Information Extraction

Event-Centric Natural Language Processing (Part I)

Manling Li, Heng Ji
Department of Computer Science
University of Illinois at Urbana-Champaign

Aug 2021

ACL Tutorials

Event-Centric Natural Language Processing

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.

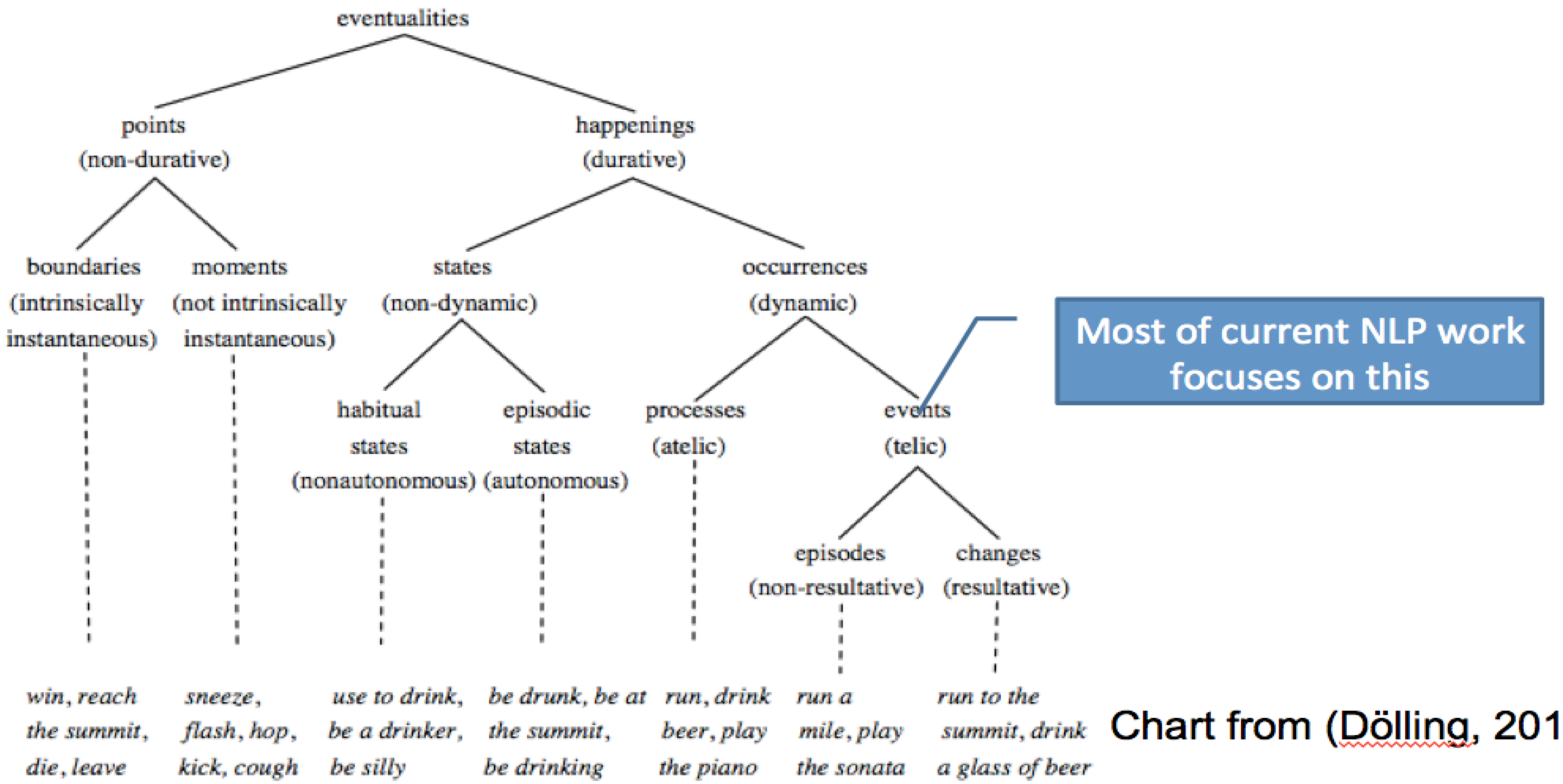
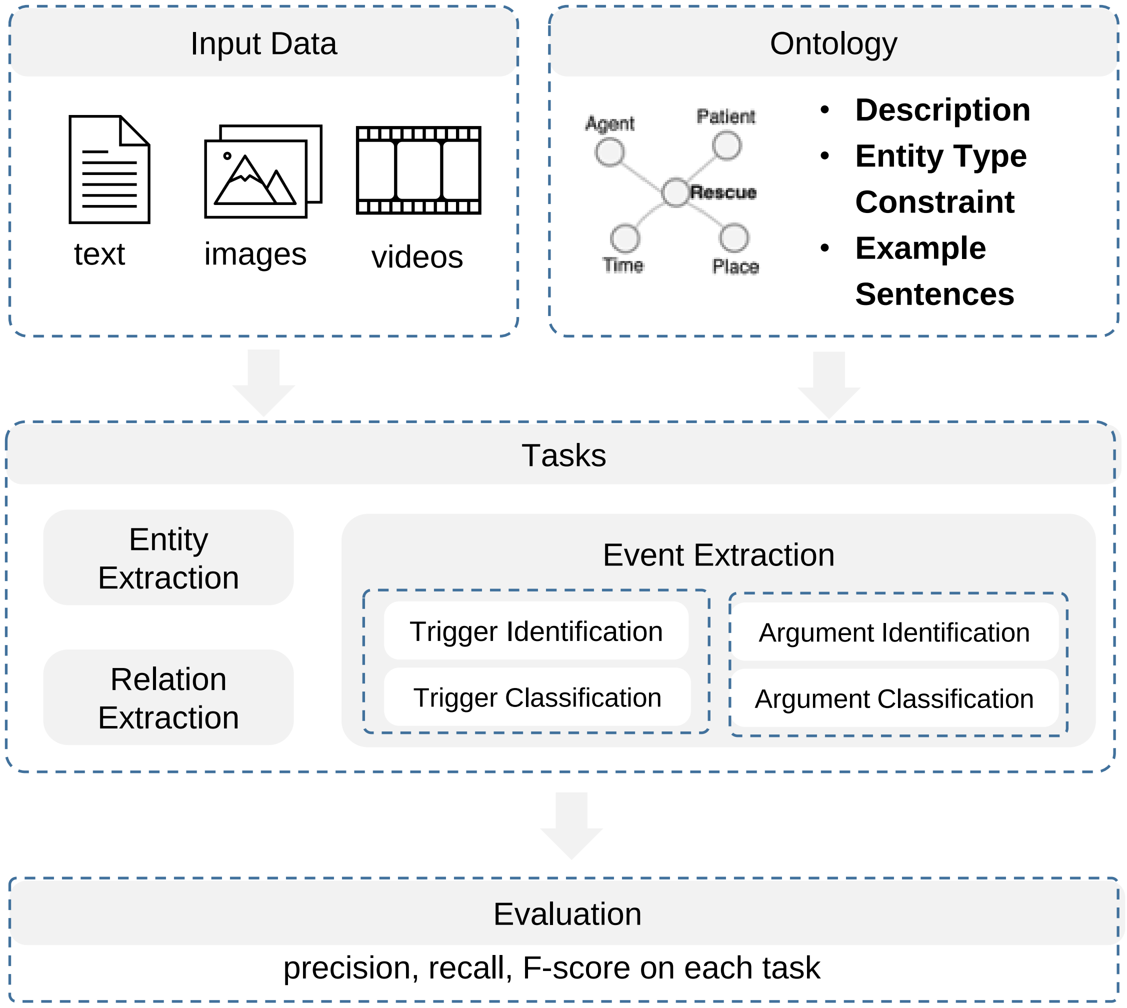


Chart from (Dölling, 2011)

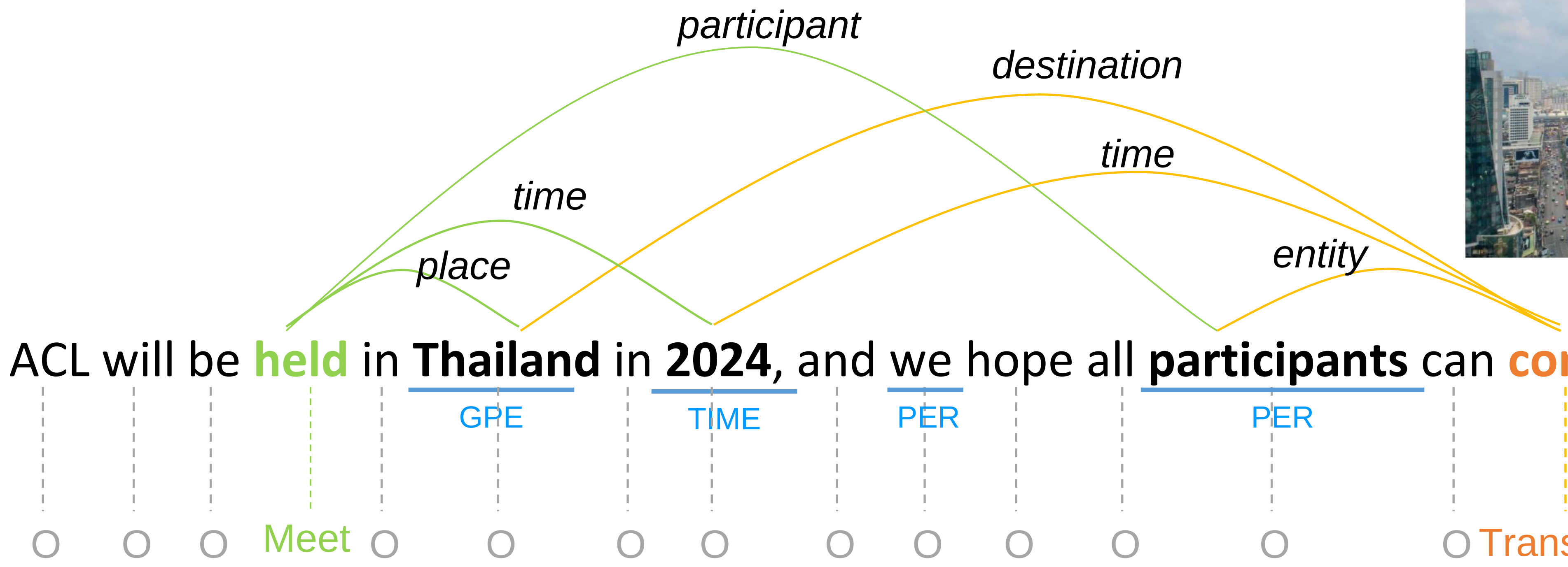
General Problem Statement

- Input …
  - A piece of text, or images, or videos
  - Ontology
    - The target event types and the argument roles for each type
    - Optional: the description of types and roles, the entity type constraint of each argument role, the example sentences, etc
- Aims to extract …
  - Events
    - Trigger identification
    - Trigger classification
  - Arguments
    - Argument identification
    - Argument classification
- Evaluated on …
  - precision and recall on each task



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

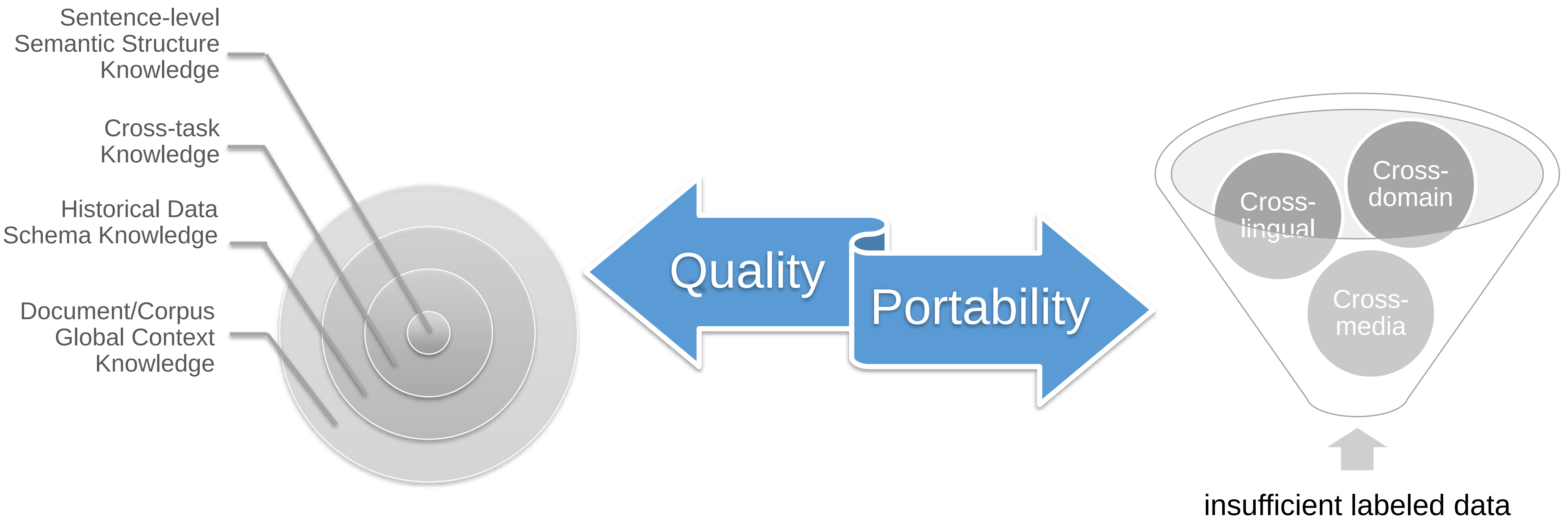


The Berkeley Hotel, Bangkok, Thailand

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

* BIO tag scheme is used, where the prefix B- marks the beginning of a mention, and I- means inside of a mention. A token not belonging to any mention is tagged with O.

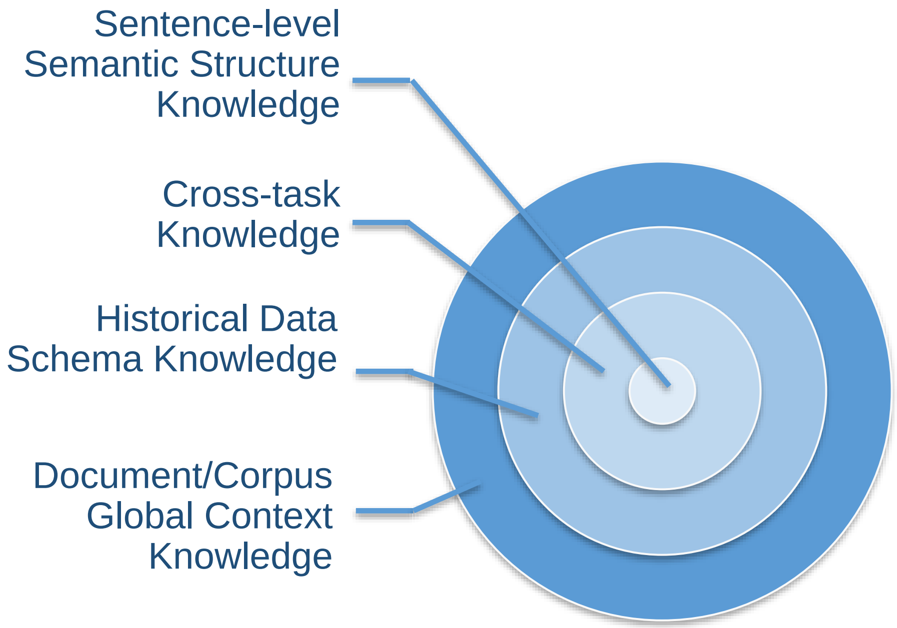
Challenges

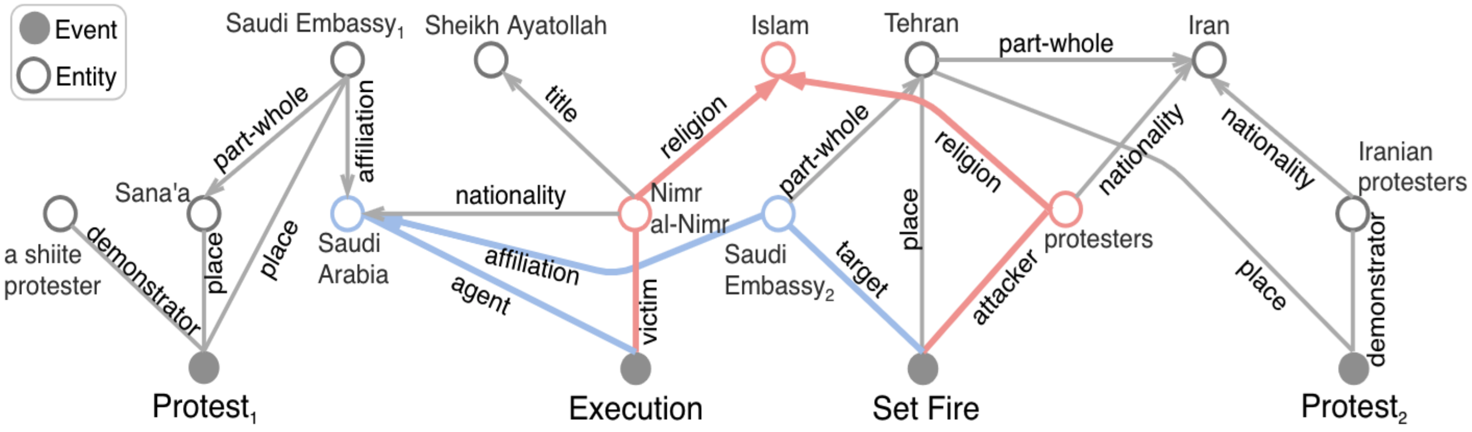


Sentence-level Semantic Structure Knowledge
Cross-task Knowledge
Historical Data Schema Knowledge
Document/Corpus Global Context Knowledge
Quality
Portability
Cross-domain
Cross-lingual
Cross-media
insufficient labeled data

Challenge 1: Quality

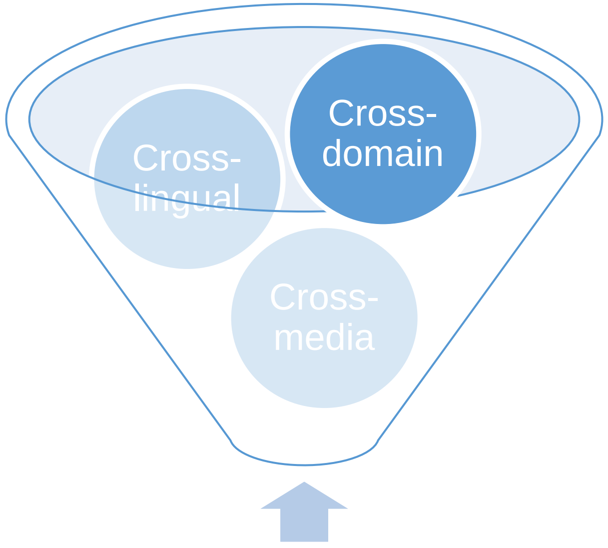
- How to encode the knowledge for better scene understanding?
  - semantic structure knowledge such as dependency graph, AMR graph, etc
  - cross-task knowledge such as the interactions between relation extraction and entity typing
  - schema knowledge induced from historical data
  - document-level or corpus-level global context knowledge



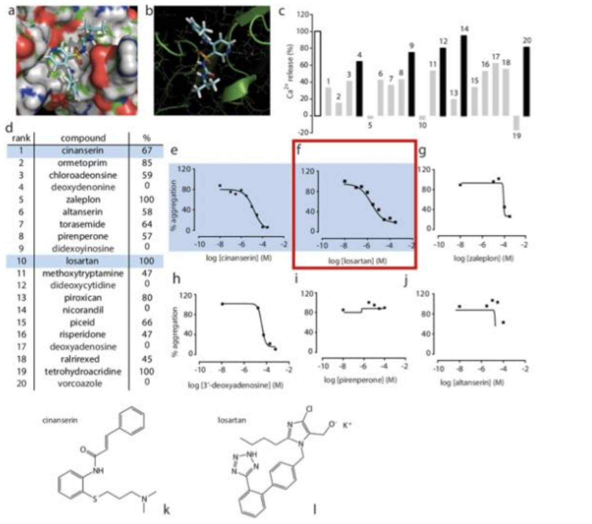


Challenge 2: Portability (cross-domain)

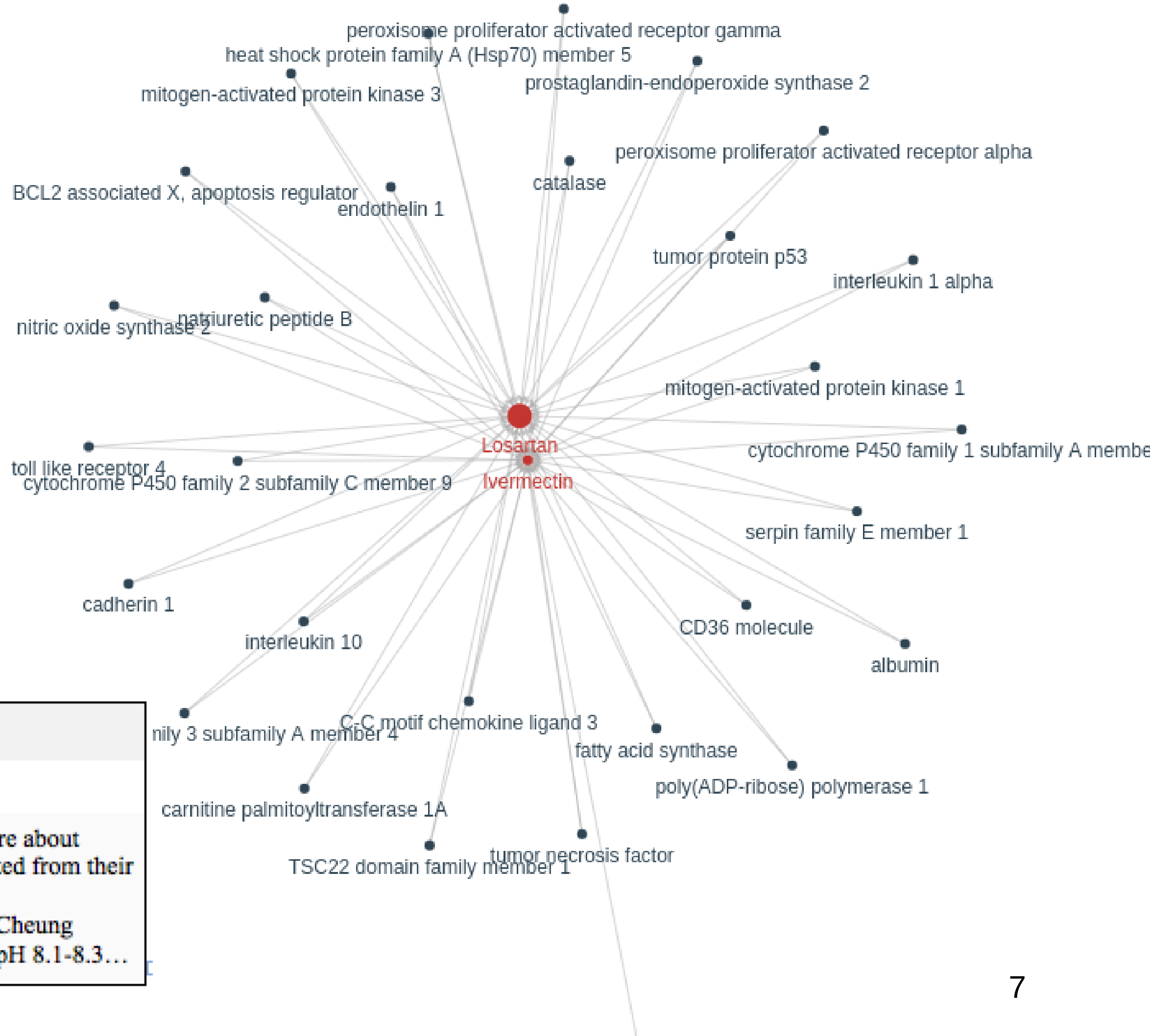
- How to deal with unseen types for a new domain?



insufficient labeled data



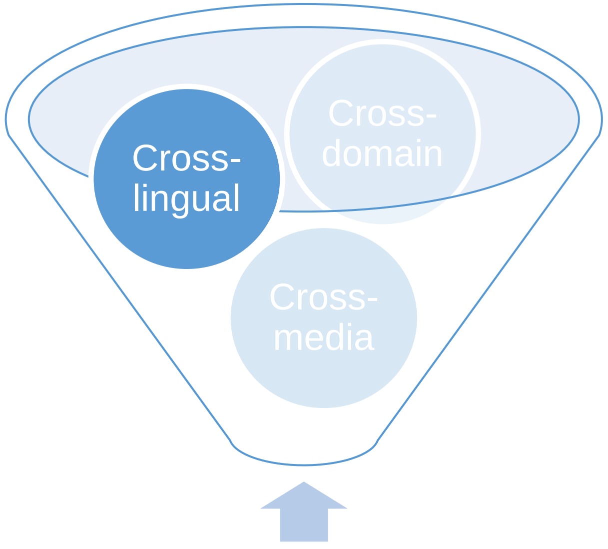
Src: PMC4074120 **Fig. 1.** In silico identifies GPVI antagonist. Representative image capture of in silico docking into GPVI using Glide, with space filling model is shown in a, and H-bonding to relevant side chains is detailed in b. The 20 highest ranking compounds were screened for effects on Ca2+ release by the GPVI-specific agonist CRP-XL (10 mg/ml) (c and d, % refers to percent inhibition of Ca2+ release). Maximum Ca2+ release is shown in white, compounds that inhibited Ca2+ release by , 50% or more are in grey, and the remainder in black. Commercially available compounds that inhibited CRP-XL-induced Ca2+ release .50% were further screened by light transmission aggregometry to identify compounds displaying dose-dependent inhibition (e-j). Examples are shown of weak antagonism (g and h) and false positives (i and j). Cinanserin (l) and losartan (k) were taken on for further study.



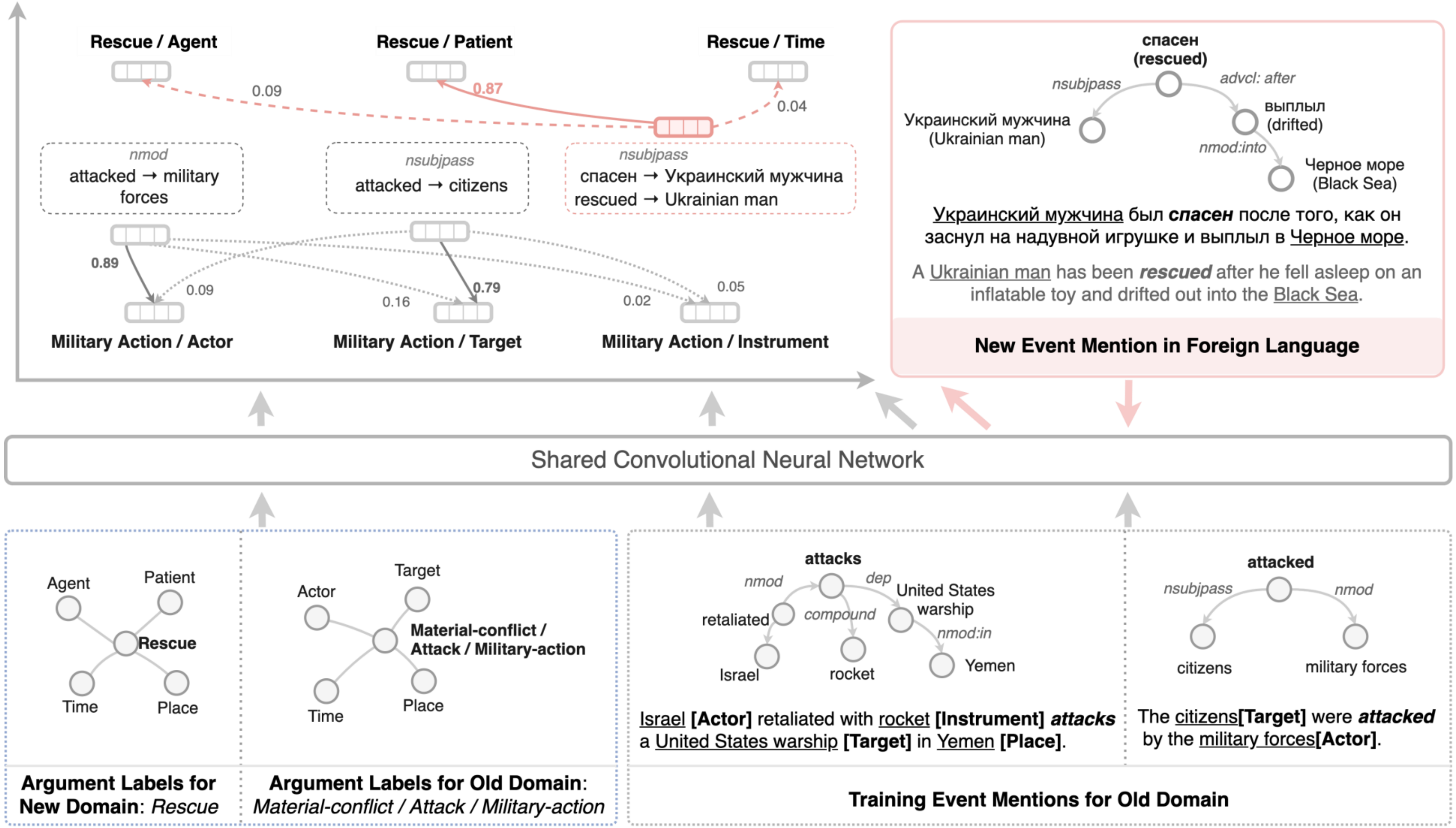
Disease	**Cardiovascular disease**
PMID, PMCID	Evidence Sentences
22800722 PMC7102827	The in vitro half-maximal inhibitory concentration (IC50) values of food-derived ACE inhibitory peptides are about 1000-fold higher than that of synthetic captopril but they have higher in vivo activities than would be expected from their in vitro activities..... Germinal ACE depends on chloride to a lesser extent compared with the C domain of sACE. Cushman and Cheung reported an optimal in vitro ACE activity of rabbit rung acetone extract in the presence of 300 mM NaCl at pH 8.1-8.3…

Challenge 2: Portability (cross-lingual)

- How to extract events from low-resource languages?

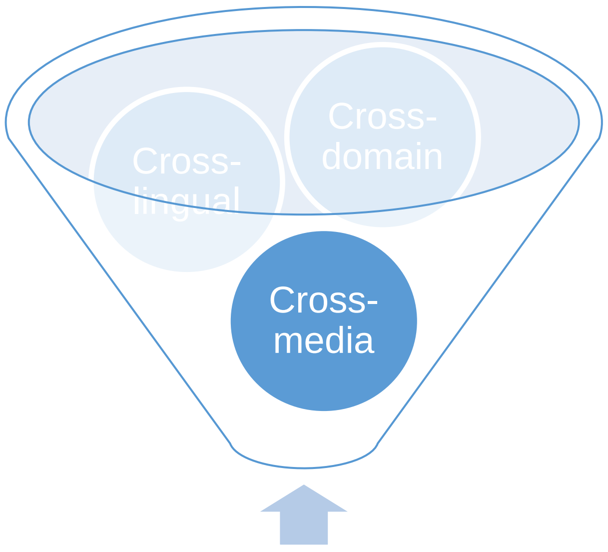


insufficient labeled data



Challenge 2: Portability (cross-modality)

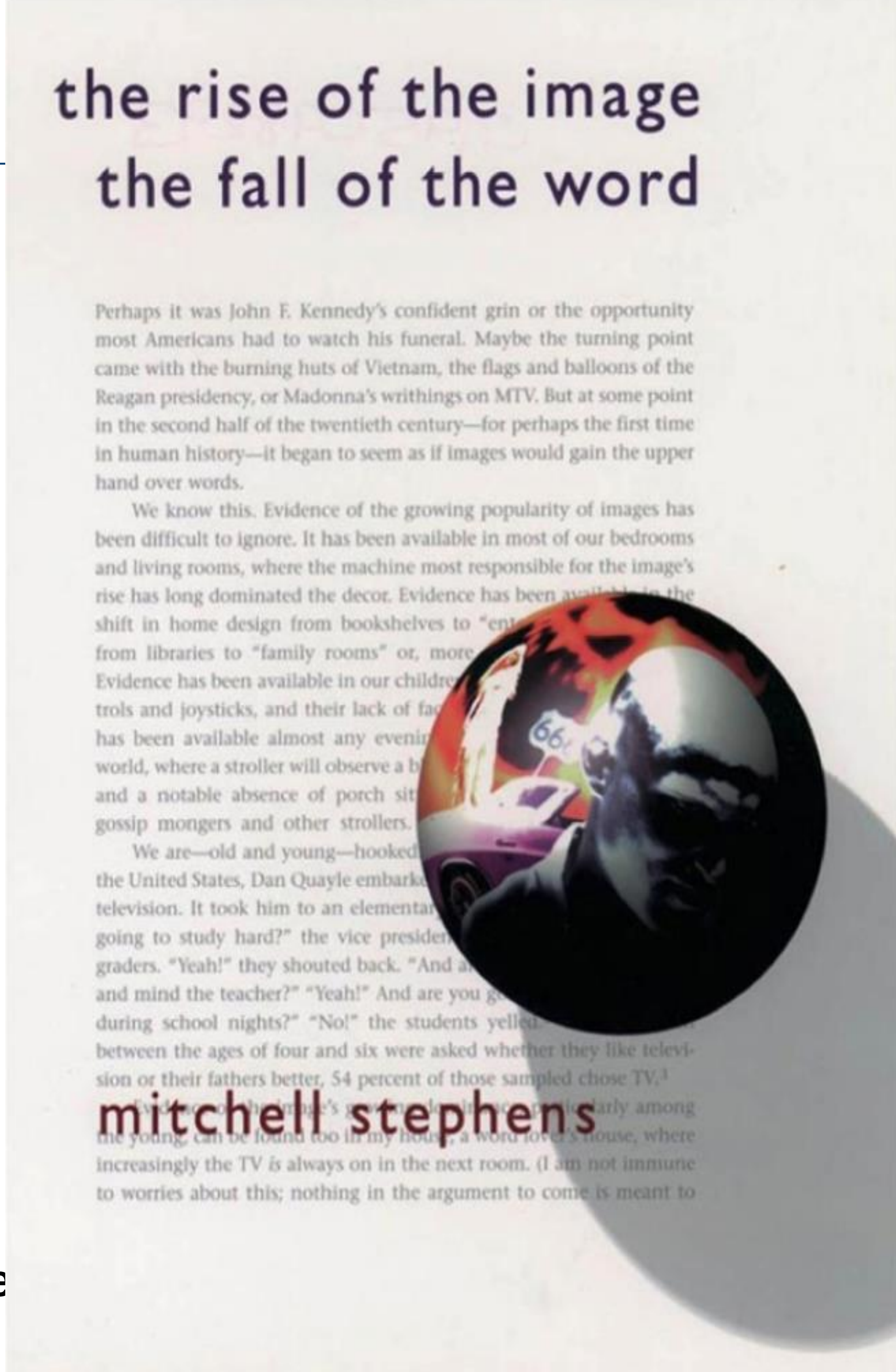
- How to jointly extract events from multimedia?



insufficient labeled data

- 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* = stretche**



Outline

Quality

wider context to extract

Joint IE

- How to capture the **graph structure** of the IE graph?

→ dependency graph, AMR graph, etc.

Schema-guided IE

- From these **graph structures of historical data**, are there general patterns to guide IE?

→ schema induction

Document-level IE

- Understanding connections between events requires **inference across sentences**

→ document-level IE (Argument linking, QA based , NLG based)

wider context as features

Portability

Cross-domain

- **Structure transfer**: AMR graph ←→ event structure
- **Common space:** To generate informative embeddings for labels, we should seek information from definition, example sentences, etc

Cross-lingual

- **Structure transfer**: universal dependency ←→ event structure
- **Common space:** Alignment should be multilevel (word-level, entity-level)

Cross-media

- **Structure transfer:** text AMR graph ←→ image situation graph
- **Common space:** Alignment should be multi-level (word-level, image-level)

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

- The key of supervised IE is to design effective features.

→ Feature engineering by experts.

feature engineering
- local context

- **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

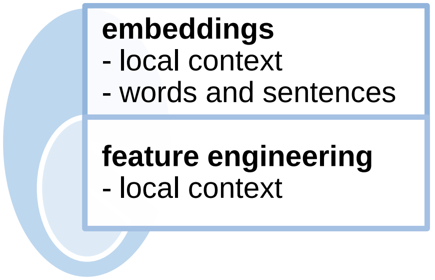
(Chen and Ji, 2009)

- Pros: Use linguistic knowledge from experts
- Cons: (1) Time-consuming (2) Not scalable to new tasks

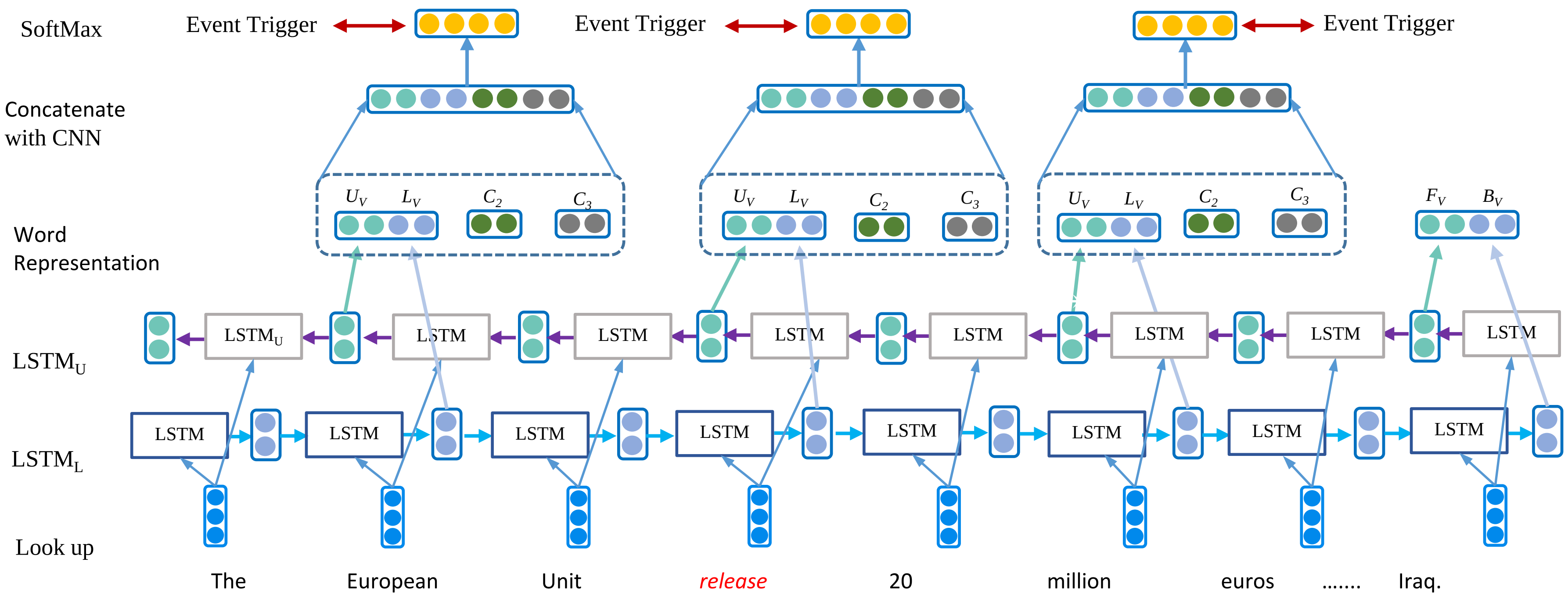
A More “Modern” Neural Event Extractor

- How to reduce the human efforts for feature engineering?

→ Embed words into semantic space



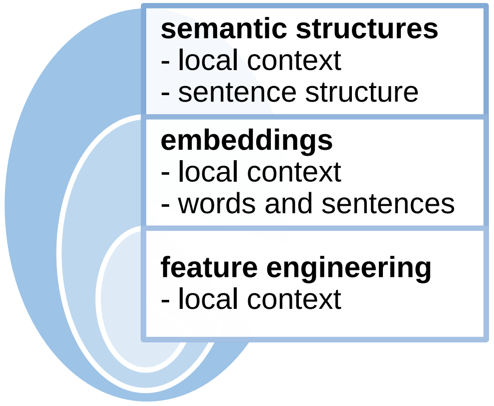
- Use word embeddings as features (Feng et al., 2016).
- Pros: Reduce feature engineering efforts to some extent
- Cons: Still rely on human annotated clean training data still fragile to noise in training data.



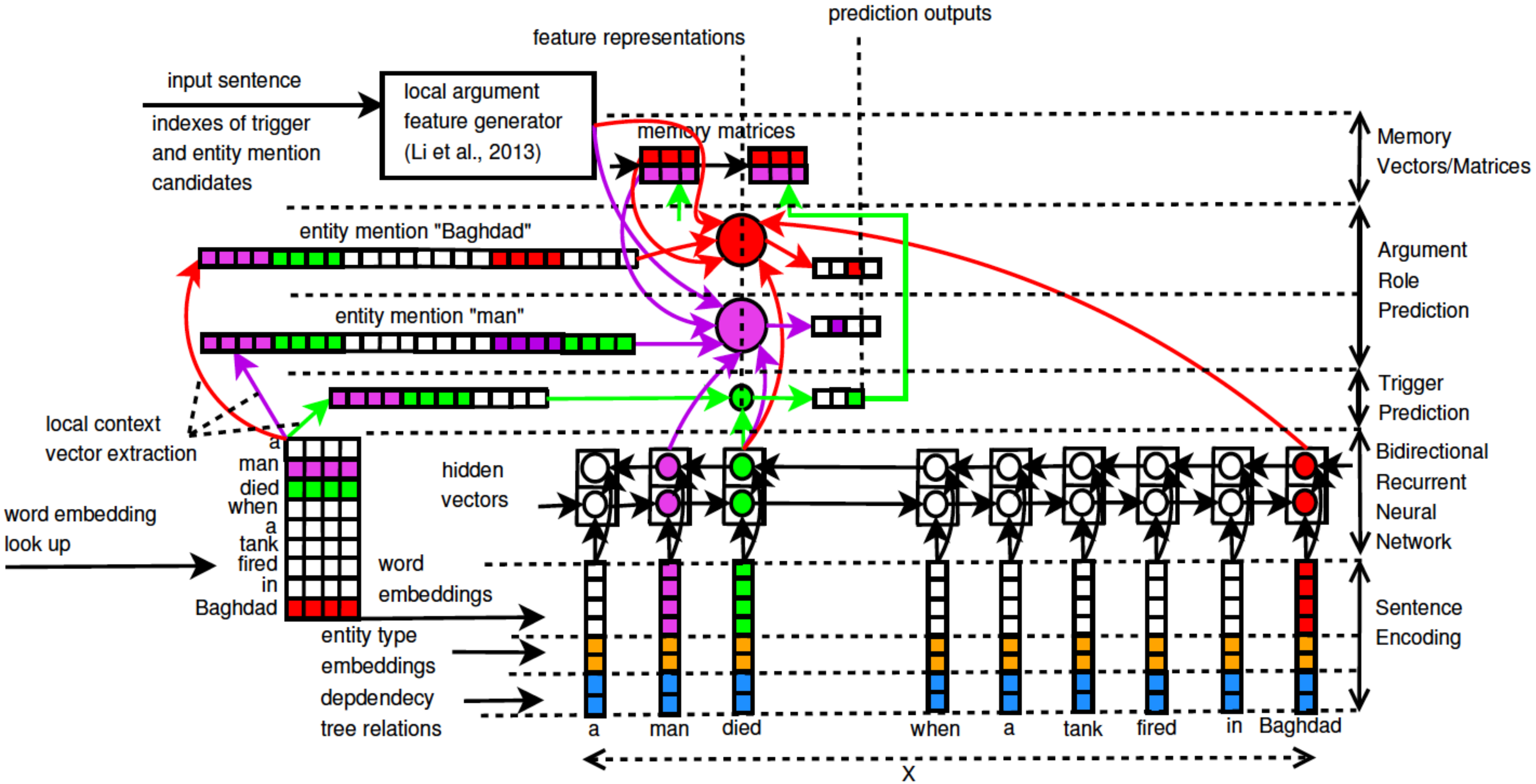
Or Put them Together…

- Embeddings can only capture local text semantics, which is fragile.

→ Add more global features targeting the dependencies between triggers and arguments



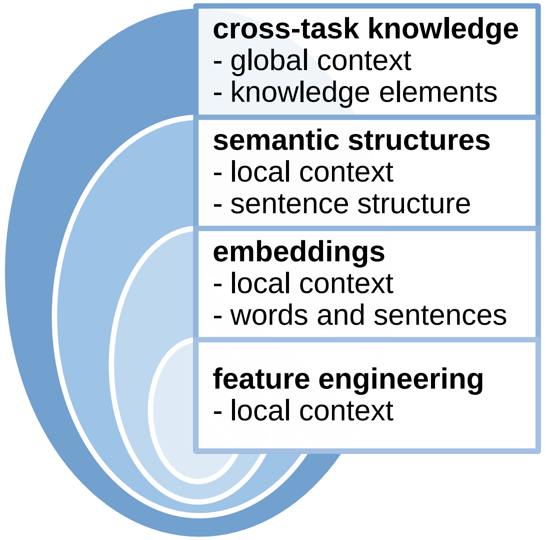
- Add symbolic features (entity type, dependency tree relations, etc) by concatenating them with embeddings (Nguyen et al., 2016)
- Pros: Capturing dependencies between triggers and arguments
- Cons: Still lack of global context, such as entity relations, etc



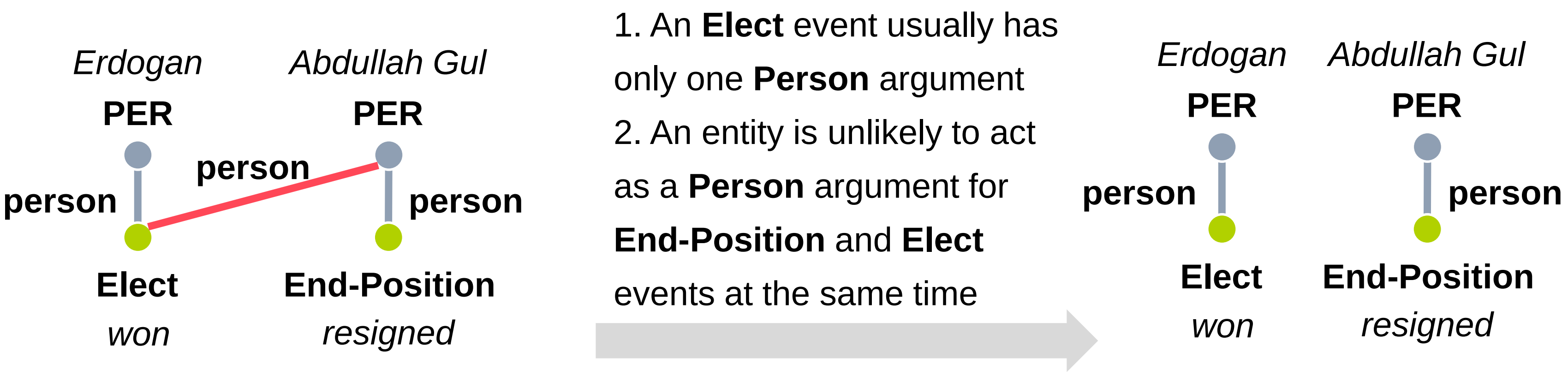
Joint Entity, Relation and Event Extraction

- However, errors can be propagated from previous tasks.

→ Jointly extracting entities, relations and events

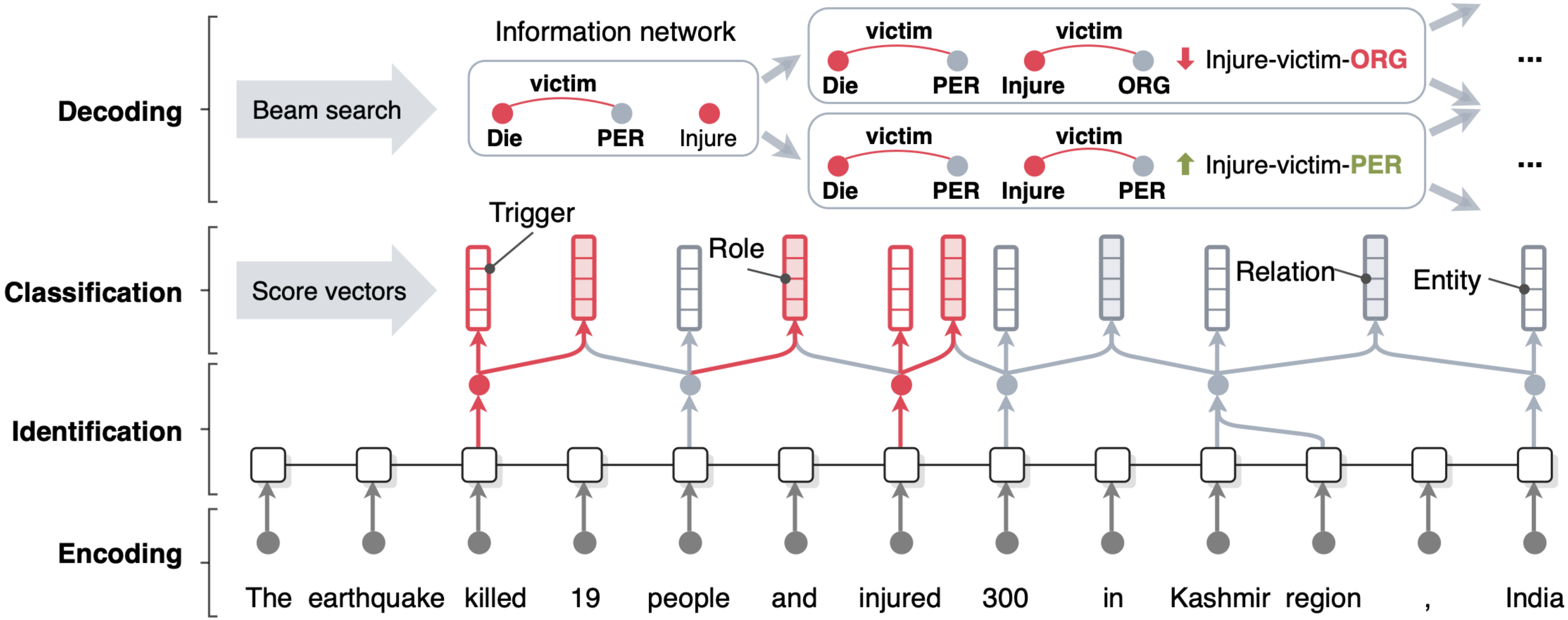


- Pipelined models suffer from the error propagation problem and disallow interactions among components
- Existing neural models do not explicitly model cross-subtask and cross-instance interactions among knowledge elements
- Example: *Prime Minister* ***Abdullah Gul*** *resigned earlier Tuesday to make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



OneIE: An End-to-end Neural Model for IE

- OneIE framework extracts the information graph (nodes: entities and events, edges: relations and arguments) from a given sentence. (Lin et al., 2020)
- Main challenge for Joint IE: How to capture interactions between knowledge elements?

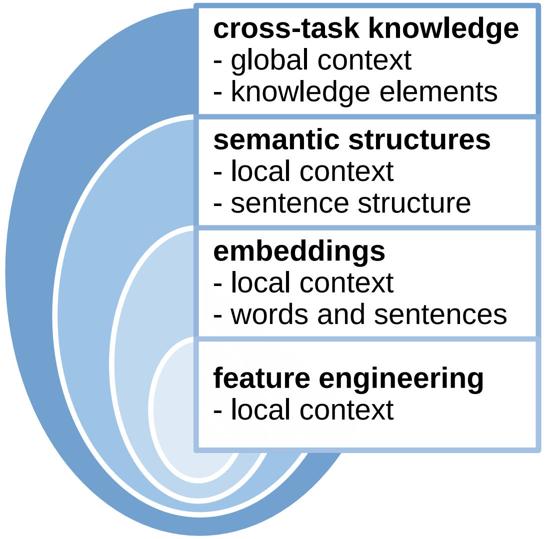


AMR-IE: An AMR-guided framework for IE

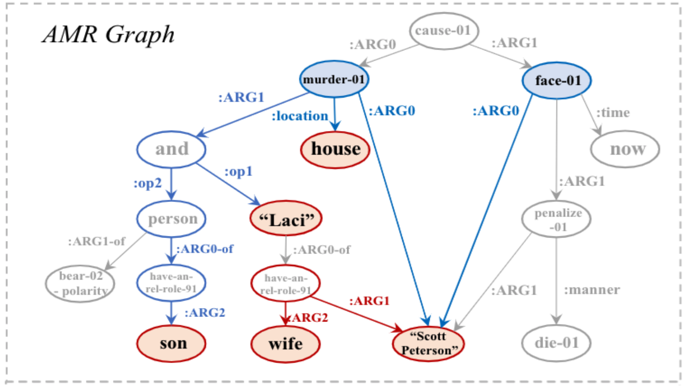
- To better capture connections between knowledge elements…

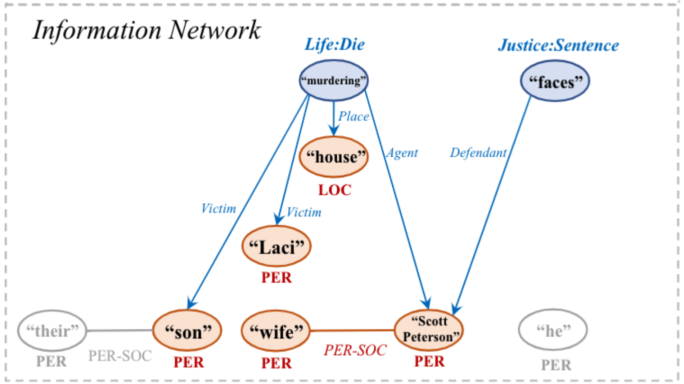
→ Adding graph structures

Graph Structure
dependency, AMR,…

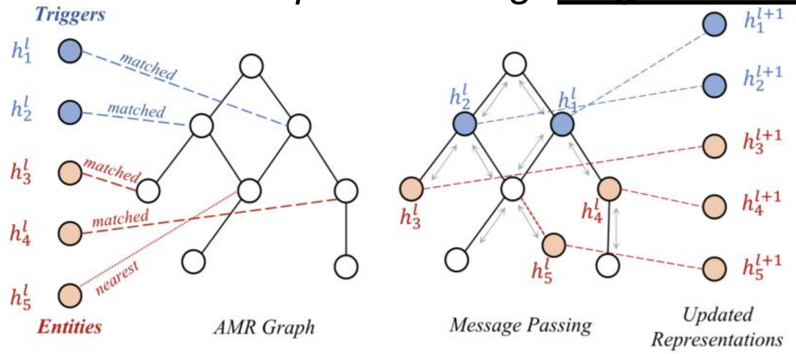


- Use a richer semantic parsing Abstract Meaning Representation (AMR) graph (Zhang et, al, 2021):

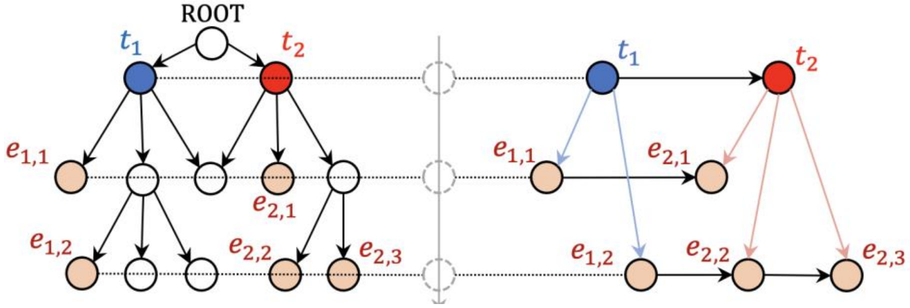




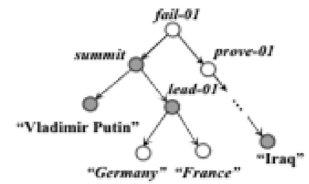
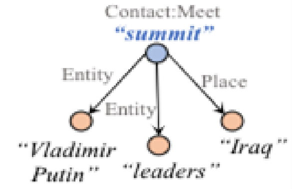
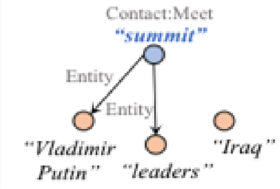
- *AMR Guided Graph Encoding*: **Edge-Conditioned GAT**



- *AMR Guided Graph Decoding*: **Ordered decoding**



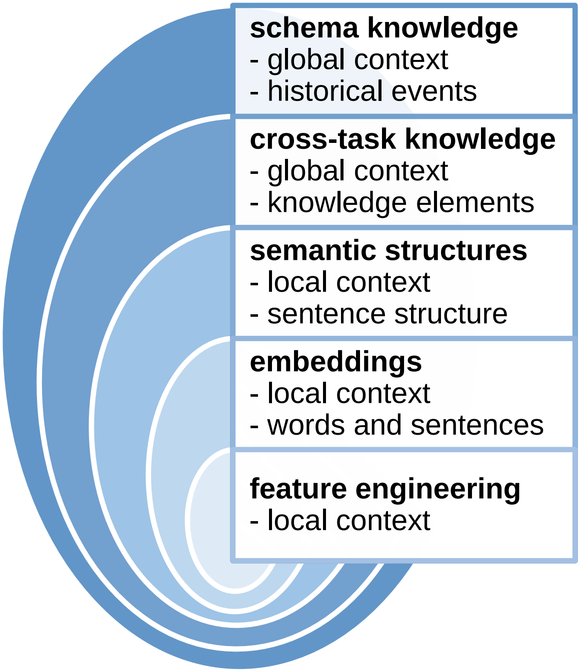
correct ordered decoding: t_1 t_2 $e_{1,1}$ $e_{2,1}$ $e_{1,2}$ $e_{2,2}$ $e_{2,3}$

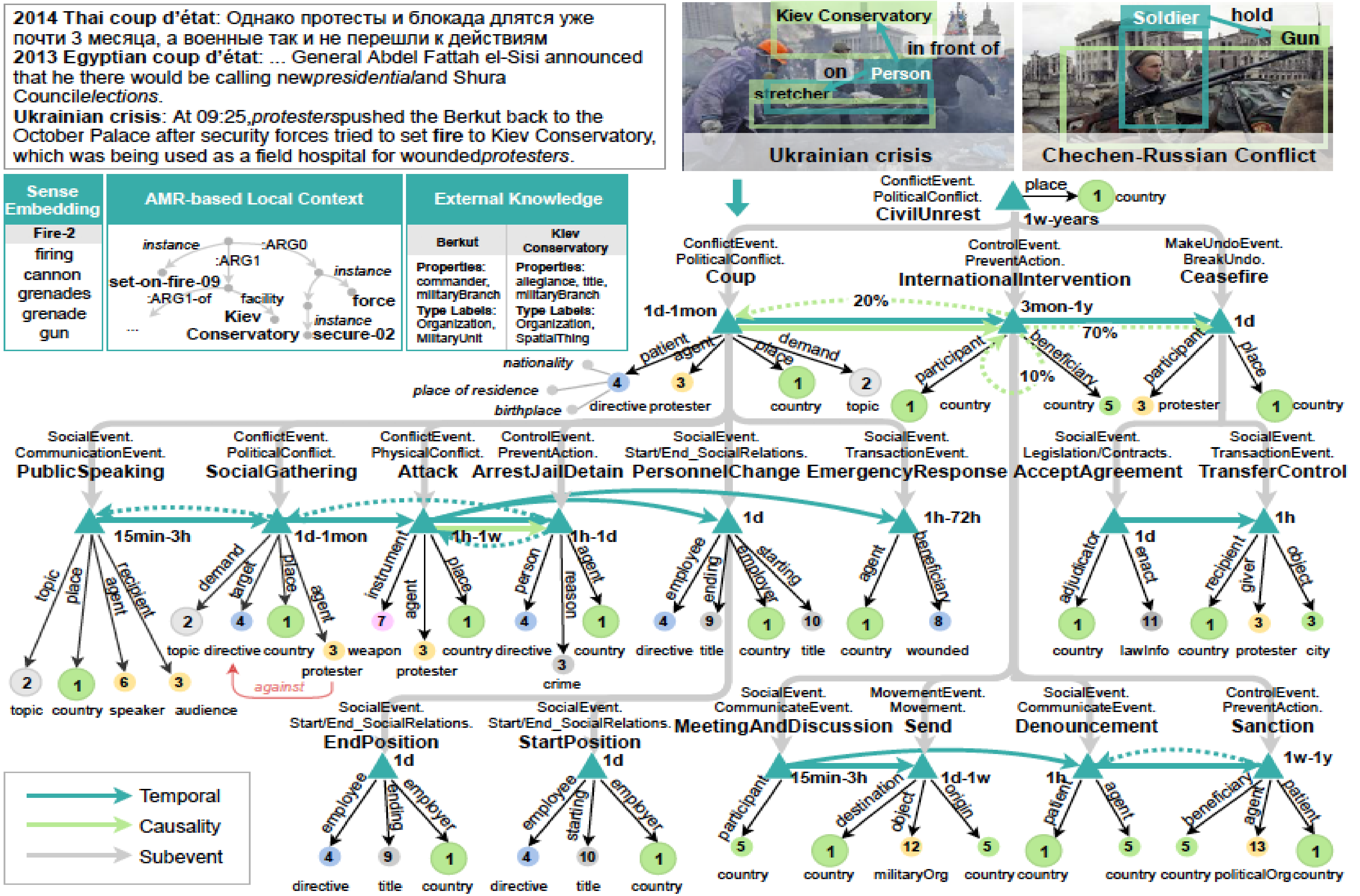
Sentence	AMR Parsing	OneIE outputs	AMR-IE outputs
Russian President **Vladimir Putin**'s **summit** with the **leaders** of Germany and France may have been a failure that proves there can be no long-term "peace camp" alliance following the end of war in **Iraq**.			

Move from Entity-Centric to Event-Centric NLU

- If we look at a wider context when capturing connections between knowledge elements…

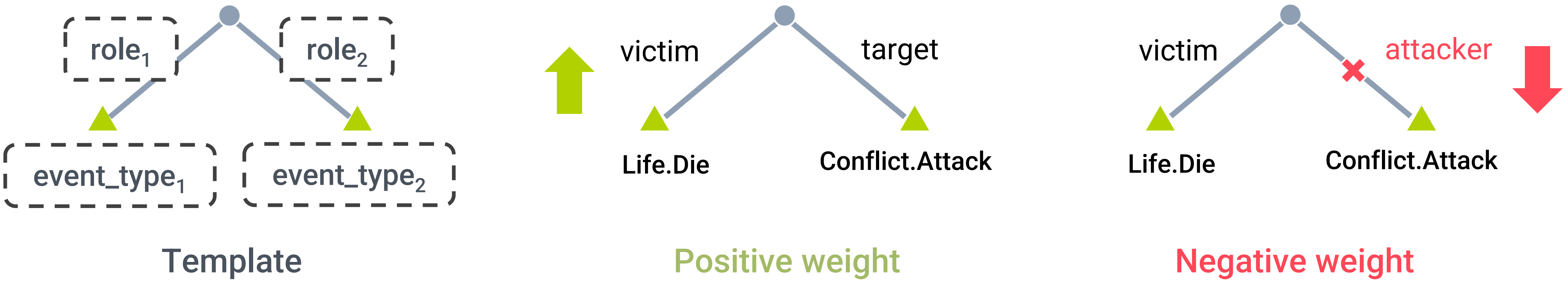
→ Schema knowledge from historical data





Event Schema Induction from Historical Data

- 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, 2020a).
- The model learns the *weight* of each feature during training



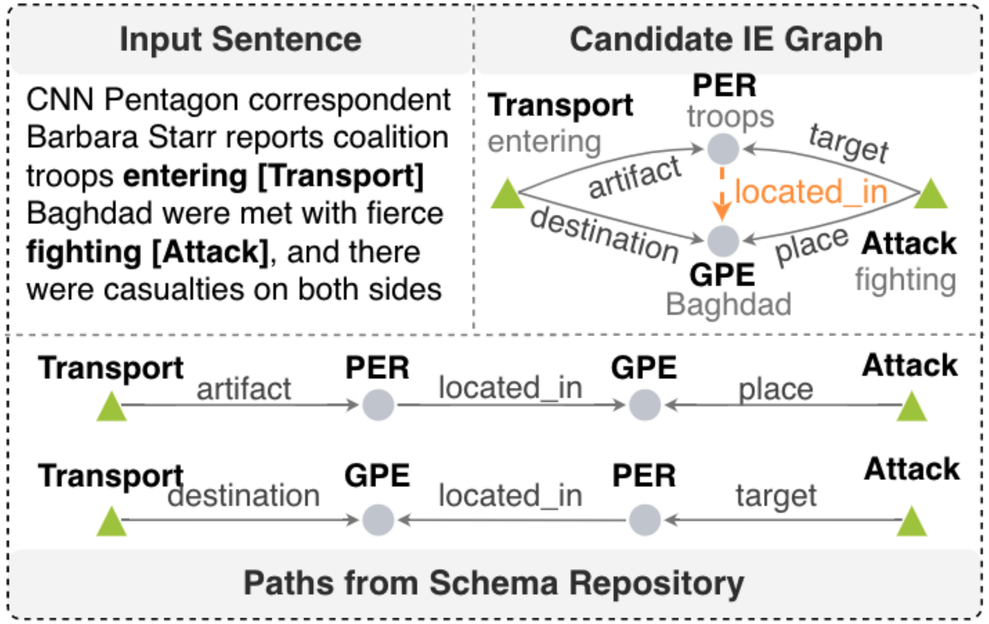
- Global score of a graph: local graph score + global feature score

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

Schema-Guided Information Extraction

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

Dataset	Entity	Event Trigger Identification	Event Trigger Classification	Event Argument Identification	Event Argument Classification	Relation
OneIE base	90.3	75.8	72.7	57.8	55.5	44.7
+PathLM	90.2	**76.0**	**73.4**	**59.0**	**56.6**	**60.9**



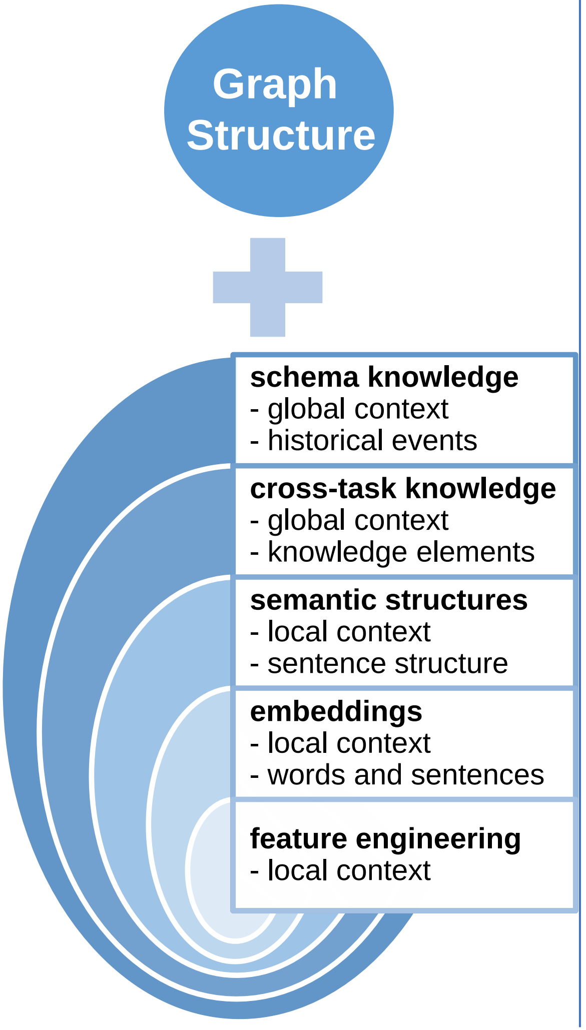
- We evaluate the portability of the proposed framework on ACE05-CN (Chinese) and ERE-ES (Spanish).

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

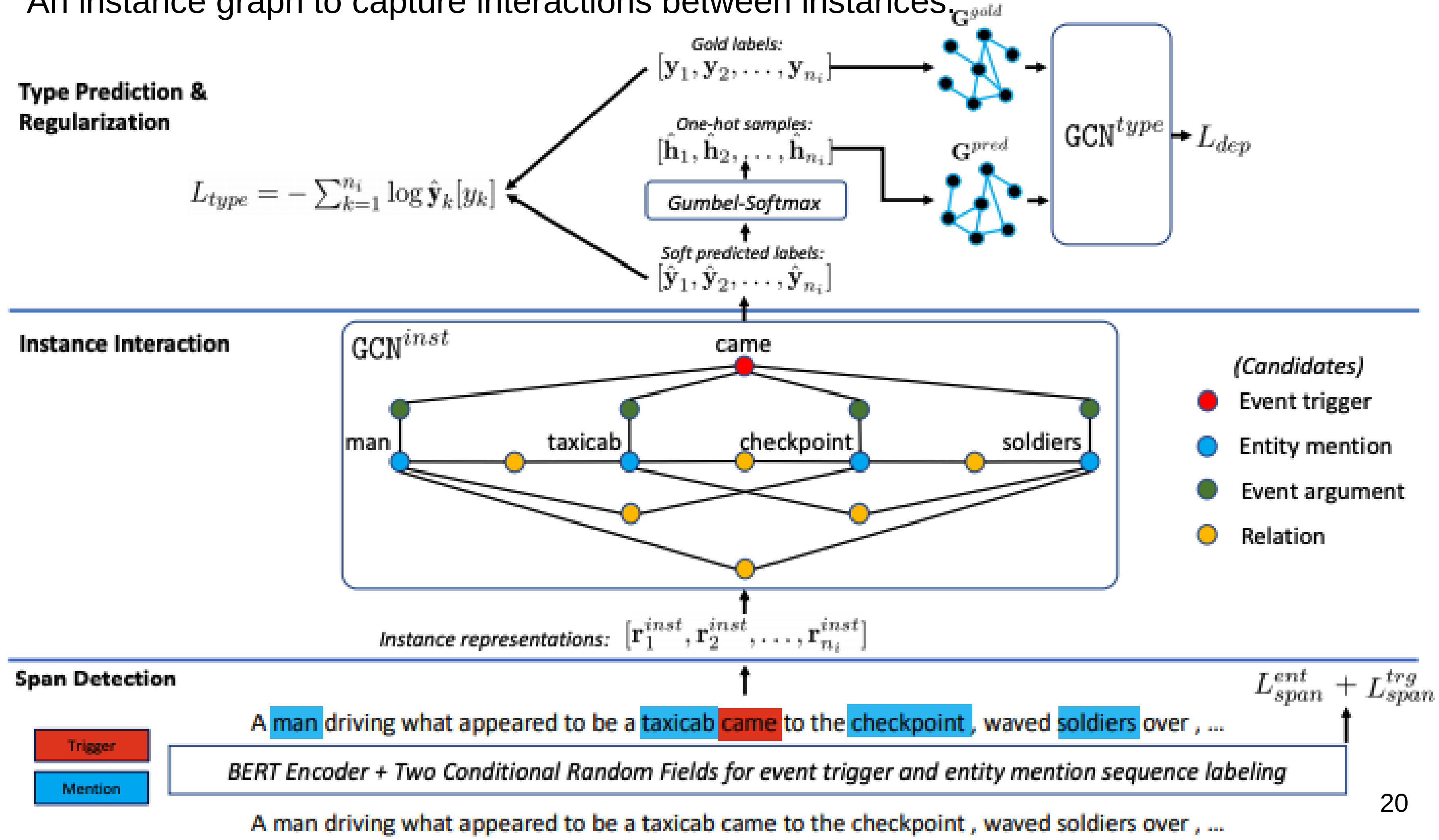
FourIE: Joint Information Extraction with GNN

- To better capture schema knowledge…

→ Adding graph structures

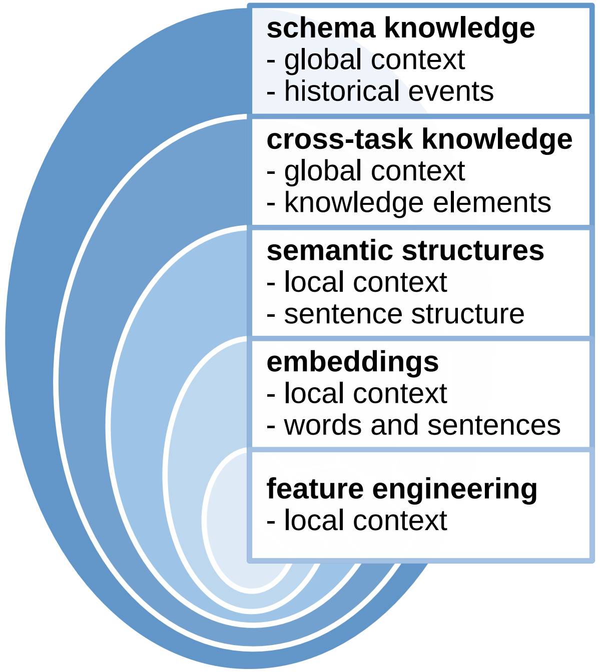


- A type dependency graph for the types in the four IE tasks: entity mention recognition, relation extraction, event trigger detection, argument extraction. (Nguyen et al, 2021)
  - e.g., the Victim of a Die event has a high chance to also be the Victim of an Attack event in the same sentence
- An instance graph to capture interactions between instances.



Summary of Supervised IE (I)

- Encoding wider context improves the IE quality (local → global)
  - A global view of event graph is introduced, to capture the global context of events
- Structure encoding is critical for IE
  - dependency graph, AMR graph, event graph structure, etc.

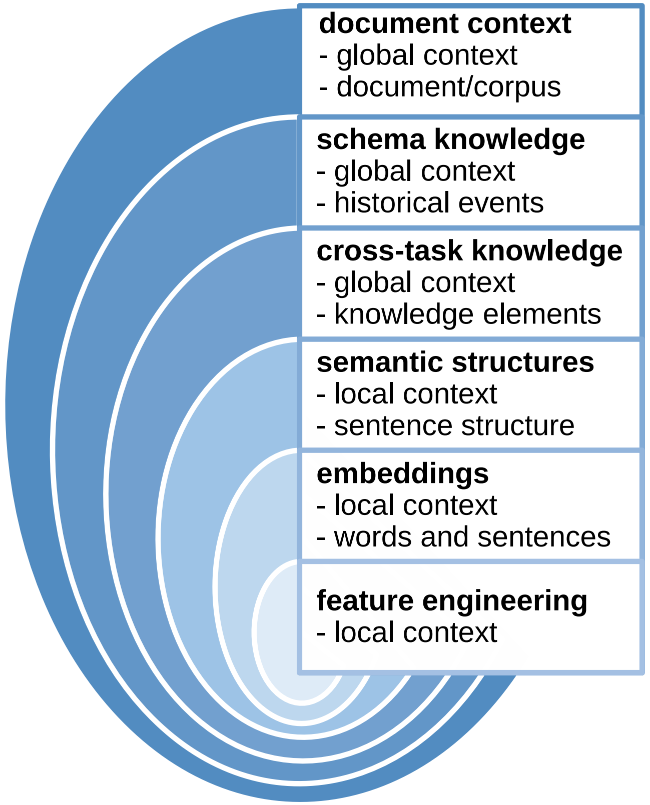


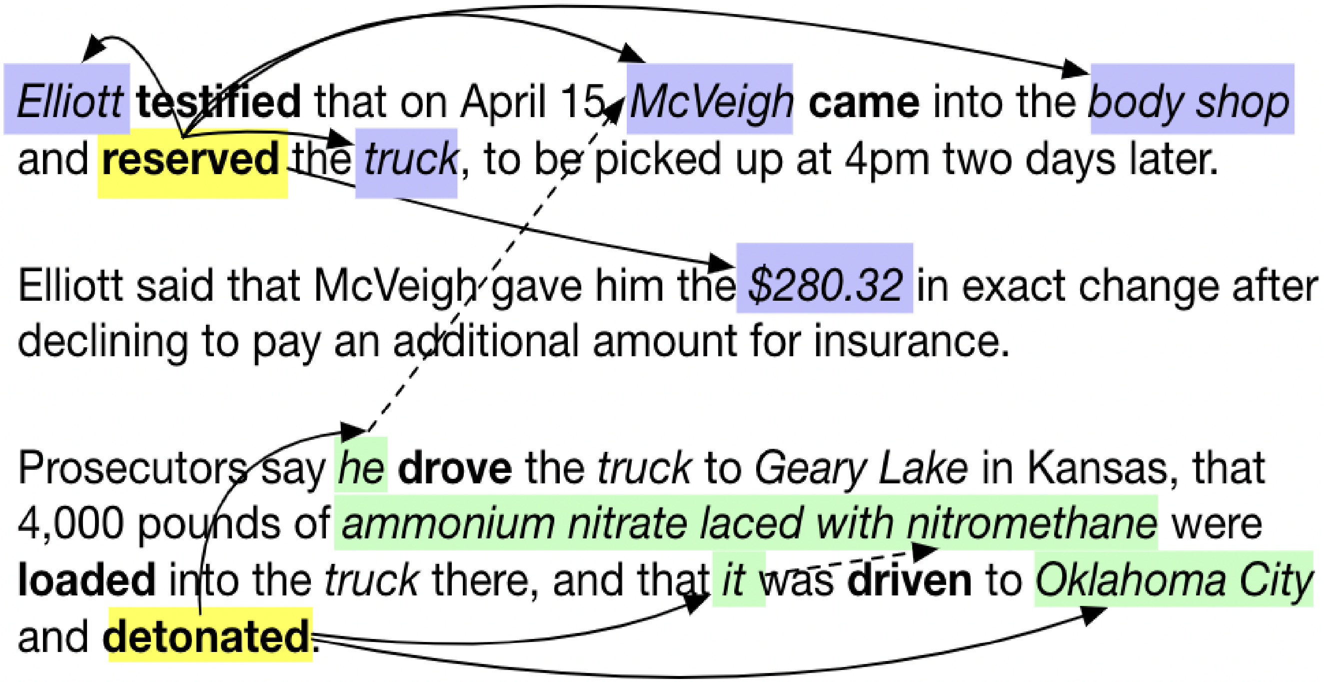
Moving forward…

- Better structure encoding: How to encode the complicated connections between events to guide IE?
  - entity paths between events (events can be walked to each other via entities)
  - horizontal: temporal relations, casual relations, etc
  - vertical: hierarchical relations
- Wider context: How to use schema knowledge?
  - extraction as schema filling task to discover salient events
  - taking advantage of schema probability

Extending from Sentence-Level to Document-Level

- Previous: capturing wider context as features to train a better model
- Is there any limitation during inference? Can we extract events from a wider context?
  → document-level IE, corpus-level IE.

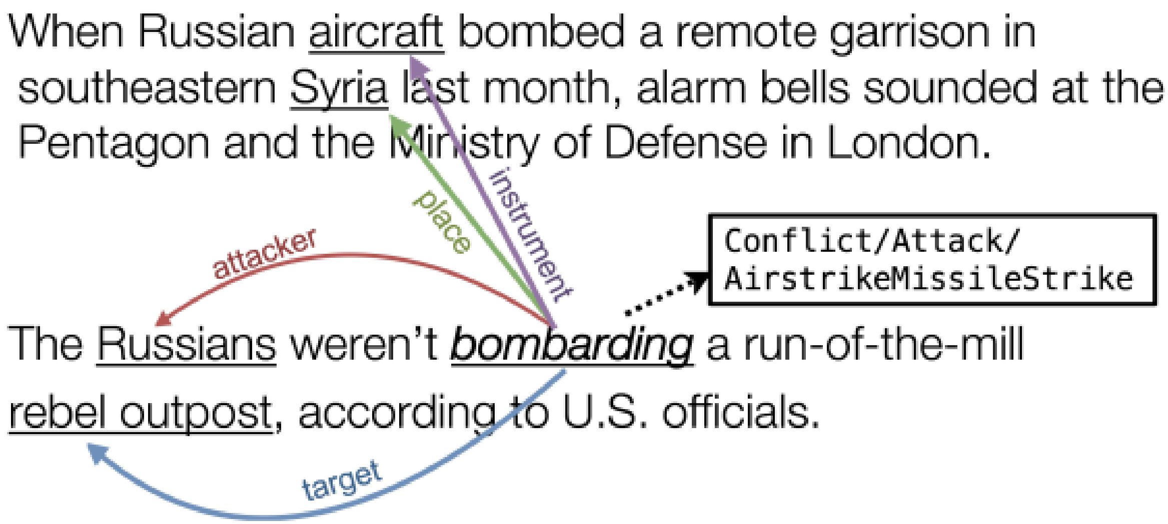


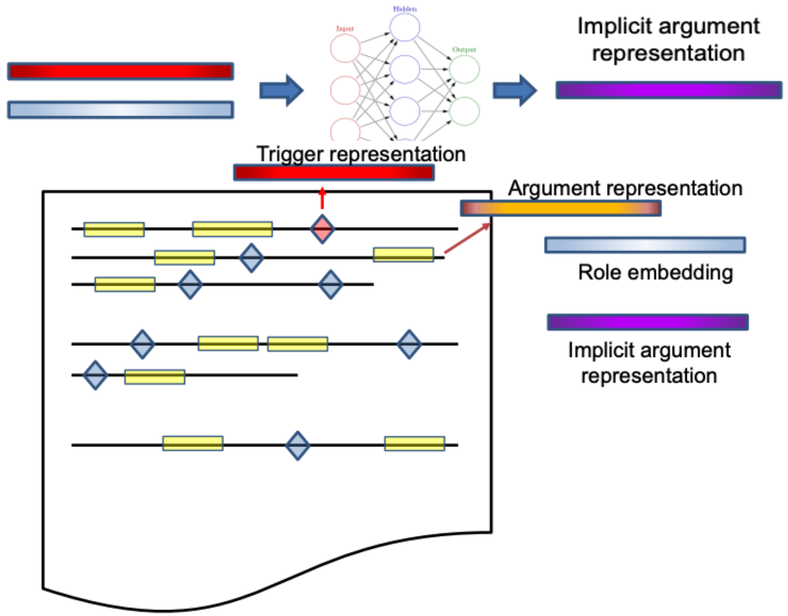


(Li, et al, 2021)

Argument Linking for Document-level IE

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





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

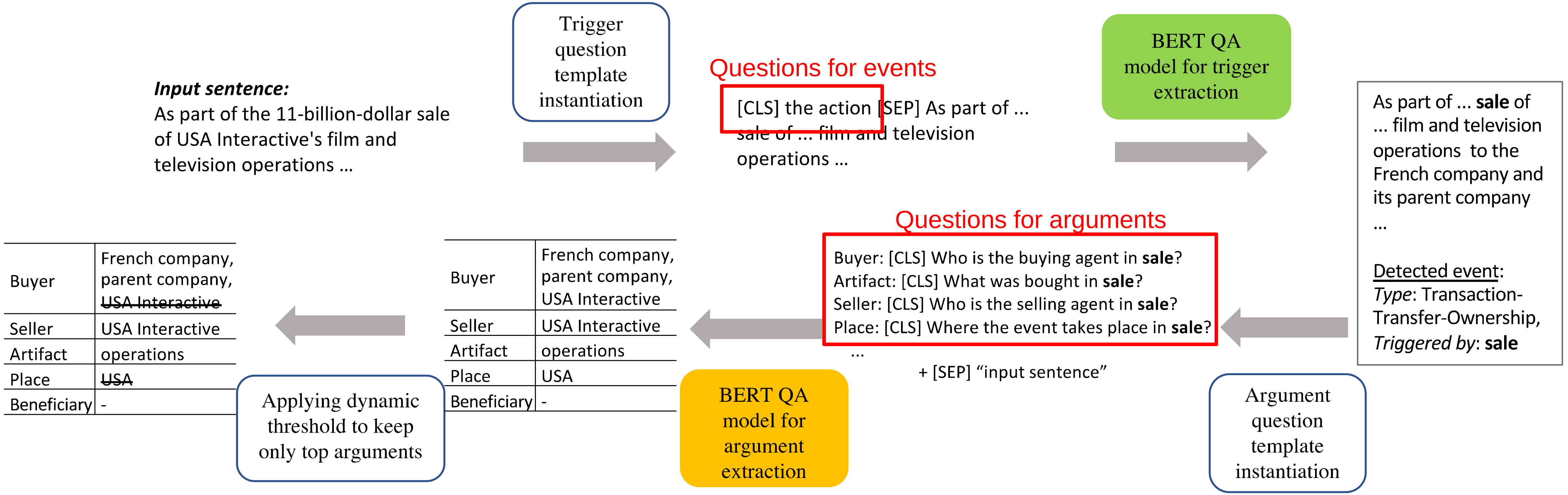
- Roles are evoked by event triggers, forming implicit arguments
- Implicit arguments (roles of each trigger) linked to explicit mentions in text
  - Representations: Learn span representations for each trigger span and candidate argument span
  - Prune: For each trigger, prune to top-K candidate arguments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

Summary of Supervised IE (II) – document-level

	Pros	Cons
Argument Linking	- good explainability based on linking score	- The representation learning for linking score relies on event annotation, which is small and costly
QA-based	- pre-trained language models	

QA-based Event Extraction

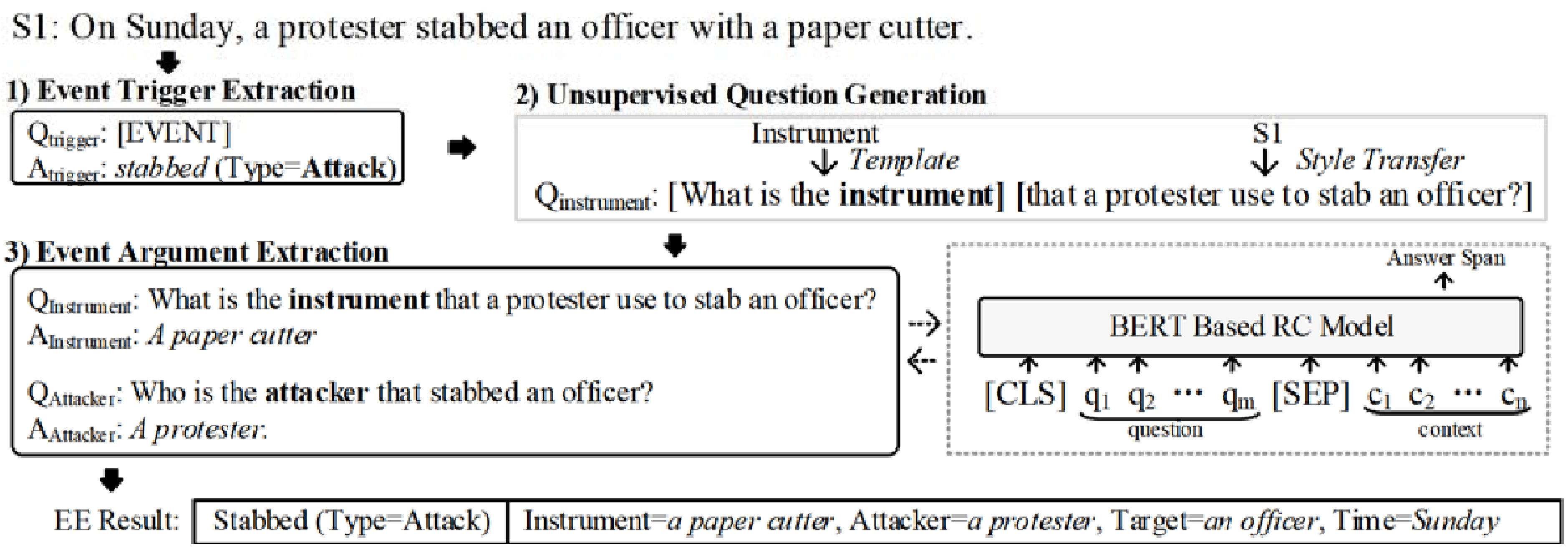
- Questions are constructed based on templates for each role and the predicted answer serves as the extracted argument (Du and Cardie, 2020).



The input sequences for the two QA models share a standard BERT-style format: **[CLS] <question> [SEP] <sentence> [SEP]**

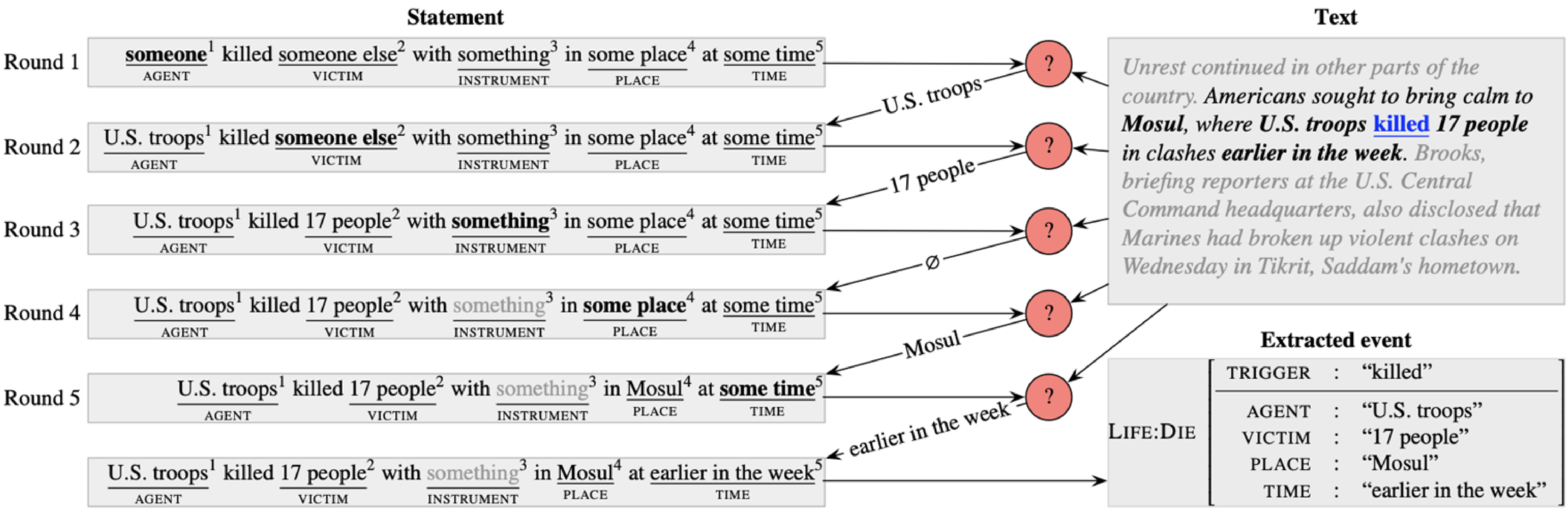
QA-based Event Extraction

- Use a style transfer to make the questions more natural (Liu et al, 2020).



Argument Extraction as Definition Comprehension

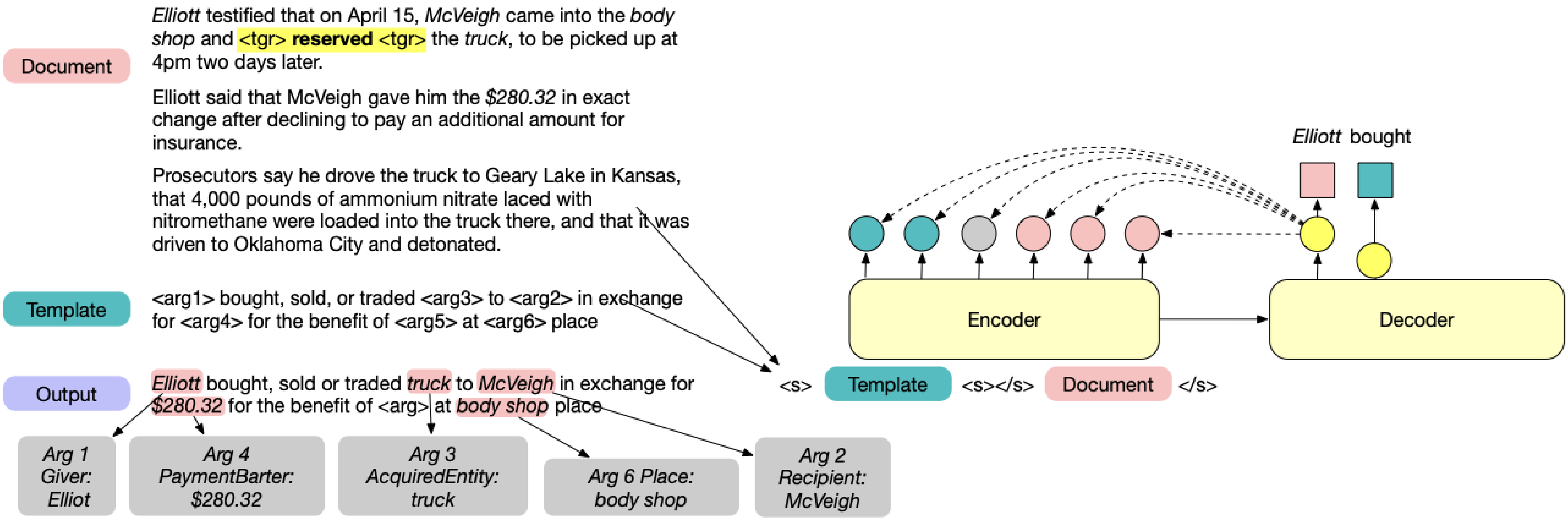
- Using the definition statement as the “question”, the statement is incrementally filled in with the predicted answers (Chen et al. 2020).



Summary of Supervised IE (II) – document-level

	Pros	Cons
Argument Linking	- good explainability based on linking score	- The representation learning relies on limited event annotation
QA-based (Different question generation: template-based, style transfer, definition-based, etc)	- pre-trained language models	- Cannot control the number of arguments for each role
NLG-based		

Argument Extraction as NLG



- Argument extraction is formulated as a **conditional generation** problem with a blank event template as part of the input and a filled template as the desired output (Li et al, 2021).
- For some datasets, this template is readily available as part of the ontology; for others, only one template is needed per event type.
- All arguments for one event can be extracted in a single pass.

Comparison between QA-based and NLG-based

- Unlike the standard QA setting, for the argument extraction task, we often face **missing arguments** and **multiple arguments** for the same role
- **Missing arguments:** output the placeholder token <arg>
- **Multiple arguments:** use "and" to connect the arguments

(" I never said that , " Trump told me . " Yes , he did , " said Smith .) The Japanese bankers **participant** with whom Trump **participant** had negotiated **contact.negotiate.correspondence** a tentative sale suddenly backed off . " The Japanese despise scandal , " one of their associates told me . Several weeks later , Donald called Ivana .

Model output: Trump and Japanese bankers communicated remotely about <arg> topic at <arg> place

Comparison between QA-based and NLG-based

Context	Role	Ours	BERT-CRF	BERT-QA
I have been in touch (E1:Contact.Contact.Correspondence) with the NDS **official** in the province and they told me that over 100 **members** of the NDS were killed(E2:Life.Die) in the big explosion , " the former provincial official said . Sharif Hotak , a member of the provincial council in Maidan Wardak said **he** saw (E3:Cognitive.IdentifyCategorize) **bodies** of 35 Afghan forces in the **hospital**."	E1: Participant	I NDS official	I official	NDS I official NDS official
	E2: Victim	members	members	members
	E3: Identifier	he	N/A	N/A
	E3: IdentifiedObject	bodies	N/A	N/A
	E3: Place	hospital	N/A	N/A

- For the Contact event E1, BERT-QA over-generates answers for the participant span.
  - QA models produce a ranking over possible answers, producing the optimal threshold is hard
- For the IdentifyCategorize event, only our model can successfully extract all arguments.
  - Sequence labeling model struggle with types with few training examples

Summary of Supervised IE (II) – document-level

	Pros	Cons
Argument Linking	- good explainability based on linking score	- The representation learning relies on event annotation, which is small and costly
QA-based (Different question generation: template-based, style transfer, definition-based, etc)	- pre-trained language models	- Cannot control the number of arguments for each role
NLG-based	- pre-trained language models - deal with missing arguments and multiple arguments	- Lack of ontological constraints - Without schema knowledge

Summary of Supervised IE (II) – document-level

Argument Linking

- linking score is based on span representations for triggers and arguments, trained on annotations only

QA based

- take advantage of pre-trained language models trained from large-scale corpus

NLG based

- is able to handle missing arguments and multiple arguments

Corpus-level IE?

Moving forward…

- How to encode ontological constraints?
- How to incorporate schema knowledge?
- How to resolve coreference during extraction?

Outline

Quality

wider context to extract

Schema-guided IE

- How to capture the **graph structure** of the IE graph?

→ dependency graph, AMR graph, etc.

Joint IE

- From these **graph structures of historical data**, are there general patterns to guide IE?

→ schema induction

- Understanding connections between events requires **inference across sentences**

→ document-level IE (Argument linking, QA based , NLG based)

Document-level IE

wider context as features

Portability

Cross-domain

- **Structure transfer**: AMR graph ←→ event structure
- **Common space:** To generate informative embeddings for labels, we should seek information from **definition, example sentences**, etc

Cross-lingual

- **Structure transfer**: universal dependency ←→ event structure
- **Common space:** Alignment should be **multi-level** (word-level, entity-level)

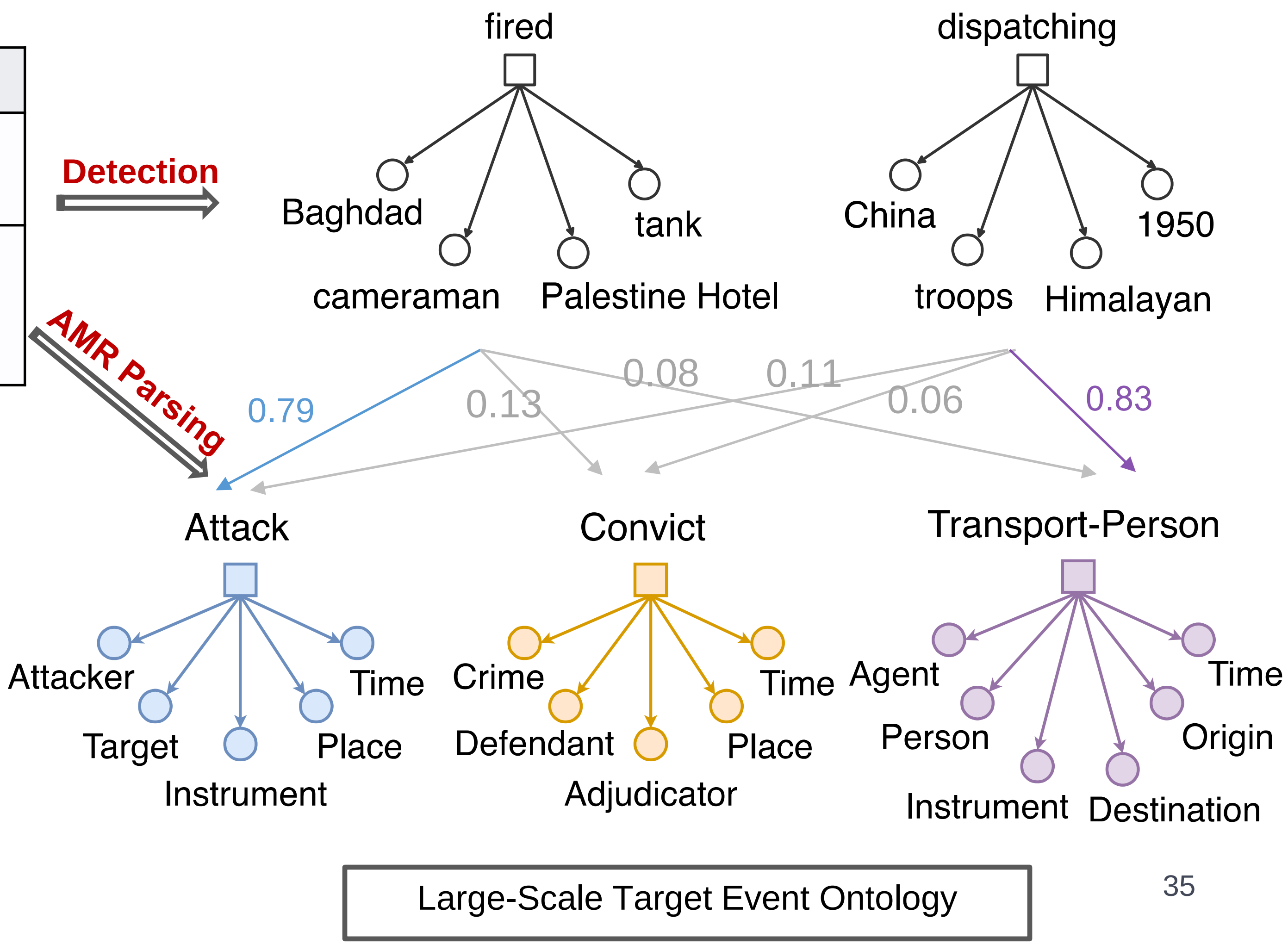
Cross-media

- **Structure transfer:** text AMR graph ←→ image situation graph
- **Common space:** Alignment should be **multi-level** (word-level, image-level)

Move to any New Types: AMR Graph Structure Transfer

- General way for transfer learning: building a shared representation (semantic common space).
- Structure transfer is the key since event extraction highly relies on structures.
  - Cross-domain structure transfer: AMR graph ←→ event structure

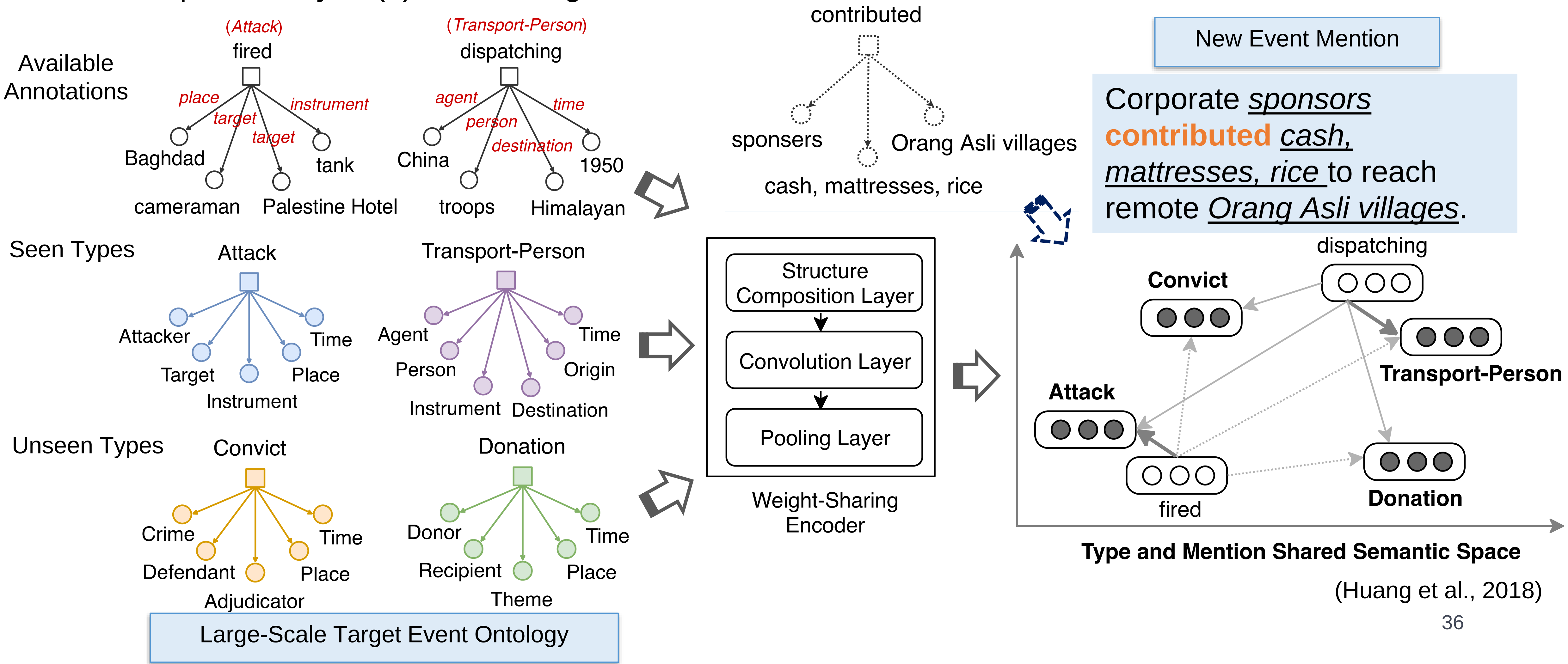
ID	Sentences
S 1	In _Baghdad_, a _cameraman_ **died** when a combat _tank_ **fired** on the _Palestine Hotel_.
S 2	The government of _China_ has ruled Tibet since 1951 after **dispatching** _troops_ to the _Himalayan_ region in _1950_.



Hypothesis: Both event mentions and types (ontology) have rich semantics and structures, which can specify their consistency and connections (Huang et al., 2018)

Zero-shot Event Extraction by embedding AMR graphs

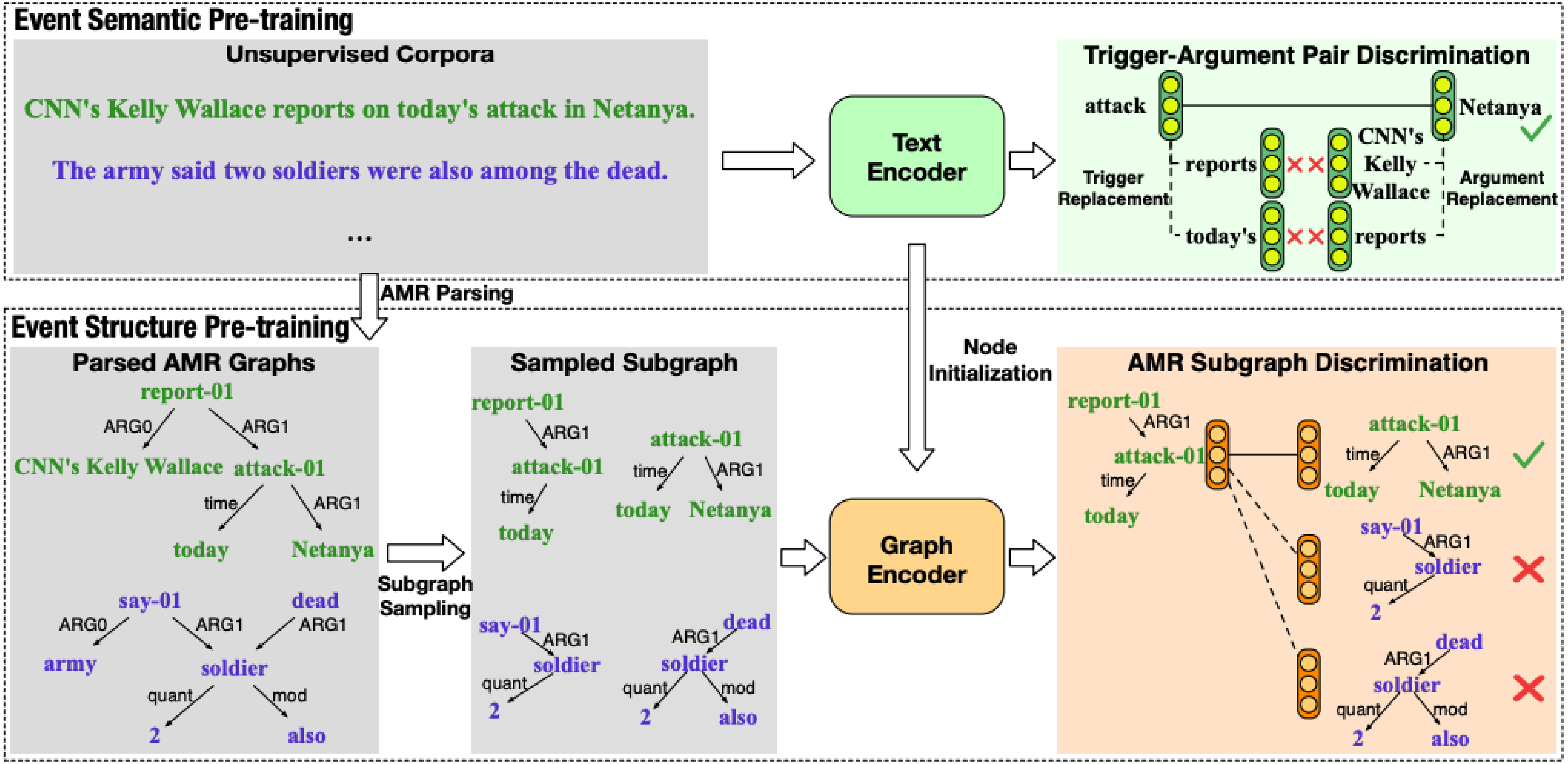
- Structure composition layer: event structure is composed by <trigger, role, argument> triples.
- Cons: (1) It can not capture longer distance between arguments, due to the simple structure composition layer. (2) The training data is limited.



(Huang et al., 2018)

Contrastive Pre-training for Event Extraction (CLEVE)

- To train a more structure-aware common space with large-scale dataset, contrastive learning is proposed to represent the words of the same events closer than those unrelated words; Graph contrastive pre-training to learn event structure representations on event related AMR structures (Wang et al, 2020).



Summary of Portability Challenges

Portability

Cross-domain

Cross-lingual

Cross-media

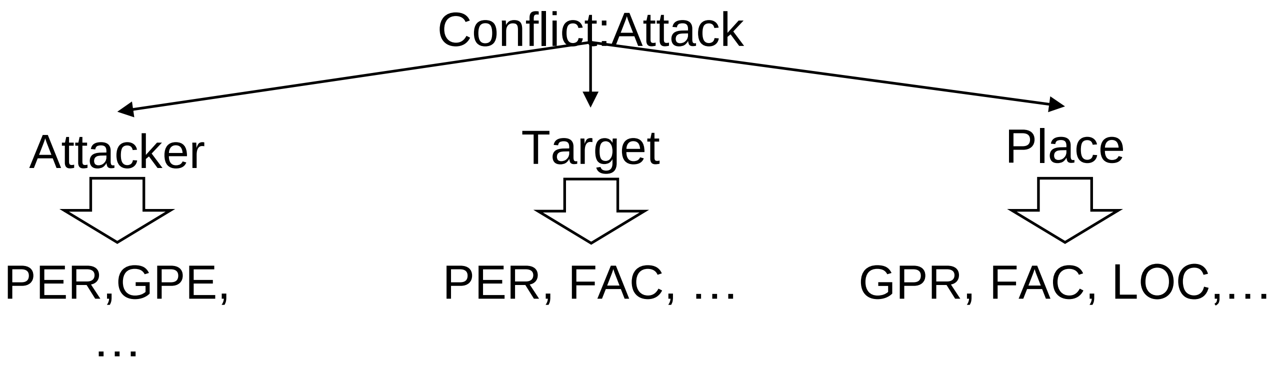
Structure Transfer

AMR Graph
|
Event Structure

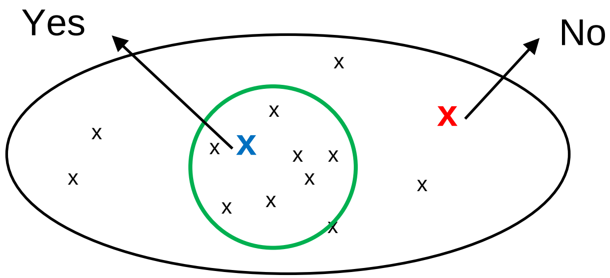
Semantic Common Space

Adding Ontology Info:
definitions, example sentences, constraints, etc.

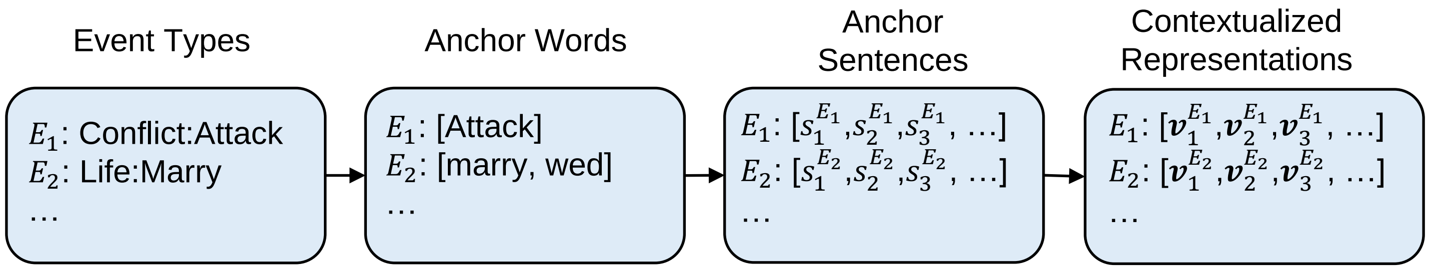
How to encode the rich information from event ontology?



- Event ontology contains rich semantics to generate more informative label (event type) embeddings (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".

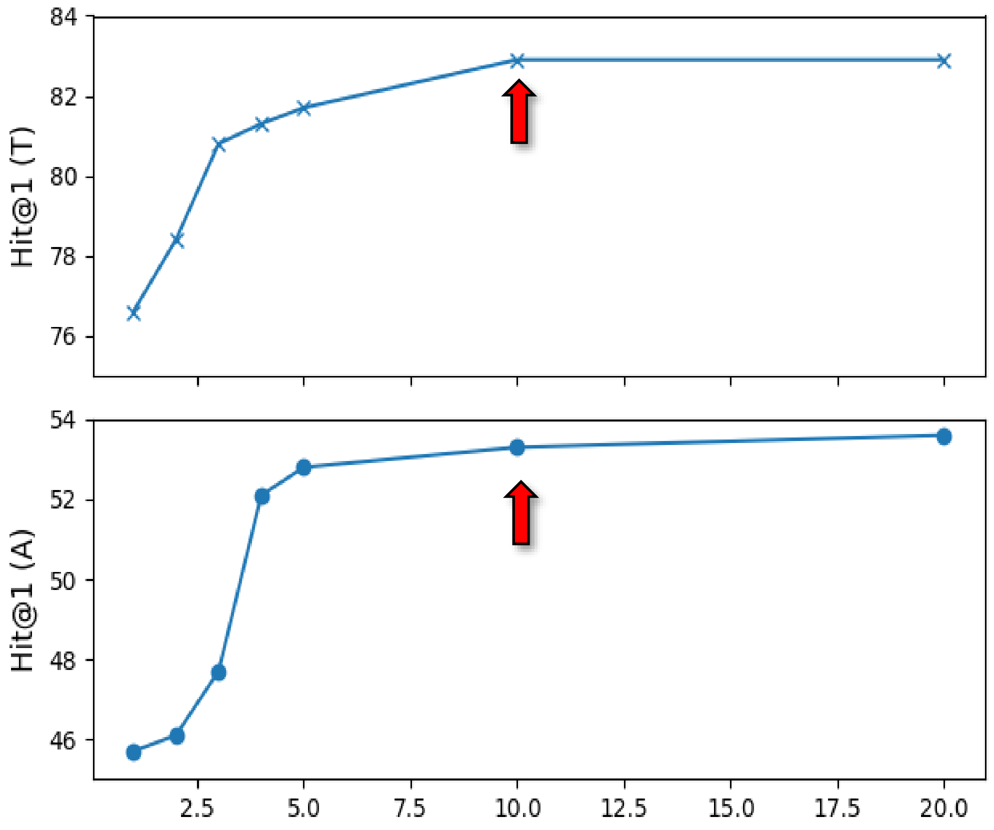


Use a cluster of contextualized embeddings to represent labels and use constraints to regularize the predictions by modeling it as an ILP problem.



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!!

Summary of Portability Challenges

Portability

Cross-domain	Cross-lingual	Cross-media

Structure Transfer

AMR Graph
|
Event Structure

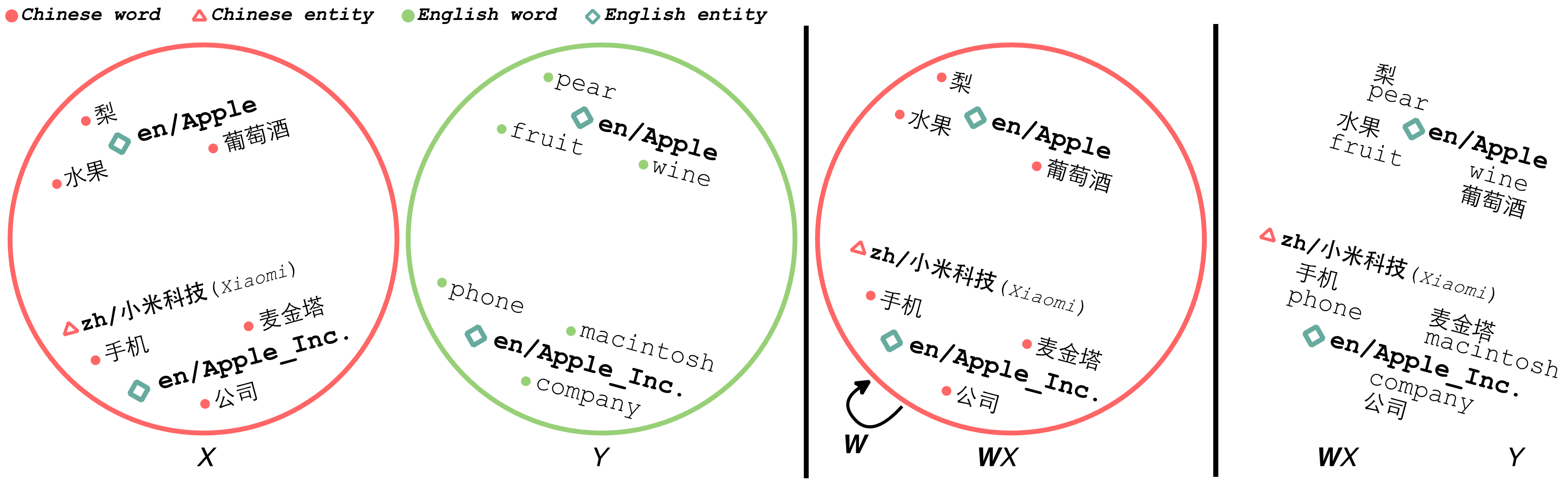
Semantic Common Space

Adding Ontology Info:
definitions, example sentences, constraints, etc.

Cross-lingual Joint Entity and Word Embedding Learning

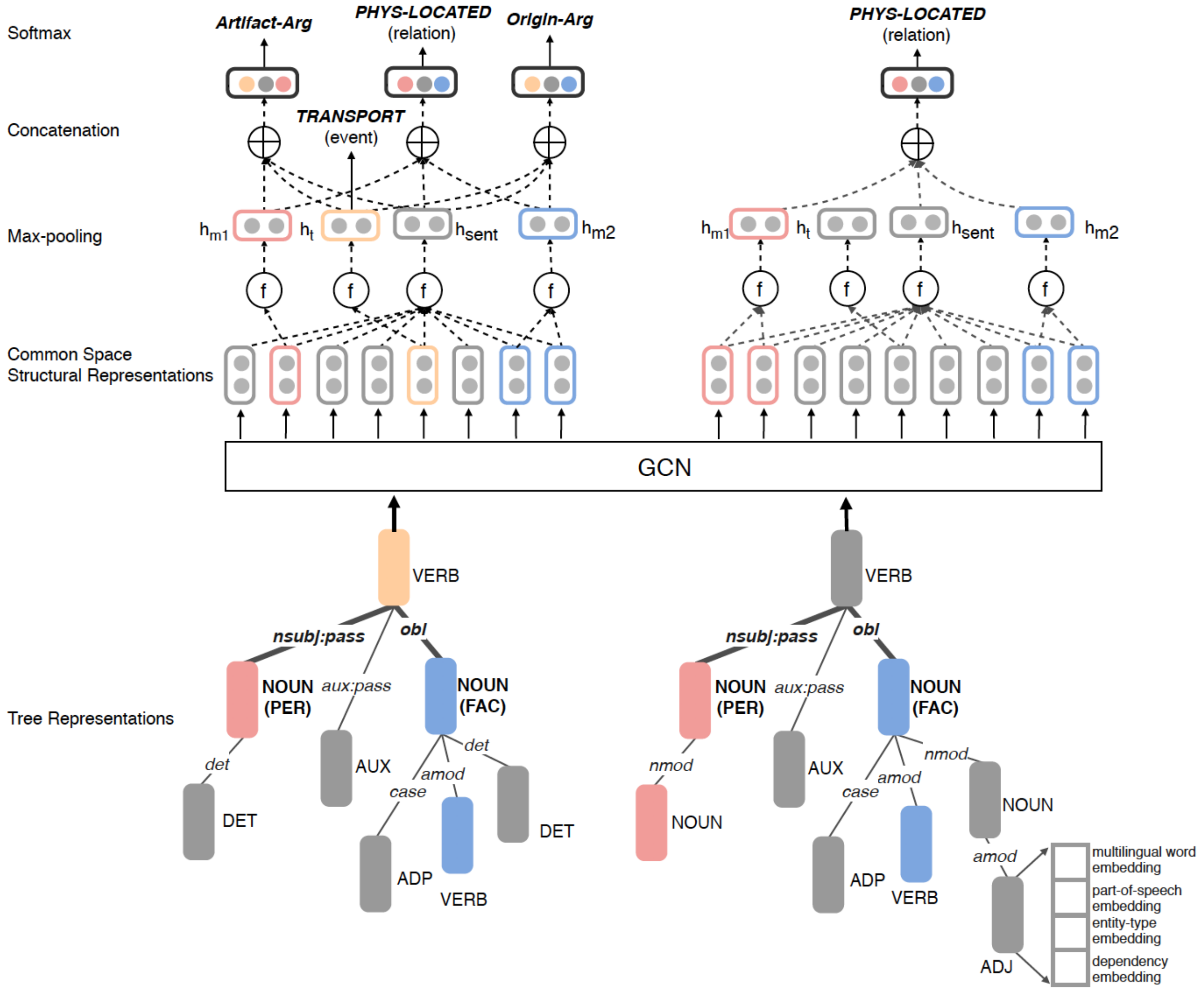
- Cross-lingual Joint Entity and Word Embedding to Improve Entity Linking and Parallel Sentence Mining (Pan et al., 2019)
  - Code-switch cross-lingual entity/word data generation

  - 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)
  - Dependency substructures covering trigger and arguments are similar across languages
  - Universal Dependency Parser (Straka and Strakova, 2017)
    - Agnostic to language word order
    - Capturing long-distance arguments
  - Cons: However, GCNs struggle to model words with long-range dependencies or are not directly connected in the dependency tree

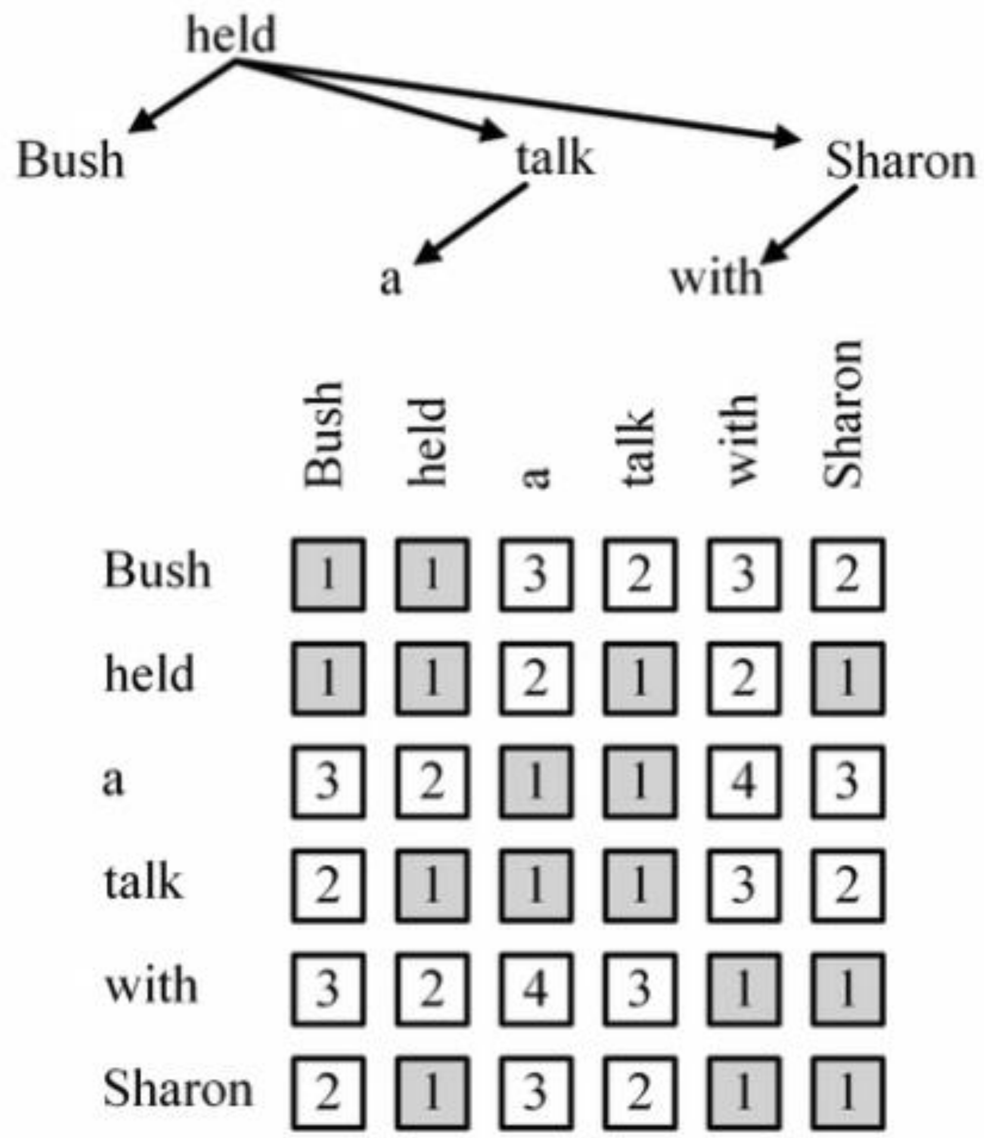


The detainees were taken to a processing center

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

Graph Attention Transformer Encoder (GATE)

- Use pairwise syntactic distances to model attentions between tokens. (Ahmad, et al, 2020).
- Distance matrix shows the shortest path distances between all pairs of words.
- Self-attention of Transformer is guided by the dependency tree distance:
  - Attend tokens that are within certain distance.



Summary of Portability Challenges

Portability

Cross-domain

Cross-lingual

Cross-media

Structure Transfer

AMR Graph
|
Event Structure

Universal Dependency Graph
|
Event Structure

Semantic Common Space

Adding Ontology Info: definitions, example sentences, constraints, etc.

Multi-level Alignment: <word-word>, <entity-entity>

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

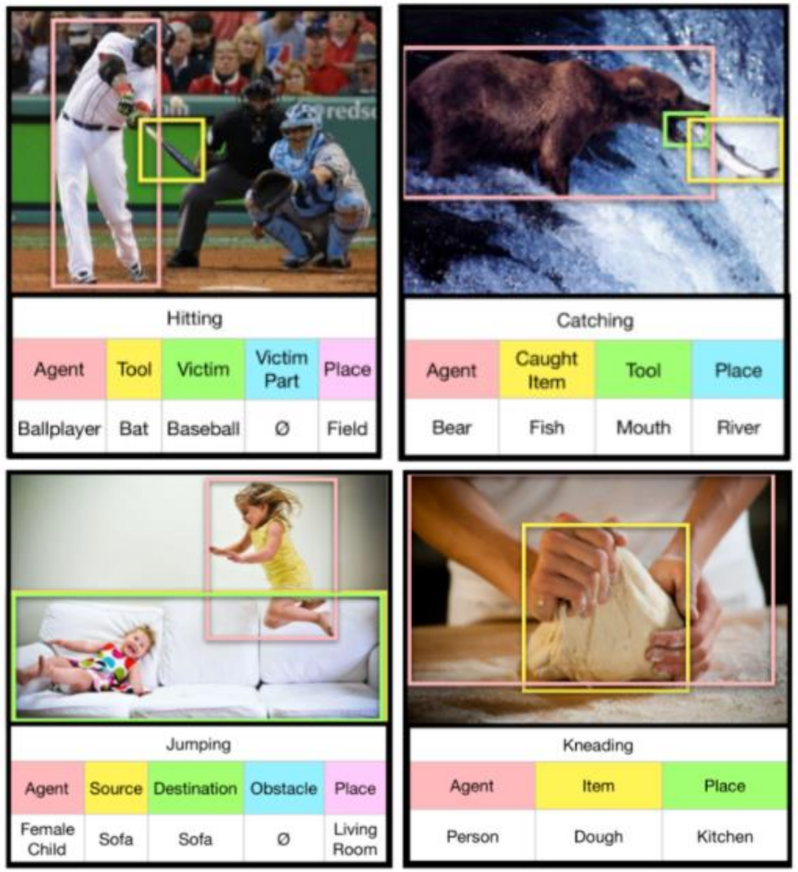
ROLE	VALUE	ROLE	VALUE
AGENT	BOY	AGENT	BEAR
SOURCE	CLIFF	SOURCE	ICEBERG
OBSTACLE	-	OBSTACLE	WATER
DESTINATION	WATER	DESTINATION	ICEBERG
PLACE	LAKE	PLACE	OUTDOOR

SPRAYING

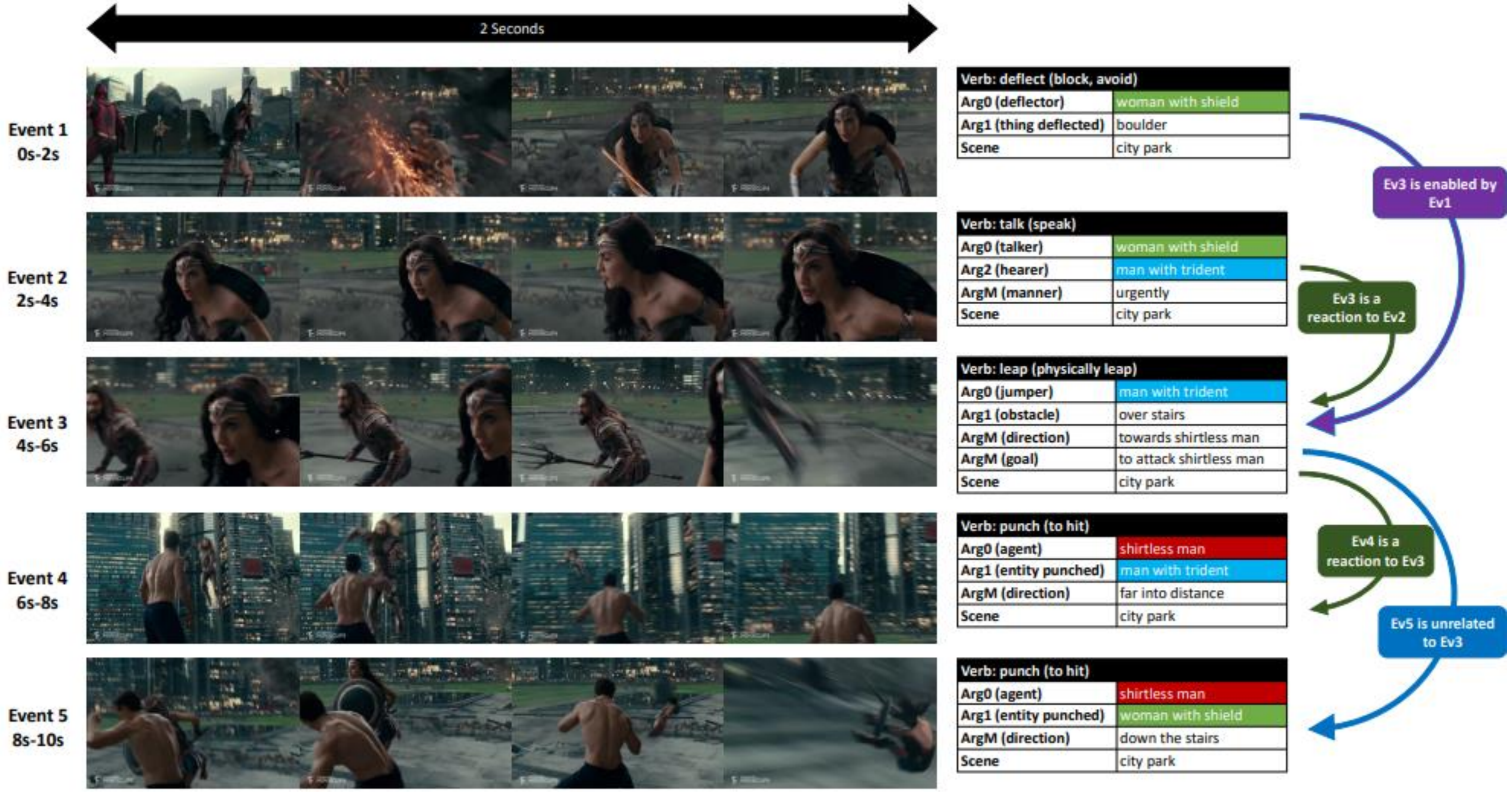
ROLE	VALUE	ROLE	VALUE
AGENT	MAN	AGENT	FIREMAN
SOURCE	SPRAY CAN	SOURCE	HOSE
SUBSTANCE	PAINT	SUBSTANCE	WATER
DESTINATION	WALL	DESTINATION	FIRE
PLACE	ALLEYWAY	PLACE	OUTSIDE

Vision-only Event and Argument Extraction

- Grounded Situation Recognition adds visual argument localization [Pratt et al, 2020]

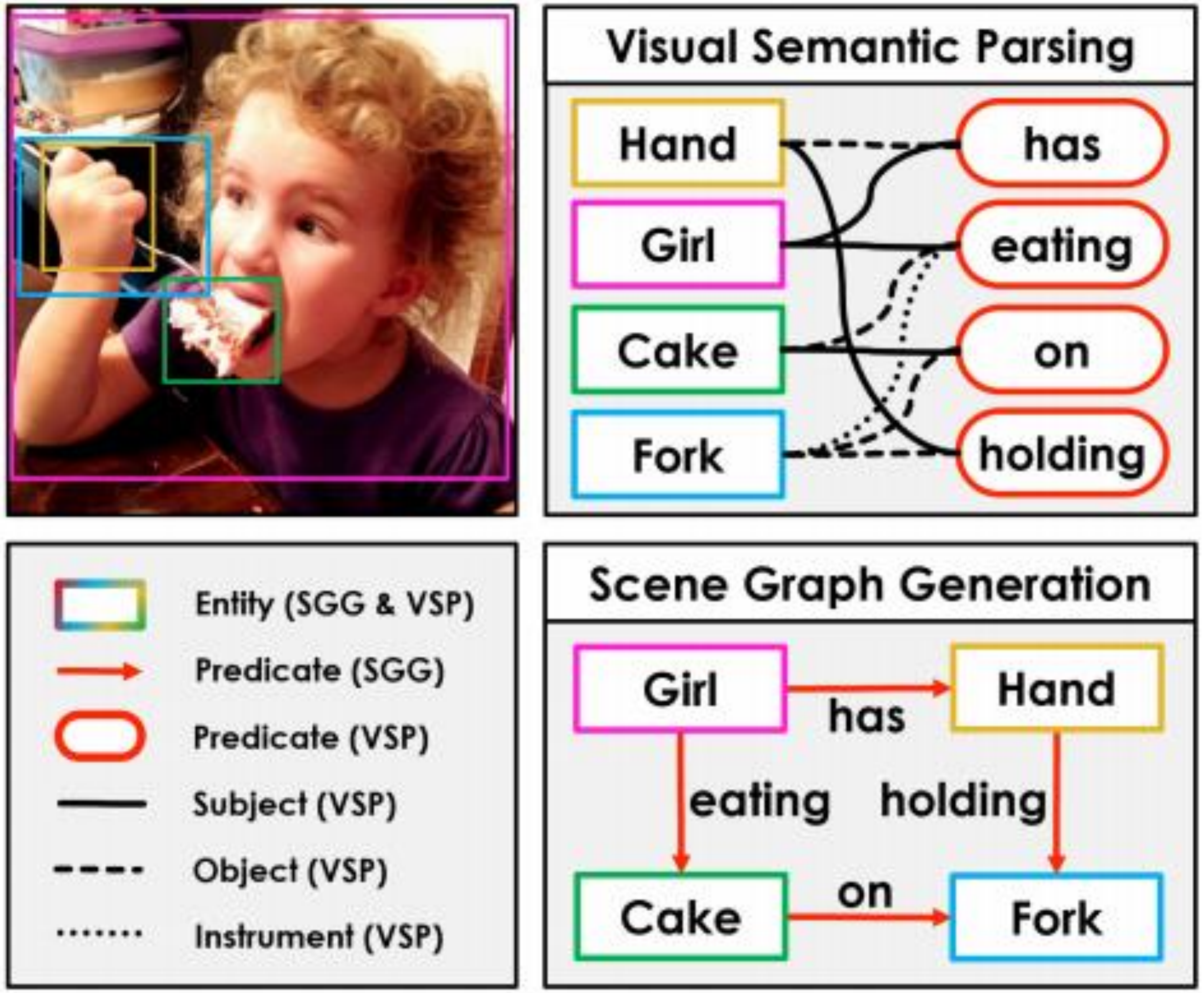


- Video Situation Recognition extends the work to videos [Sadhu et al, 2021]



Vision-only Event and Argument Extraction

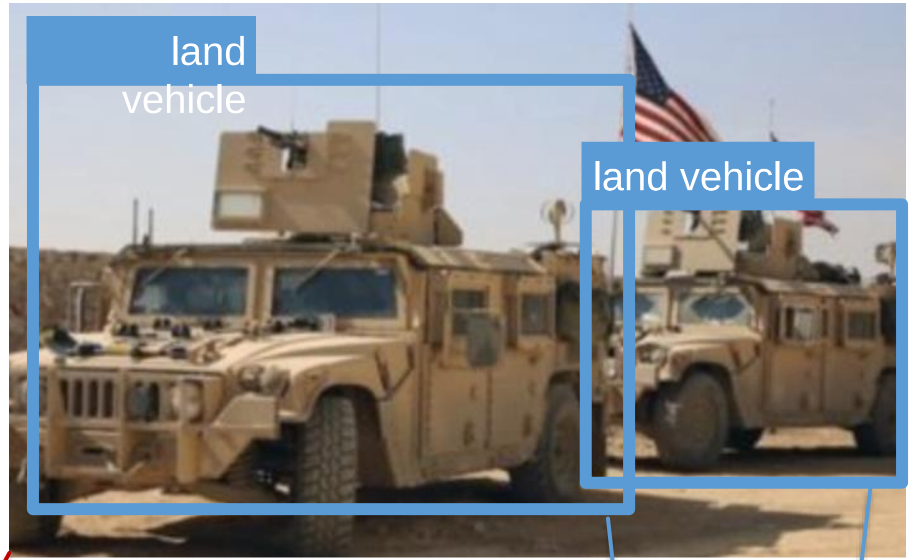
- Another line of work is based on scene graphs [Xu et al, 2017; Li et al, 2017; Yang et al, 2018; Zellers et al, 2018].
  - extracting <subject, predicate, object>
  - structure is simpler than the aforementioned multi-argument event
- Visual Semantic Parsing is using predicate as event, and subject, object, instrument as arguments [Zareian el al, 2020]
  - Add bounding box grounding



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

Input: News Article Text and Image (The first task to take both modalities as input)

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	

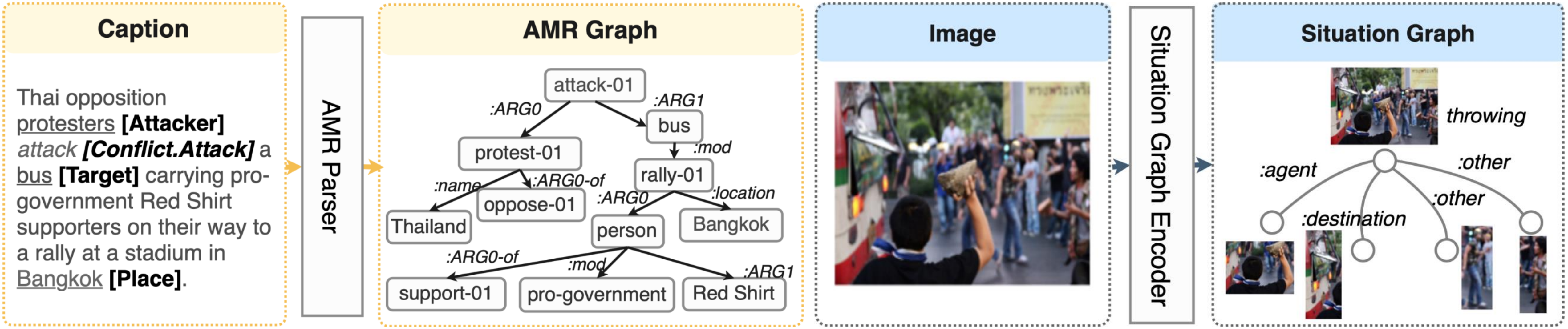
Cross-media Structured Common Space

- 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

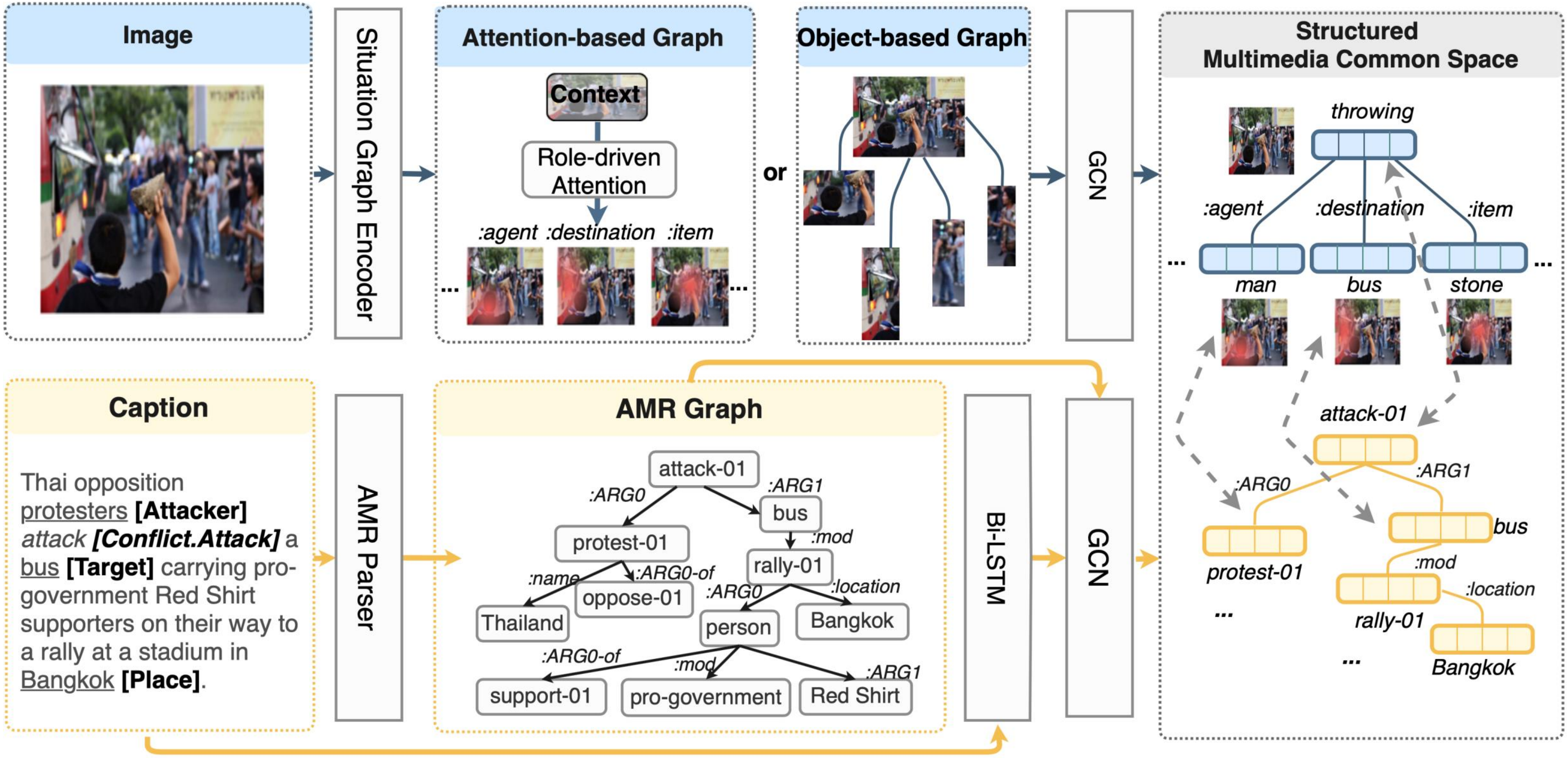
- Treat Image/Video as a foreign language
  - Represent it with a structure that is similar to AMR graph in text



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

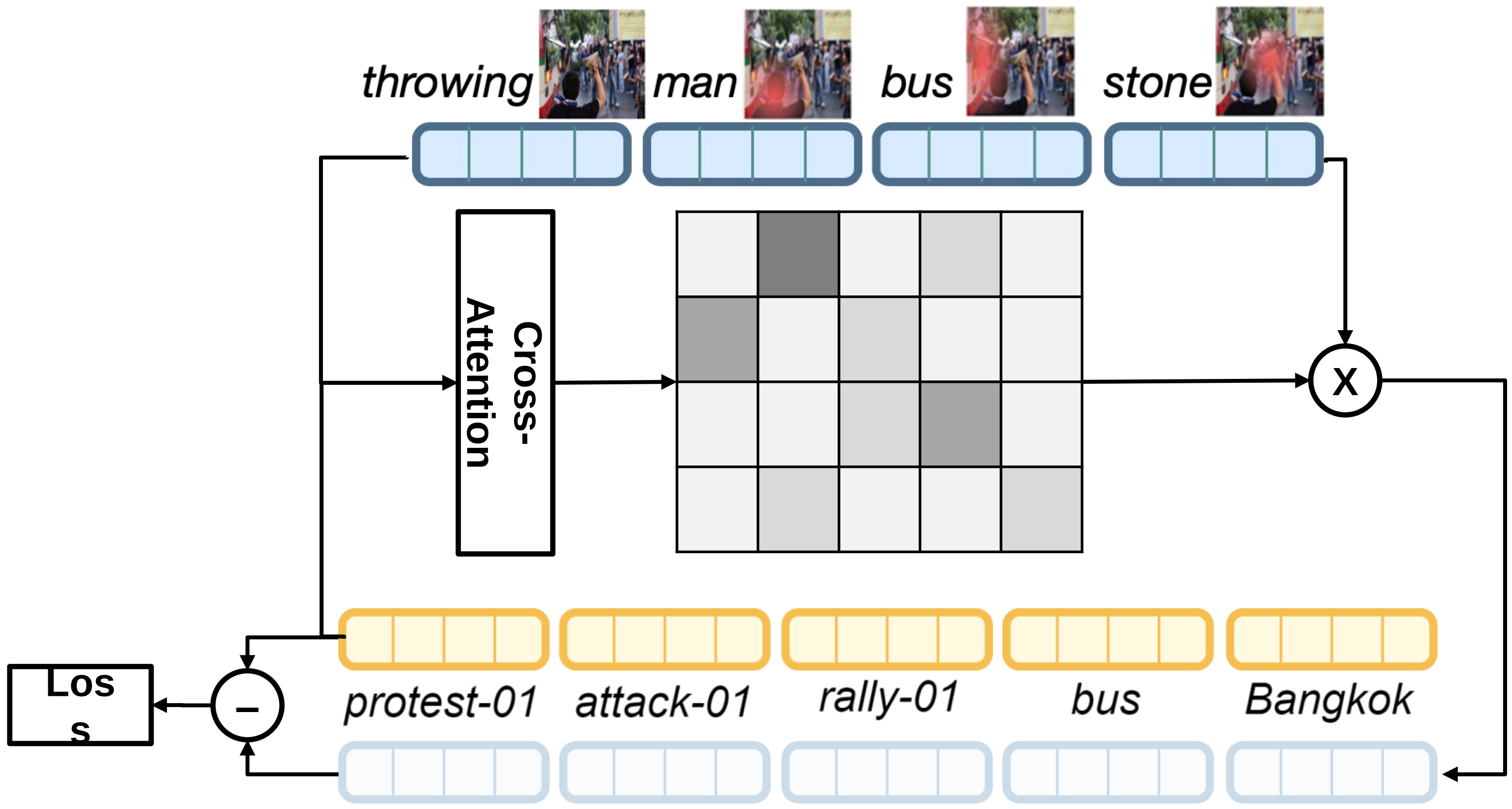
Weakly Aligned Structured Embedding (WASE)

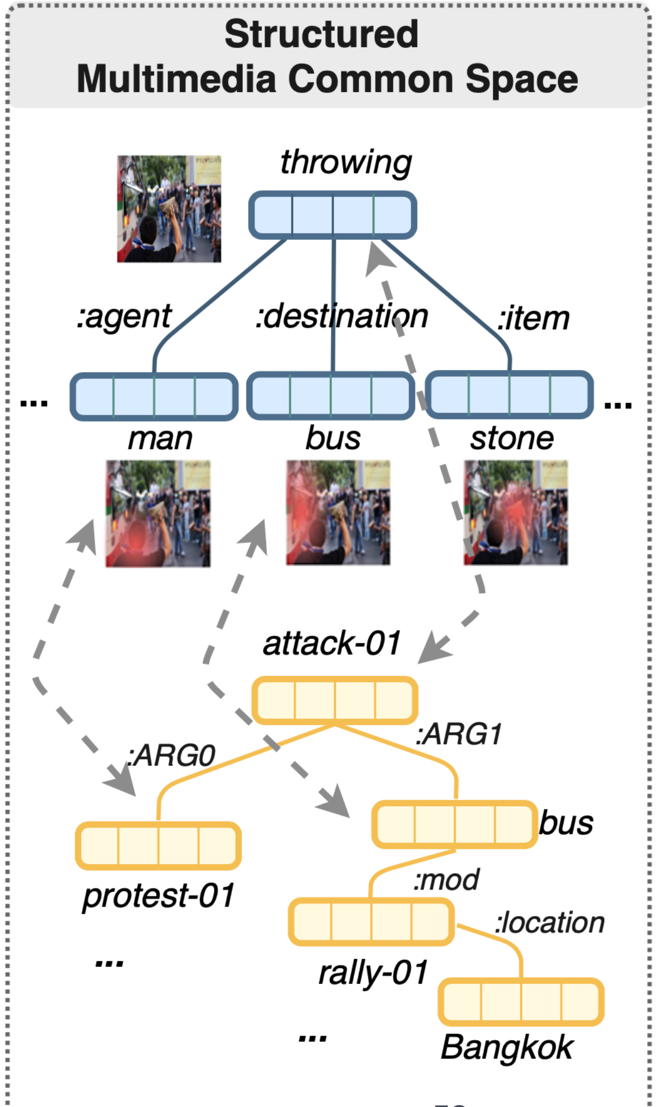
-- Training Phase (Common Space Construction)



How to align the two modalities?

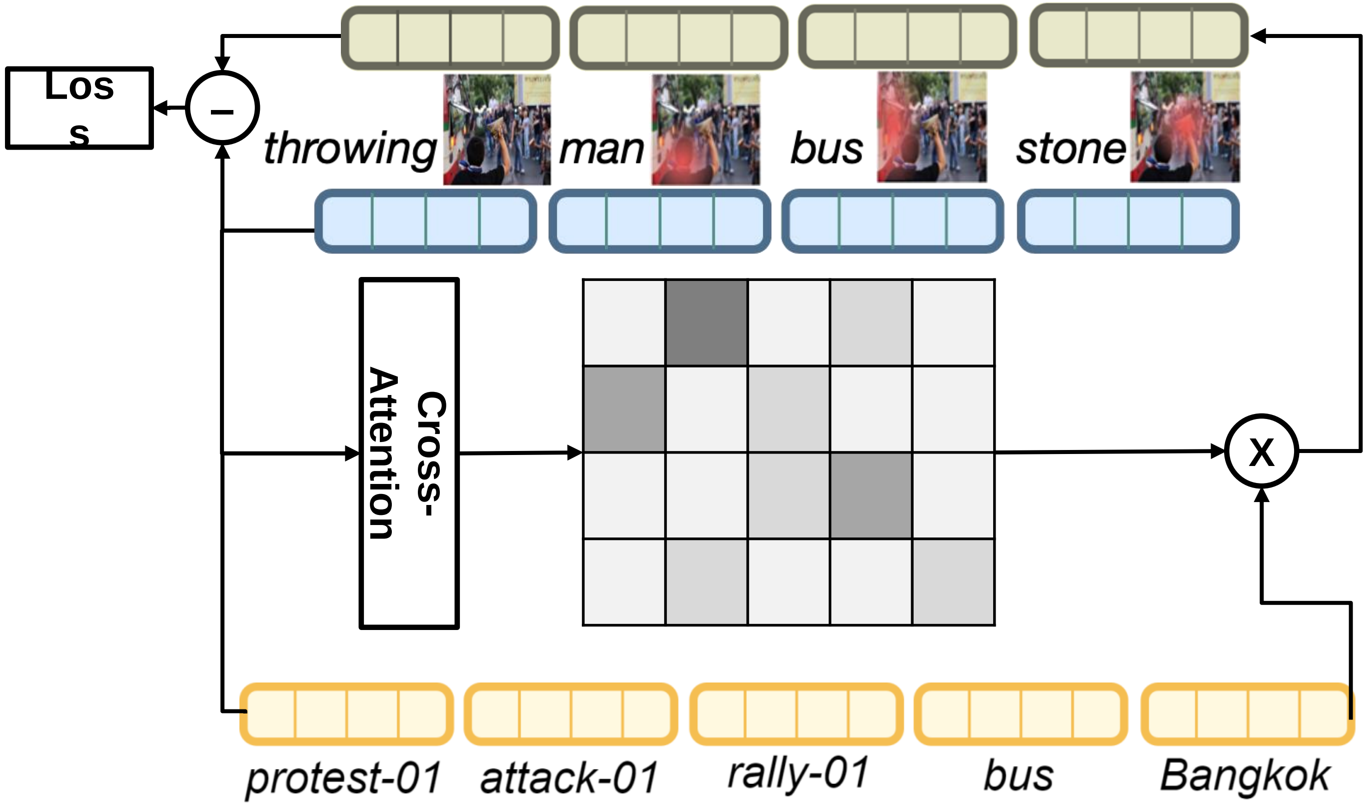
- Prior work aligns image-caption vectors by triplet loss.
- We want to align two graphs, not just single vectors.
- Ontology is shared so the nodes carry similar semantics.

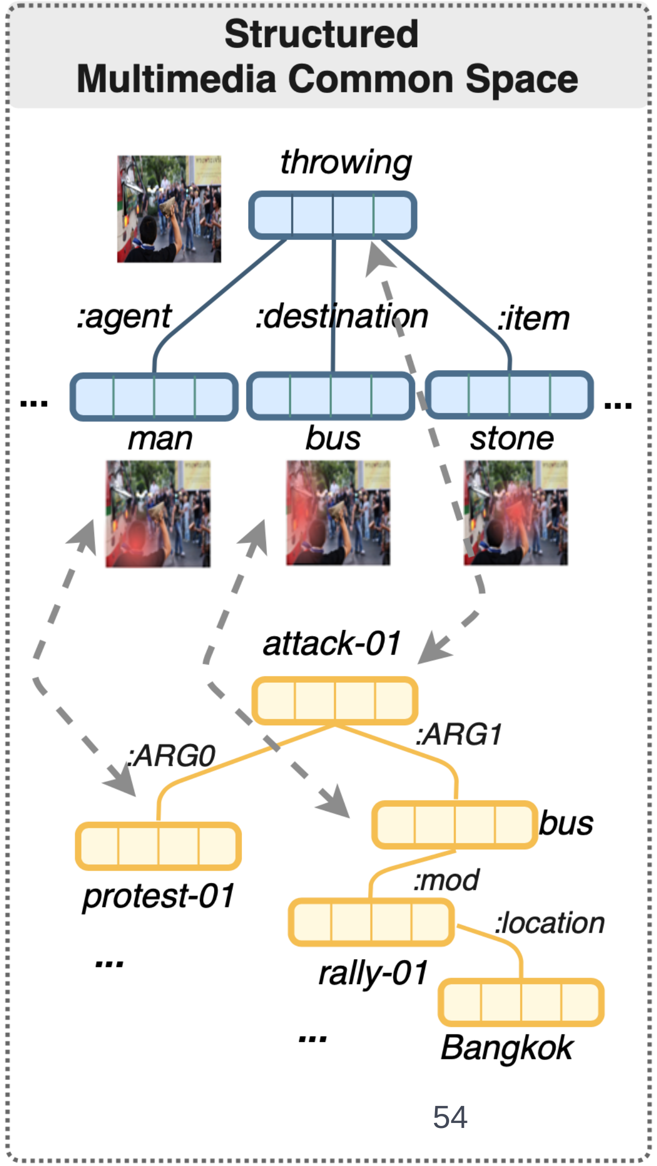




How to align the two modalities?

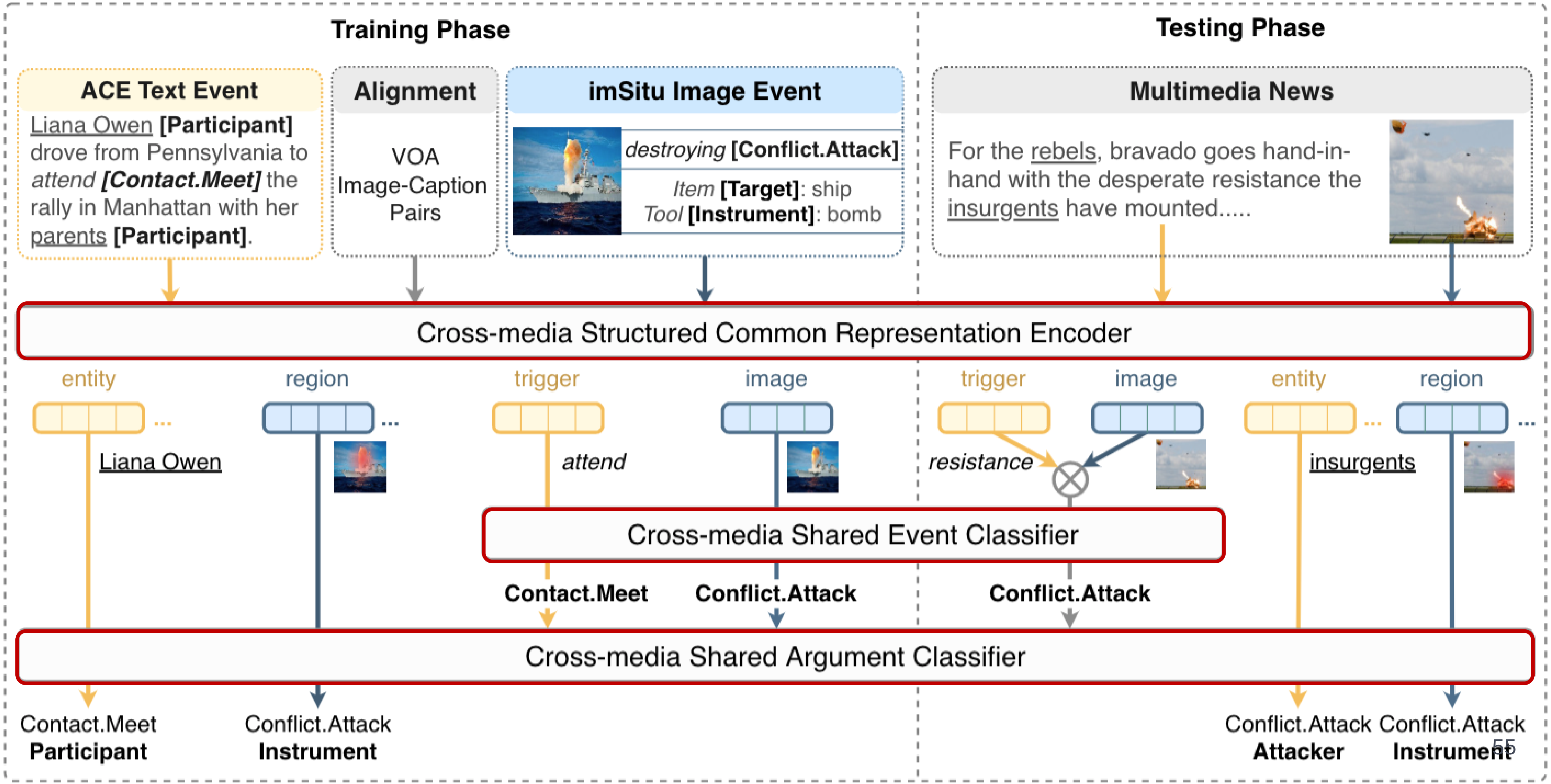
- Prior work aligns image-caption vectors by triplet loss.
- We want to align two graphs, not just single vectors.
- Ontology is shared so the nodes carry similar semantics.





Weakly Aligned Structured Embedding (WASE)

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



Compare to Single Data Modality Extraction

- Surrounding sentence helps visual event extraction.

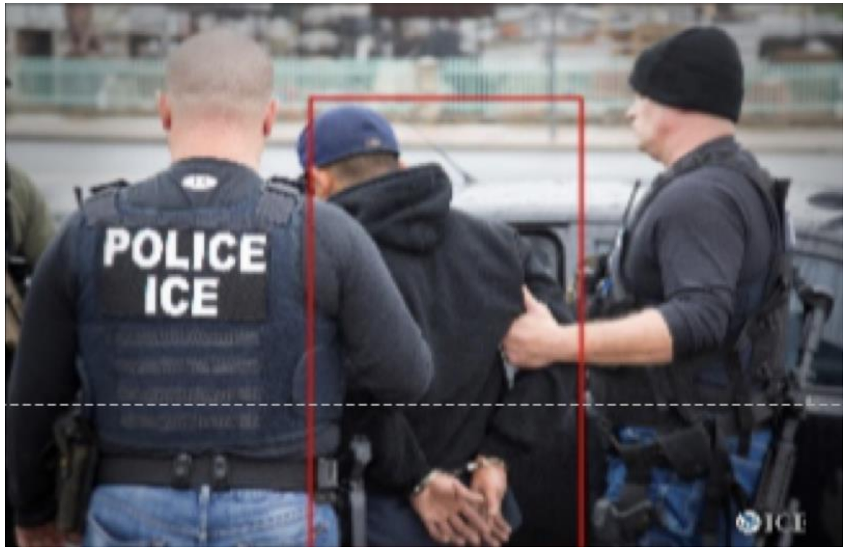
People celebrate Supreme Court ruling on Same Sex Marriage in front of the Supreme Court in Washington.

- Image helps textual event extraction.

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

Compare to Cross-media Flat Representation

- Compared to cross-media **flat** embedding, our structured common space can capture the connections between visual **objects**



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 Portability Challenges

Portability

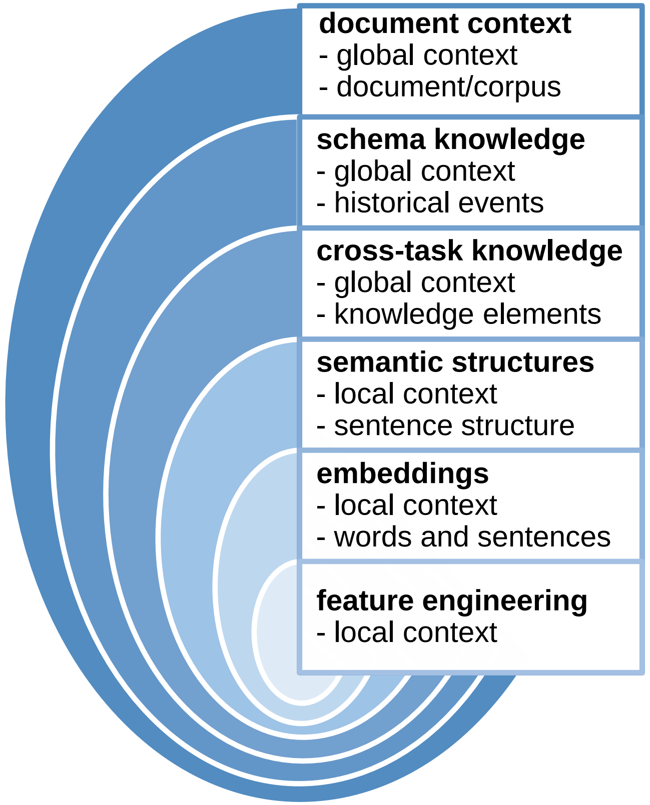
Cross-domain	Cross-lingual	Cross-media
Structure Transfer		
AMR Graph | Event Structure	Universal Dependency Graph | Event Structure	Text AMR Graph | Image Situation Graph
Semantic Common Space		
Adding Ontology Info: definitions, example sentences, constraints, etc.	**Multi-level Alignment**: <word-word>, <entity-entity>	**Multi-level Alignment**: <entity-object>, <image-caption>

Moving forward…

- **Wider input:** How to jointly extract across modalities and languages
  - e.g., if two languages share the same visual event, their events should be related
- **Wider context:** How to use schema knowledge for IE?
  - induce cross-media event schema from historical multimedia events
- **Better structure encoding:** How to make pretrained language model aware of structures?
  - especially for vision

Summary of Quality Challenges

- Encoding wider context improves the IE quality (local → global)
  - A global view of event graph is introduced, to capture the global context of events
- Structure encoding is critical for IE
  - dependency graph, AMR graph, event graph structure, etc.



Moving forward…

- Wider input
  - corpus-level IE, coreference resolution, etc
- Wider context
  - ontological constraints, schema knowledge, etc
- Better structure encoding
  - encoding more complicated schema knowledge (horizontal & vertical)

Reference

[1] Johannes Dölling, Tatjana Heyde-Zybatow, and Martin Schäfer, eds. Event structures in linguistic form and interpretation. 2011.
[2] Zheng Chen, and Heng Ji. "Towards High Performance Multilingual Event Extraction: Language Specific Issue and Feature Exploration." 2009.
[3] Xiaocheng Feng, et al. "A Language-Independent Neural Network for Event Detection." ACL. 2016.
[4] Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. "Joint event extraction via recurrent neural networks." NAACL. 2016.
[5] Ying Lin, Heng Ji, Fei Huang, Lingfei Wu. A Joint End-to-End Neural Model for Information Extraction with Global Features. ACL. 2020.
[6] Minh Van Nguyen, Viet Dac Lai, and Thien Huu Nguyen. "Cross-task instance representation interactions and label dependencies for joint information extraction with graph convolutional networks." NAACL. 2021.
[7] Zixuan Zhang and Heng Ji. "Abstract Meaning Representation Guided Graph Encoding and Decoding for Joint Information Extraction". NAACL. 2021
[8] Manling Li, Qi Zeng, Ying Lin, Kyunghyun Cho, Heng Ji, Jonathan May, Nathanael Chambers and Clare Voss. "Connecting the Dots: Event Graph Schema Induction with Path Language Modeling". EMNLP. 2020a.
[9] Sha Li, Heng Ji, Jiawei Han. "Document-Level Argument Extraction by Conditional Generation". NAACL. 2021.
[10] Seth Ebner, Patrick Xia, Ryan Culkin, Kyle Rawlins and Benjamin Van Durme. Multi-Sentence Argument Linking. ACL. 2020.
[11] Xinya Du and Claire Cardie. Event Extraction by Answering (Almost) Natural Questions. ACL. 2020
[12] Jian Liu, et al. "Event extraction as machine reading comprehension.". EMNLP. 2020.
[13] Yunmo Chen, Tongfei Chen, Seth Ebner, Aaron Steven White, and Benjamin Van Durme. "Reading the manual: Event extraction as definition comprehension." ACL Workshop SPNLP. 2020.
[14] Lifu Huang, Heng Ji, Kyunghyun Cho, Ido Dagan, Sebastian Riedel and Clare Voss. 2018. "Zero-shot transfer learning for event extraction". ACL. 2018.
[15] Ziqi Wang, et al. "CLEVE: Contrastive Pre-training for Event Extraction". ACL. 2020
[16] Hongming Zhang, Haoyu Wang, Dan Roth. 2020. Unsupervised Label-aware Event Trigger and Argument Classification. ACL Findings. 2021.
[17] Pan, Xiaoman, et al. "Cross-lingual joint entity and word embedding to improve entity linking and parallel sentence mining." EMNLP Workshop DeepLo. 2019.
[18] Ananya Subburathinam, et al. "Cross-lingual structure transfer for relation and event extraction." EMNLP. 2019.
[19] Wasi Uddin Ahmad, Nanyun Peng, and Kai-Wei Chang. "GATE: Graph Attention Transformer Encoder for Cross-lingual Relation and Event Extraction." AAAI. 2021.
[20] M. Straka and J. Strakova. "Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe." CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, 88–99. 2017
[21] Sarah Pratt, Mark Yatskar, Luca Weihs, Ali Farhadi, Aniruddha Kembhavi. "Grounded Situation Recognition". ECCV. 2020
[22] Mark Yatskar, Luke Zettlemoyer, and Ali Farhadi. "Situation recognition: Visual semantic role labeling for image understanding." CVPR. 2016.
[23] Sadhu, Arka, et al. "Visual Semantic Role Labeling for Video Understanding." CVPR. 2021.
[24] Alireza Zareian, Svebor Karaman, Shih-Fu Chang. "Weakly Supervised Visual Semantic Parsing". CVPR. 2020
[25] Manling Li, Alireza Zareian, Qi Zeng, Spencer Whitehead, Di Lu, Heng Ji, Shih-Fu Chang. 2020. Cross-media Structured Common Space for Multimedia Event Extraction. Proc. ACL2020.
[26] Heng Ji and Clare Voss. 2020. Low-resource Event Extraction based on Share-and-Transfer Invited Book Chapter for Events and Stories, Cambridge Press.

Thank You