CLEVE: Contrastive Pre-training for Event Extraction

Ziqi Wang*[1], Xiaozhi Wang*[1], Xu Han[1], Yankai Lin[2],
Lei Hou[1], Zhiyuan Liu[1], Peng Li[2], Juanzi Li[1], Jie Zhou[2]

1 清華大學 Tsinghua University

2 WeChat AI

Introduction

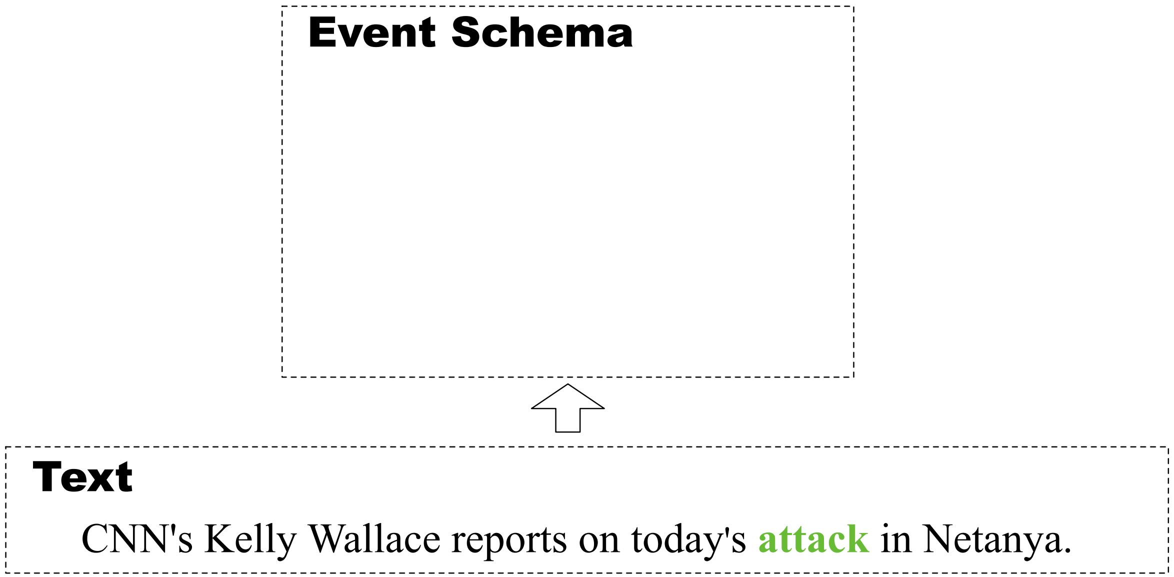
Event Extraction:

1. Event Detection
2. Event Argument Extraction

Introduction

Event Extraction:

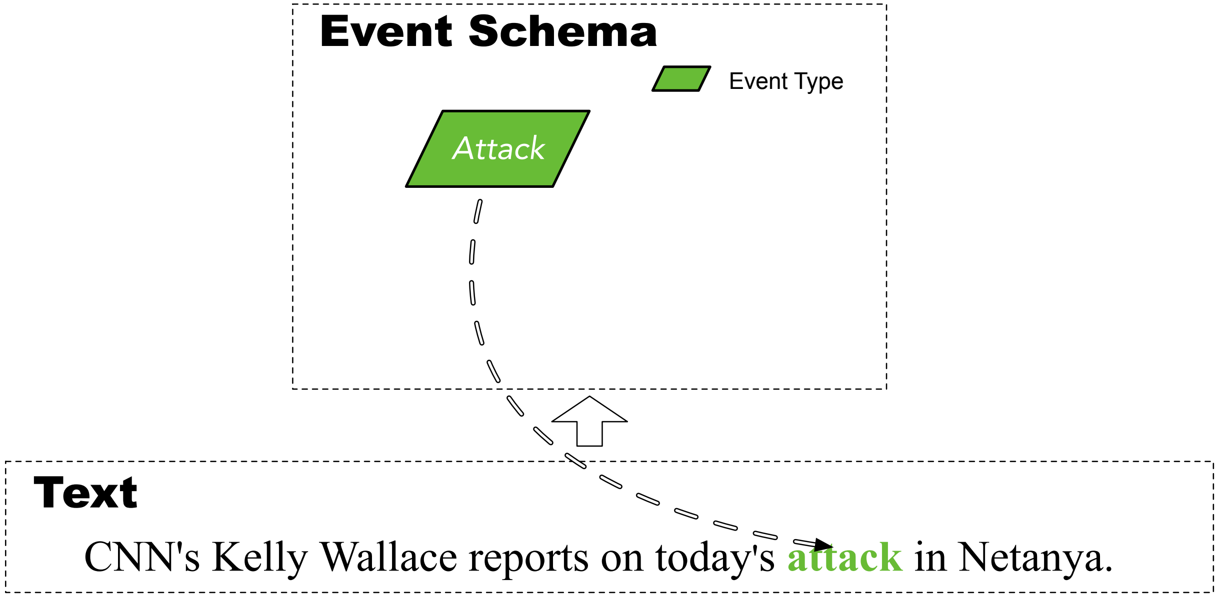
1. Event Detection
   - Identify trigger words that indicating the occurrence of an event
   - Classify their event types
2. Event Argument Extraction



Introduction

Event Extraction:

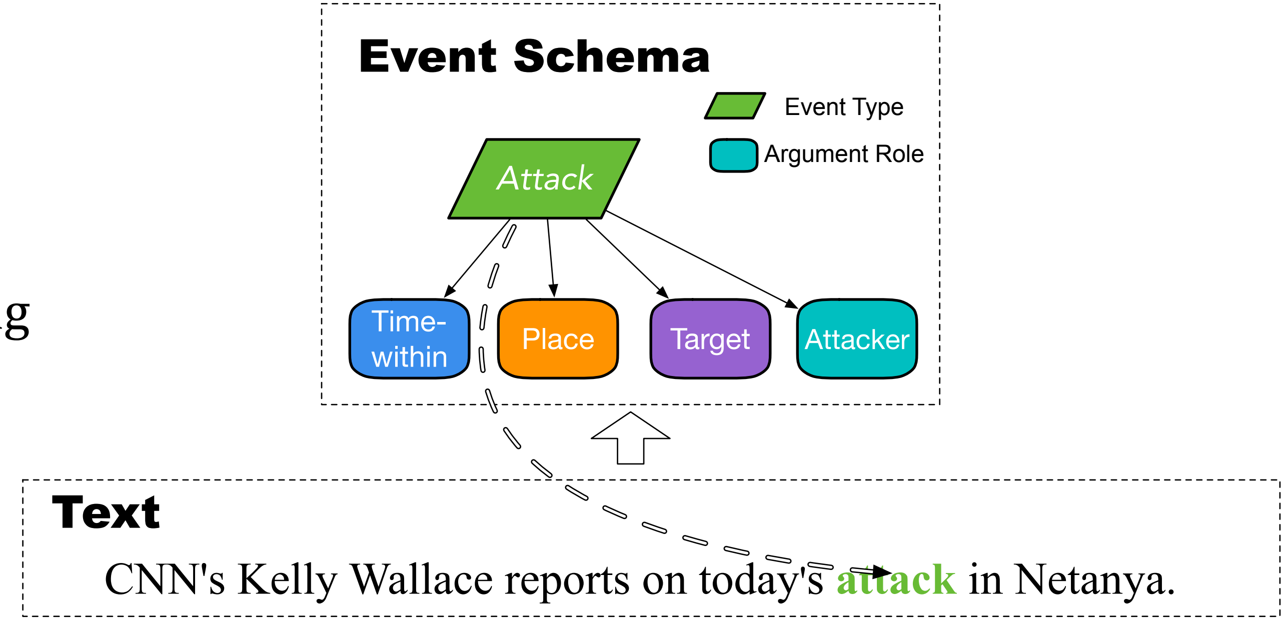
1. Event Detection
    - Identify trigger words that indicating the occurrence of an event
    - Classify their event types
2. Event Argument Extraction



Introduction

Event Extraction:

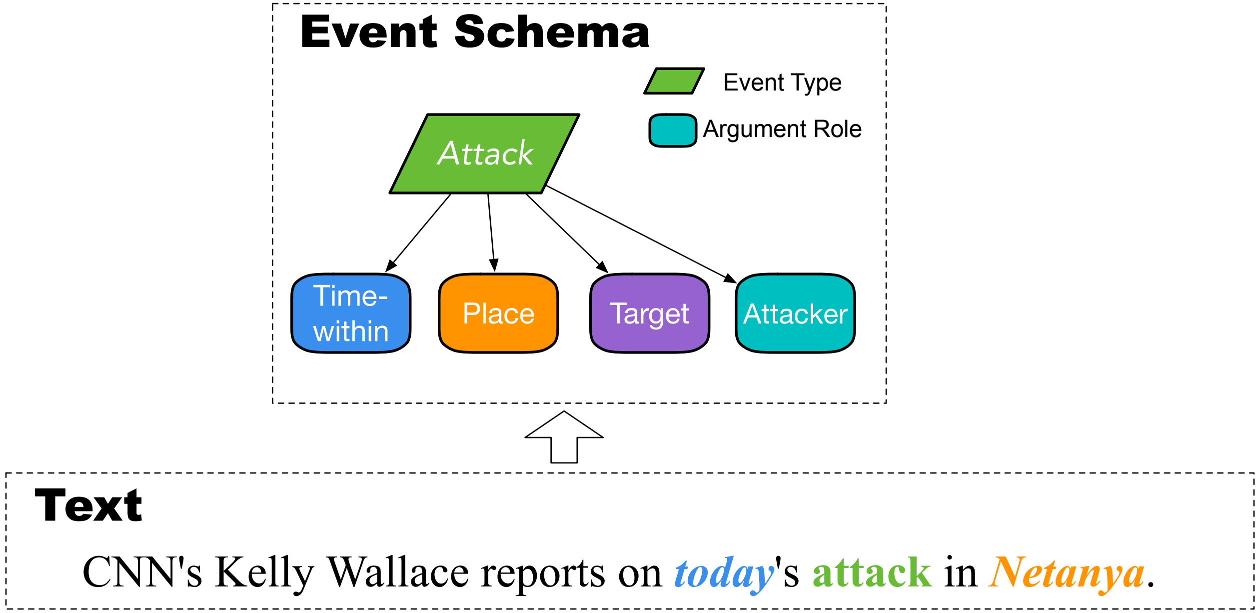
1. Event Detection
   - Identify trigger words that indicating the occurrence of an event
   - Classify their event types
2. Event Argument Extraction



Introduction

Event Extraction:

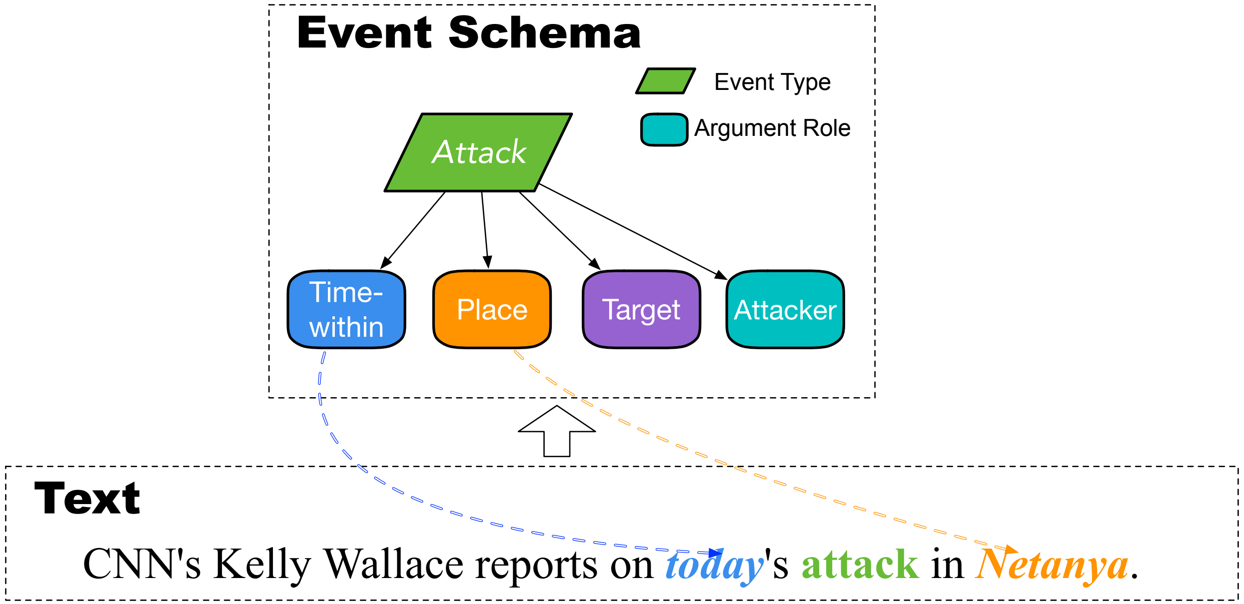
1. Event Detection
2. Event Argument Extraction
   - Identify entities serving as event arguments
   - Classify their roles



Introduction

Event Extraction:

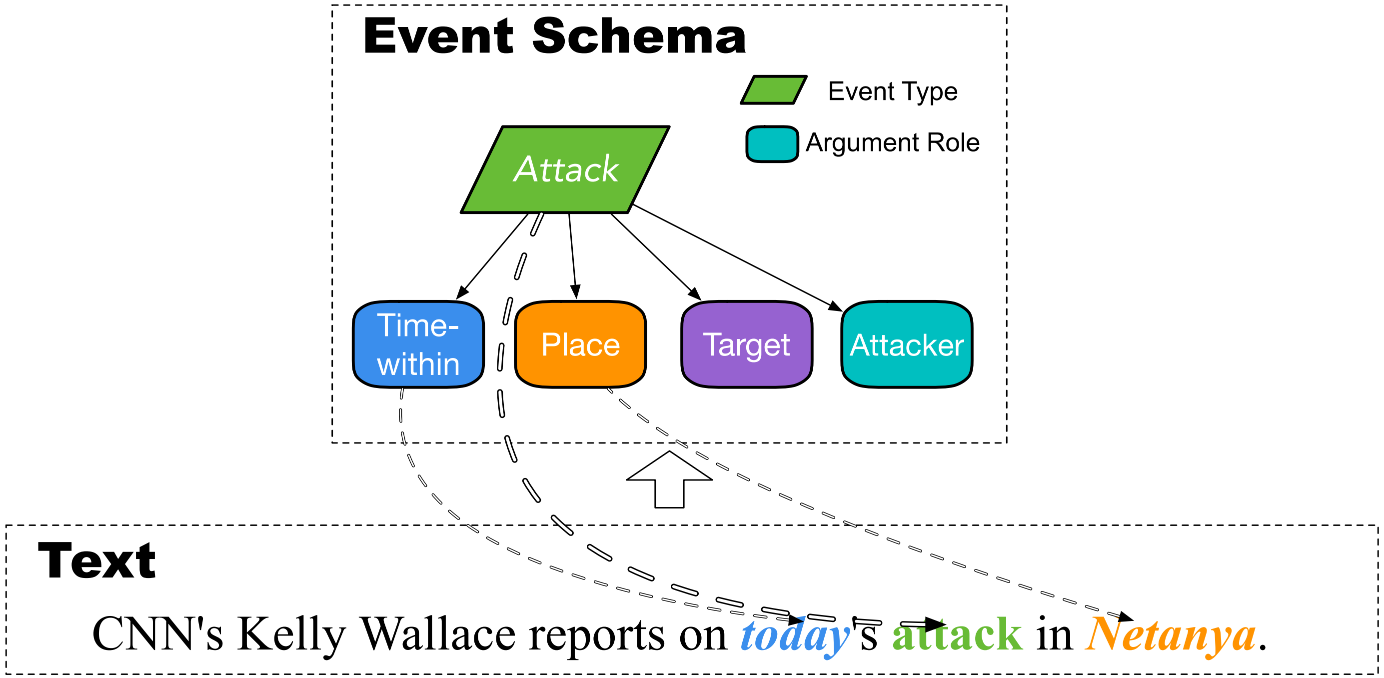
1. Event Detection
2. Event Argument Extraction
   - Identify entities serving as event arguments
   - Classify their roles



Introduction

Event Extraction:

1. Event Detection
2. Event Argument Extraction

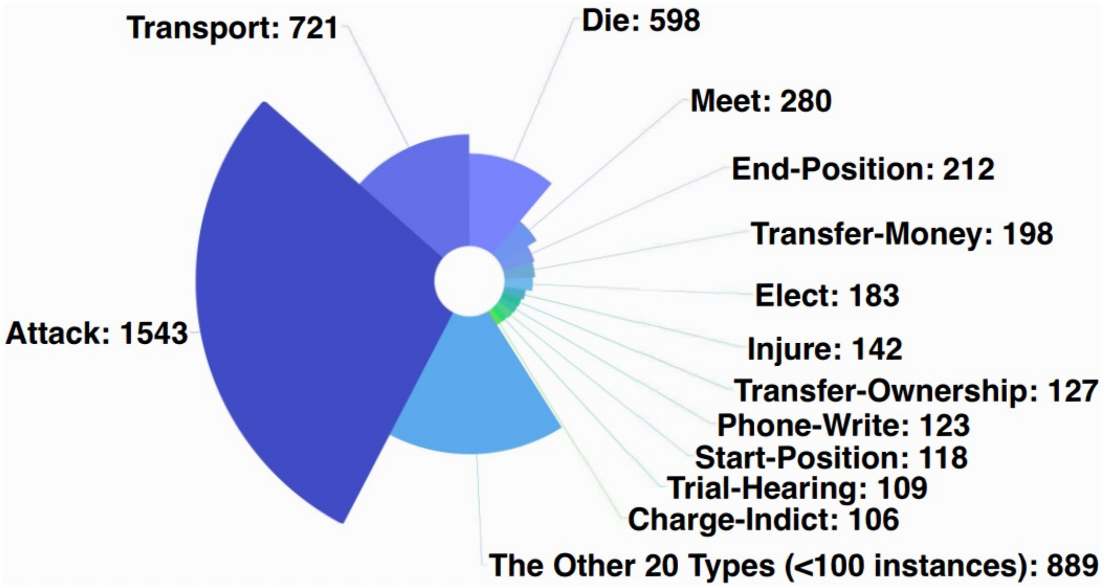


Motivation

- Current methods suffer from:

Motivation

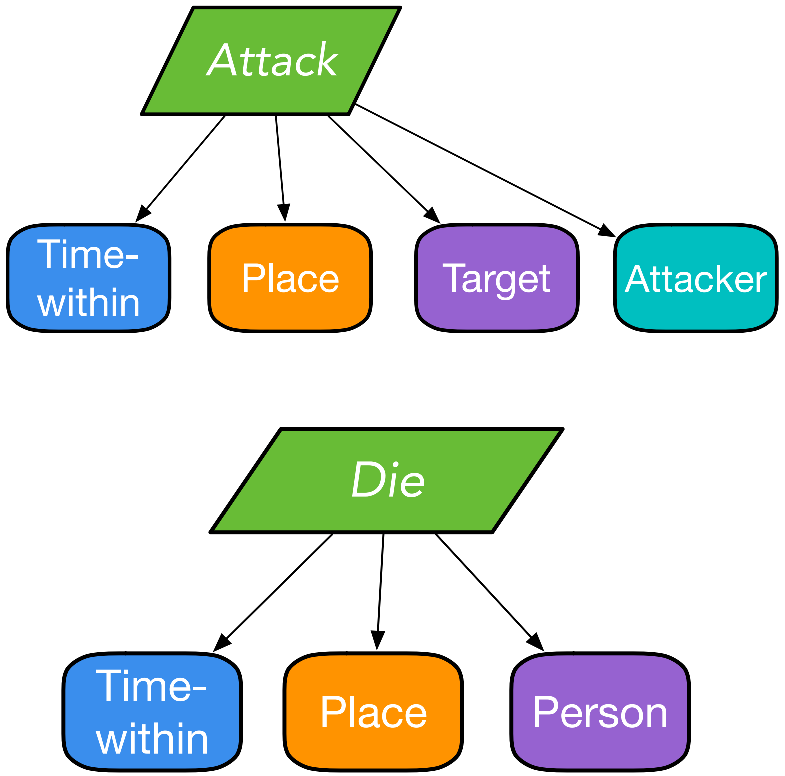
- Current methods suffer from:
  - Data scarcity:



ACE data distributions

Motivation

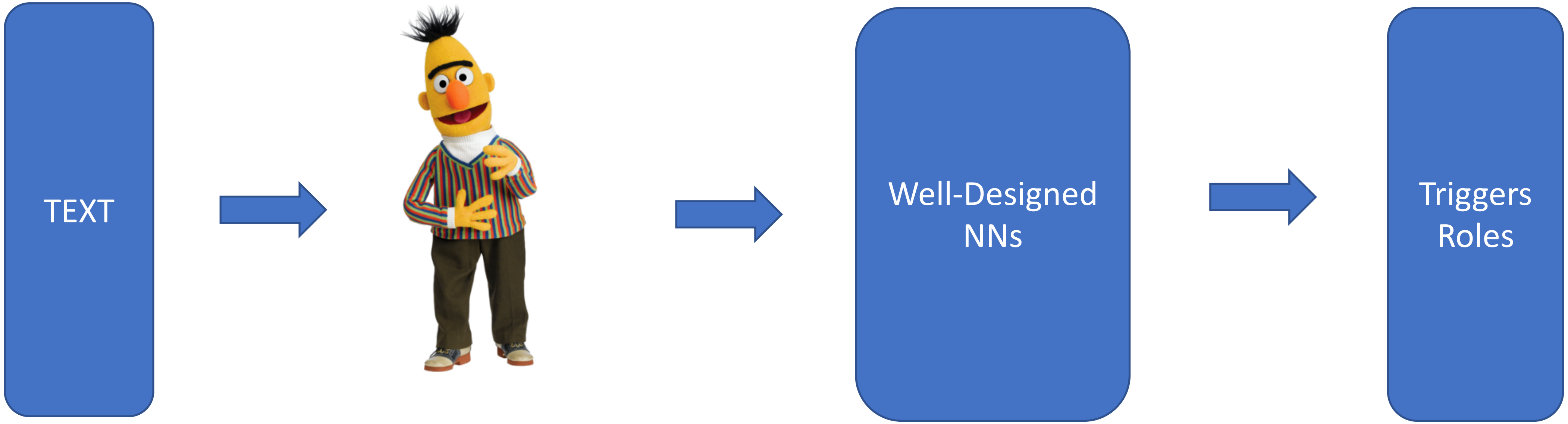
- Current methods suffer from:
  - Data scarcity
  - Limited Generalizability:



Motivation

Pre-trained Language Models
for Event Extraction

Motivation



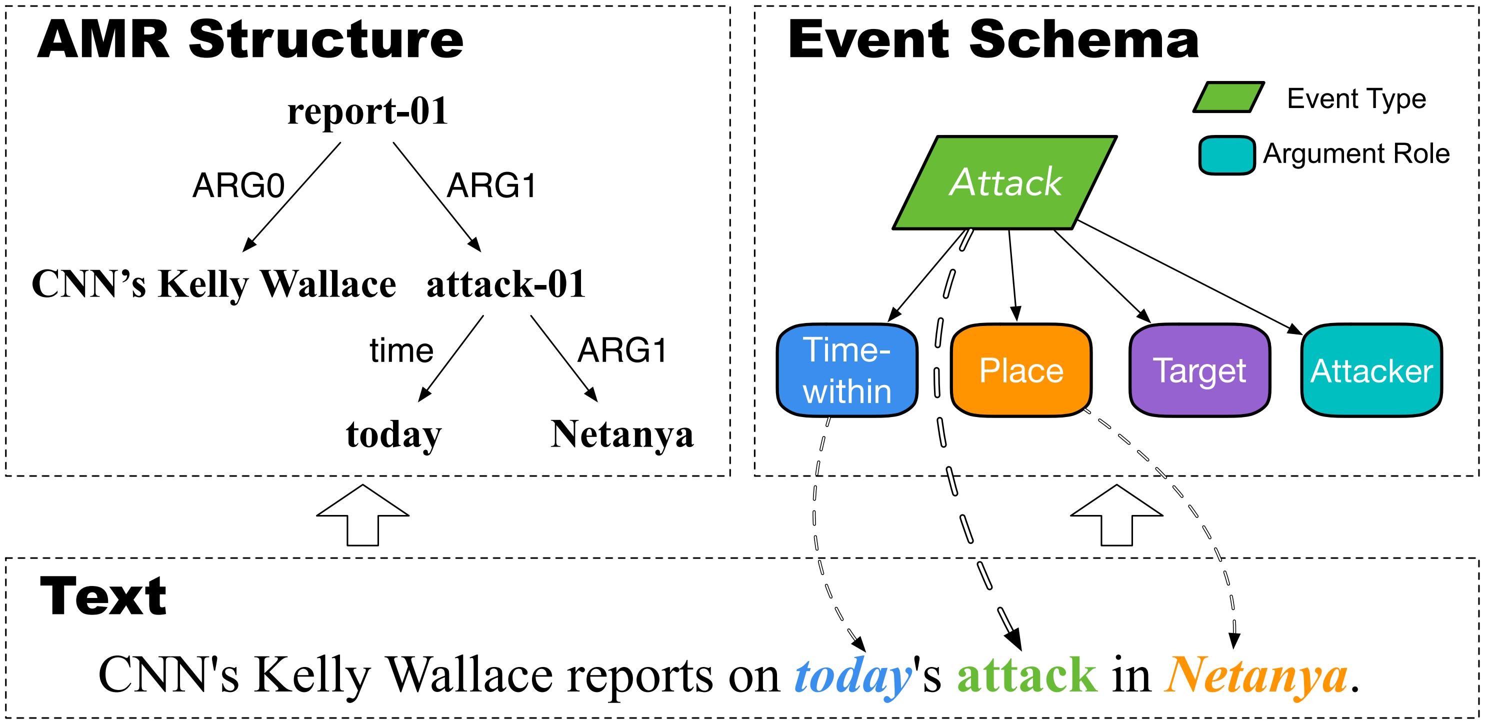
Pre-trained Language Models

Finetune PLMs

Pretrain?

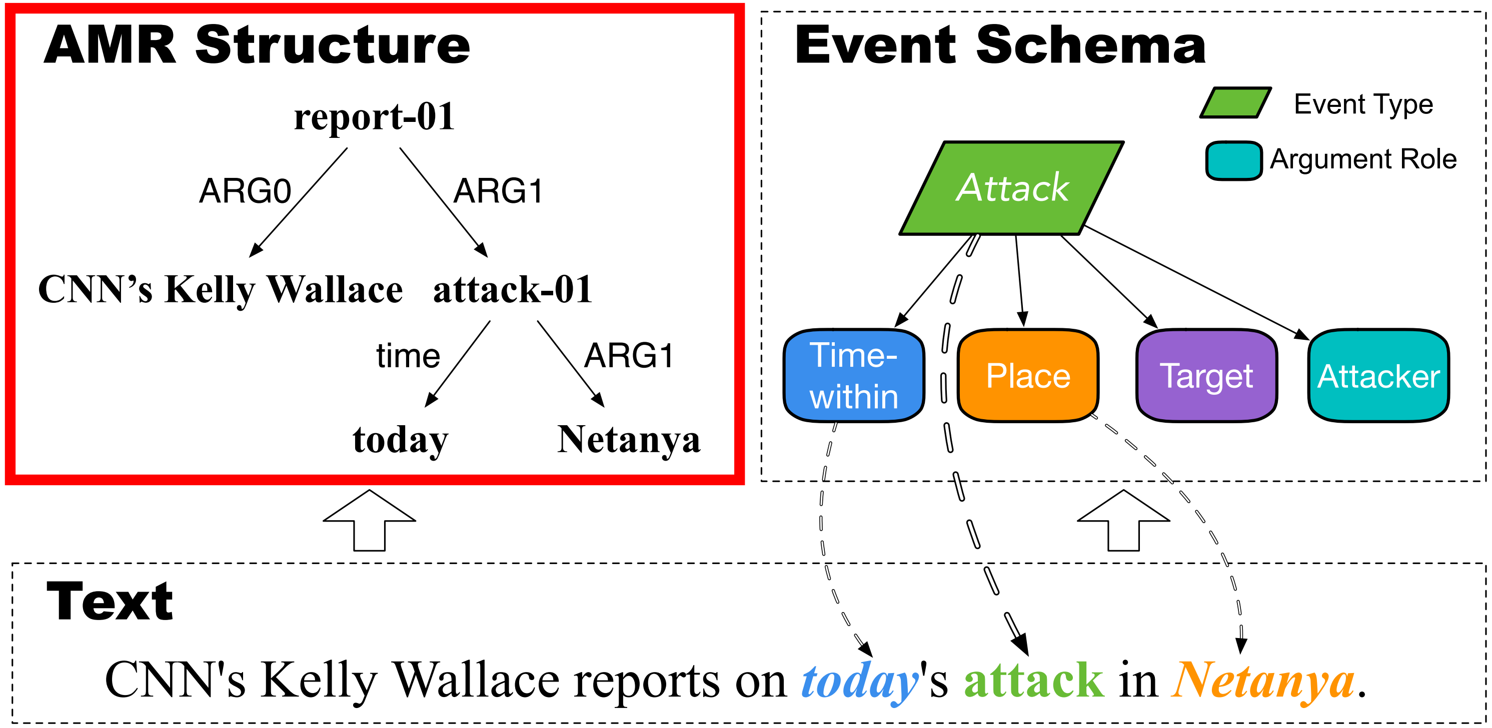
Motivation

- Signals from unsupervised data:
  - AMR structure



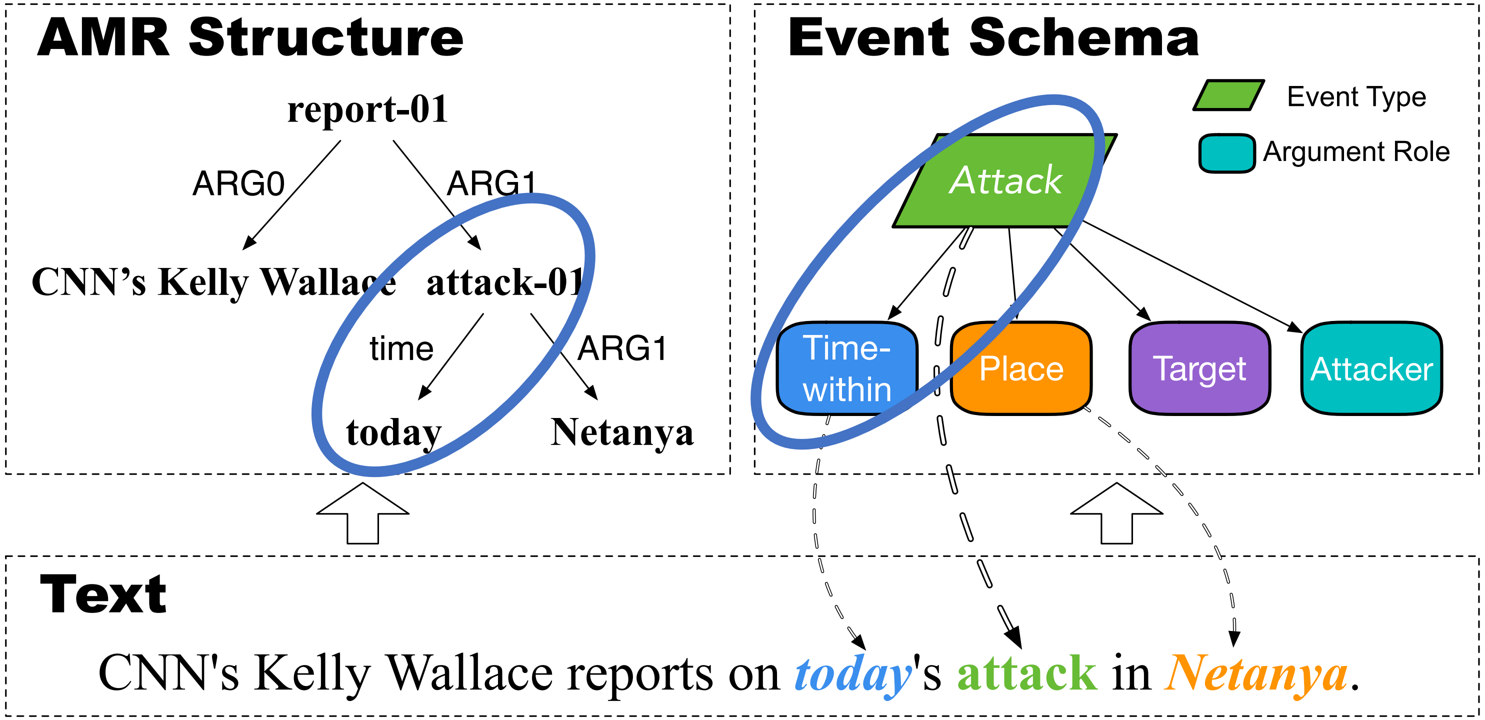
Motivation

- Signals from unsupervised data:
  - AMR structure



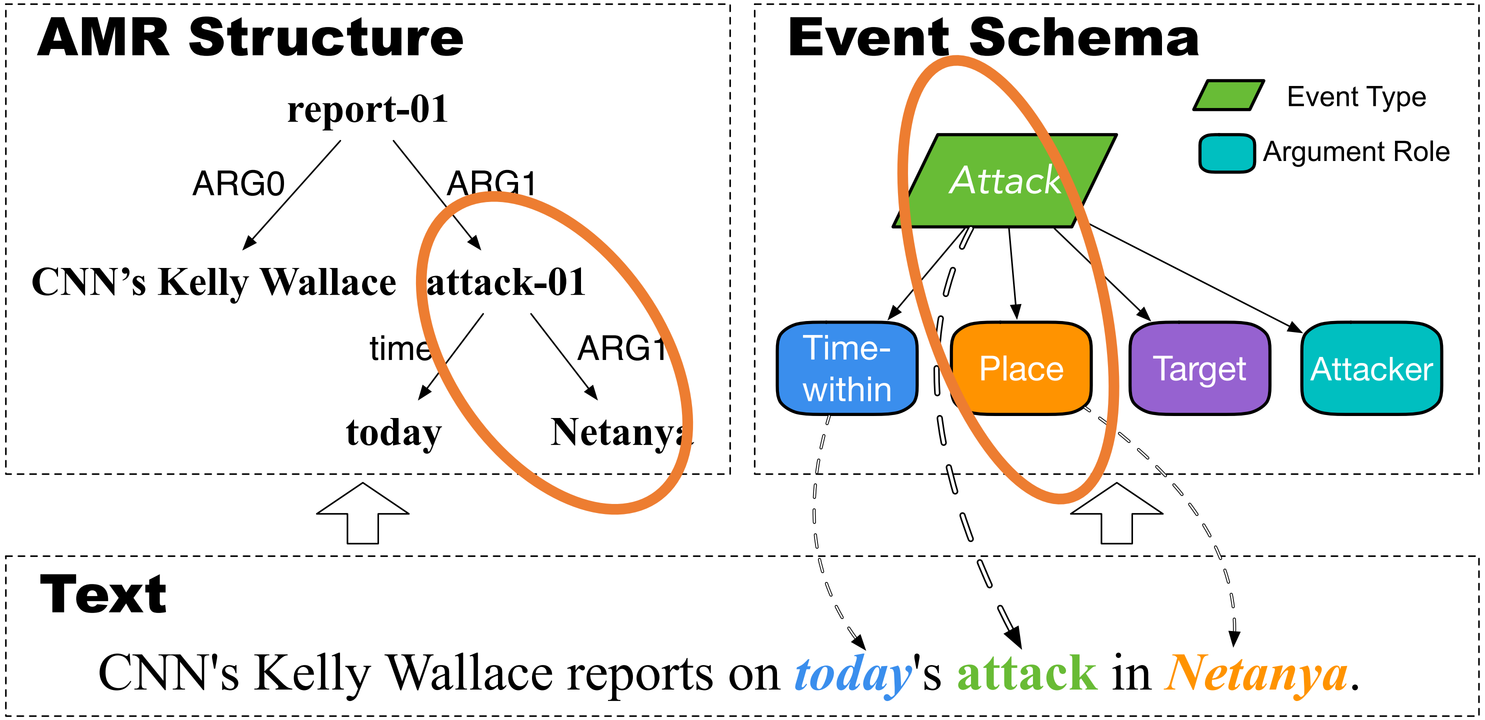
Motivation

- Signals from unsupervised data:
  - AMR structure



Motivation

- Signals from unsupervised data:
  - AMR structure



Motivation

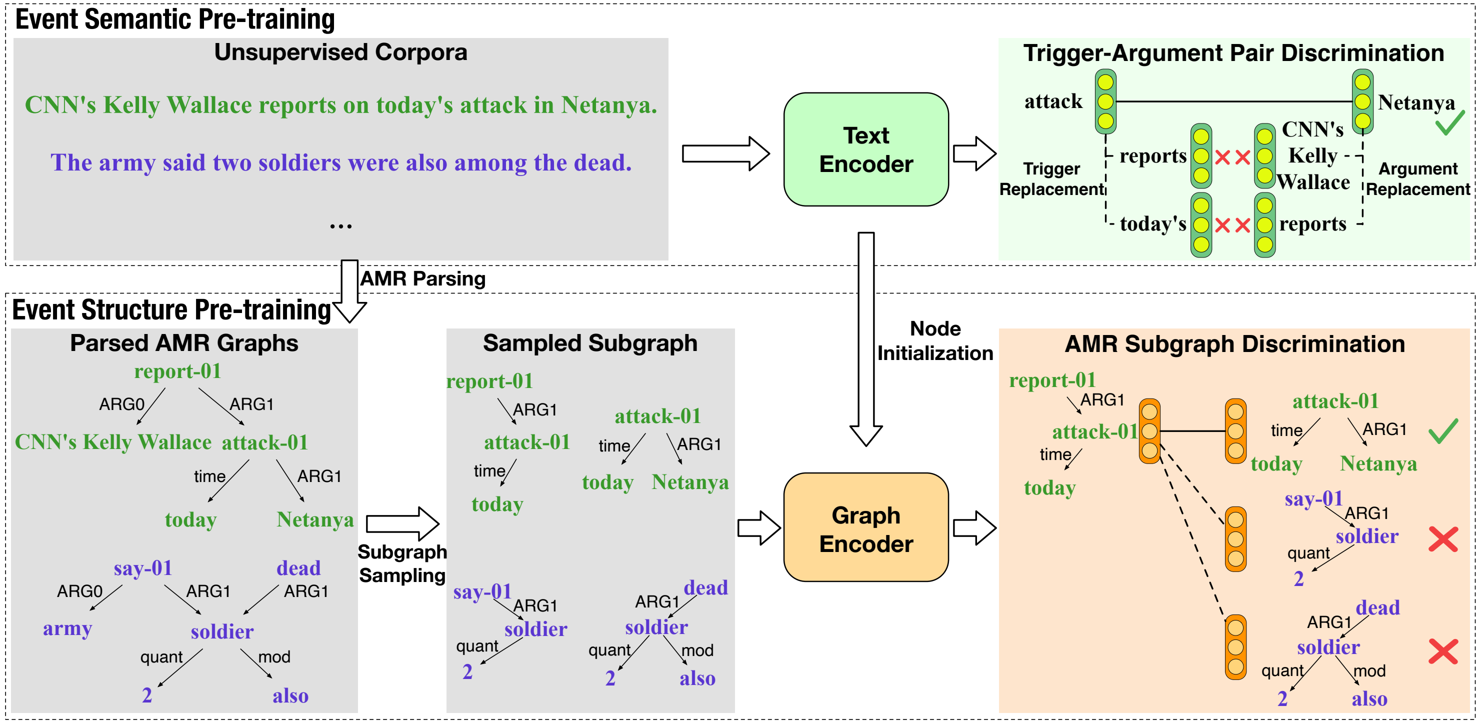
How to use AMR structures to pre-train PLMs?

Method

Framework:

- Event Semantic Pre-training
- Event Structure Pre-training

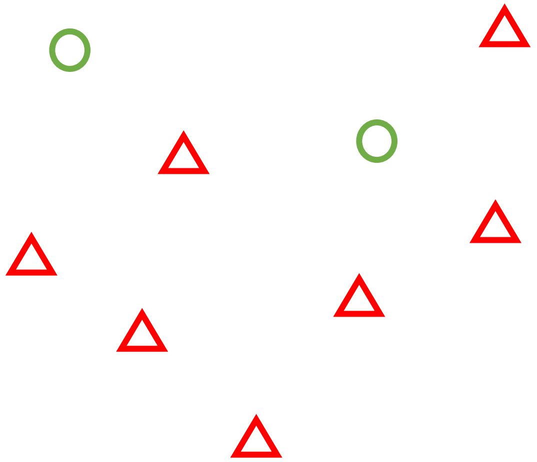
} Contrastive Learning



Method

Contrastive Learning:

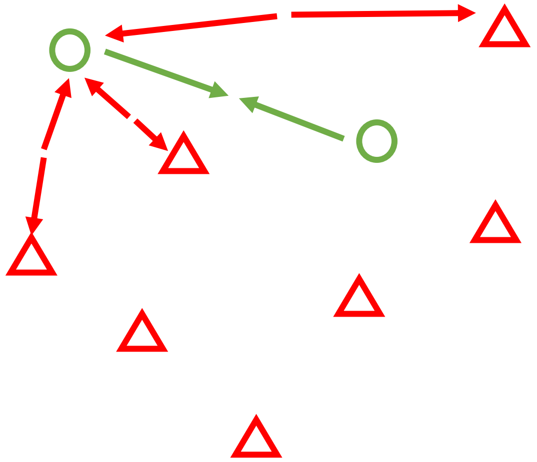
- Positive instances, Negative instances



Method

Contrastive Learning:

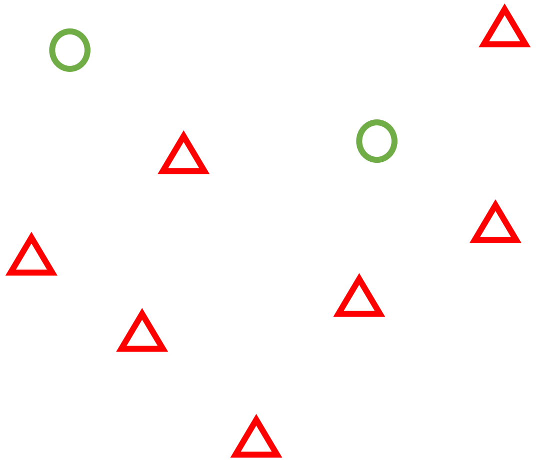
- Positive instances, Negative instances



Method

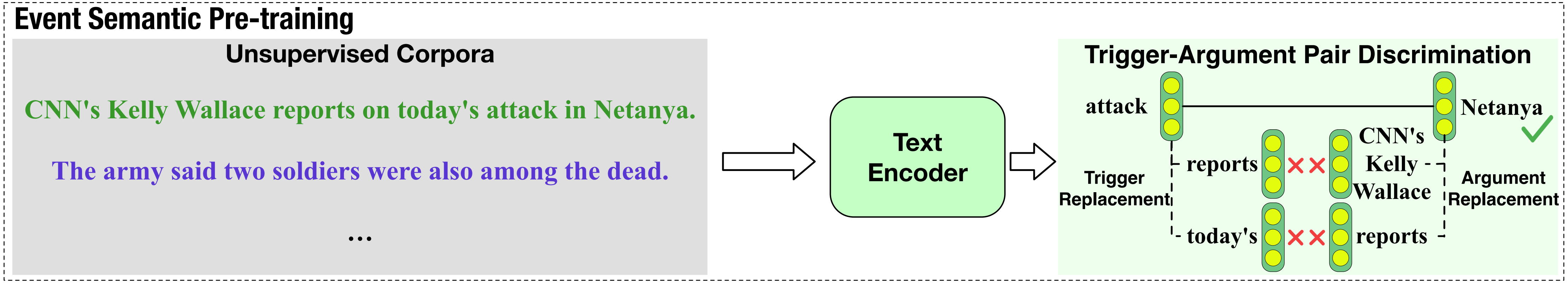
Contrastive Learning:

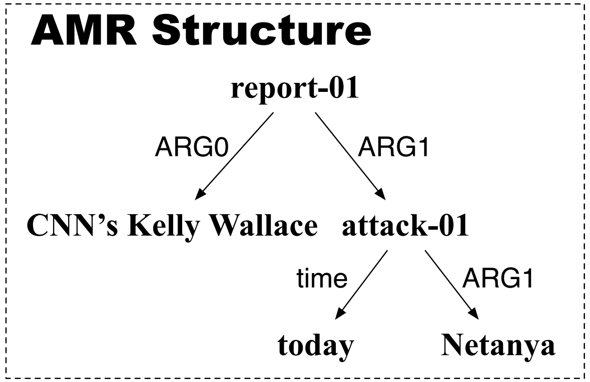
- Positive instances, Negative instances
- Follow SimCLR (Chen et al. 2020)



Method

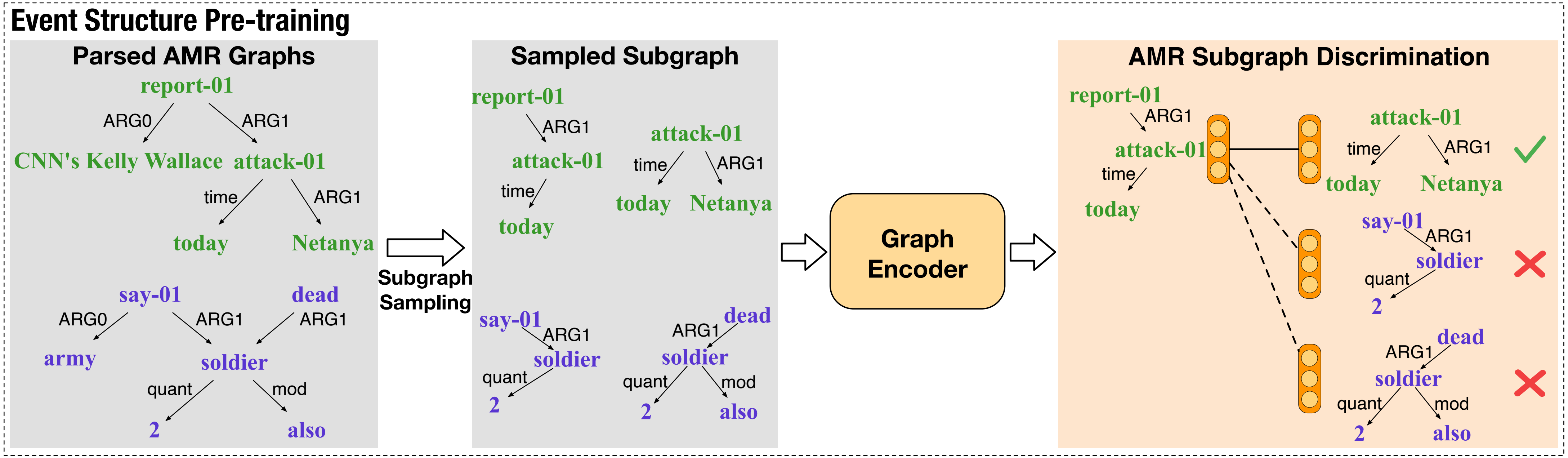
- Event Semantic Pre-training:
  - Positive instances: Node pairs linked by edge Arg, Time, Location
  - Negative instances: Other Node pairs





Method

- Event Structure Pre-training:
  - Positive instances: subgraphs sampled from the same AMR structure
  - Negative instances: subgraphs sampled from different AMR structures



Supervised Experiments

	ED			EAE		
Metric	**P**	**R**	**F1**	**P**	**R**	**F1**
JointBeam	73.7	62.3	67.5	64.7	44.4	52.7
DMCNN	75.6	63.6	69.1	62.2	46.9	53.5
dbRNN	74.1	69.8	71.9	66.2	52.8	58.7
GatedGCN	**78.8**	76.3	77.6	–	–	–
SemSynGTN	–	–	–	**69.3**	55.9	61.9
RCEE_ER	75.6	74.2	74.9	63.0	64.2	**63.6**
RoBERTa	75.1	79.2	77.1	53.5	66.8	59.4
CLEVE	78.1	**81.5**	**79.8**	55.4	**68.0**	61.1
w/o semantic	75.3	79.7	77.4	53.8	67.0	59.7
w/o structure	78.0	81.1	79.5	55.1	67.6	60.7
on ACE (golden)	76.2	79.8	78.0	54.2	67.5	60.1
on ACE (AMR)	75.7	79.5	77.6	53.6	66.9	59.5

Table 1: Supervised EE performance (%) of various models on ACE 2005.

	ED		
Metric	**P**	**R**	**F1**
DMCNN	**66.3**	55.9	60.6
BiLSTM	59.8	67.0	62.8
BiLSTM+CRF	63.4	64.8	64.1
MOGANED	63.4	64.1	63.8
DMBERT	62.7	72.3	67.1
BERT+CRF	65.0	70.9	67.8
RoBERTa	64.3	72.2	68.0
CLEVE	64.9	**72.6**	**68.5**
w/o semantic	64.5	72.4	68.2
w/o structure	64.7	72.5	68.4

Table 2: Supervised EE performance (%) of various models on MAVEN.

Supervised Experiments

Metric	ED P	ED R	ED F1	EAE P	EAE R	EAE F1
JointBeam	73.7	62.3	67.5	64.7	44.4	52.7
DMCNN	75.6	63.6	69.1	62.2	46.9	53.5
dbRNN	74.1	69.8	71.9	66.2	52.8	58.7
GatedGCN	**78.8**	76.3	77.6	–	–	–
SemSynGTN	–	–	–	**69.3**	55.9	61.9
RCEE_ER	75.6	74.2	74.9	63.0	64.2	**63.6**
RoBERTa	75.1	79.2	77.1	53.5	66.8	59.4
CLEVE	78.1	**81.5**	**79.8**	55.4	**68.0**	61.1
w/o semantic	75.3	79.7	77.4	53.8	67.0	59.7
w/o structure	78.0	81.1	79.5	55.1	67.6	60.7
on ACE (golden)	76.2	79.8	78.0	54.2	67.5	60.1
on ACE (AMR)	75.7	79.5	77.6	53.6	66.9	59.5

Table 1: Supervised EE performance (%) of various models on ACE 2005.

Metric	ED P	ED R	ED F1
DMCNN	**66.3**	55.9	60.6
BiLSTM	59.8	67.0	62.8
BiLSTM+CRF	63.4	64.8	64.1
MOGANED	63.4	64.1	63.8
DMBERT	62.7	72.3	67.1
BERT+CRF	65.0	70.9	67.8
RoBERTa	64.3	72.2	68.0
CLEVE	64.9	**72.6**	**68.5**
w/o semantic	64.5	72.4	68.2
w/o structure	64.7	72.5	68.4

Table 2: Supervised EE performance (%) of various models on MAVEN.

Unsupervised Experiments

	ED			EAE		
Metric (B-Cubed)	**P**	**R**	**F1**	**P**	**R**	**F1**
LiberalEE	55.7	45.1	49.8	36.2	26.5	30.6
RoBERTa	44.3	24.9	31.9	24.2	17.3	20.2
RoBERTa+VGAE	47.0	26.8	34.1	25.6	17.9	21.1
CLEVE	**62.0**	**47.3**	**53.7**	**41.6**	**30.3**	**35.1**
w/o semantic	60.6	46.2	52.4	40.9	29.8	34.5
w/o structure	45.7	25.6	32.8	25.0	17.9	20.9
on ACE (AMR)	61.1	46.7	52.9	41.5	30.1	34.9

Table 3: Unsupervised "liberal" EE performance (%) of various models on ACE 2005.

	ED		
Metric (B-Cubed)	**P**	**R**	**F1**
RoBERTa	32.1	25.2	28.2
RoBERTa+VGAE	37.7	28.5	32.5
CLEVE	**55.6**	**46.4**	**50.6**
w/o semantic	53.2	44.8	48.6
w/o structure	32.8	26.1	29.1

Table 4: Unsupervised "liberal" EE performance (%) of various models on MAVEN.

Unsupervised Experiments

	ED			EAE		
Metric (B-Cubed)	**P**	**R**	**F1**	**P**	**R**	**F1**
LiberalEE	55.7	45.1	49.8	36.2	26.5	30.6
RoBERTa	44.3	24.9	31.9	24.2	17.3	20.2
RoBERTa+VGAE	47.0	26.8	34.1	25.6	17.9	21.1
CLEVE	**62.0**	**47.3**	**53.7**	**41.6**	**30.3**	**35.1**
w/o semantic	60.6	46.2	52.4	40.9	29.8	34.5
w/o structure	45.7	25.6	32.8	25.0	17.9	20.9
on ACE (AMR)	61.1	46.7	52.9	41.5	30.1	34.9

Table 3: Unsupervised "liberal" EE performance (%) of various models on ACE 2005.

	ED		
Metric (B-Cubed)	**P**	**R**	**F1**
RoBERTa	32.1	25.2	28.2
RoBERTa+VGAE	37.7	28.5	32.5
CLEVE	**55.6**	**46.4**	**50.6**
w/o semantic	53.2	44.8	48.6
w/o structure	32.8	26.1	29.1

Table 4: Unsupervised "liberal" EE performance (%) of various models on MAVEN.

Analysis

- Performance vs. Different AMR parsers

	AMR 1.0	ACE 2005		MAVEN
	Parsing	ED	EAE	ED
Wang et al. (2015)	62.0	79.8	61.1	68.5
Xu et al. (2020)	79.1	80.6	61.5	69.0

Analysis

- Performance vs. Different pre-training corpus

	ACE 2005		MAVEN
	ED	EAE	ED
NYT	**79.8**	**61.1**	68.5
w/o semantic	77.4	59.7	68.2
w/o structure	79.5	60.7	68.4
Wikipedia	79.1	60.4	**68.8**
w/o semantic	77.3	59.5	68.4
w/o structure	78.8	60.0	68.6

Summary & Future Work

- Propose a contrastive pre-training framework for EE task
- Utilize rich event knowledge lying in large unsupervised data

- Other kinds of semantic structures
- Overcome noises

THANKS!

Paper

Code