Information Extraction

Event-Centric Natural Language Understanding (Part I)

Manling Li

Department of Computer Science

University of Illinois at Urbana-Champaign

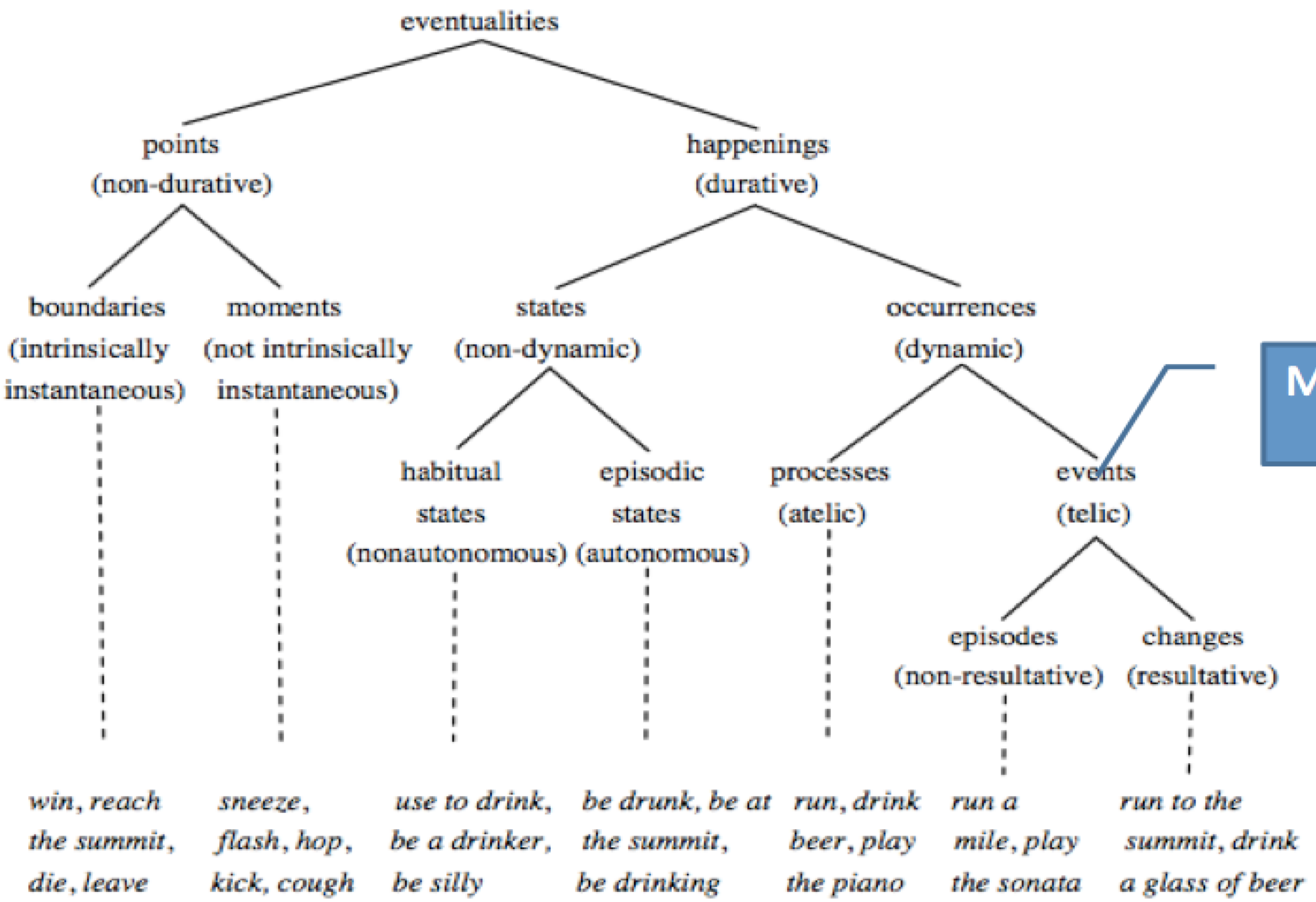
Feb 2020

AAAI Tutorials

Recent Advances in Transferable Representation Learning

What is an event?

- An Event is a specific occurrence involving participants.
- An Event is something that happens.
- An Event can frequently be described as a change of state.



Most of current NLP work focuses on this

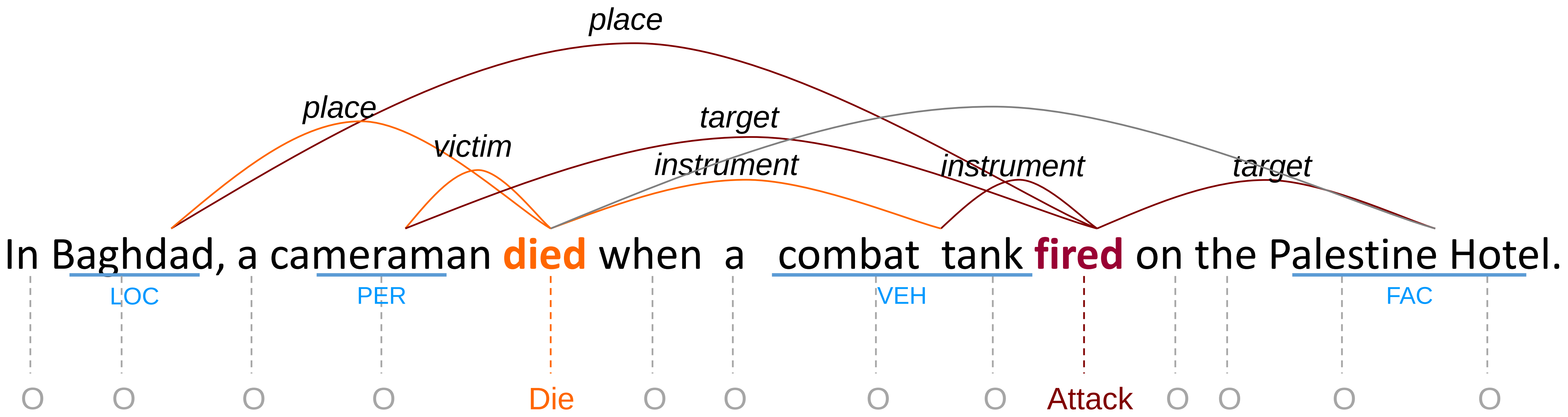
Chart from (Dölling, 2011)

Event Extraction

- Supervised Event Extraction
  - Schema-guided Event Extraction
  - Document-level Event Extraction
- Cross-domain Zero-shot Transfer for Event Extraction
- Cross-lingual Transfer for Multi-lingual Event Extraction
- Cross-media Structured Common Space for Multimedia Event Extraction

What is Information Extraction (IE)?

- Extract structured information and knowledge from unstructured data of heterogeneous data types, in various domains, genres, languages, and data modalities



- It's naturally a structure prediction task! Convert unstructured sequences to graphs

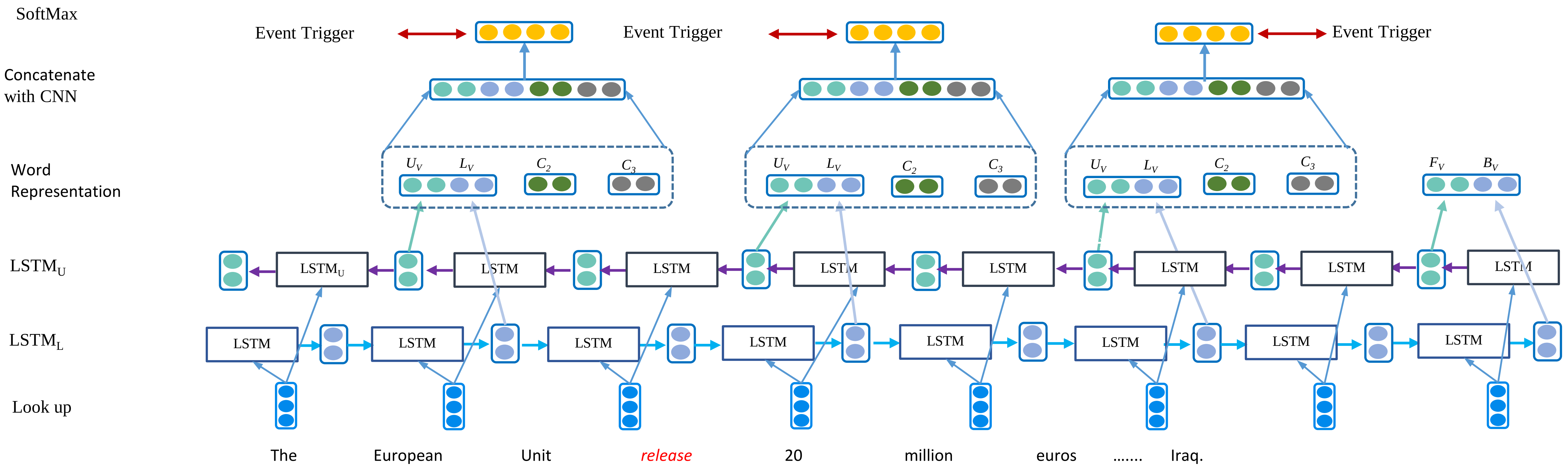
“Old” Days: Supervised Learning with Hand-crafed Features

- Trigger Labeling
  - Lexical
    - Tokens and POS tags of candidate trigger and context words
  - Dictionaries
    - Trigger list, synonym gazetteers
  - Syntactic
    - the depth of the trigger in the parse tree
    - the path from the node of the trigger to the root in the parse tree
    - the phrase structure expanded by the parent node of the trigger
    - the phrase type of the trigger
  - Entity
    - the entity type of the syntactically nearest entity to the trigger in the parse tree
    - the entity type of the physically nearest entity to the trigger in the sentence

- Argument Labeling
  - Event type and trigger
    - Trigger tokens
    - Event type and subtype
  - Entity
    - Entity type and subtype
    - Head word of the entity mention
  - Context
    - Context words of the argument candidate
  - Syntactic
    - the phrase structure expanding the parent of the trigger
    - the relative position of the entity regarding to the trigger (before or after)
    - the minimal path from the entity to the trigger
    - the shortest length from the entity to the trigger in the parse tree

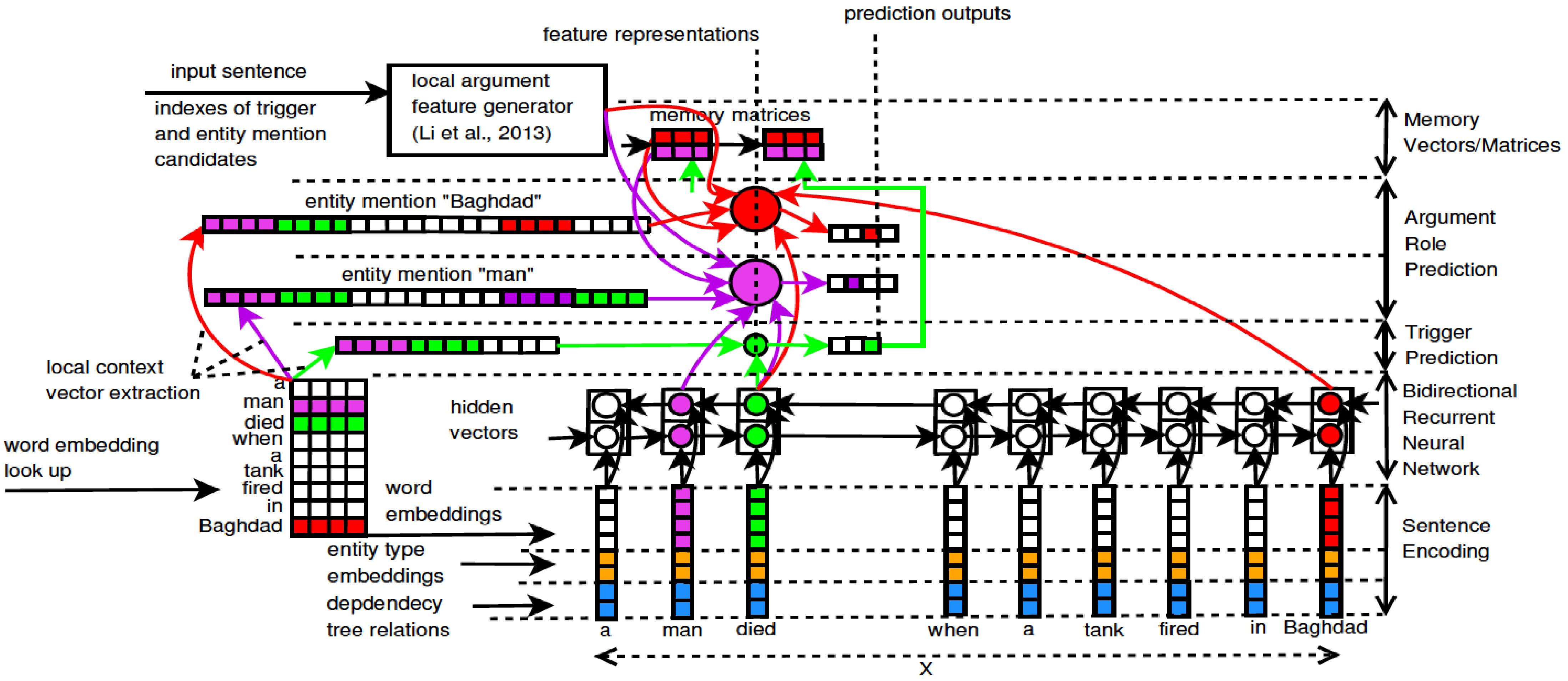
(Chen and Ji, 2009)

A More “Modern” Neural Event Extractor



- Reduce feature engineering efforts to some extent (Feng et al., 2016)
- But still rely on human annotated clean training data still fragile to noise in training data

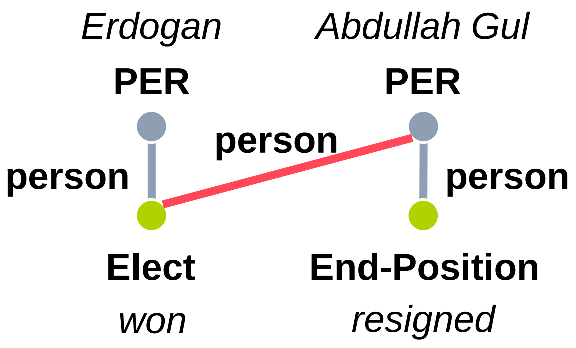
Or Put them Together…



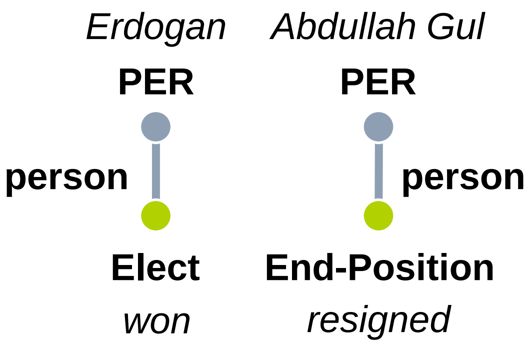
- Add symbolic features by concatenating them with embeddings (Nguyen et al., 2016)

Joint Entity, Relation and Event Extraction

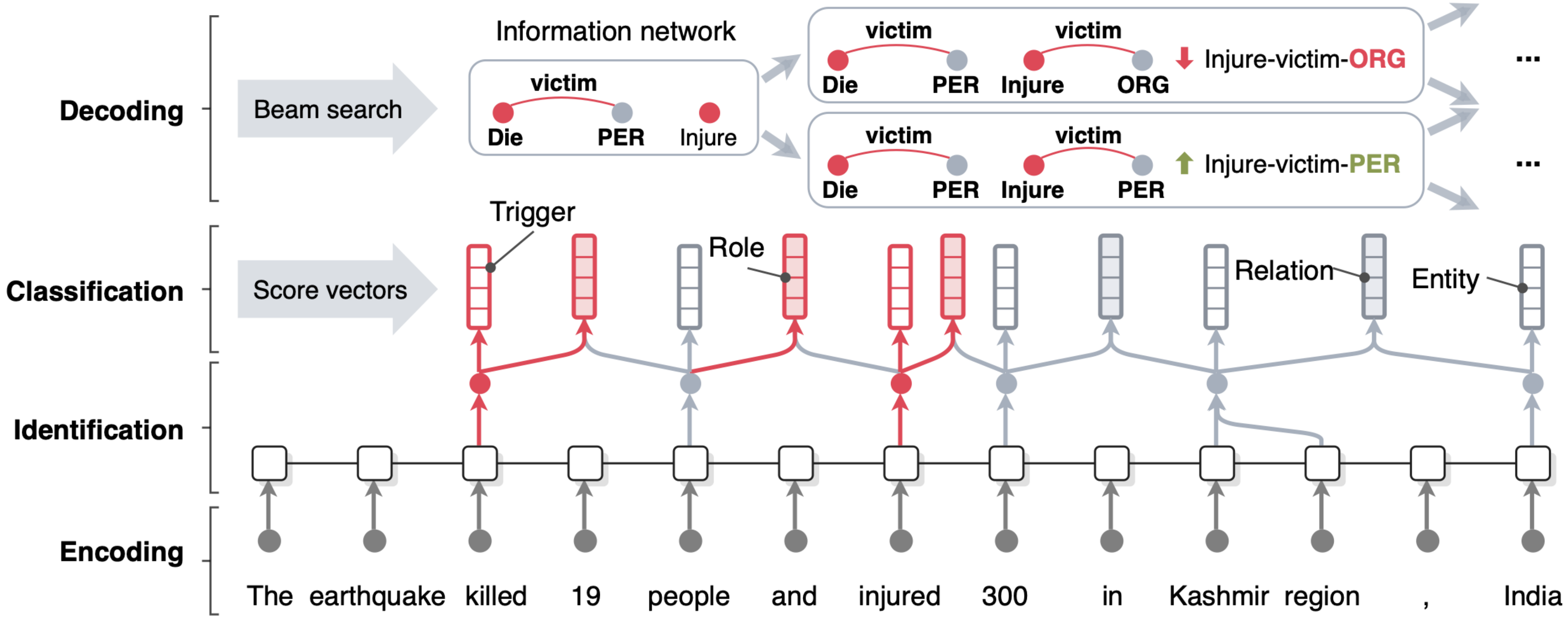
- Pipelined models suffer from the error propagation problem and disallow interactions among components
- Existing neural models do not explicitly model cross-subtask and cross-instance interactions among knowledge elements
- Example: *Prime Minister* ***Abdullah Gul*** *<u>resigned</u> earlier Tuesday to make way for* ***Erdogan****, <u>who</u> won a parliamentary seat in by-elections Sunday.*



1. An **Elect** event usually has only one **Person** argument
2. An entity is unlikely to act as a **Person** argument for **End-Position** and **Elect** events at the same time

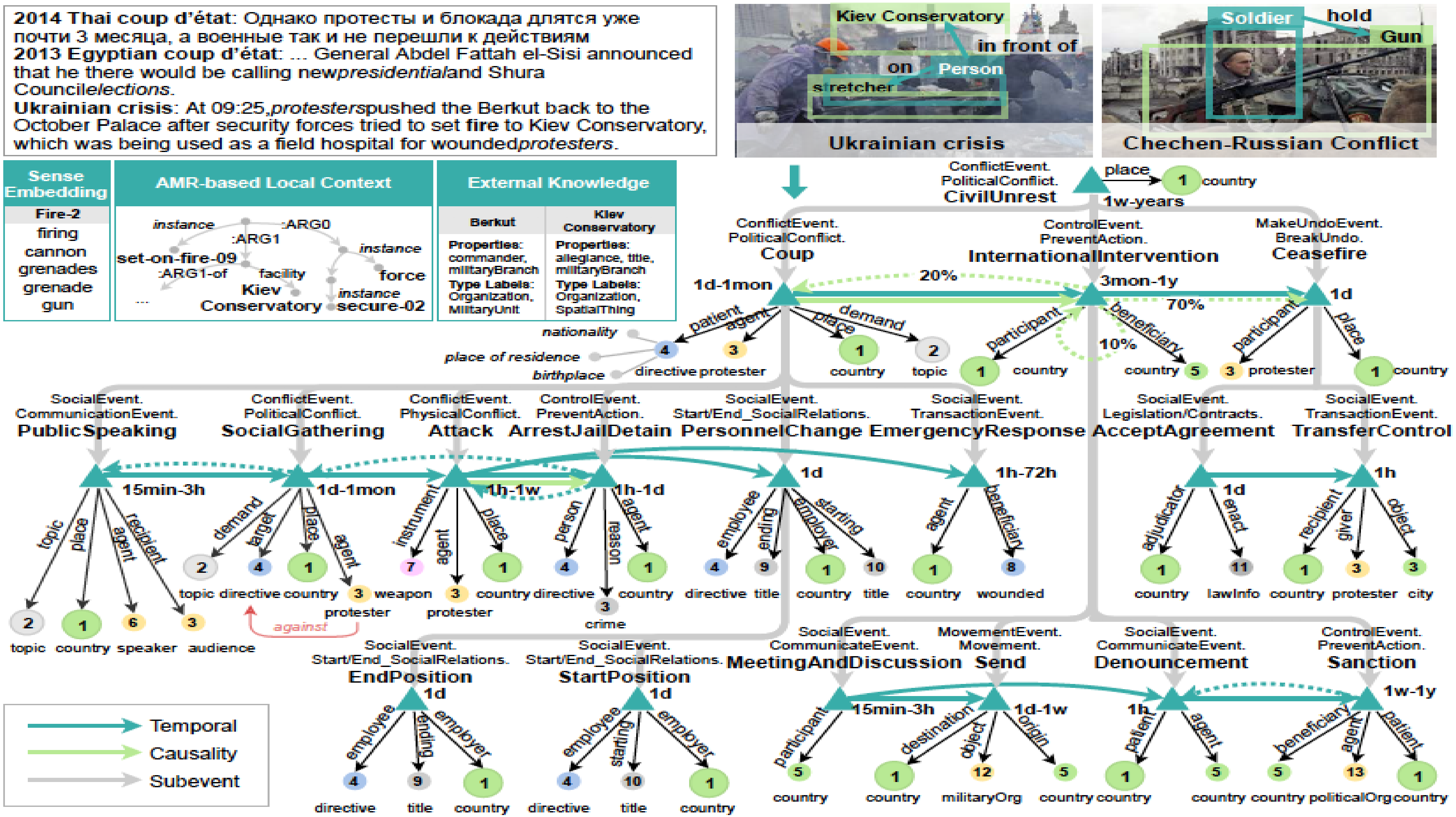


OneIE: An End-to-end Neural Model for IE (Lin et al., 2020)



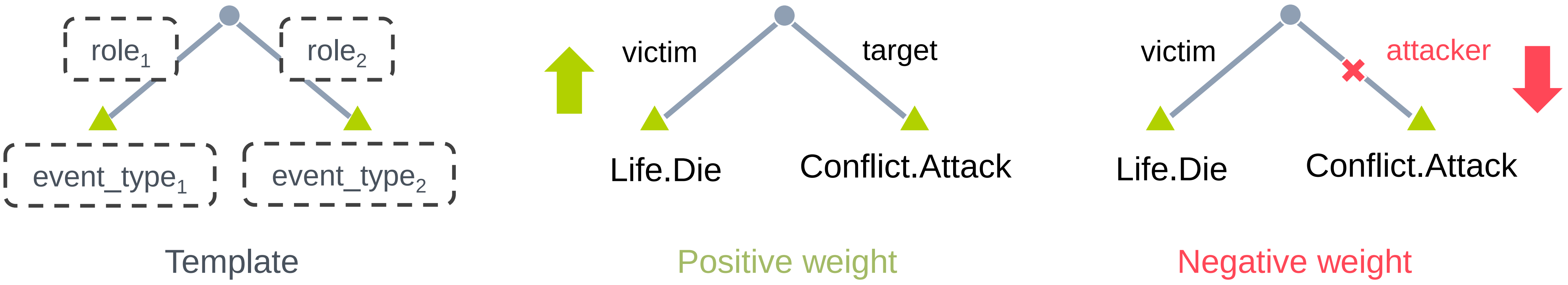
- Our OneIE framework extracts the information graph from a given sentence in four steps: encoding, identification, classification, and decoding

Move from Entity-Centric to Event-Centric NLU



Event Schema Induction

- We design a set of *global feature templates* (e.g., $event\_type_1$ – $role_1$ – $role_2$ – $event\_type_2$: an entity acts a $role_1$ argument for an $event\_type_1$ event and a $role_2$ argument for an $event\_type_2$ event in the same sentence). A more comprehensive event schema library is inducted following (Li et al, 2020).
- The model learns the *weight* of each feature during training



Incorporating Global Features

- Given a graph G, we generate its global feature vector as $f(G)$, where f is a function that evaluates a certain feature and returns a scalar. For example,

$$f_i(G) = \begin{cases} 1, & G \text{ has multiple ATTACK events} \\ 0, & \text{otherwise.} \end{cases}$$

- Next, we learn a weight vector and calculate the global feature score of as the dot production of and .
- **Global score** of a graph: local graph score + global feature score:

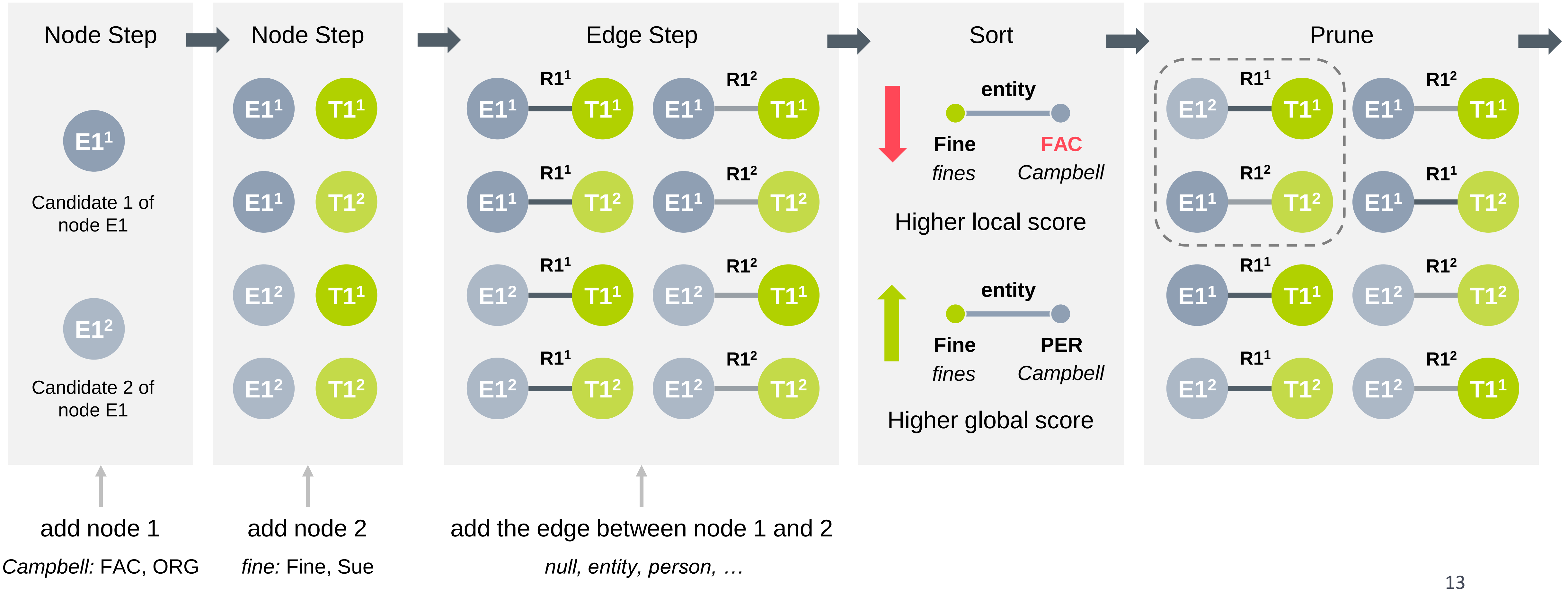
$$s(G) = s'(G) + \boldsymbol{u}\boldsymbol{f}_G$$

- We assume that the gold-standard graph for a sentence should achieve the highest global score and minimize the following loss function:

$$\mathcal{L}^{\mathrm{G}} = s(\hat{G}) - s(G)$$

Decoding

- We use beam search to decode the information graph
- Example: *He also brought a check from* ***Campbell*** *to pay the* ***fines*** *and fees.*



Experiment Results

- We conduct our experiments on ACE (Automatic Content Extraction) 2005 (F-score, %)

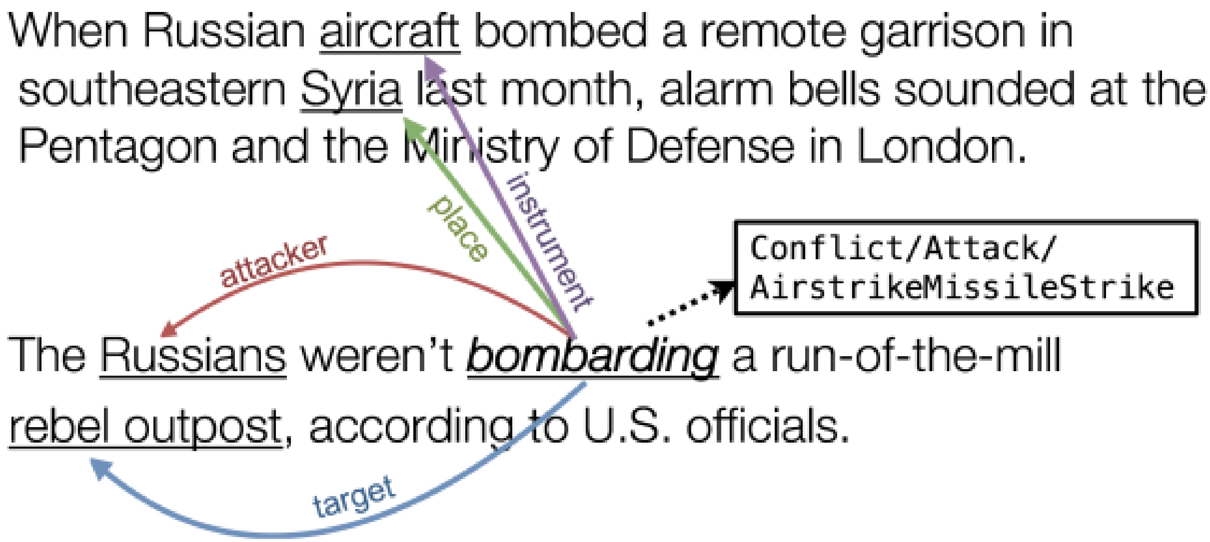
Model	ACE05-R		ACE05-E				
	Entity	Relation	Entity	Trigger Identification	Trigger Classification	Argument Identification	Argument Classification
DyGIE++	88.6	63.4	89.7	-	69.7	53.0	48.8
DyGIE++*	-	-	**90.7**	76.5	73.6	55.4	52.5
OneIE	**88.8**	**67.5**	90.2	**78.2**	**74.7**	**59.2**	**56.8**

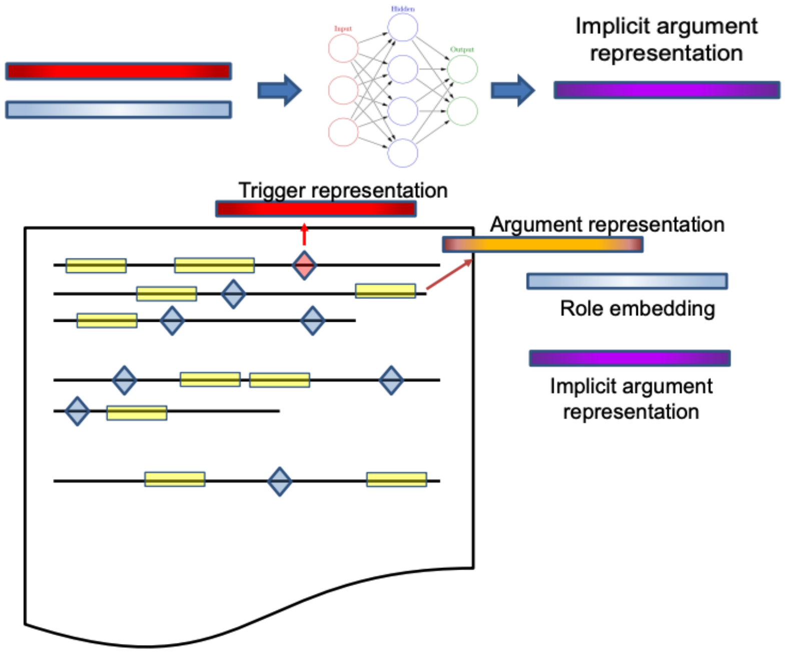
- We evaluate the portability of the proposed framework on ACE05-CN (Chinese) and ERE-ES (Spanish).

Dataset	Training	Entity	Relation	Trigger	Argument
ACE05-CN	CN	88.5	62.4	65.6	52.0
	CN+EN	89.8	62.9	67.7	53.2
ERE-ES	ES	81.3	48.1	56.8	40.3
	ES+EN	81.8	52.9	59.1	42.3

Extending from Sentence-Level to Document-Level

- Multi-Sentence Argument Linking (Ebner et al., 2020)



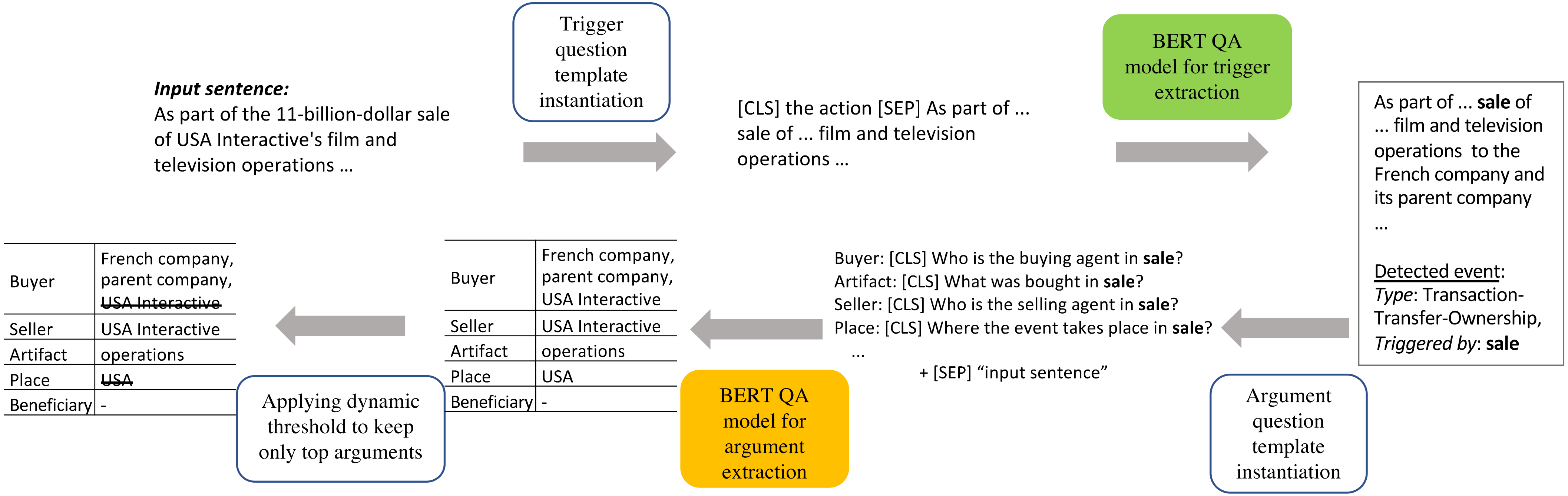


$$l(a, \tilde{a}_{e,r}) = s_{E,R}(e,r) + s_{A,R}(a,r) + s_l(a, \tilde{a}_{e,r}) + s_c(e,a), \quad a \neq \epsilon$$

- Roles are evoked by event triggers, forming implicit arguments
- Implicit arguments linked to explicit mentions in text
  - Representations: Learn span representations for each trigger span and candidate argument span
  - Prune: For each trigger, prune to top-K candidate arguments
  - Linking score: Score representations of implicit arguments with representations of explicit arguments using a decomposable scoring function

Extending from Sentence-Level to Document-Level

- Event Extraction by Answering (Almost) Natural Questions (Du and Cardie, 2020)



The input sequences for the two QA models share a standard BERT-style format

[CLS] <question> [SEP] <sentence> [SEP]

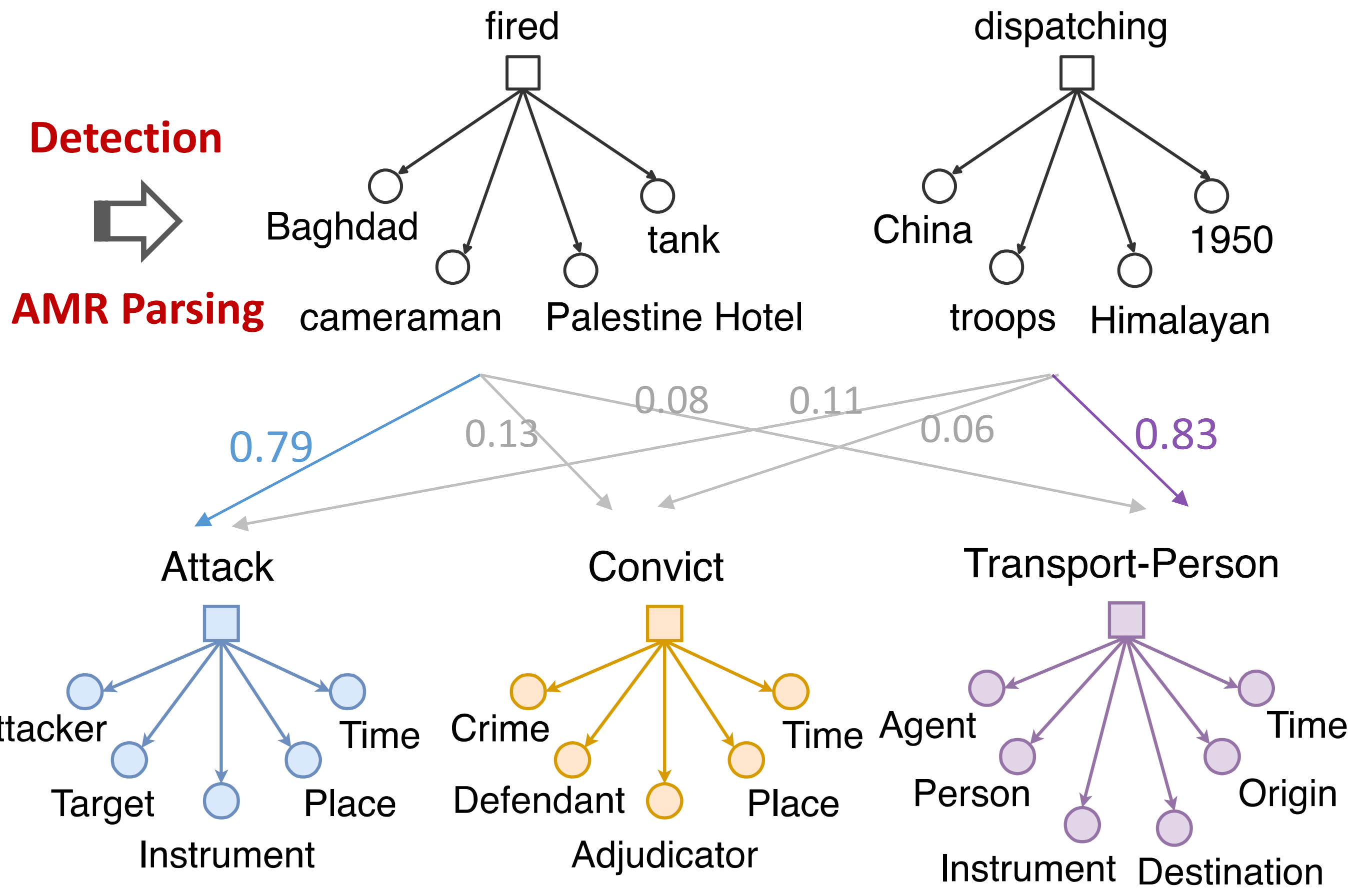
Event Extraction

- Supervised Event Extraction
  - Schema-guided Event Extraction
  - Document-level Event Extraction
- Cross-domain Zero-shot Transfer for Event Extraction
- Cross-lingual Transfer for Multi-lingual Event Extraction
- Cross-media Structured Common Space for Multimedia Event Extraction

Move to any New Types: Zero-shot Event Extraction (Huang et al., 2018)

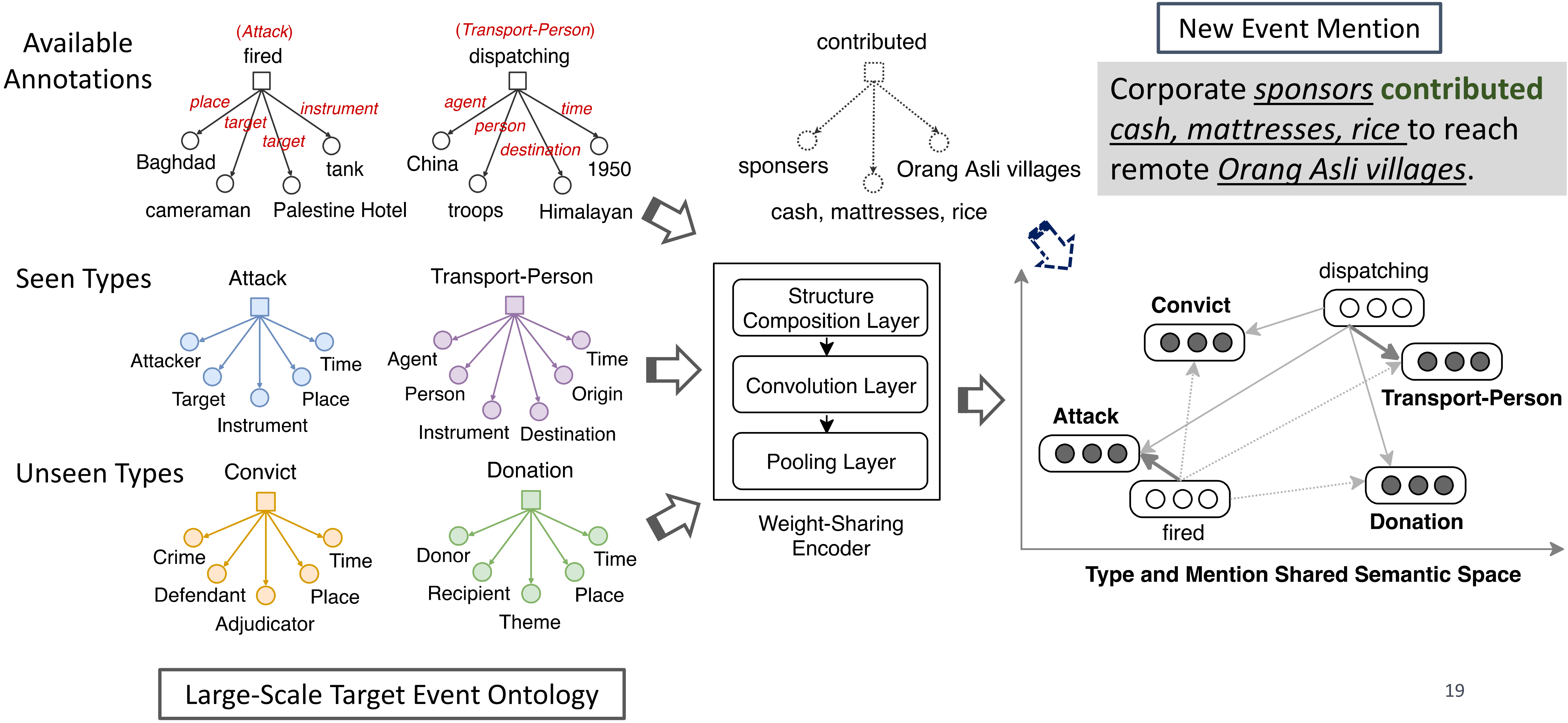
ID	Sentences
S1	In _Baghdad_, a _cameraman_ **died** when a combat _tank_ **fired** on the _Palestine Hotel_.
S2	The government of _China_ has ruled Tibet since 1951 after **dispatching** _troops_ to the _Himalayan_ region in _1950_.

Hypothesis: Both event mentions and types have rich semantics and structures, which can specify their consistency and connections

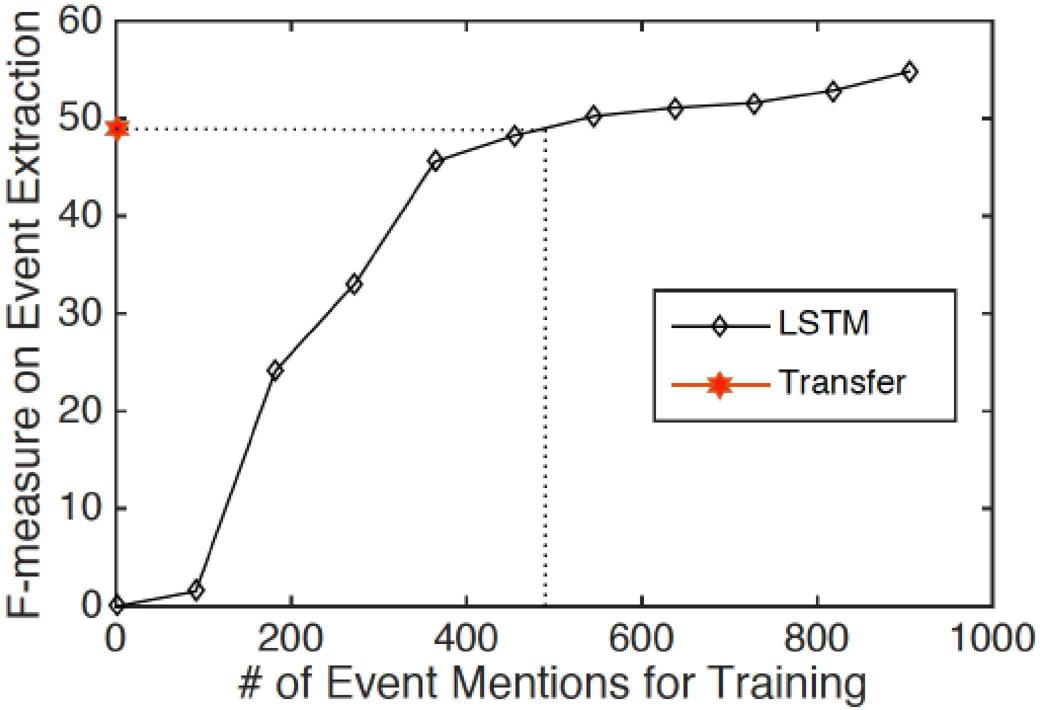


Large-Scale Target Event Ontology

Zero-shot Event Extraction



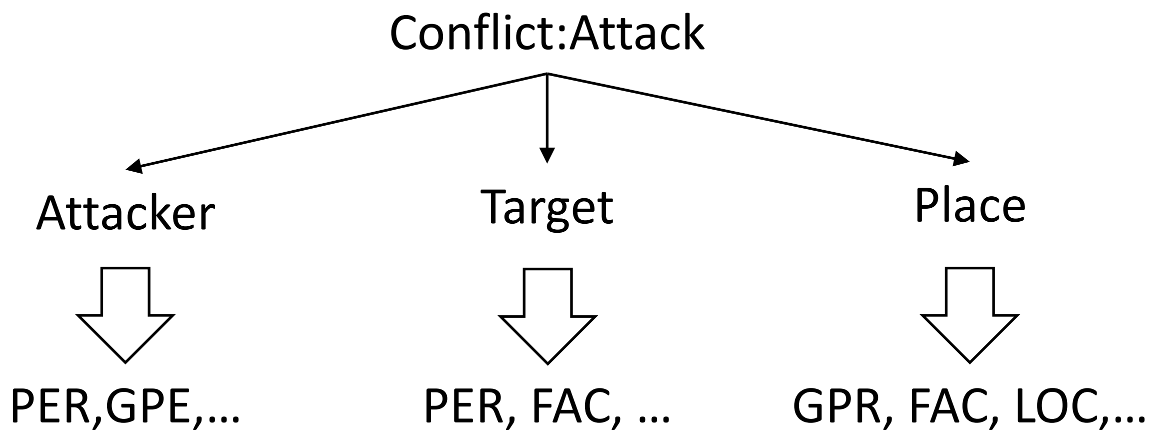
How Much Human Effort Can We Save?



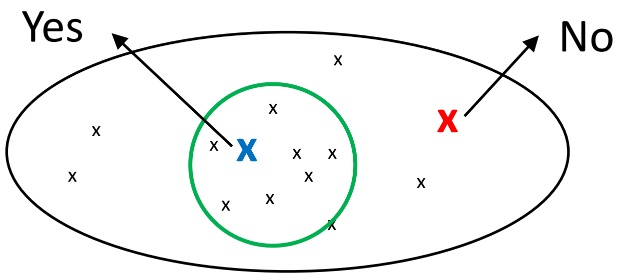
Achieved **comparable** performance as a supervised system when it's trained on **500** event mentions from **3000** sentences

- Target Event Ontology: ACE(33 types) + FrameNet (1161 frames) = 1194 types
- Seen types for training: 10 most popular ACE types
- Unseen type: 23 remaining ACE types

Label-aware Classification (Zhang et al., 2020)

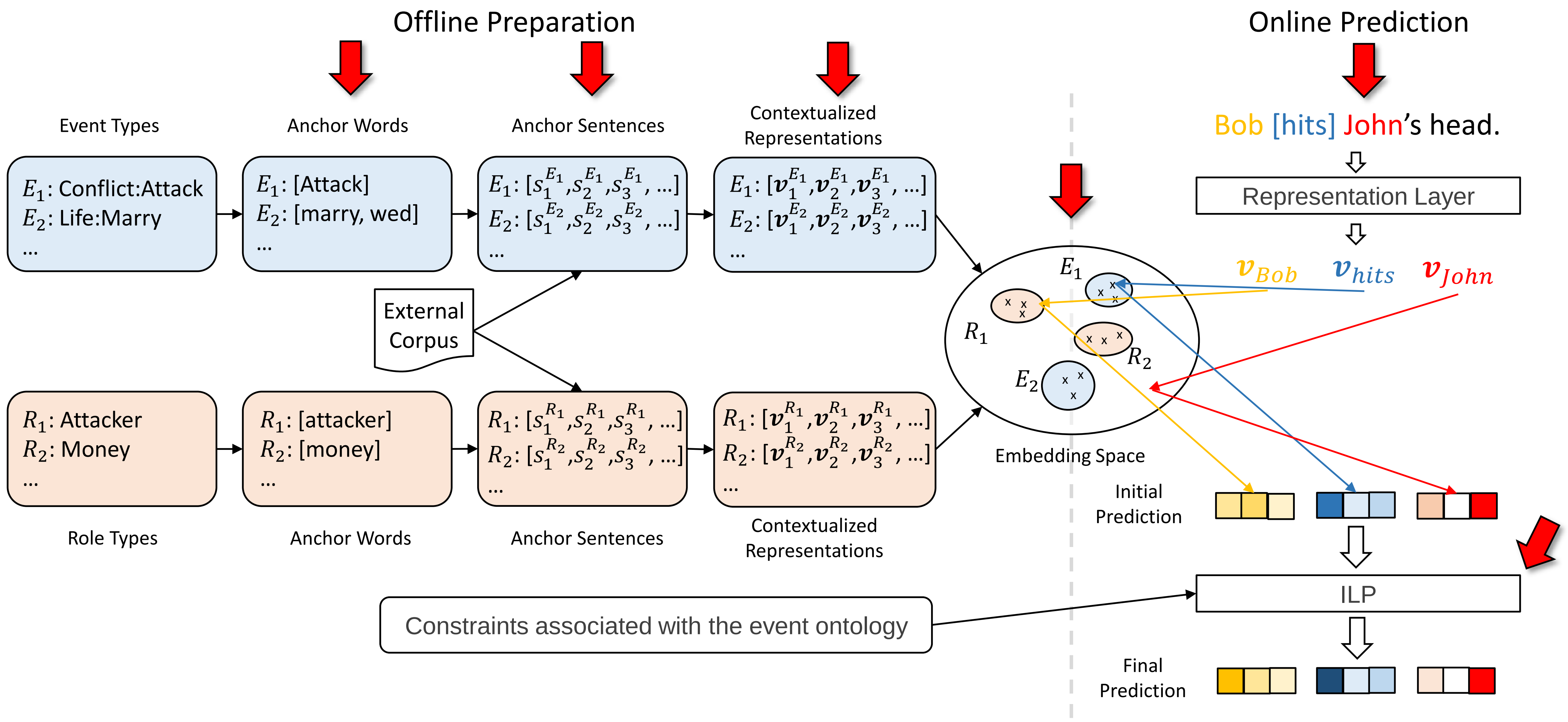


- Label semantics
  - We select “attack” as the label because we assume that it can represent the overall meaning of this event type.
- Constraints
  - “Attacker” can only be the argument of “Conflict:Attack” rather than “Life:Marry”.



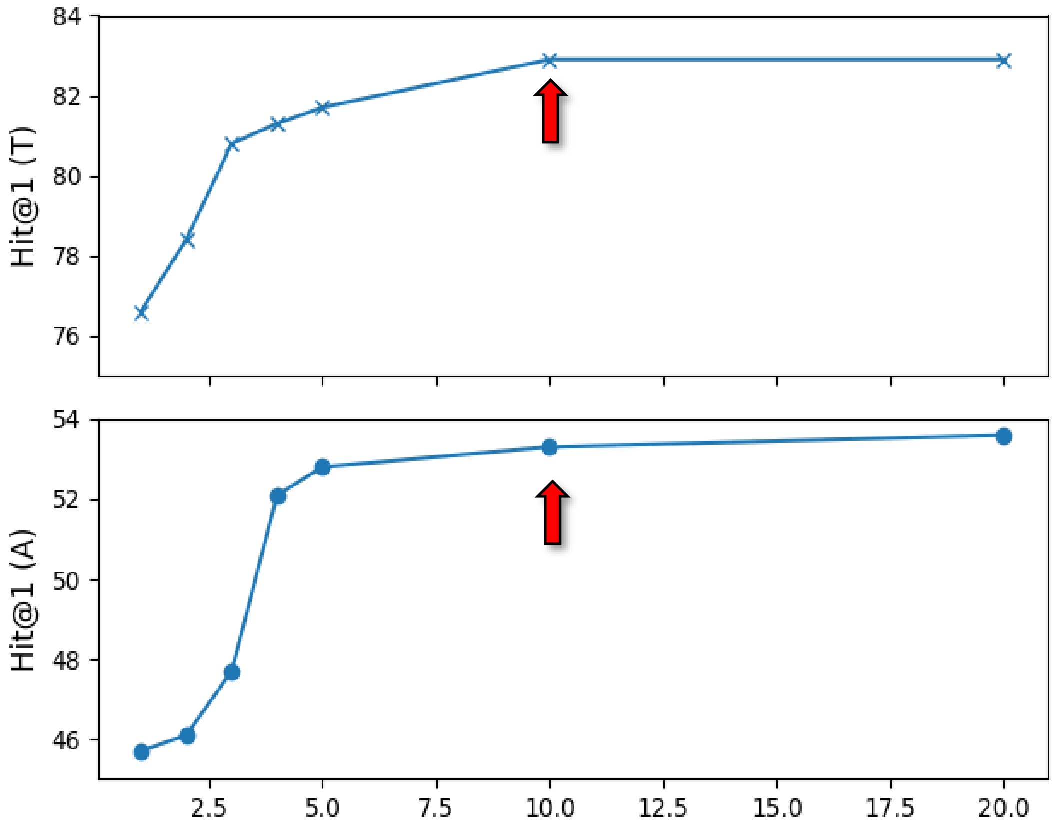
Use a cluster of contextualized embeddings to represent labels and use constraints to regularize the predictions by modeling it as an ILP problem.

The Proposed Framework



How many anchor sentences do we need?

	Model	Train types	Test types	Trig Hit@1	Trig Hit@3	Trig Hit@5	Arg Hit@1	Arg Hit@3	Arg Hit@5
Unseen types only	Frequency	0	23	9.6	27.2	42.5	25.9	63.4	80.6
	WSD	0	23	1.7	13.0	22.8	2.4	2.8	2.8
	Transfer-learning (A)	1	23	4.0	23.8	32.5	1.3	3.4	3.6
	Transfer-learning (B)	3	23	7.0	12.5	36.8	3.5	6.0	6.3
	Transfer-learning (C)	5	23	20.1	34.7	46.5	9.6	14.7	15.7
	Transfer-learning (D)	10	23	33.5	51.4	68.3	14.7	26.5	27.7
	Our Approach	0	23	**80.5**	**88.9**	**93.2**	**68.5**	**94.2**	**96.8**
Entire dataset	Frequency	0	33	28.9	53.6	62.7	13.8	33.8	51.0
	Our Approach	0	33	**82.9**	**93.1**	**96.2**	**53.6**	**87.9**	**92.4**



Ten sentences are good enough!!

Event Extraction

- Supervised Event Extraction
  - Schema-guided Event Extraction
  - Document-level Event Extraction
- Cross-domain Zero-shot Transfer for Event Extraction
- ➢ Cross-lingual Transfer for Multi-lingual Event Extraction
- Cross-media Structured Common Space for Multimedia Event Extraction

Cross-lingual Joint Entity and Word Embedding Learning

- Cross-lingual Joint Entity and Word Embedding to Improve Entity Linking and Parallel Sentence Mining (Pan et al., 2019)
  - Code-switch cross-lingual entity/word data generation

Example Chinese Wikipedia Sentence:

[[小米科技|小米]] 被 誉为 中国的 [[苹果公司|苹果]] 。

link ↓ langlink link ↓ langlink

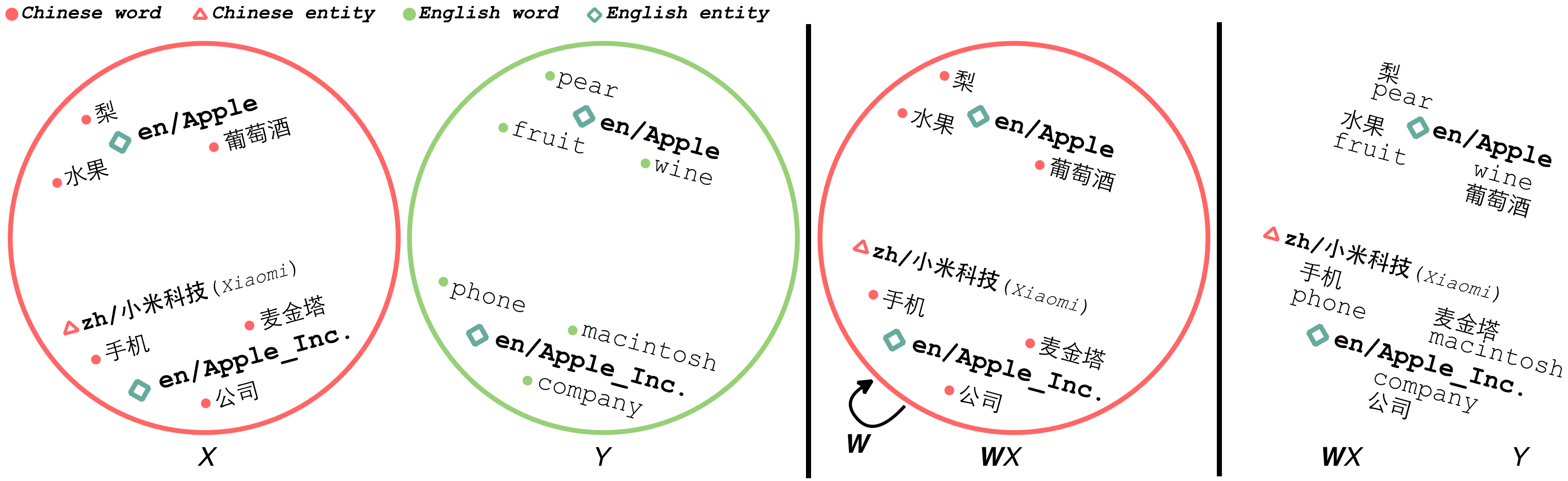
zh/小米科技 ~~→~~ zh/苹果公司 → en/Apple_Inc.

Our Approach:

zh/小米科技 *被 誉为 中国的* en/Apple_Inc. 。

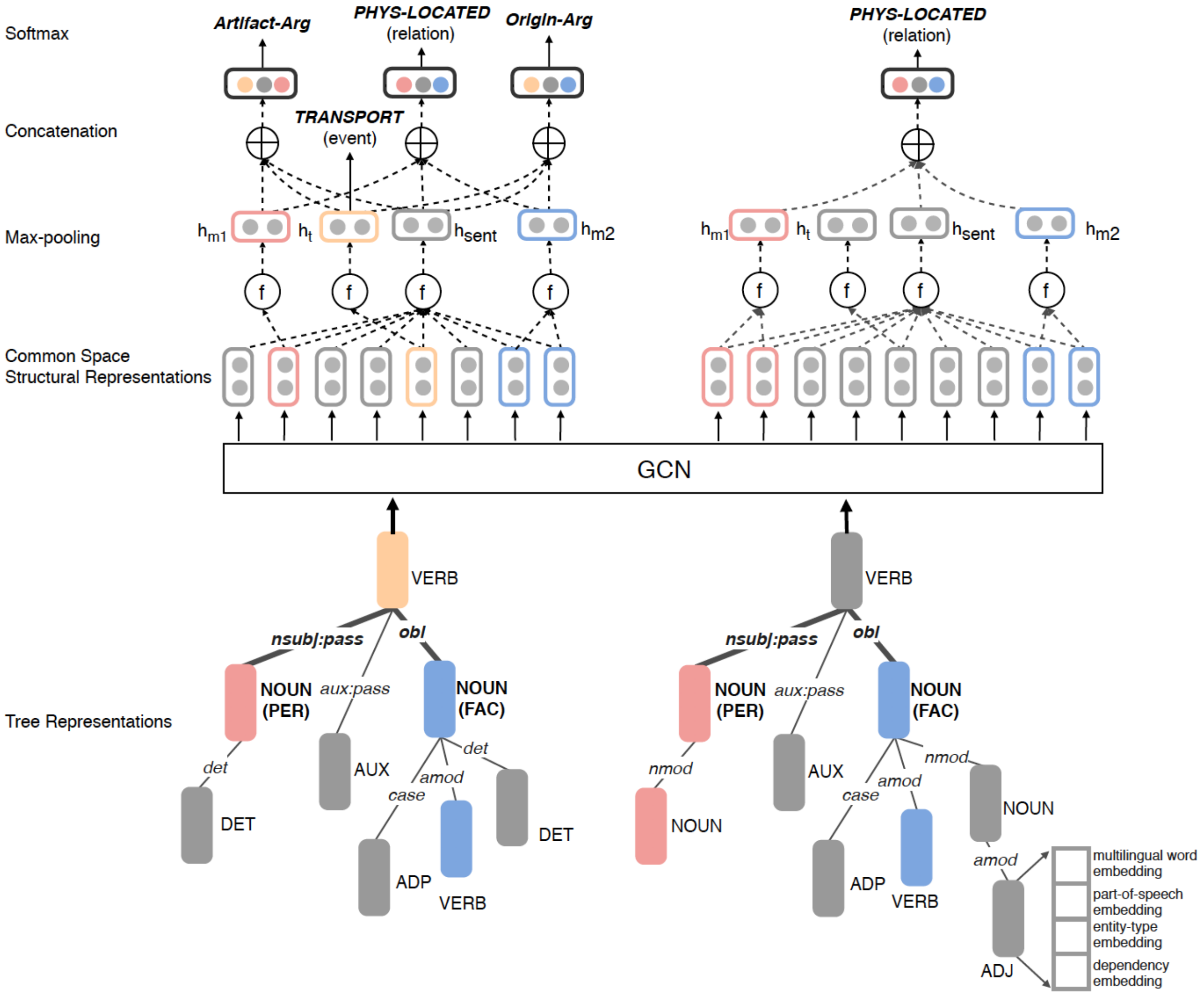
(*Xiaomi*) (*is*) (*known as*) (*Chinese*)

  - Use English entities as anchor points to learn a mapping (rotation matrix) *W* which aligns distributions in IL and English



Cross-lingual Structure Transfer Event Extraction

- Cross-lingual Structure Transfer for Relation and Event Extraction (Subburathinam et al., 2019)



The detainees were taken to a processing center

Команды врачей были замечены в упакованных отделениях скорой помощи
(teams of doctors were seen in packed emergency rooms)

Graph Convolutional Networks (GCN) Encoder

- Extend the monolingual design (Zhang et al., 2018) to cross-lingual
  - Convert a sentence with N tokens into N*N adjacency matrix *A*
  - Node: token, each edge is a directed dependency edge
- Initialization of each node's representation

$$\boldsymbol{h}_i^{(0)} = \underbrace{\boldsymbol{x}_i^w}_{\text{Word embedding}} \oplus \underbrace{\boldsymbol{x}_i^p}_{\text{POS tag}} \oplus \underbrace{\boldsymbol{x}_i^d}_{\text{Dependency relation}} \oplus \underbrace{\boldsymbol{x}_i^e}_{\text{Entity type}}$$

- At the k^th layer, derive the hidden representation of each node from the representations of its neighbors at previous layer

$$\boldsymbol{h}_i^{(k)} = \mathrm{ReLU}\left(\sum_{j=0}^{N} \frac{\boldsymbol{A}_{ij}\boldsymbol{W}^{(k)}\boldsymbol{h}_j^{(k-1)}}{d_i + b^{(k)}}\right)$$

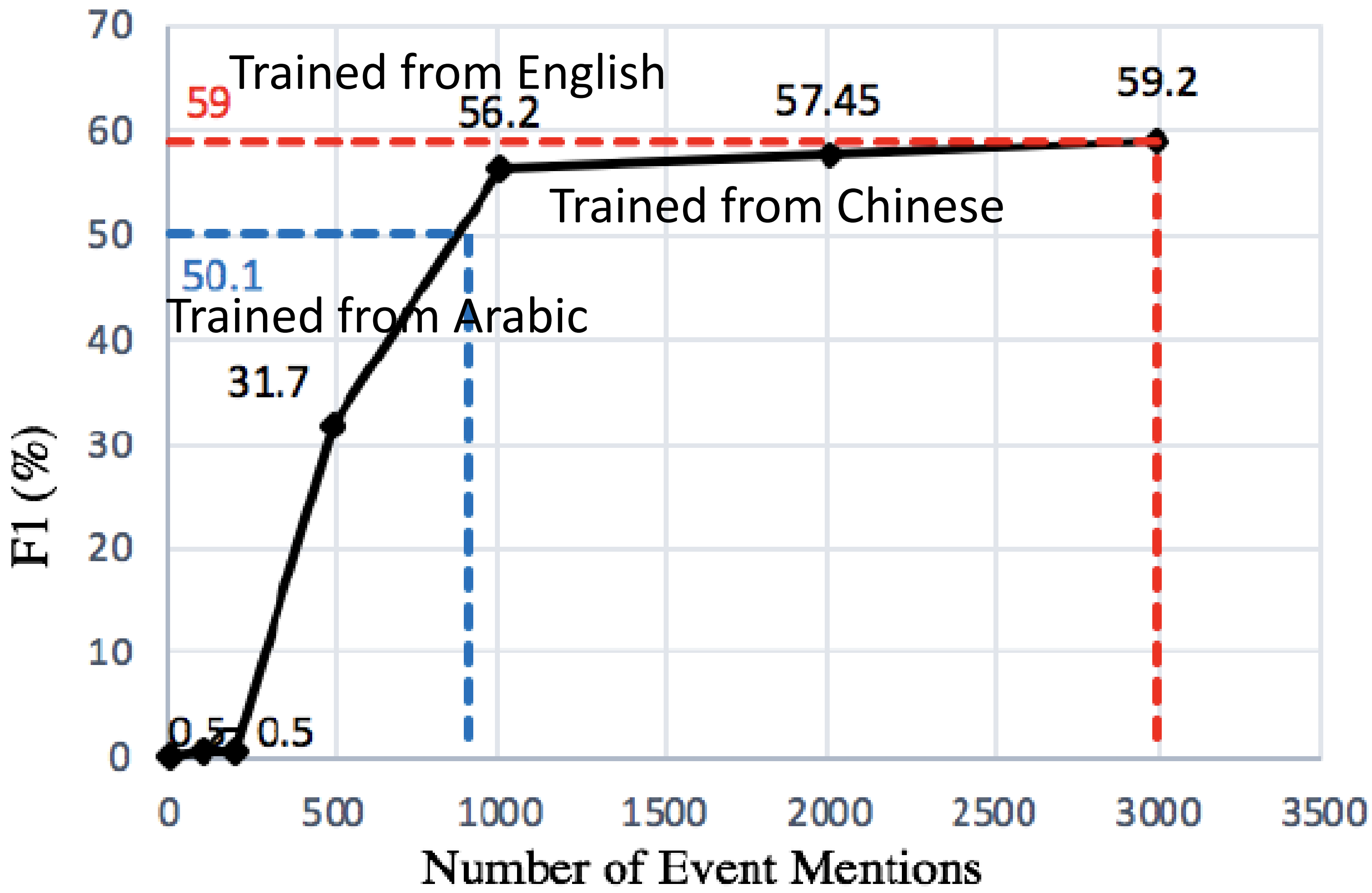
Application on Event Argument Extraction

- Task: Classify each pair of event trigger and entity mentions into one of pre-defined event argument roles or NONE
- Max-pooling over the final node representations to obtain representations for sentence, trigger and argument candidate, and concatenate them
- A softmax output layer for argument role labeling

$$L^a = \sum_{i=1}^{N} \sum_{j=1}^{L_i} y_{ij} \log(\sigma(\boldsymbol{U}^a \cdot [\boldsymbol{h}_i^t; \boldsymbol{h}_{ij}^s; \boldsymbol{h}_j^a]))$$

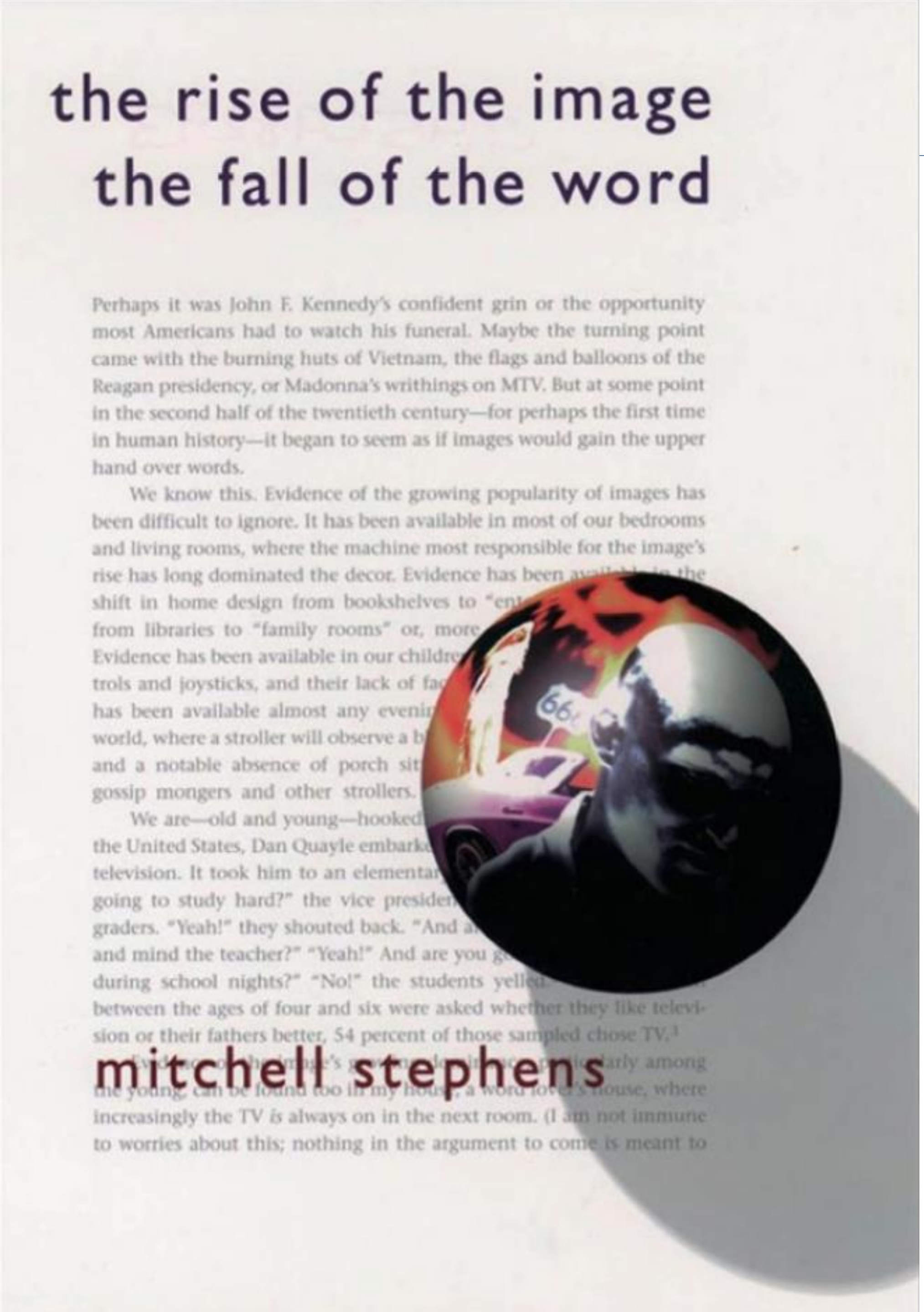
Cross-lingual Event Transfer Performance

- Chinese Event Argument Extraction (Subburathinam et al., EMNLP2019)



Event Extraction

- Supervised Event Extraction
  - Schema-guided Event Extraction
  - Document-level Event Extraction
- Cross-domain Zero-shot Transfer for Event Extraction
- Cross-lingual Transfer for Multi-lingual Event Extraction
- ➢ Cross-media Structured Common Space for Multimedia Event Extraction



Knowledge is Beyond Just Text

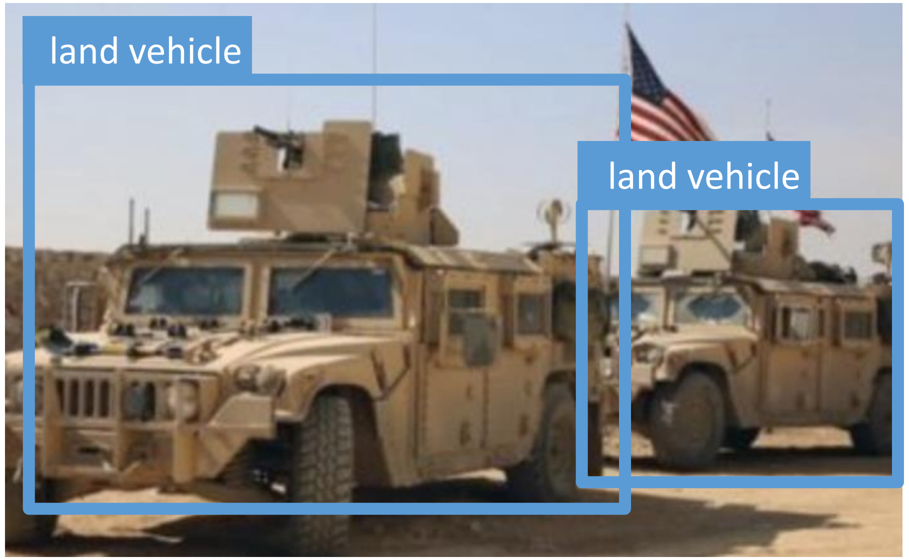
- Multimedia Event Extraction (Li et al., ACL2020)
- We produce and consume news content through multimedia, 33% of news images contain event arguments not mentioned in surrounding texts

TransportPerson_Instrument **= stretcher**

A New Task: Multimedia Event Extraction (M^2E^2)

Input: News Article Text and Image

Last week , U.S . Secretary of State Rex Tillerson visited Ankara, the first senior administration official to visit Turkey, to try to seal a deal about the battle for Raqqa and to overcome President Recep Tayyip Erdogan's strong objections to Washington's backing of the Kurdish Democratic Union Party (PYD) militias. Turkish forces have attacked SDF forces in the past around Manbij, west of Raqqa, forcing the **United States** to **deploy** dozens of **soldiers** on the **outskirts** of the town in a mission to prevent a repeat of clashes, which risk derailing an assault on Raqqa.



Output: Events & Argument Roles

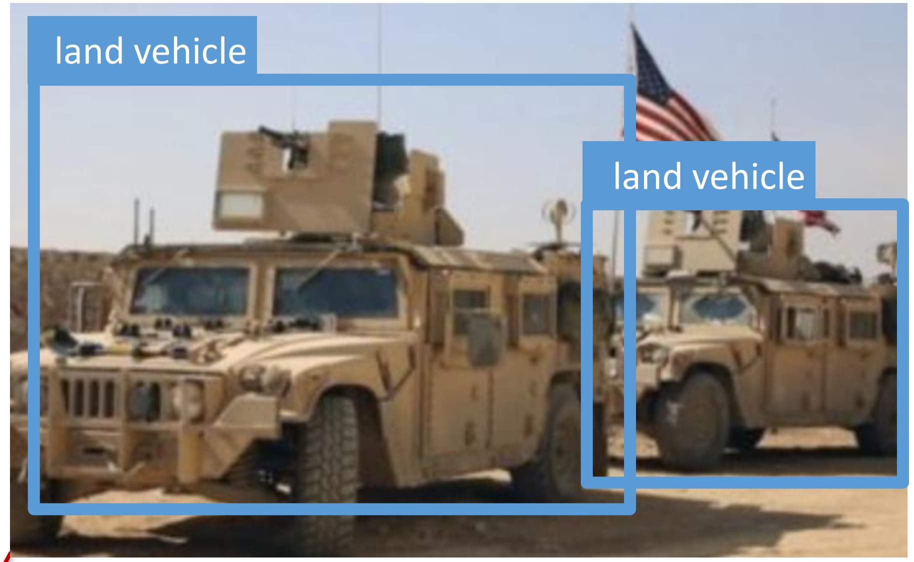
Event Type	Movement.Transport	
Event	**Text Trigger**	deploy
	Image	

Arguments	**Agent**	United States
	Destination	outskirts
	Artifact	soldiers
	Vehicle	
	Vehicle	

A New Task: Multimedia Event Extraction (M^2E^2)

Input: News Article Text and Image

Last week , U.S . Secretary of State Rex Tillerson visited Ankara, the first senior administration official to visit Turkey, to try to seal a deal about the battle for Raqqa and to overcome President Recep Tayyip Erdogan's strong objections to Washington's backing of the Kurdish Democratic Union Party (PYD) militias. Turkish forces have attacked SDF forces in the past around Manbij, west of Raqqa, forcing the **United States** to **deploy** dozens of **soldiers** on the **outskirts** of the town in a mission to prevent a repeat of clashes, which risk derailing an assault on Raqqa.



Output: Multimedia Events & Argument Roles

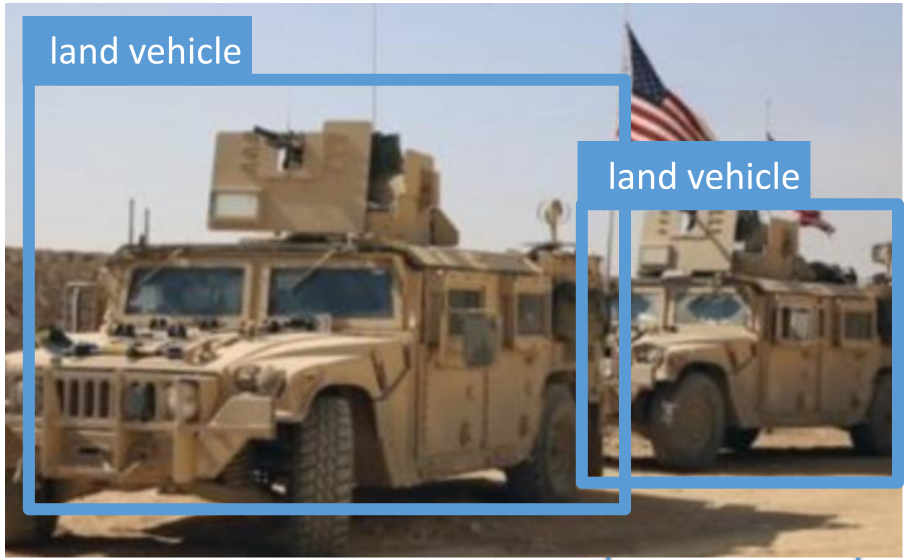
Event Type	Movement.Transport	
Event	**Text Trigger**	deploy
	Image	

Arguments	**Agent**	United States
	Destination	outskirts
	Artifact	soldiers
	Vehicle	
	Vehicle	

A New Task: Multimedia Event Extraction (M^2E^2)

Input: News Article Text and Image

Last week , U.S . Secretary of State Rex Tillerson visited Ankara, the first senior administration official to visit Turkey, to try to seal a deal about the battle for Raqqa and to overcome President Recep Tayyip Erdogan's strong objections to Washington's backing of the Kurdish Democratic Union Party (PYD) militias. Turkish forces have attacked SDF forces in the past around Manbij, west of Raqqa, forcing the **United States** to **deploy** dozens of **soldiers** on the **outskirts** of the town in a mission to prevent a repeat of clashes, which risk derailing an assault on Raqqa.



Output: Multimedia Events & Argument Roles

Event Type	Movement.Transport	
Event	**Text Trigger**	deploy
	Image	

Arguments	**Agent**	United States
	Destination	outskirts
	Artifact	soldiers
	Vehicle	
	Vehicle	

Vision vs. NLP for Event Extraction

- Vision does not study newsworthy, complex events
  - Focusing on daily life and sports (Perera et al., 2012; Chang et al., 2016; Zhang et al., 2007; Ma et al., 2017)
  - Without localizing a complete set of arguments for each event (Gu et al., 2018; Li et al., 2018; Duarte et al., 2018; Sigurdsson et al., 2016; Kato et al., 2018; Wu et al., 2019a)
- Most related: Situation Recognition (Yatskar et al., 2016)
  - Classify an image as one of 500+ FrameNet verbs
  - Identify 192 generic semantic roles via a 1-word description

CLIPPING

ROLE	VALUE	ROLE	VALUE
AGENT	MAN	AGENT	VET
SOURCE	SHEEP	SOURCE	DOG
TOOL	SHEARS	TOOL	CLIPPER
ITEM	WOOL	ITEM	CLAW
PLACE	FIELD	PLACE	ROOM

JUMPING

ROLE	VALUE	ROLE	VALUE
AGENT	BOY	AGENT	BEAR
SOURCE	CLIFF	SOURCE	ICEBERG
OBSTACLE	-	OBSTACLE	WATER
DESTINATION	WATER	DESTINATION	ICEBERG
PLACE	LAKE	PLACE	OUTDOOR

SPRAYING

ROLE	VALUE	ROLE	VALUE
AGENT	MAN	AGENT	FIREMAN
SOURCE	SPRAY CAN	SOURCE	HOSE
SUBSTANCE	PAINT	SUBSTANCE	WATER
DESTINATION	WALL	DESTINATION	FIRE
PLACE	ALLEYWAY	PLACE	OUTSIDE

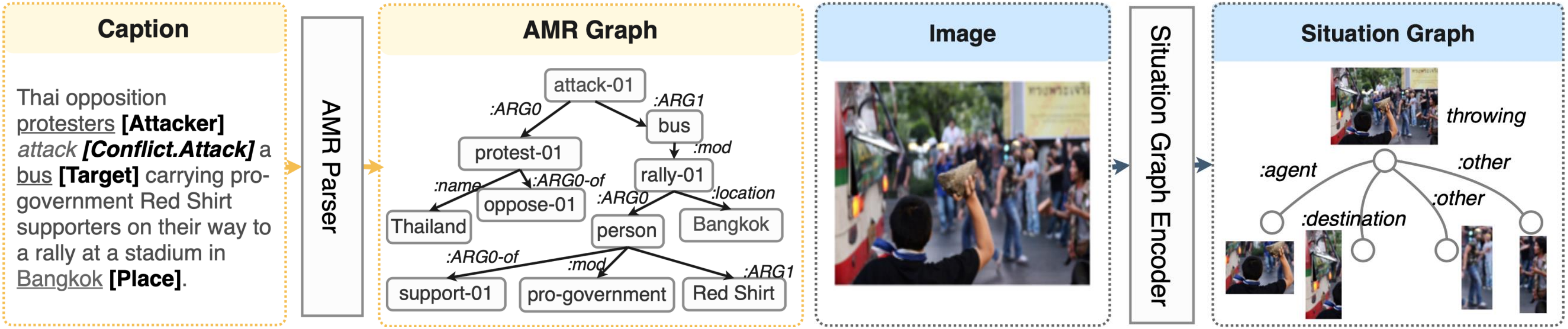
Cross-media Structured Common Space

- Treat Image/Video as a foreign language

Text	Image / Video Frame
Word	Image Region
Entity	Visual Object
Relation	Visual Relation
Entity-Relation Graph	Visual Scene Graph
Event Trigger	Visual Activity
Linguistic Structure	**Situation Graph**

Cross-media Structured Common Space

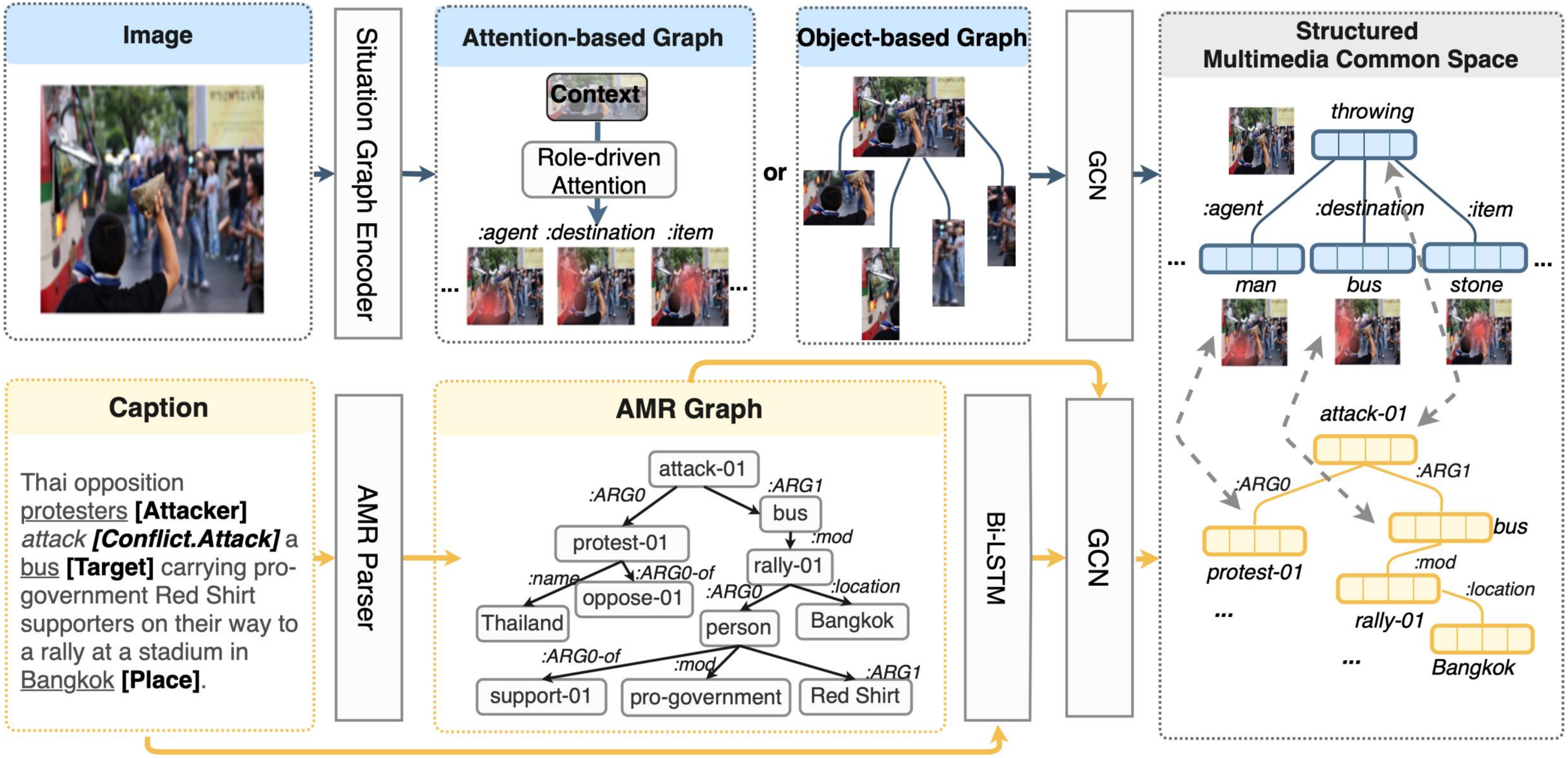
- Treat Image/Video as a foreign language
  - Represent it with a structure that is similar to AMR graph in text



Linguistic Structure, e.g., Dependency Tree Abstract Meaning Representation (AMR)	Situation Graph

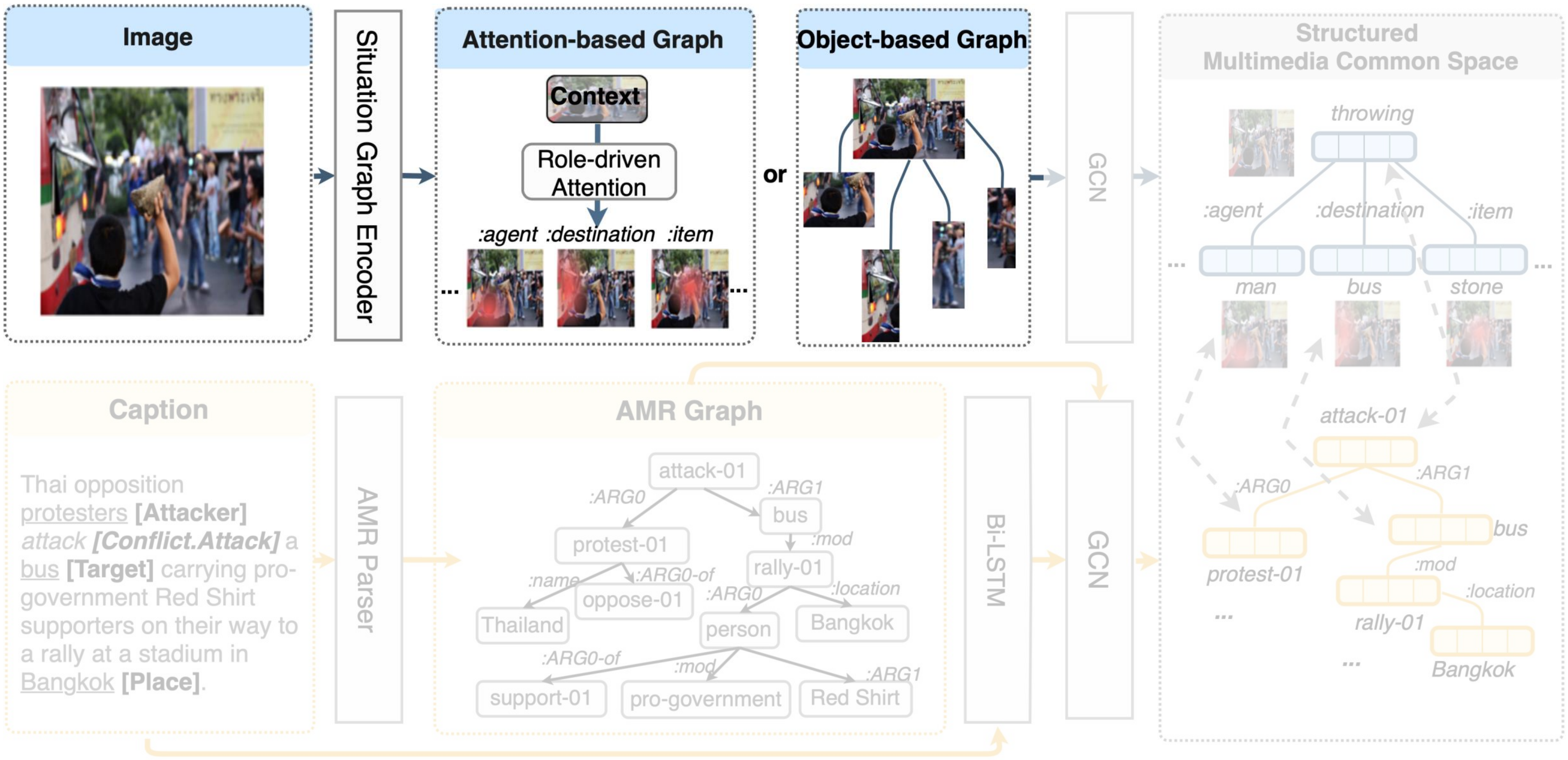
Weakly Aligned Structured Embedding (WASE)

-- Training Phase (Common Space Construction)



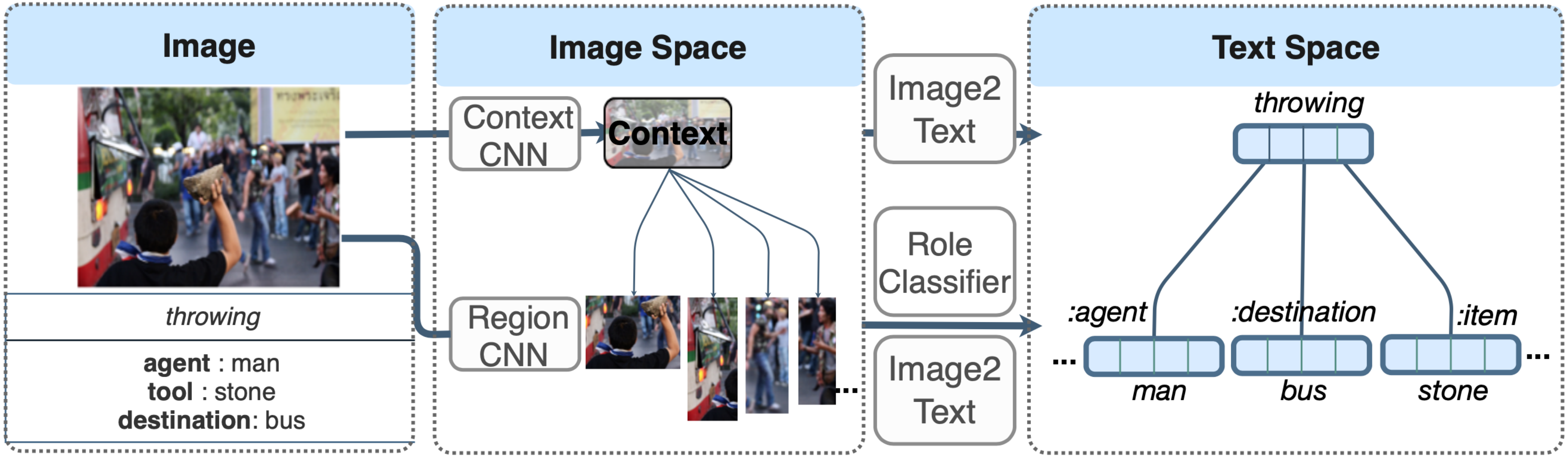
Weakly Aligned Structured Embedding (WASE)

-- Training Phase (Common Space Construction)



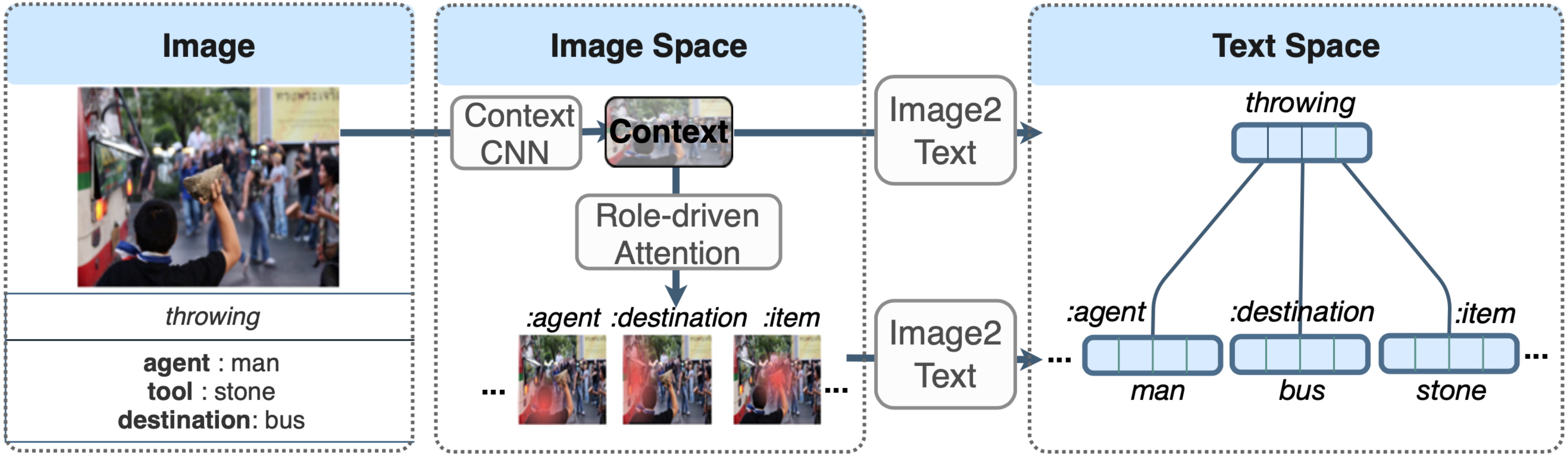
How to generate situation graph?

- Method 1: Object-based Graph Training
  - Learn to project image to verb embedding, and object to noun
  - Learn to classify each object-image pair to a semantic role



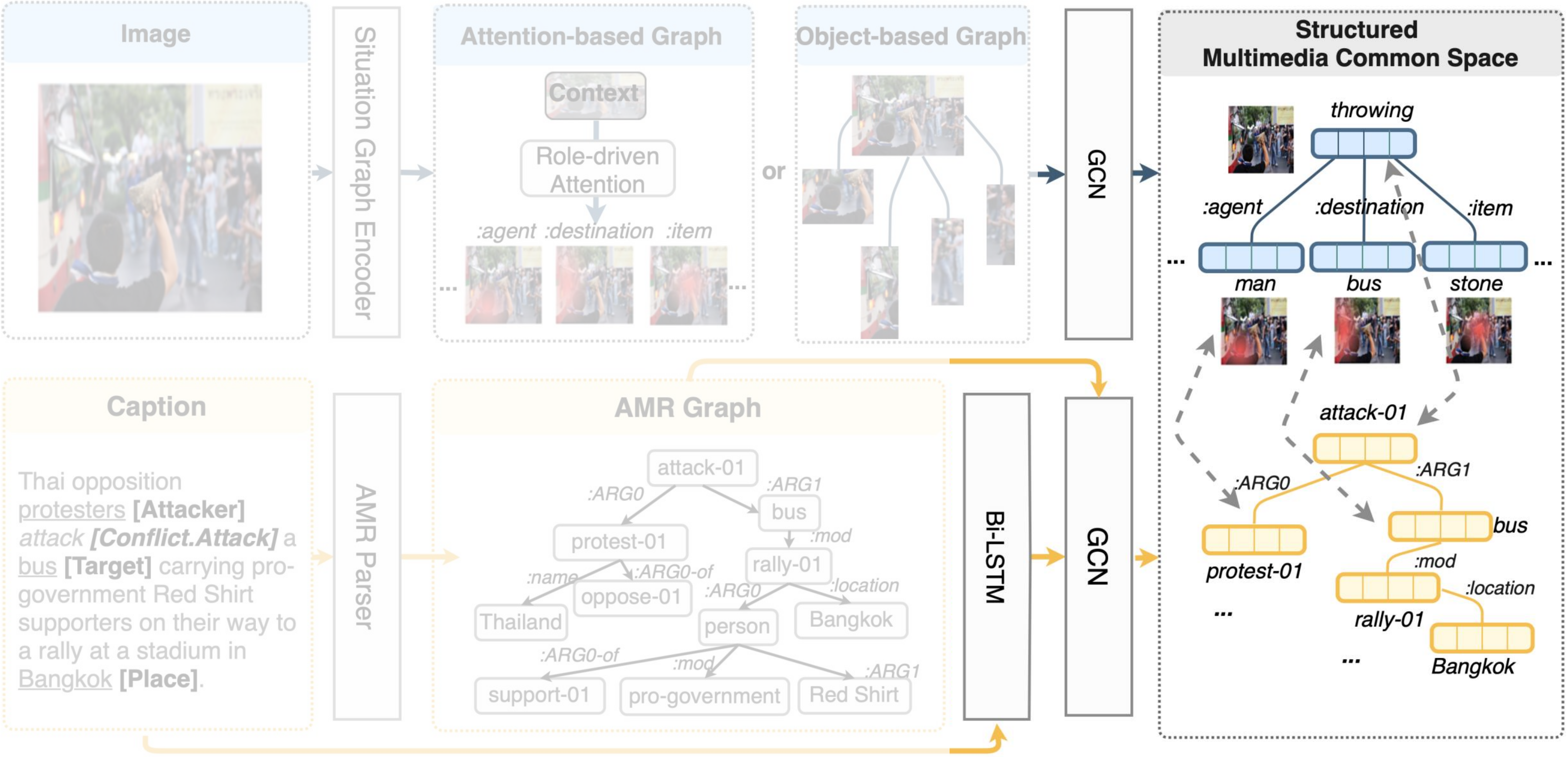
How to generate situation graph?

- Method 2: Role-driven Attention Graph
  - Learn to project image embedding to verb embedding
  - Learn a spatial attention on image for each role
  - Learn to project attended role region to noun embedding



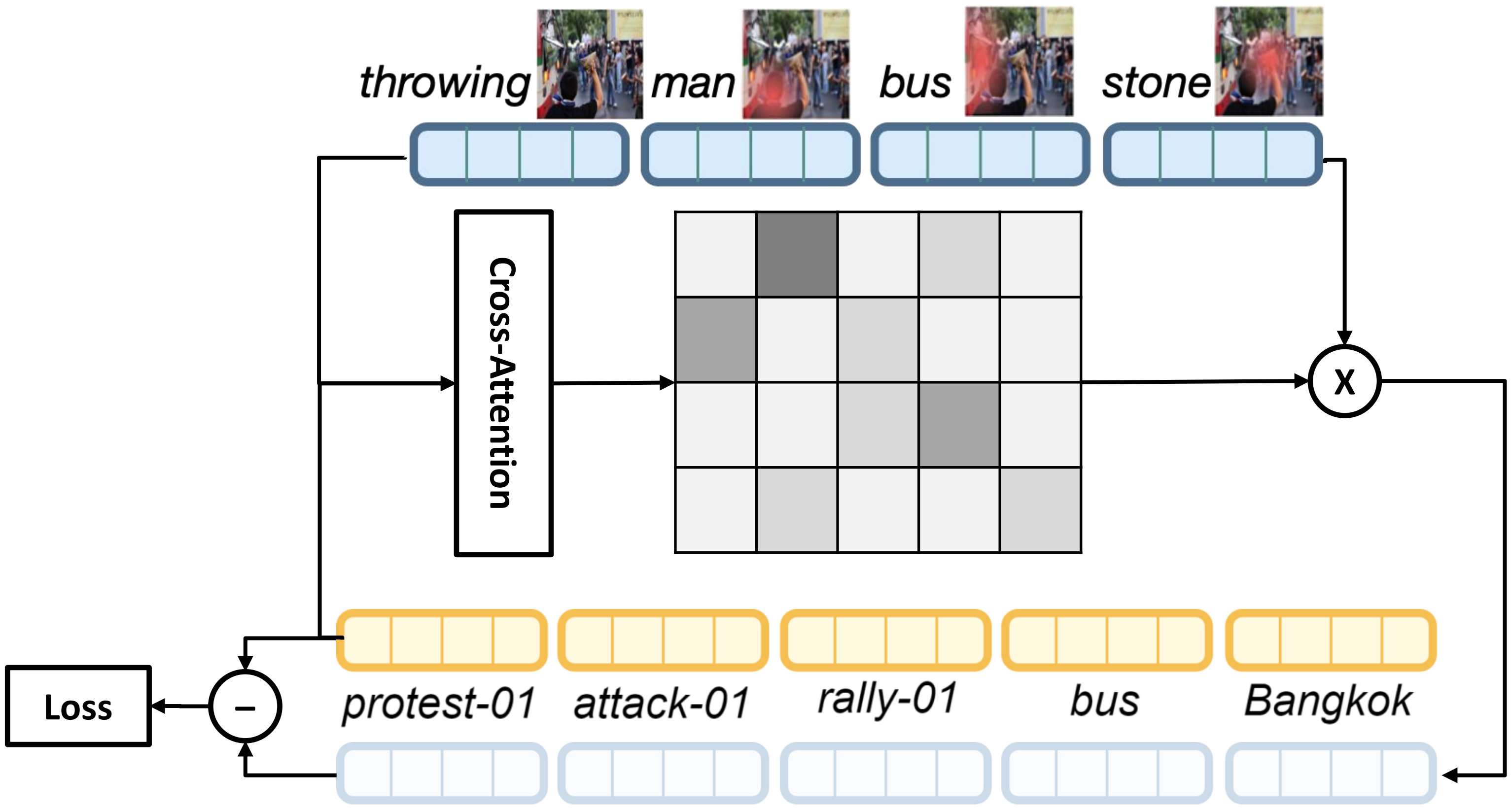
Weakly Aligned Structured Embedding (WASE)

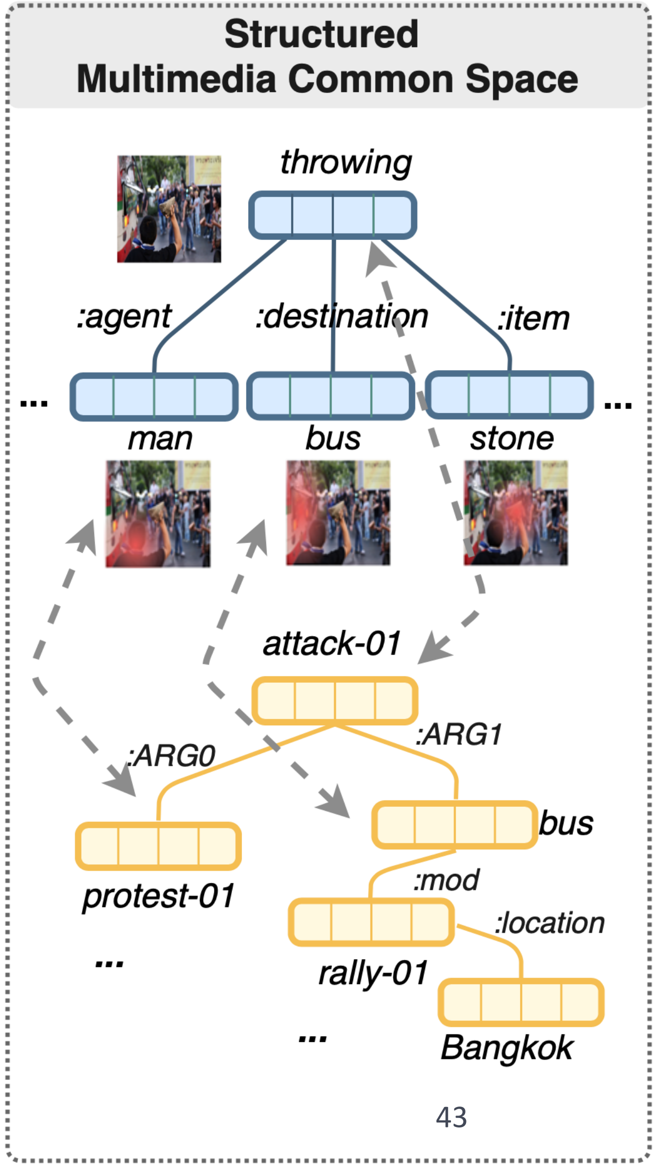
-- Training Phase (Common Space Construction)



How to align the two modalities?

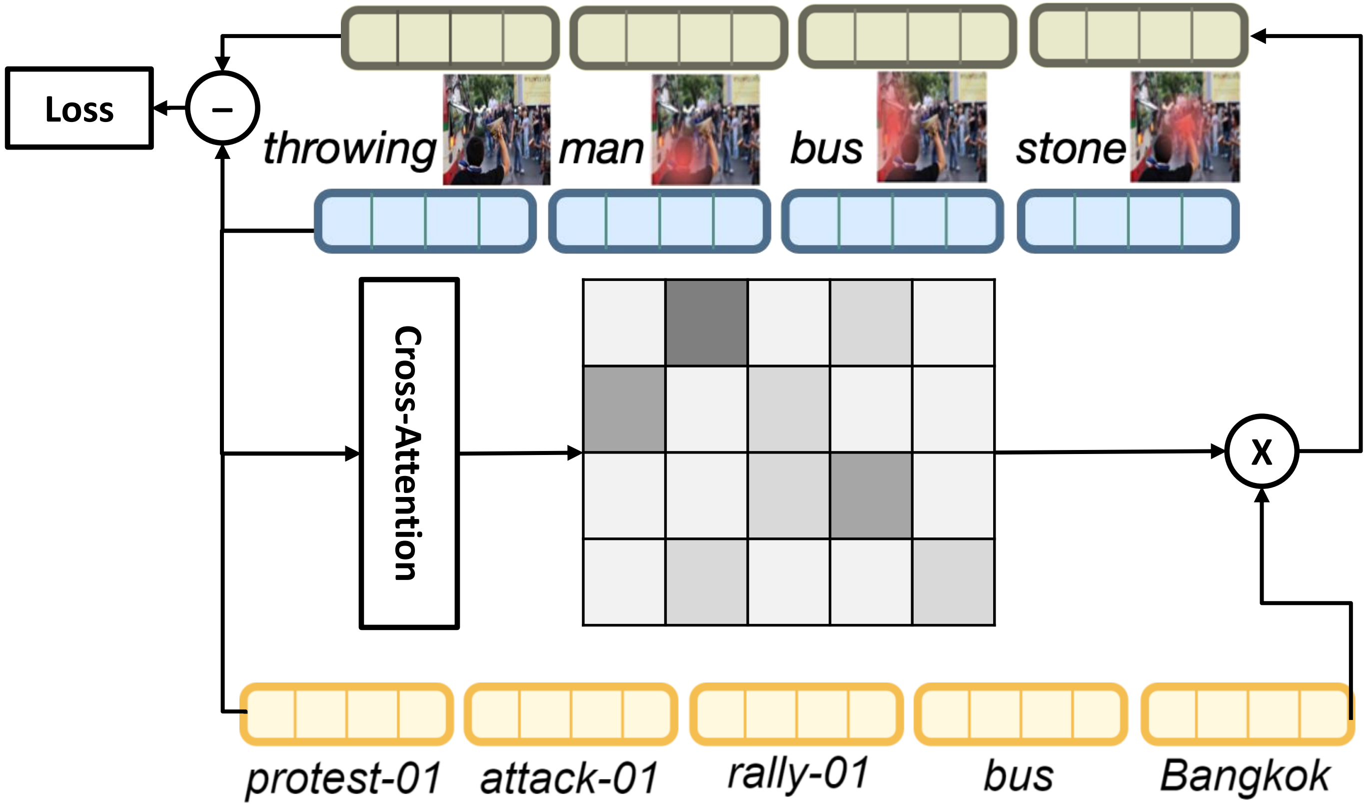
- Prior work aligns image-caption vectors by triplet loss.
- We want to align two graphs, not just single vectors.
- Ontology is shared so the nodes carry similar semantics.

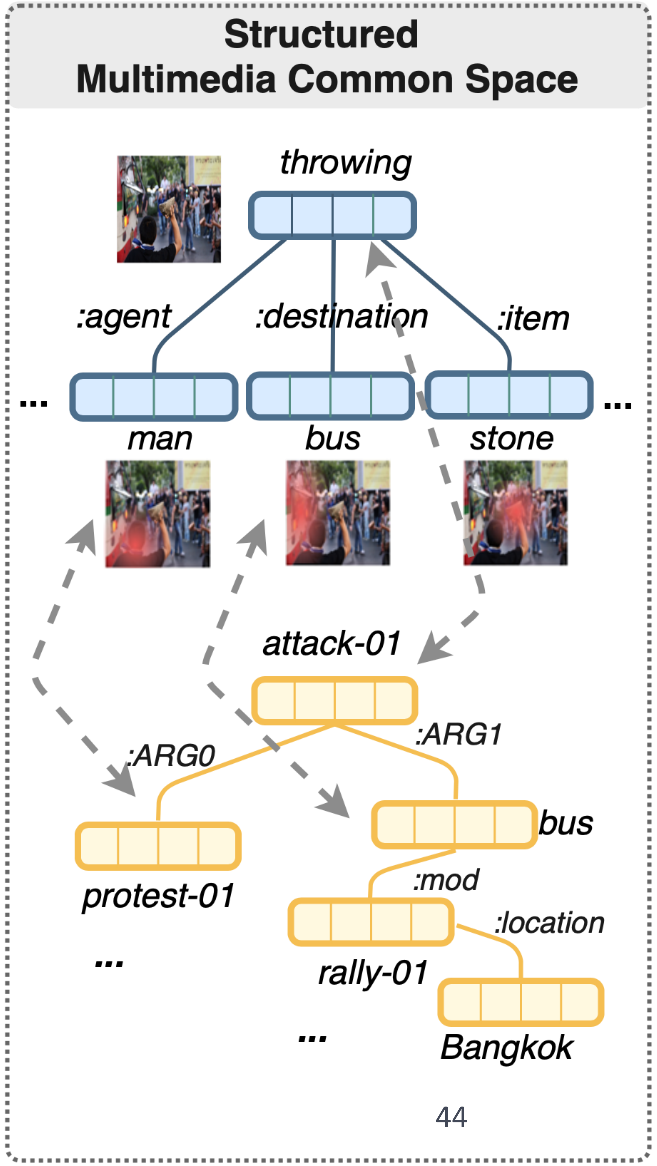




How to align the two modalities?

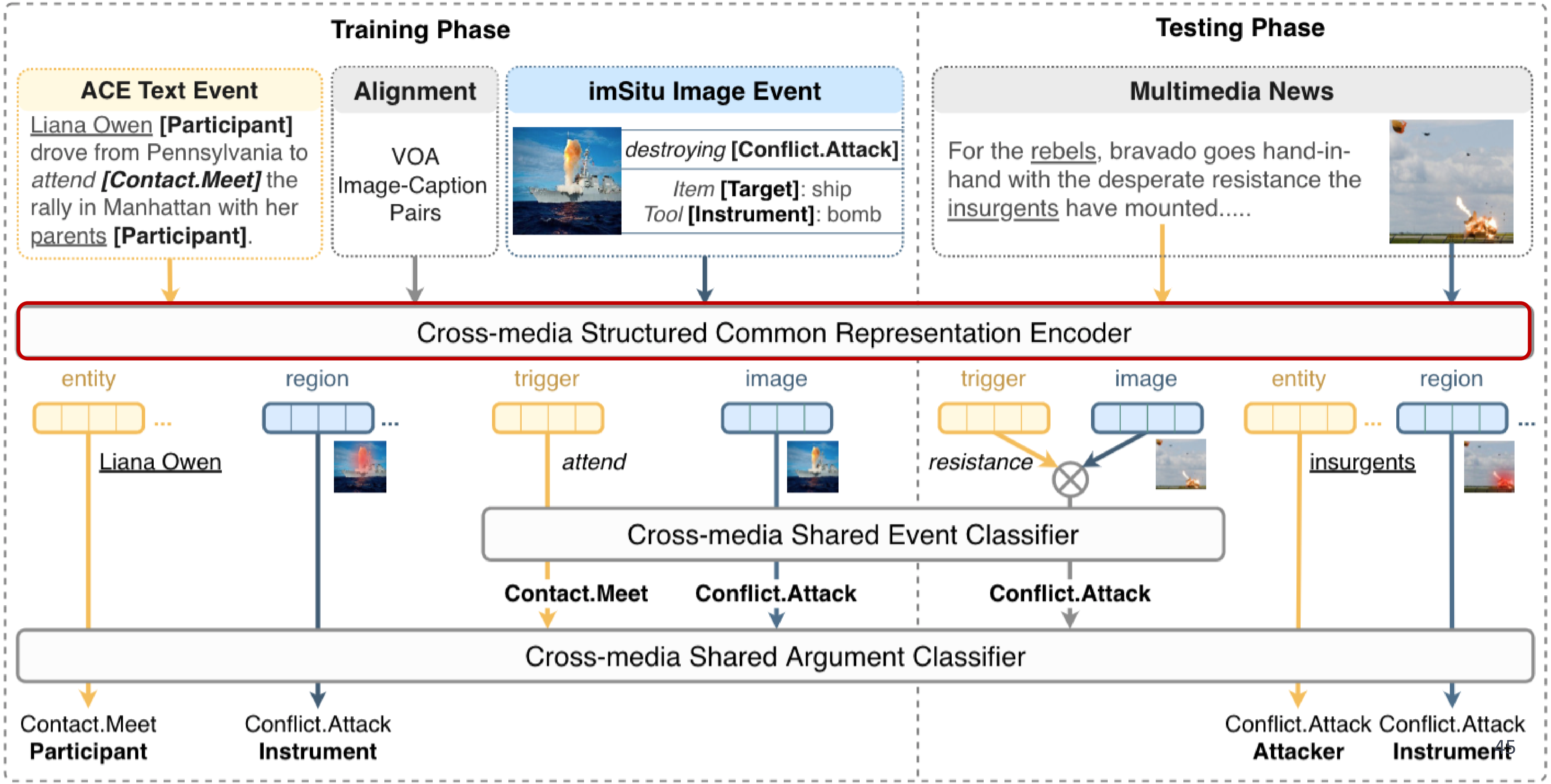
- Prior work aligns image-caption vectors by triplet loss.
- We want to align two graphs, not just single vectors.
- Ontology is shared so the nodes carry similar semantics.





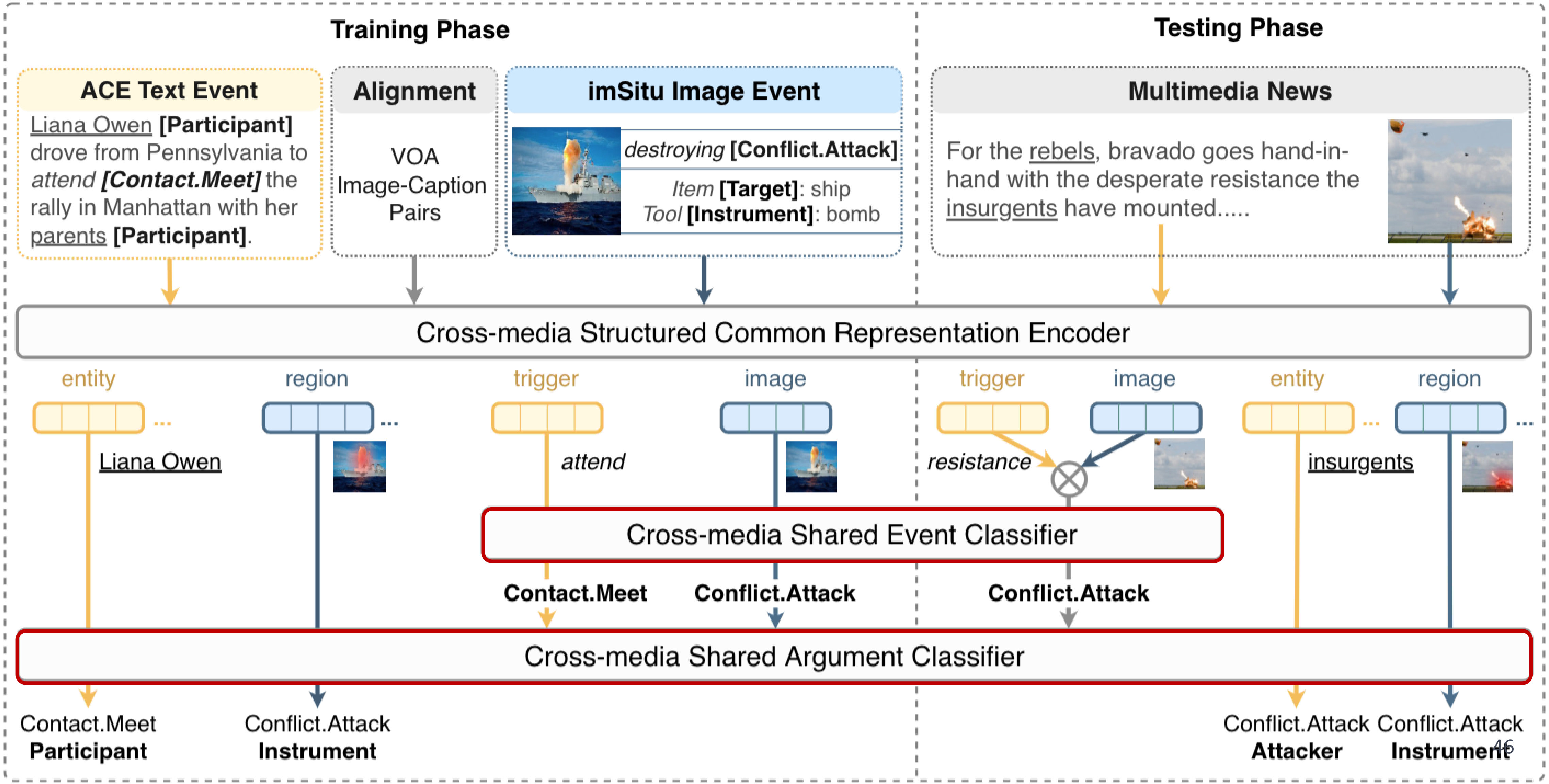
Weakly Aligned Structured Embedding (WASE)

-- Training and Testing Phase (Cross-media shared classifiers)



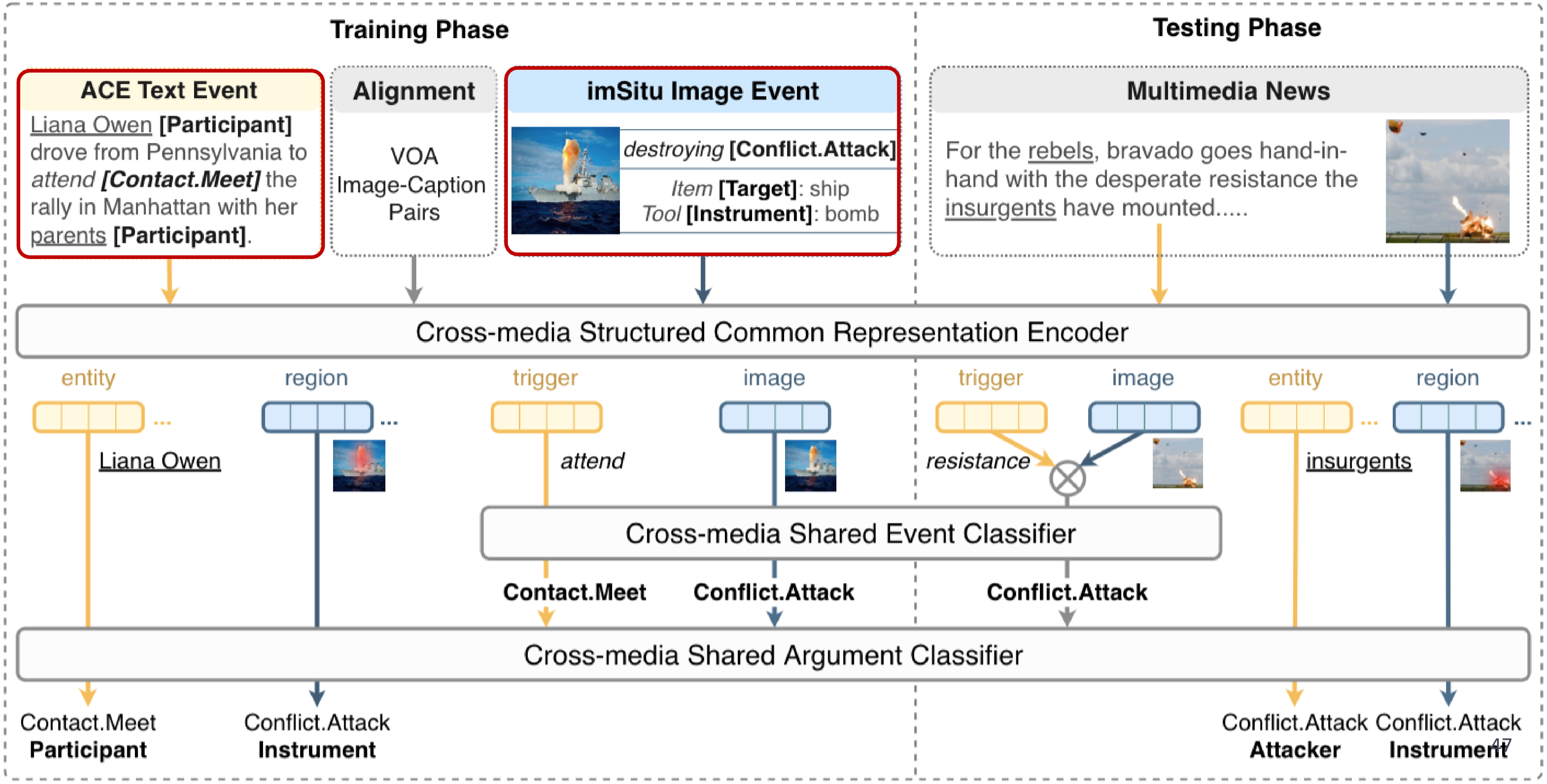
Weakly Aligned Structured Embedding (WASE)

-- Training and Testing Phase (Cross-media shared classifiers)



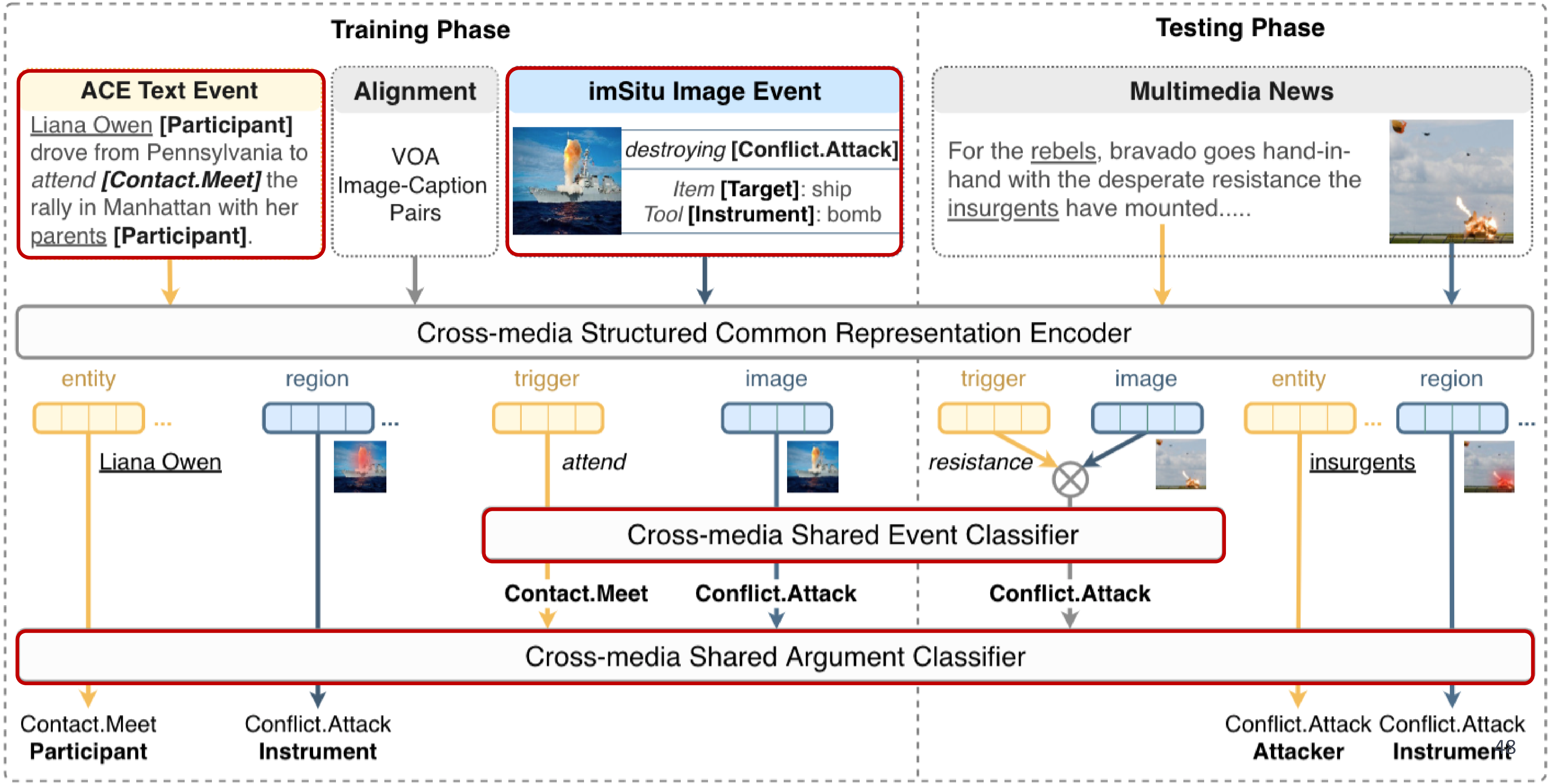
Weakly Aligned Structured Embedding (WASE)

-- Training and Testing Phase (Cross-media shared classifiers)



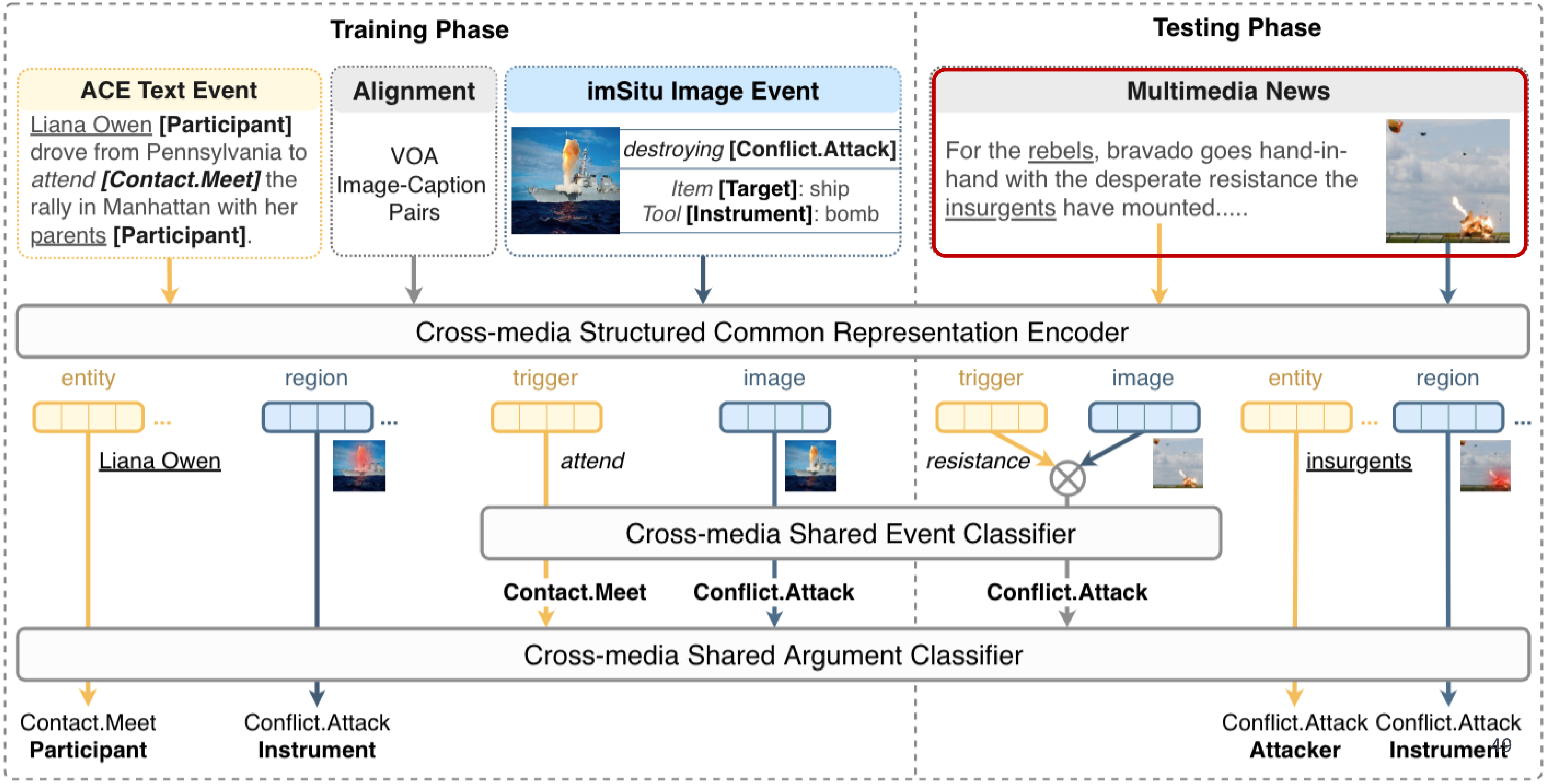
Weakly Aligned Structured Embedding (WASE)

-- Training and Testing Phase (Cross-media shared classifiers)



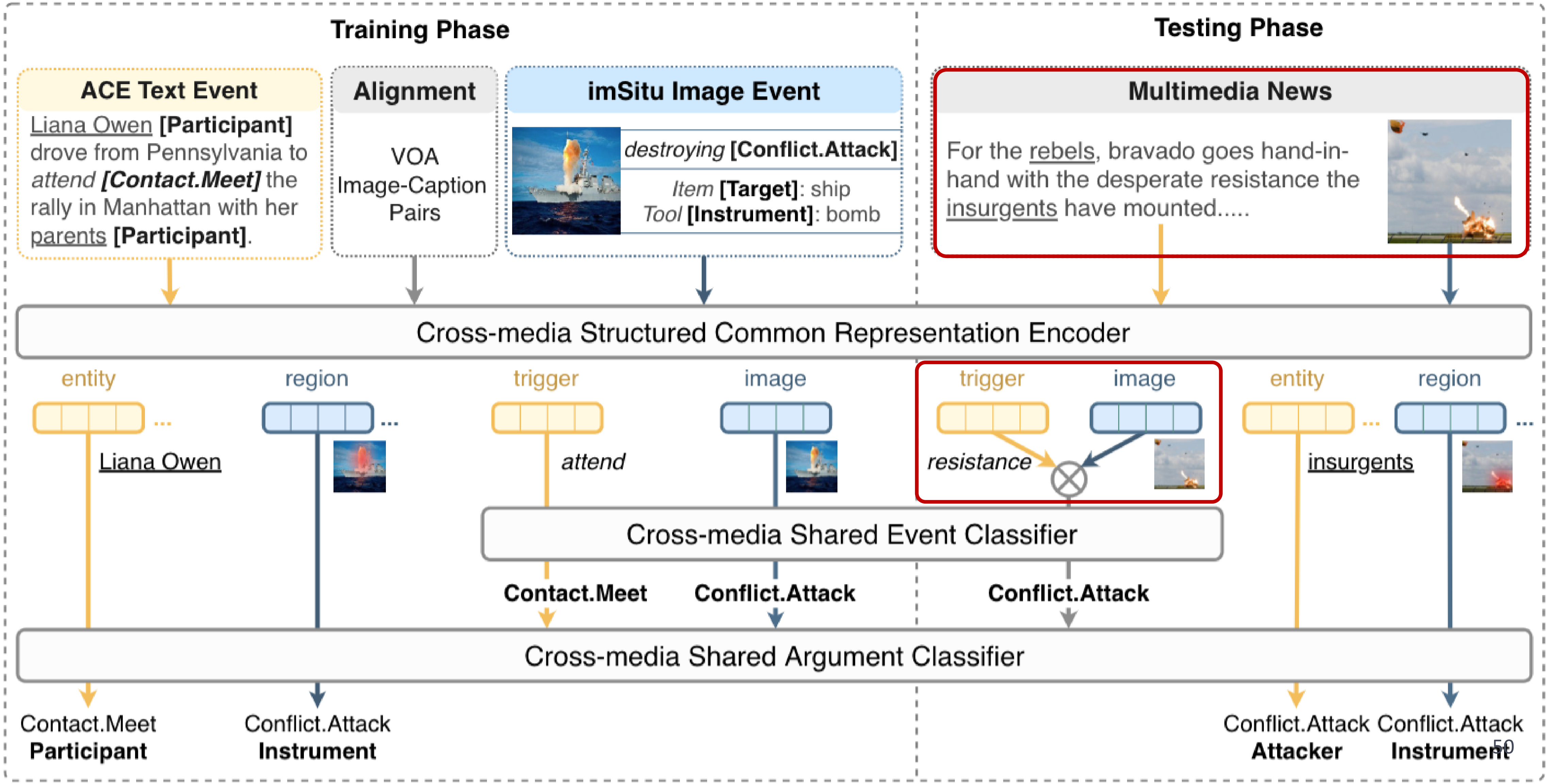
Weakly Aligned Structured Embedding (WASE)

-- Training and Testing Phase (Cross-media shared classifiers)



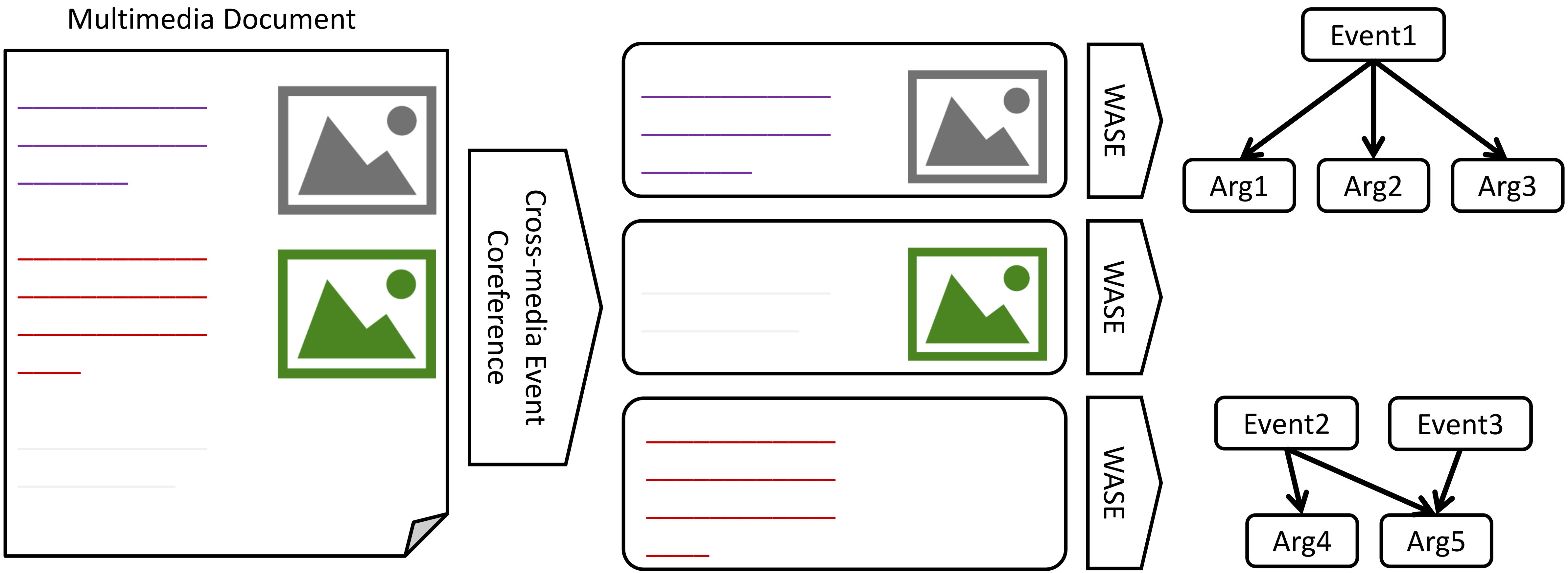
Weakly Aligned Structured Embedding (WASE)

-- Training and Testing Phase (Cross-media shared classifiers)



Weakly Aligned Structured Embedding (WASE)

-- System Diagram



Experiment Results

Training	Model	Text-Only Evaluation						Image-Only Evaluation						Multimedia Evaluation					
		Event Mention			Argument Role			Event Mention			Argument Role			Event Mention			Argument Role		
		P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Text	JMEE	42.5	58.2	48.7	22.9	28.3	25.3	-	-	-	-	-	-	42.1	34.6	38.1	21.1	12.6	15.8
	GAIL	43.4	53.5	47.9	23.6	29.2	26.1	-	-	-	-	-	-	44.0	32.4	37.3	22.7	12.8	16.4
	$\text{WASE}^{\mathbb{T}}$	42.3	58.4	48.2	21.4	30.1	24.9	-	-	-	-	-	-	41.2	33.1	36.7	20.1	13.0	15.7
Image	$\text{WASE}^{\mathbb{I}}_{att}$	-	-	-	-	-	-	29.7	61.9	40.1	9.1	10.2	9.6	28.3	23.0	25.4	2.9	6.1	3.8
	$\text{WASE}^{\mathbb{I}}_{obj}$	-	-	-	-	-	-	28.6	59.2	38.7	13.3	9.8	11.2	26.1	22.4	24.1	4.7	5.0	4.9
Multimedia	VSE-C	33.5	47.8	39.4	16.6	24.7	19.8	30.3	48.9	26.4	5.6	6.1	5.7	33.3	48.2	39.3	11.1	14.9	12.8
	Flat_{att}	34.2	63.2	44.4	20.1	27.1	23.1	27.1	57.3	36.7	4.3	8.9	5.8	33.9	59.8	42.2	12.9	17.6	14.9
	Flat_{obj}	38.3	57.9	46.1	21.8	26.6	24.0	26.4	55.8	35.8	9.1	6.5	7.6	34.1	56.4	42.5	16.3	15.9	16.1
	WASE_{att}	37.6	66.8	48.1	27.5	33.2	**30.1**	32.3	63.4	42.8	9.7	11.1	10.3	38.2	67.1	49.1	18.6	21.6	**19.9**
	WASE_{obj}	42.8	61.9	**50.6**	23.5	30.3	26.4	43.1	59.2	**49.9**	14.5	10.1	**11.9**	43.0	62.1	**50.8**	19.5	18.9	19.2

Experiment Results

Training	Model	Text-Only Evaluation						Image-Only Evaluation						Multimedia Evaluation					
		Event Mention			Argument Role			Event Mention			Argument Role			Event Mention			Argument Role		
		P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Text	JMEE	42.5	58.2	48.7	22.9	28.3	25.3	-	-	-	-	-	-	42.1	34.6	38.1	21.1	12.6	15.8
	GAIL	43.4	53.5	47.9	23.6	29.2	26.1	-	-	-	-	-	-	44.0	32.4	37.3	22.7	12.8	16.4
	$\text{WASE}^{\mathbb{T}}$	42.3	58.4	48.2	21.4	30.1	24.9	-	-	-	-	-	-	41.2	33.1	36.7	20.1	13.0	15.7
Image	$\text{WASE}^{\mathbb{I}}_{\text{att}}$	-	-	-	-	-	-	29.7	61.9	40.1	9.1	10.2	9.6	28.3	23.0	25.4	2.9	6.1	3.8
	$\text{WASE}^{\mathbb{I}}_{\text{obj}}$	-	-	-	-	-	-	28.6	59.2	38.7	13.3	9.8	11.2	26.1	22.4	24.1	4.7	5.0	4.9
Multimedia	VSE-C	33.5	47.8	39.4	16.6	24.7	19.8	30.3	48.9	26.4	5.6	6.1	5.7	33.3	48.2	39.3	11.1	14.9	12.8
	Flat_{att}	34.2	63.2	44.4	20.1	27.1	23.1	27.1	57.3	36.7	4.3	8.9	5.8	33.9	59.8	42.2	12.9	17.6	14.9
	Flat_{obj}	38.3	57.9	46.1	21.8	26.6	24.0	26.4	55.8	35.8	9.1	6.5	7.6	34.1	56.4	42.5	16.3	15.9	16.1
	WASE_{att}	37.6	66.8	48.1	27.5	33.2	**30.1**	32.3	63.4	42.8	9.7	11.1	10.3	38.2	67.1	49.1	18.6	21.6	**19.9**
	WASE_{obj}	42.8	61.9	**50.6**	23.5	30.3	26.4	43.1	59.2	**49.9**	14.5	10.1	**11.9**	43.0	62.1	**50.8**	19.5	18.9	19.2

Experiment Results

Training	Model	Text-Only Evaluation						Image-Only Evaluation						Multimedia Evaluation					
		Event Mention			Argument Role			Event Mention			Argument Role			Event Mention			Argument Role		
		P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Text	JMEE	42.5	58.2	48.7	22.9	28.3	25.3	-	-	-	-	-	-	42.1	34.6	38.1	21.1	12.6	15.8
	GAIL	43.4	53.5	47.9	23.6	29.2	26.1	-	-	-	-	-	-	44.0	32.4	37.3	22.7	12.8	16.4
	$\text{WASE}^{\mathbb{T}}$	42.3	58.4	48.2	21.4	30.1	24.9	-	-	-	-	-	-	41.2	33.1	36.7	20.1	13.0	15.7
Image	$\text{WASE}^{\mathbb{I}}_{\text{att}}$	-	-	-	-	-	-	29.7	61.9	40.1	9.1	10.2	9.6	28.3	23.0	25.4	2.9	6.1	3.8
	$\text{WASE}^{\mathbb{I}}_{\text{obj}}$	-	-	-	-	-	-	28.6	59.2	38.7	13.3	9.8	11.2	26.1	22.4	24.1	4.7	5.0	4.9
Multimedia	VSE-C	33.5	47.8	39.4	16.6	24.7	19.8	30.3	48.9	26.4	5.6	6.1	5.7	33.3	48.2	39.3	11.1	14.9	12.8
	Flat_{att}	34.2	63.2	44.4	20.1	27.1	23.1	27.1	57.3	36.7	4.3	8.9	5.8	33.9	59.8	42.2	12.9	17.6	14.9
	Flat_{obj}	38.3	57.9	46.1	21.8	26.6	24.0	26.4	55.8	35.8	9.1	6.5	7.6	34.1	56.4	42.5	16.3	15.9	16.1
	WASE_{att}	37.6	66.8	48.1	27.5	33.2	**30.1**	32.3	63.4	42.8	9.7	11.1	10.3	38.2	67.1	49.1	18.6	21.6	**19.9**
	WASE_{obj}	42.8	61.9	**50.6**	23.5	30.3	26.4	43.1	59.2	**49.9**	14.5	10.1	**11.9**	43.0	62.1	**50.8**	19.5	18.9	19.2

Cross-Media Coreference Accuracy

Model	P (%)	R (%)	F_1 (%)
rule_based	10.1	100	18.2
VSE	31.2	74.5	44.0
Flat_{att}	33.1	73.5	45.6
Flat_{obj}	34.3	76.4	47.3
WASE_{att}	39.5	73.5	51.5
WASE_{obj}	40.1	75.4	52.4

Compare to Single Data Modality Extraction

- Surrounding sentence helps visual event extraction.

People celebrate Supreme Court ruling on Same Sex Marriage in front of the Supreme Court in Washington.

- Image helps textual event extraction.

Iraqi security forces *search* ***[Justice.Arrest]*** a civilian in the city of Mosul.

Why Does Vision Help NLP?

- Various triggers and context can be coherent in visual space.
- Cross-media Common space pushes scattered sentences towards the visual cluster.

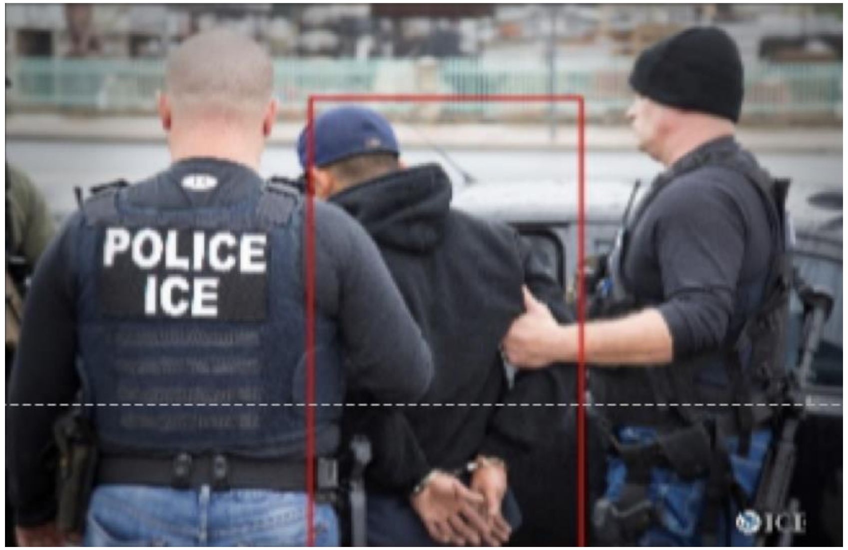
Berlin police tweeted that six people were arrested after a joint operation with the Berlin's prosecutor's office.

The man in Kosovo is an ethnic Albanian arrested south of the capital, Pristina.

He was asleep in a suburban Seattle house last week morning when immigration agents showed up to arrest his father.

But shortly after the round table began, Marko Djuric, head of the Serbian government office on Kosovo, was detained by police.

Compare to Cross-media Flat Representation



Model	Event Type	Argument Role	
Flat	Justice.ArrestJail	Agent =	man
Ours	Justice.ArrestJail	Entity =	man

Model	Event Type	Argument Role	
Flat	Movement.Transport	Artifact =	none
Ours	Movement.Transport	Artifact =	man

Summary of Event Extraction Methods

IE Methods		Supervised Learning	Bootstrapping	Distant Supervision	Open IE/ Zero-shot	Schema/ Discovery
Approach Overview		Learn rules or supervised model from labeled data	Send seeds to extract common patterns from unlabeled data	Project large database entries into unlabeled data to obtain annotations	Open-domain IE based on syntactic patterns	Automatically discover scenarios, event types and templates
Requirement of labeled data		Large unstructured labeled data	Small seeds	Large seeds	Small unstructured labeled data	Little labeled data
Quality	Precision	High	Moderate	Low	Moderate	Moderate
	Recall	High	Difficult to measure	Moderate	Low	Moderate
Portability		Poor	Moderate	Moderate	Good	Good
Scalability		Poor	Moderate	Moderate	Good	Good