



Open Temporal Relation Extraction for Question Answering

Chao Shang, Peng Qi, Guangtao Wang, Jing Huang, Youzheng Wu, Bowen Zhou

JD AI Research

Problem Definition

Q1: What happened right after the election?

A: frozen, took, victory

Q2: What failed to happen after the election?

A: agreement, meet, lift, end

Q3: What happened right before the election?

A: No answers

- How to understand the nuanced difference in textual expressions that indicate different temporal relations.
- Existing QA techniques are usually ill-equipped to tackle the problem of temporal question answering.

Our Task

Given a piece of text, the model needs to answer **temporal relation questions**.

- **Explicit timestamps** are not available in texts and questions.
- Need to identify the events and **time of these events**.
- The result includes **multiple events** for one temporal relation question.

Context

The European Union and the United States have **frozen aid** to the Palestinian Authority ever since the Hamas-led government **took** power in March two months after its upset parliamentary **election victory**. Abbas's Fatah faction and Hamas **labored** for months to **reach** a power-sharing **agreement** that would **meet** international conditions to **lift** the **siege** and **end** the spiraling **crisis**, but those **talks failed** late last month.

Questions & Answers

Q1: What happened right after the election?

A: frozen, took, victory

Q2: What failed to happen after the election?

A: agreement, meet, lift, end

Q3: What happened right before the election?

A: No answers

Q4: What happened before the talks?

A: took, election, victory

Q5: What happened during the talks?

A: labored

Problem Reformulation

Context

The European Union and the United States have **frozen aid** to the Palestinian Authority ever since the Hamas-led government **took** power in March two months after its upset parliamentary **election victory**. Abbas's Fatah faction and Hamas **labored** for months to **reach** a power-sharing **agreement** that would **meet** international conditions to **lift** the **siege** and **end** the spiraling **crisis**, but those **talks failed** late last month.

Blue words are the events

Questions & Answers

Q1: **What happened right after** the election?

A: frozen, took, victory

Q2: **What failed to happen after** the election?

A: agreement, meet, lift, end

Q3: **What happened right before** the election?

A: No answers

Q4: **What happened before** the talks?

A: took, election, victory

Q5: **What happened during** the talks?

A: labored

Q6: **What happened after** the victory?

A: frozen, took, labored, talks, failed

Q7: **What happened right after** the victory?

A: frozen, took, labored

More questions

1. What are open temporal relations (OTRs)?

- The relations in questions are diverse so that they cannot be specified in advance.

2. What is the format of OTRs?

- OTRs are expressed in natural language.

Question event	Open temporal relation	Answer events
election	What happened right after	frozen, took, victory
election	What failed to happen after	agreement, meet, lift, end
election	What happened right before	No answer events
talks	What happened before	took, election, victory
talks	What happened during	labored
victory	What happened after	frozen, took, labored, talks, failed
victory	What happened right after	frozen, took, labored

Non-predefined Relations

- Reformulate the problem of temporal question answering as one of **open temporal relation extraction**.

Why use the OTRs?

Context

The European Union and the United States have **frozen aid** to the Palestinian Authority ever since the Hamas-led government **took** power in March two months after its upset parliamentary **election victory**. Abbas's Fatah faction and Hamas **labored** for months to **reach** a power-sharing **agreement** that would **meet** international conditions to **lift** the **siege** and **end** the spiraling **crisis**, but those **talks failed** late last month.

Blue words are the events

Questions & Answers

Q1: **What happened right after** the election?

A: frozen, took, victory

Q2: **What failed to happen after** the election?

A: agreement, meet, lift, end

Q3: **What happened right before** the election?

A: No answers

Q4: **What happened before** the talks?

A: took, election, victory

Q5: **What happened during** the talks?

A: labored

Q6: **What happened after** the victory?

A: frozen, took, labored, talks, failed

Q7: **What happened right after** the victory?

A: frozen, took, labored

More questions

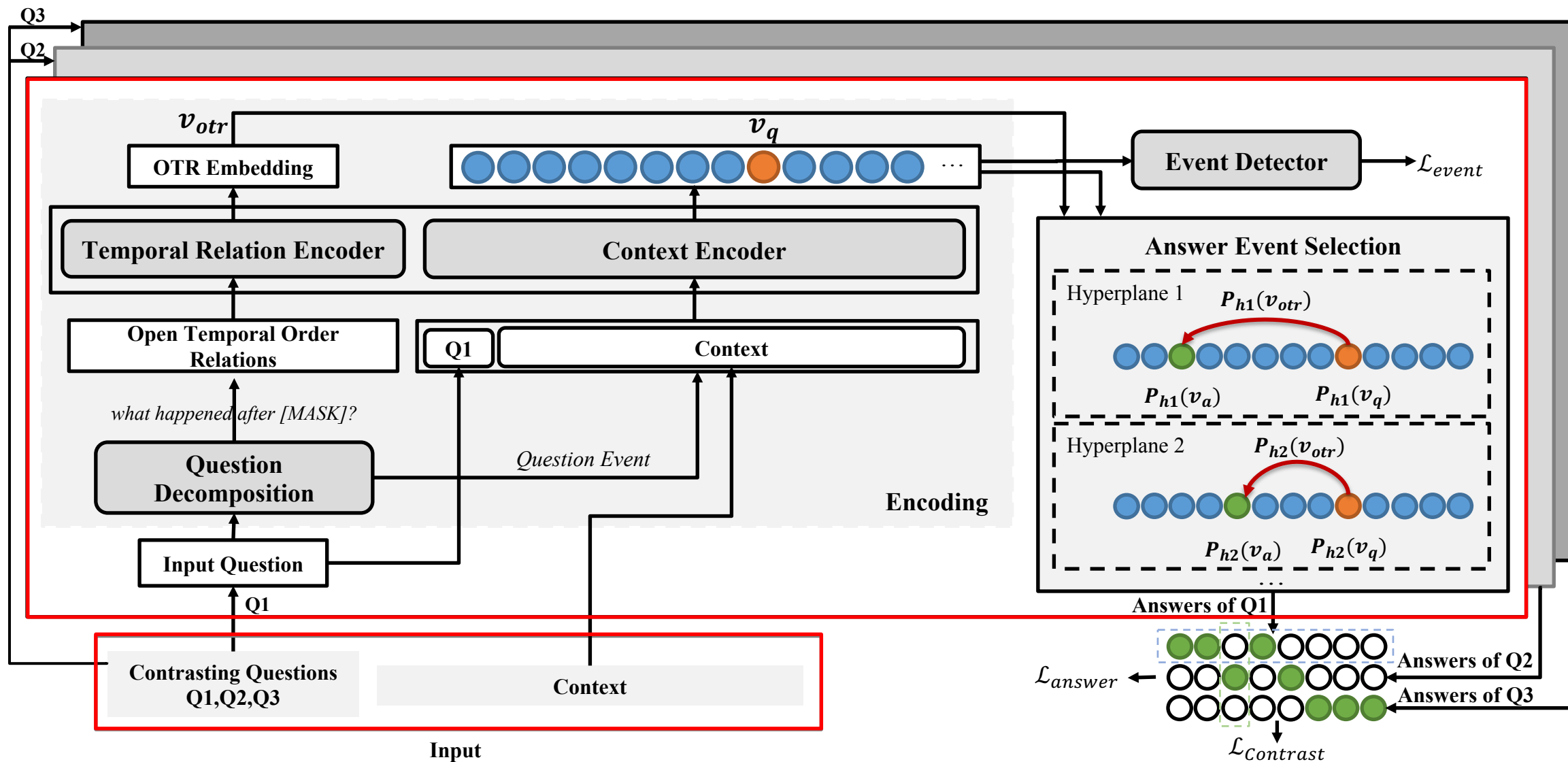
Temporal Relation Extraction

Question event	Open temporal relation	Answer events
election	What happened right after	frozen, took, victory
election	What failed to happen after	agreement, meet, lift, end
election	What happened right before	No answer events
talks	What happened before	took, election, victory
talks	What happened during	labored
victory	What happened after	frozen, took, labored, talks, failed
victory	What happened right after	frozen, took, labored

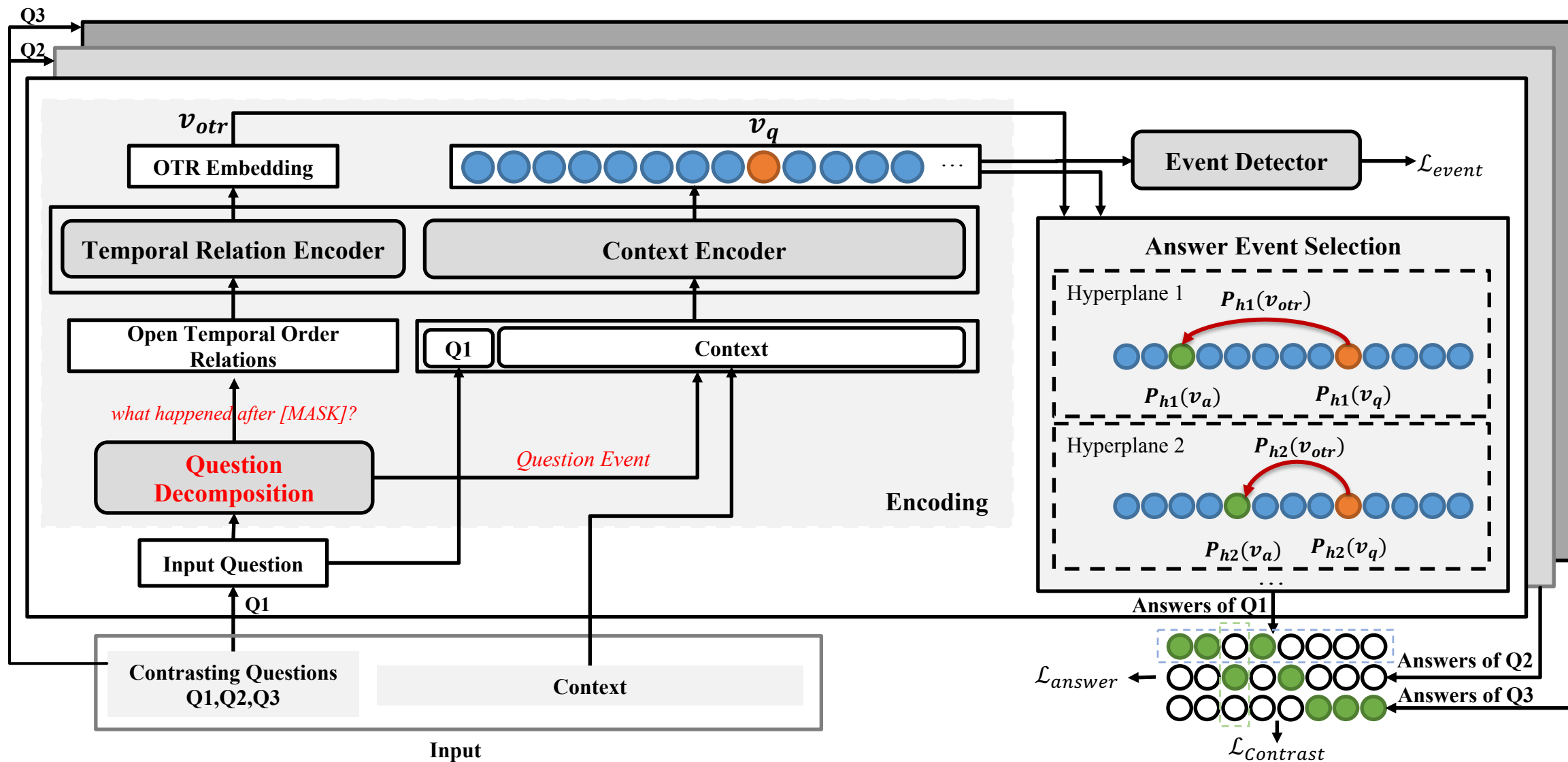
Non-predefined Relations

- Allow us to model temporal relations in a context-agnostic manner, which **shares supervision signal** from different contexts and events to the same underlying open temporal relation.
- Allow us to explicitly model the differences in temporal relations with a contrastive loss function, which helps **capture mutually exclusive relations**.

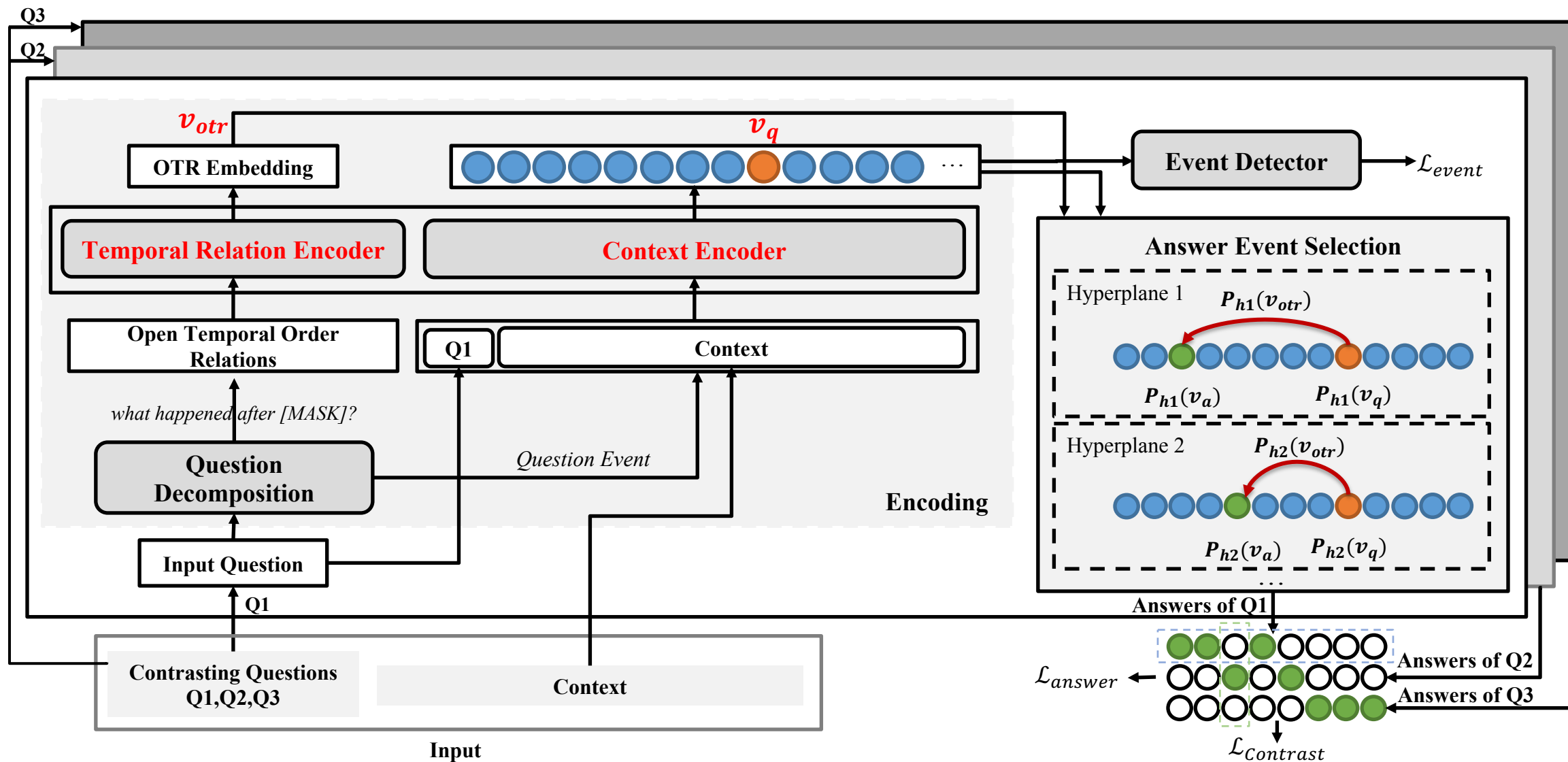
OTR-QA model



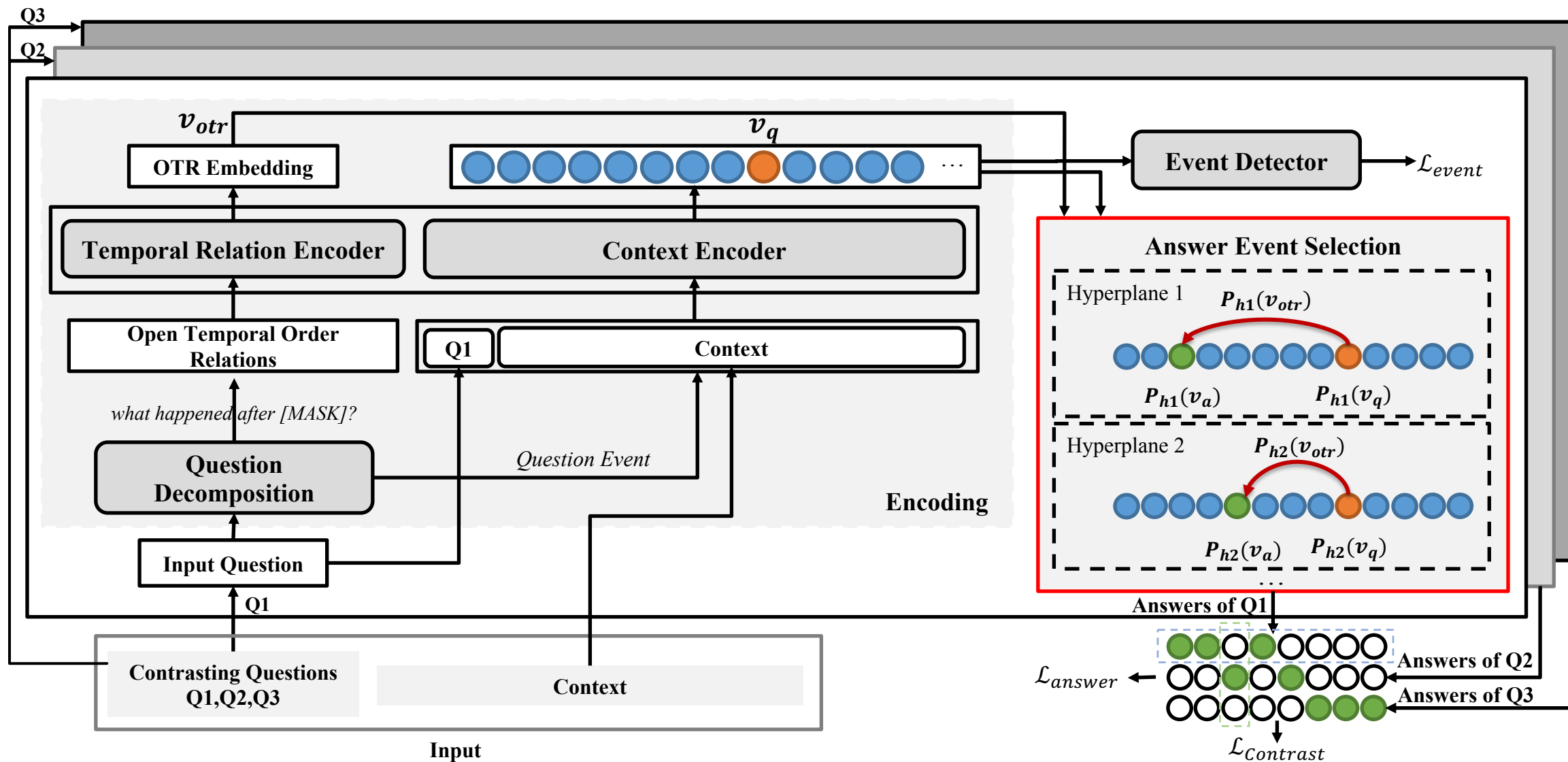
OTR-QA model



OTR-QA model



OTR-QA model



- Model Comparisons

Model	Val			Test		
	F1	EM	C	F1	EM	C
BERT-base [Ning et al., 2020]	67.6	39.6	24.3	67.2	39.8	23.6
BERT-large [Ning et al., 2020]	72.8	46.0	30.7	71.9	45.9	29.1
RoBERTa-base [Ning et al., 2020]	72.2	44.5	28.7	72.6	45.7	29.9
RoBERTa-large [Ning et al., 2020]	75.7	50.4	36.0	75.2	51.1	34.5
OTR-QA (RoBERTa-base)	75.2	49.2	36.1	73.4	47.1	32.7
OTR-QA (RoBERTa-large)	77.1	51.6	40.6	76.3	50.9	37.5

Table 1: Results on TORQUE dev set and test set.

- Data Efficiency
 - This proves that the OTR-QA is much more data-efficient, because it is designed to learn temporal relations in a context-agnostic manner, and thus needs less training data.

Model	F1	EM	C
RoBERTa-base (10% training data)	57.3	33.3	13.8
RoBERTa-base (20% training data)	66.8	39.8	24.1
RoBERTa-base (50% training data)	69.7	44.3	27.8
OTR-QA (RoBERTa-base) (10% training data)	69.0	40.7	25.0
OTR-QA (RoBERTa-base) (20% training data)	71.2	43.3	29.1
OTR-QA (RoBERTa-base) (50% training data)	73.4	47.2	32.4

Table 2: Comparison of the baseline models and OTR-QA models trained using different percentages of training data. Results are on the validation data.

———— Q&A ————

THANKS