened after the snow started?

Q10: What happened before the snow started?

Group of contrast questions

3.2k passages from the 2.8k articles in TempEval3

- Two sentences per passage
- 50 tokens per passage

24.9k events

For each passage, we have

- 3 hard-coded questions

- on average, 7 user-generated questions

- Grouped as contrast sets

- \$15k on MTurk (\$0.7 per QA pair)

Heavy snow is causing disruption to transport across the UK, with heavy rainfall bringing flooding to the south-west of England. Rescuers searching for a woman trapped in a landslide at her home said they had found a body.

Q1: What event has already finished?

A: searching trapped landslide said found

Q2: What event has begun but has not finished?

A: snow causing disruption rainfall bringing flooding

Q3: What will happen in the future?

A: No answers.

Hard-coded questions

Q4: What happened before a woman was trapped?

A: landslide

Q5: What had started before a woman was trapped?

A: snow rainfall landslide

Q6: What happened while a woman was trapped?

A: searching

Q7: What happened after a woman was trapped?

A: searching said found

Group of contrast questions

Q8: What happened at about the same time as the snow?

A: rainfall

Q9: What happened after the snow started?

A: causing disruption bringing flooding searching trapped landslide said found

Q10: What happened before the snow started?

A: No answers.

Group of contrast questions

Questions querying temporal relations

- ❑ A temporal relation can be represented as a triplet: (Event A, Relation r, Event B)
- ❑ A valid question is: (? , r, B)
 - ❑ What happened after B?
- ❑ Perturbation: (? , r' , B)
 - ❑ What happened before/while ...?
 - ❑ What had begun before B?
 - ❑ What happened before B finished?
 - ❑ ...

How to Answer These Questions

1. Must choose from the list of events
2. Must choose all the correct answers
3. Answer from the time point when the passage was written.

TORQUE: Data collection process

Given a passage, an annotator needs to

1. Label all the events
2. Repeat the following steps
 - ▣ Ask a question querying temporal relations
 - ▣ Select all the correct answers from the event list
 - ▣ Slightly perturb the question and answer it

Please check the paper & the actual annotation guidelines at <https://qatmr-qualification.github.io/>

Annotation interface: <https://qatmr.github.io/>

Advantages of QA as a format

1. Handle fuzzy relations

Fuzzy relations

Heavy snow is causing disruption to transport across the UK, with heavy rainfall bringing flooding to the south-west of England.

Q: What happens at about the same time as the disruption?

A: flooding

Advantages of QA as a format

2. Handle events in different modalities

Events in different **modes**

The lion had a large meal and slept for 24 hours.

[**Negated**] The lion **didn't** sleep after having a large meal.

[**Uncertain**] The lion **may** have had a large meal before sleeping.

[**Hypothetical**] If the lion has a large meal, it will sleep for 24 hours.

[**Repetitive**] The lion **used to** sleep for 24 hours after having large meals.

[**Generic**] After having a large meal, **lions** may sleep longer.

Advantages of QA as a format

2. Handle events in different modalities

Questions that query events in different **modes**

[**Negated**] What didn't the lion do after a large meal?

[**Uncertain**] What might the lion do before sleeping?

[**Hypothetical**] What will the lion do if it has a large meal?

[**Repetitive**] What did the lion use to do after large meals?

[**Generic**] What do lions do after a large meal?

Advantages of QA as a format

3. Prior formalisms had a major issue with defining when two events should have a relation.

□ e.g., *TimeBank*, *TimeBank-dense*, *VerbClause*, *RED*, and *MATRES*

In TORQUE, the format of QA naturally bypasses this issue – we don't need to teach annotators these linguistic formalisms.

Experiment

❑ Setup:

- ❑ Given a passage and a question,
- ❑ model looks at every token in the passage
- ❑ makes a binary classification of whether this token is an answer to the question or not

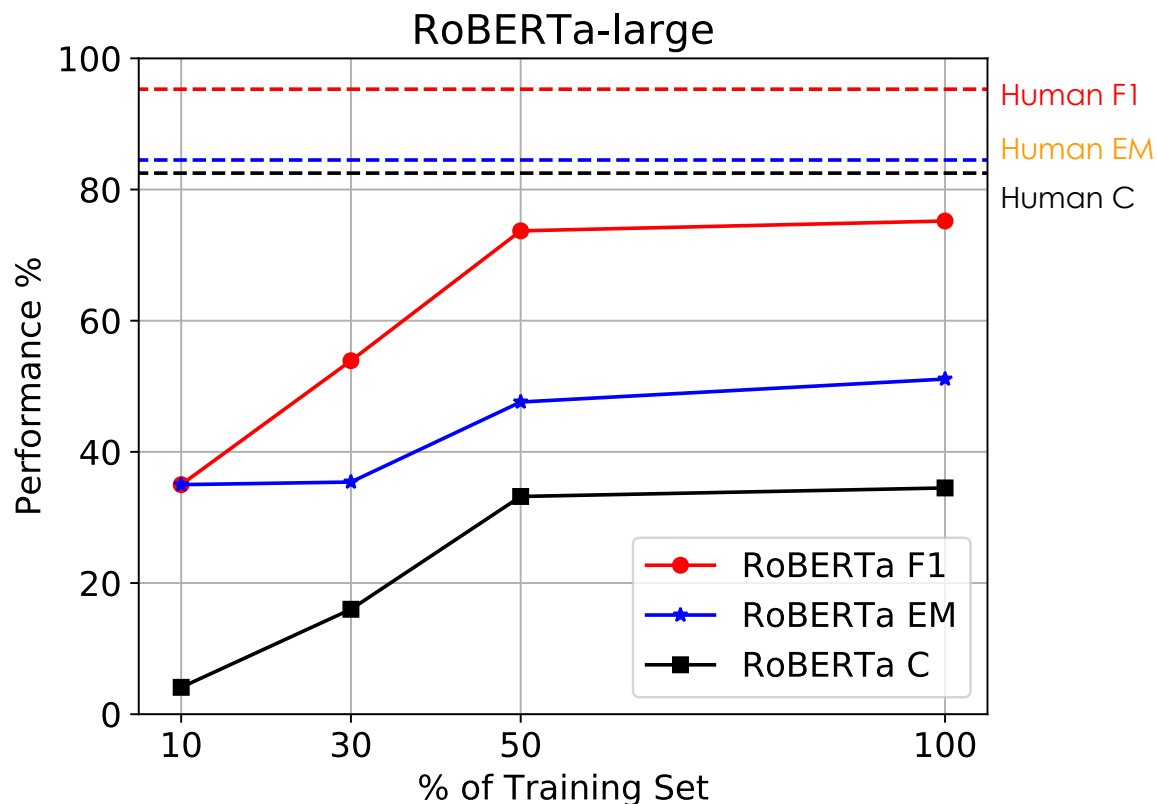
❑ Model:

- ❑ One-layered perceptron on top of RoBERTa

❑ Split:

- ❑ Train 80%, dev 5%, test 15%
- ❑ Dev & test are provided by 5/3 different annotators on their events/answers

Experiment



F1 & EM: standard metrics when there're multiple reference answers

C (consistency) is the percentage of contrast groups for which a model's predictions have $F1 \geq 80\%$ for all questions in a group

Summary

- ❑ Time is important in stories
 - ❑ but existing machine reading comprehension datasets don't cover these phenomena.
- ❑ TORQUE is a new crowdsourced dataset
 - ❑ on 3k English news snippets
 - ❑ with 10k hard-coded questions and 21k human-generated questions
- ❑ TORQUE is challenging and RoBERTa is behind human performance by a large margin
- ❑ Please visit <https://allennlp.org/torque.html> for dataset, leaderboard, and our annotation tools.