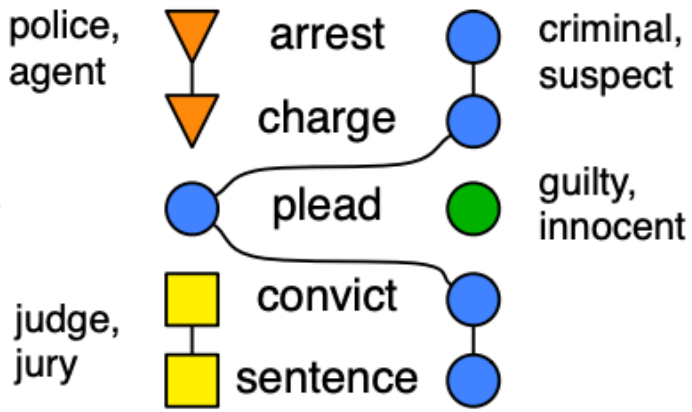
Connecting the Dots: Event Graph Schema Induction with Path Language Modeling

Manling Li, Qi Zeng, Ying Lin, Kyunghyun Cho, Heng Ji,
Jonathan May, Nathanael Chambers, Clare Voss

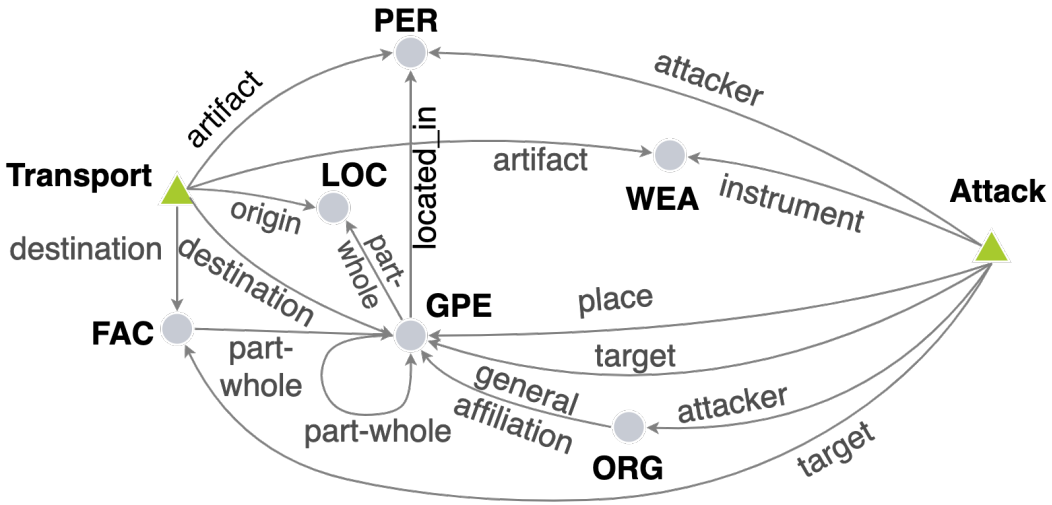
ILLINOIS NYU USC Viterbi School of Engineering UNITED STATES NAVAL ACADEMY HONOR... COURAGE... COMMITMENT ARL

Event Schema Induction Related Work

- How to capture complex connections among events?
  - Temporal relations exist between almost all events, even those that are not semantically related
  - Causal relations have been hobbled by low inter-annotator agreement (Hong et al., 2016).
- Two events are connected through entities and their relations



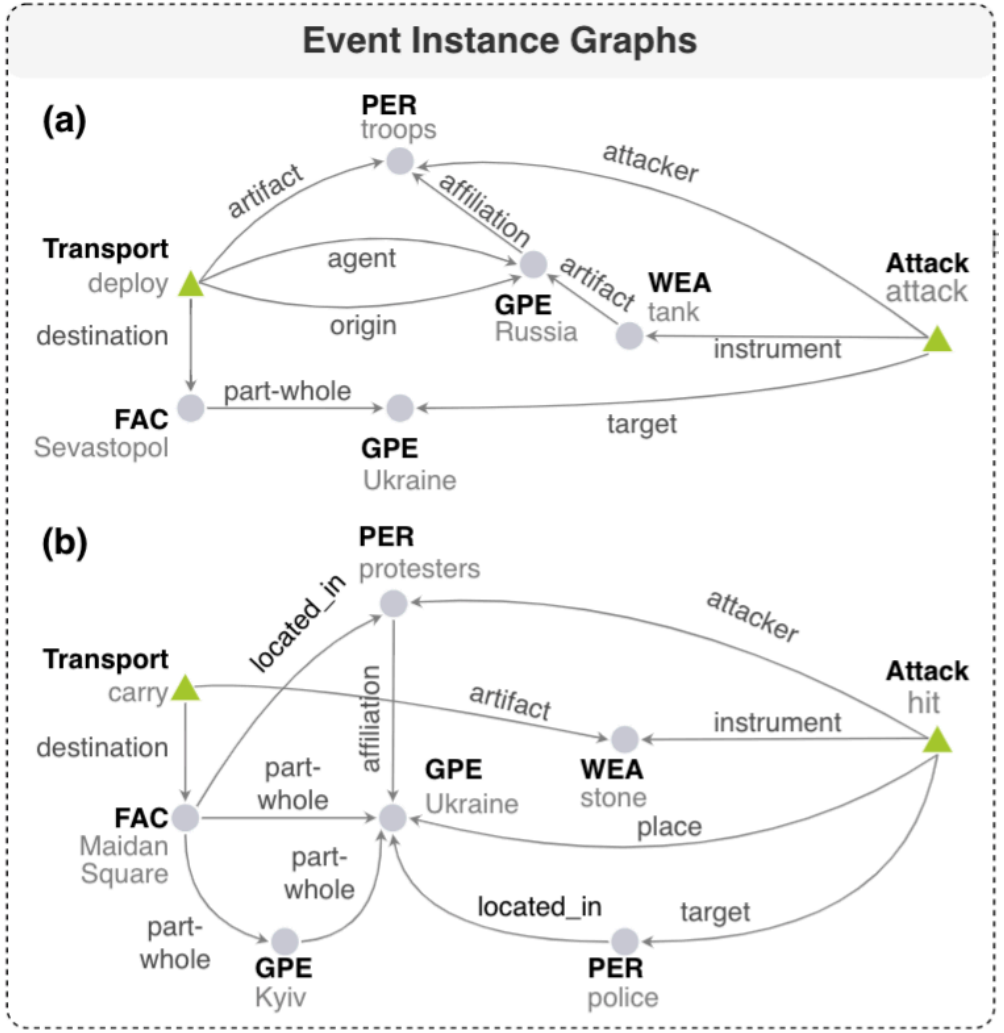
Previous Work:
Event Narrative Chain (Chambers and Jurafsky, 2008, 2009, 2010)

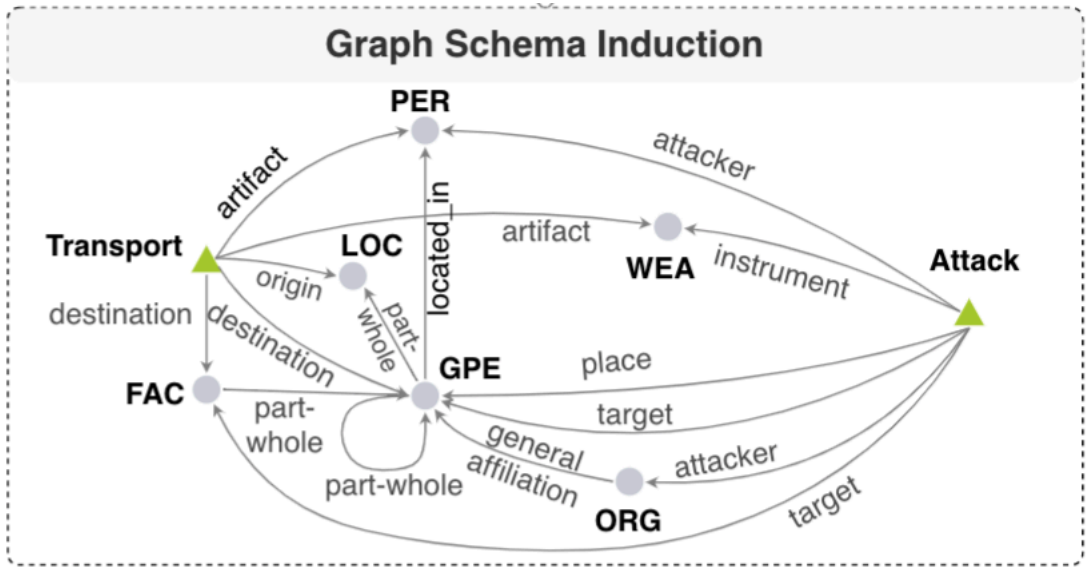


Our Paper:
Event Narrative Graph Schema

Problem Formulation

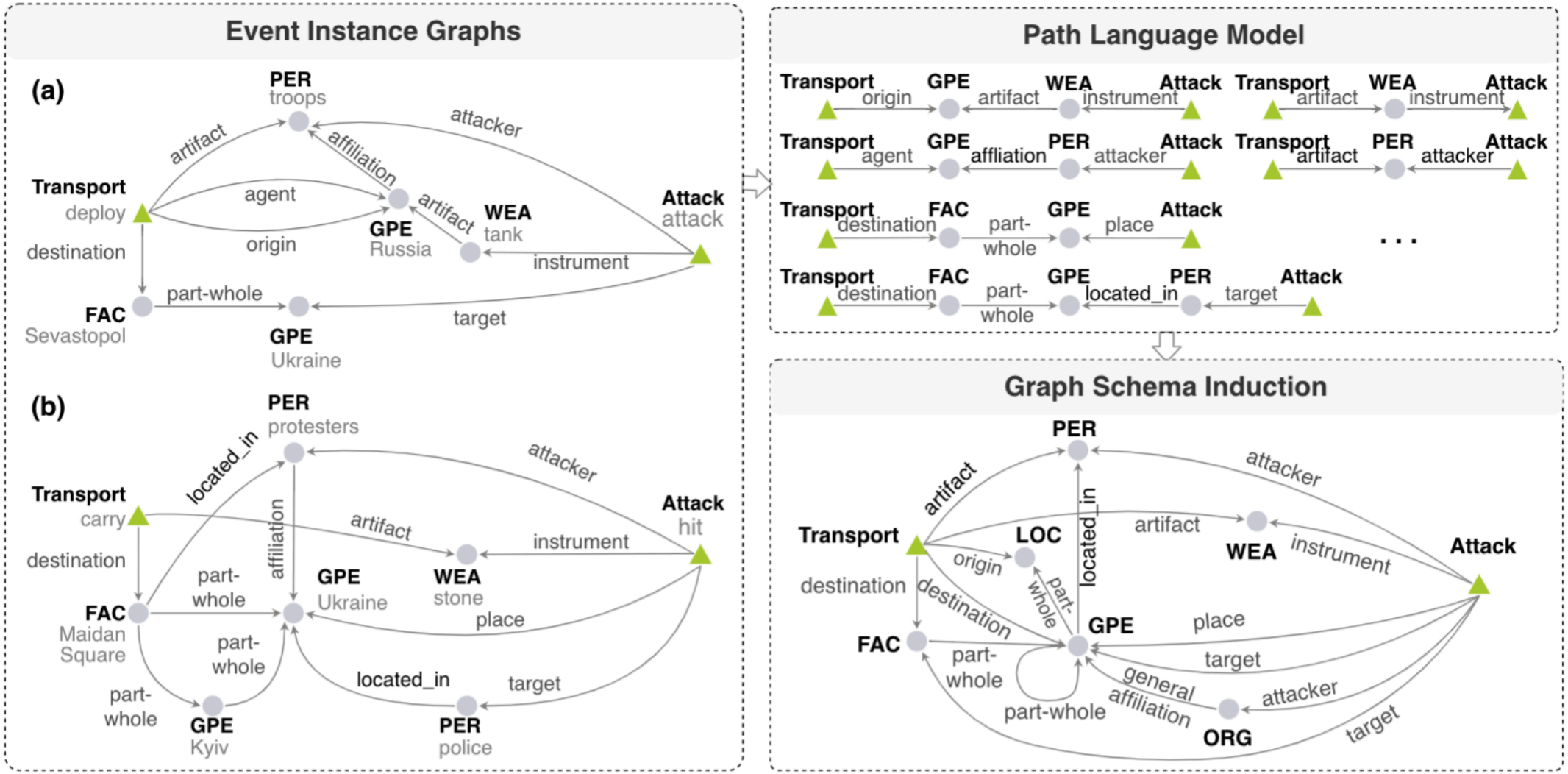
- **Input**: Instance graphs, where each node is an entity or event, each edge is an argument role or relation.
- **Output**: Graph schemas between two event types, where each node is an entity type or event type.
- Instance graphs (a) and (b) refer to very different event instances, but they both illustrate a same scenario.





Event Graph Schema Induction Framework

- We select **salient** and **coherent** paths based on Path Language Model, and merge them into graph schemas.



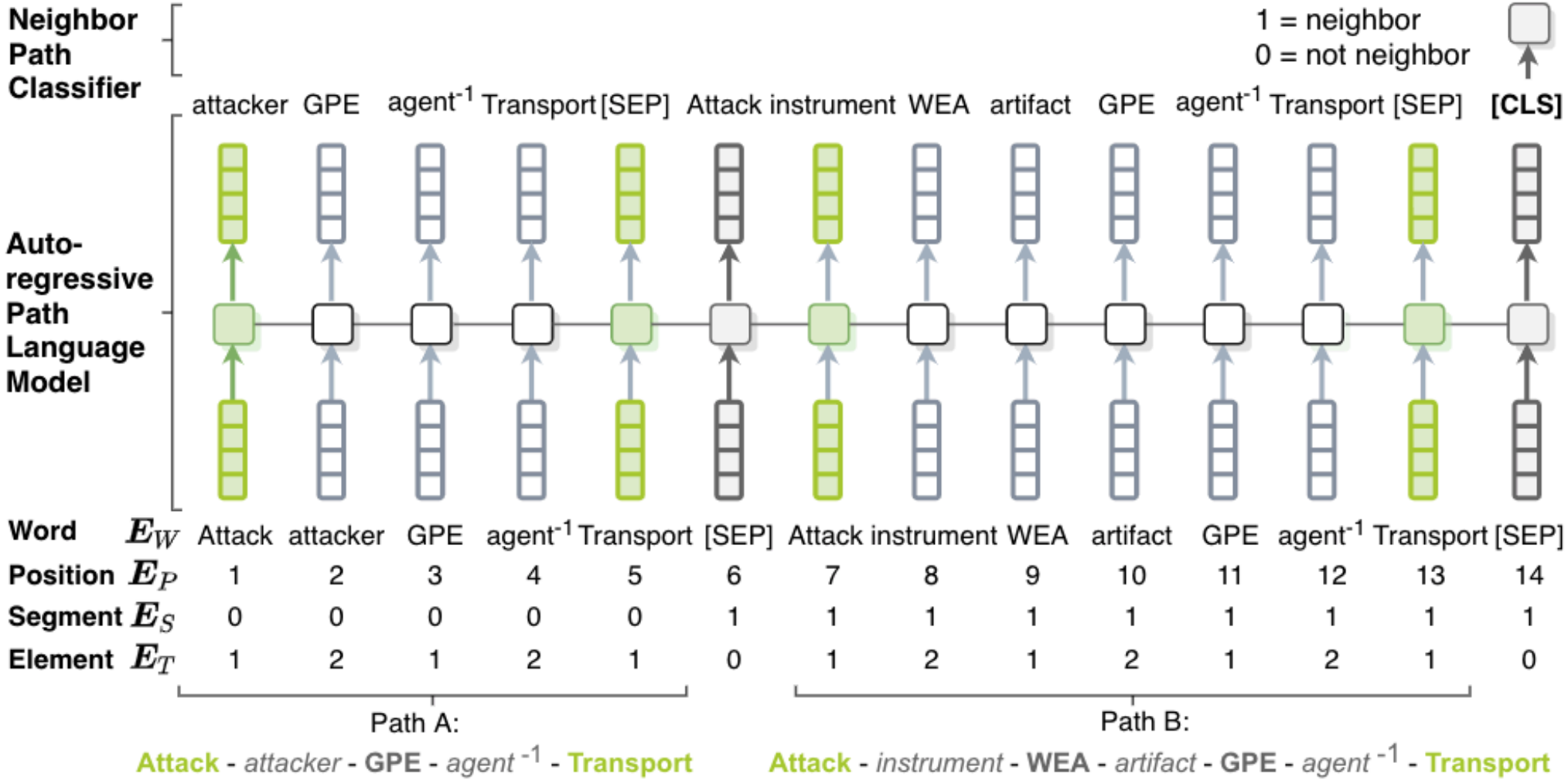
What is a good graph schema?

- A good graph schema for two event types consists of **salient** and **coherent** paths between them.
  - Salience: recurring event-event connection patterns
  - Coherence: semantically coherent

Criteria			Examples	Frequency
Single Path	Salience	High	Transport —agent→ GPE ←affiliation— PER ←attacker— Attack	31
		Low	Transport —destination→ GPE ←affiliation— PER ←attacker— Attack	2
	Semantic Coherence	High	Transport —origin→ FAC —part-whole→ LOC ←part-whole— GPE ←affiliation— PER ←attacker— Attack	9
		Low	Transport —agent→ GPE ←affiliation— PER —affiliation→ GPE ←resident— PER ←target— Attack	24
Multiple Paths	Semantic Consistency	High	Transport —destination→ GPE —place→ Attack Transport —artifact→ PER —located_in→ GPE ←place— Attack	20
		Low	Transport —destination→ GPE —place→ Attack Transport —origin→ GPE ←place— Attack	0

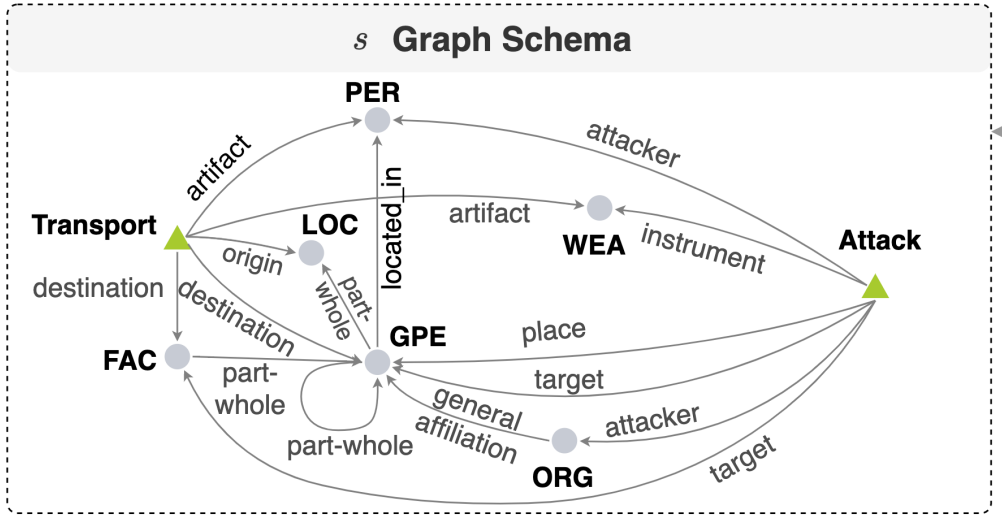
Path Language Model

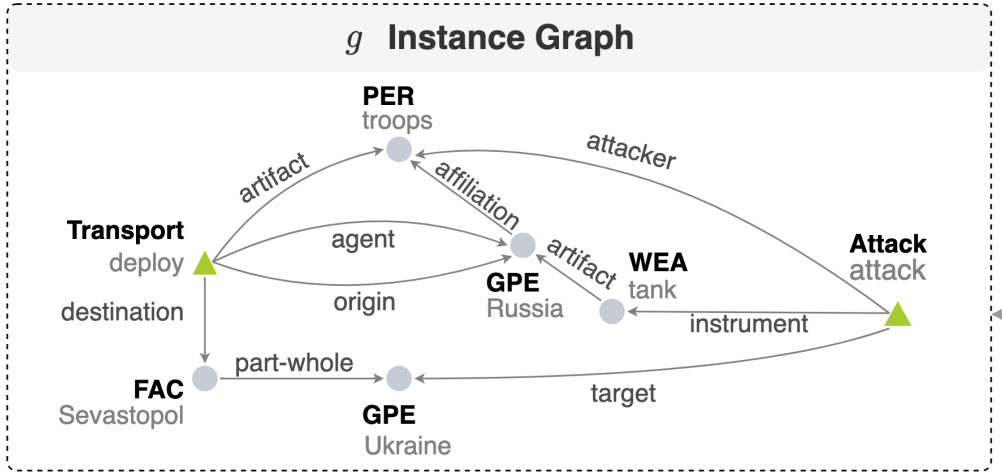
- Path Language Model is trained on two tasks
  - Autoregressive Language Model Loss: capturing the frequency and coherence of a single path
  - Neighbor Path Classification Loss: capturing co-occurrence of two paths



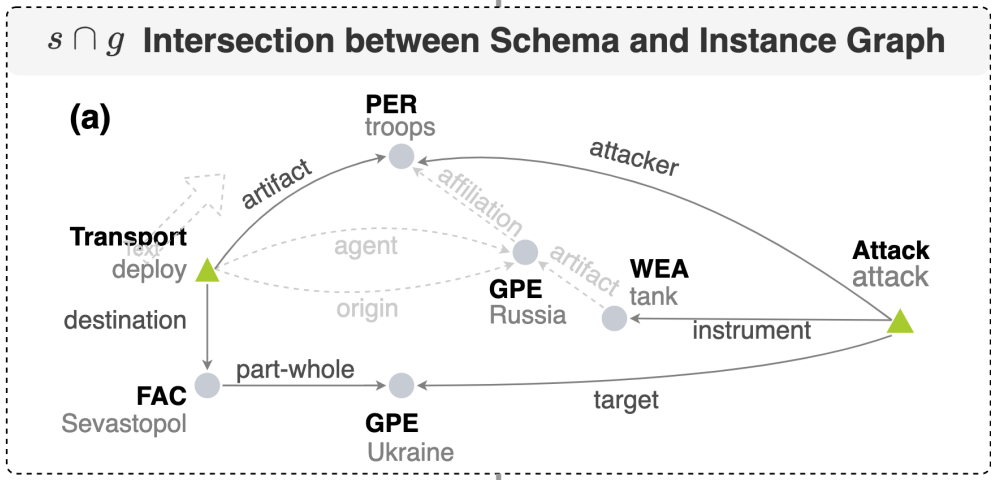
Experiment: Instance Coverage

- A salient schema can serve as a skeleton to recover instance graphs
  - We use each graph schema to match back to each ground-truth instance graph and evaluate their intersection in terms of Precision and Recall.





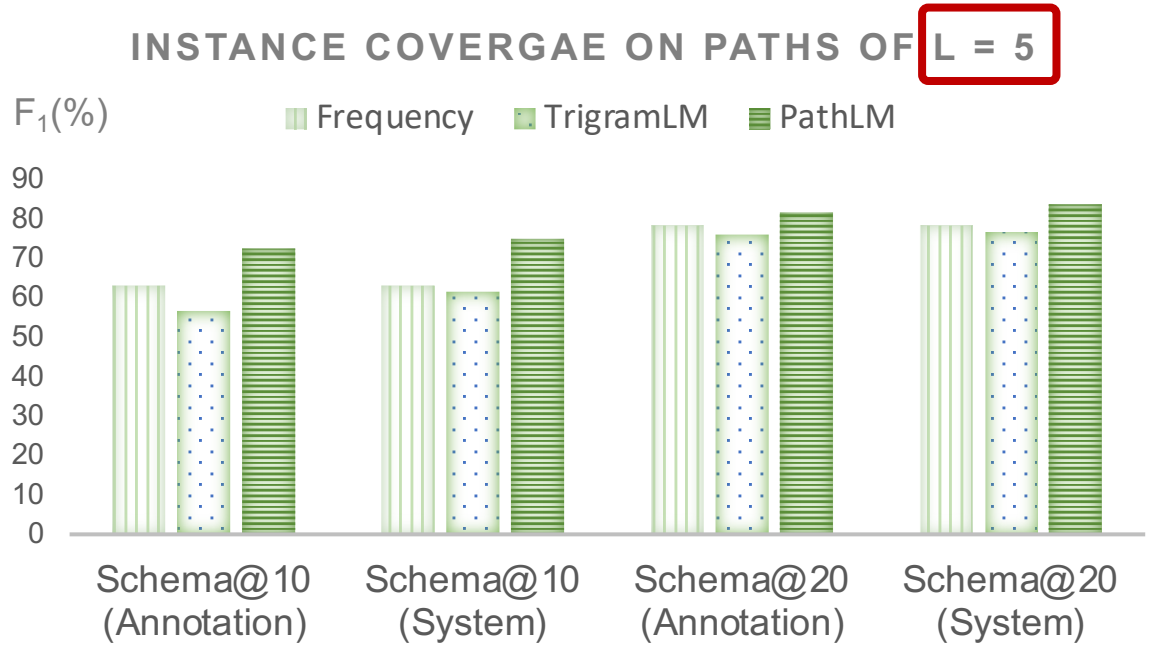
$$\text{Precision} = \frac{\sum_{s \in \mathcal{S}} \sum_{g \in \mathcal{G}} |g \cap s|_{\mathbb{S}}}{\sum_{s \in \mathcal{S}} |s|_{\mathbb{S}}}$$

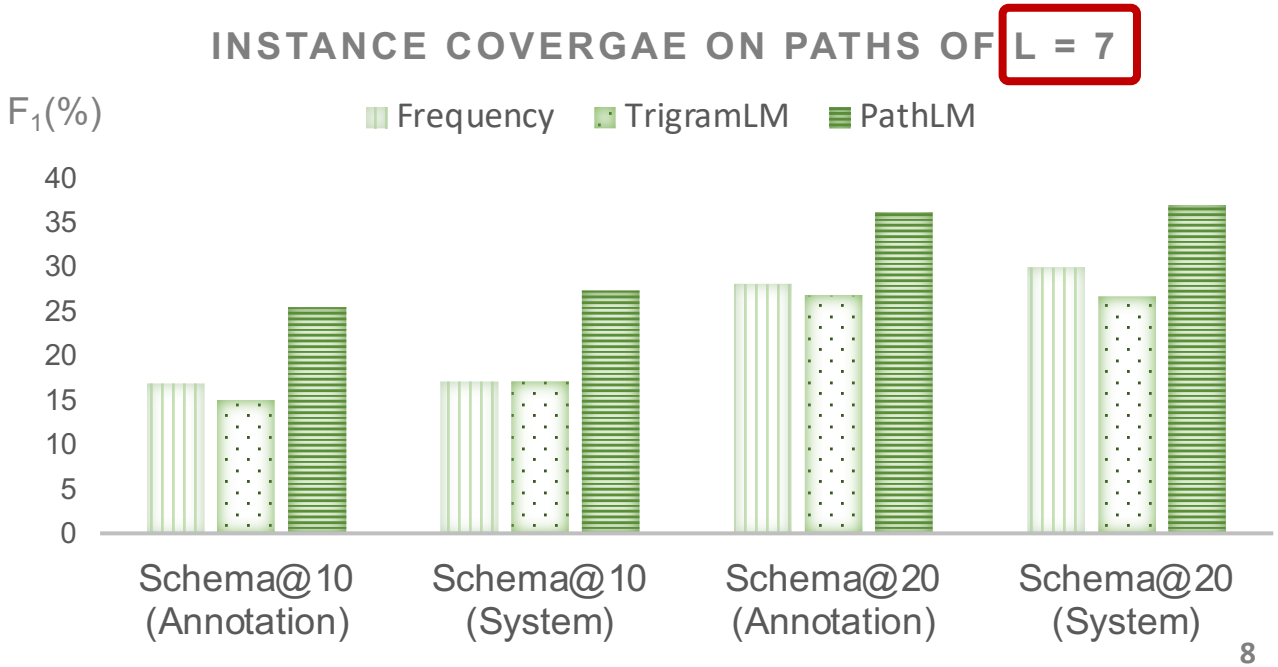


$$\text{Recall} = \frac{\sum_{s \in \mathcal{S}} \sum_{g \in \mathcal{G}} |g \cap s|_{\mathbb{I}}}{\sum_{g \in \mathcal{G}} |g|_{\mathbb{I}}}$$

Experiment: Instance Coverage

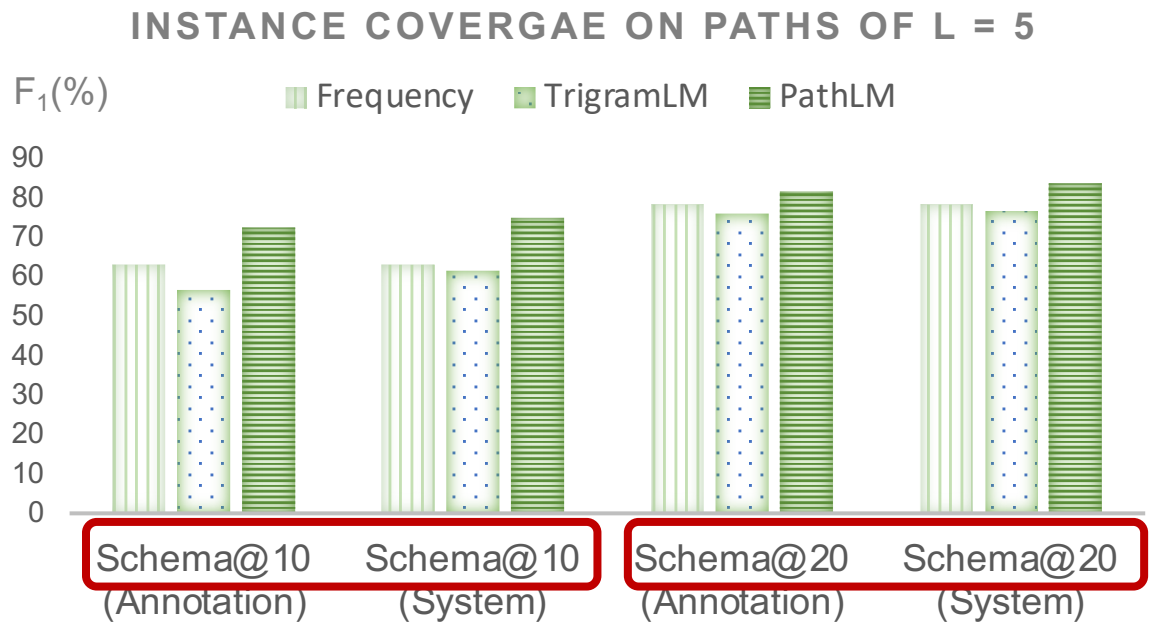
- We evaluate the intersection by substructures of the schema graph, i.e., paths of different lengths
- Proves ability of PathLM:
  - Higher gains on longer path queries (e.g. l = 7) → able to capture complex graph structures involving long distance between related events.
  - Larger gains compared to baselines on Schema@10 than Schema@20, demonstrating the effectiveness of our ranking approach, especially on top ranked ones.
  - Schemas induced from automatically constructed and manually constructed event graph instances have comparable performance → robust to extraction noise

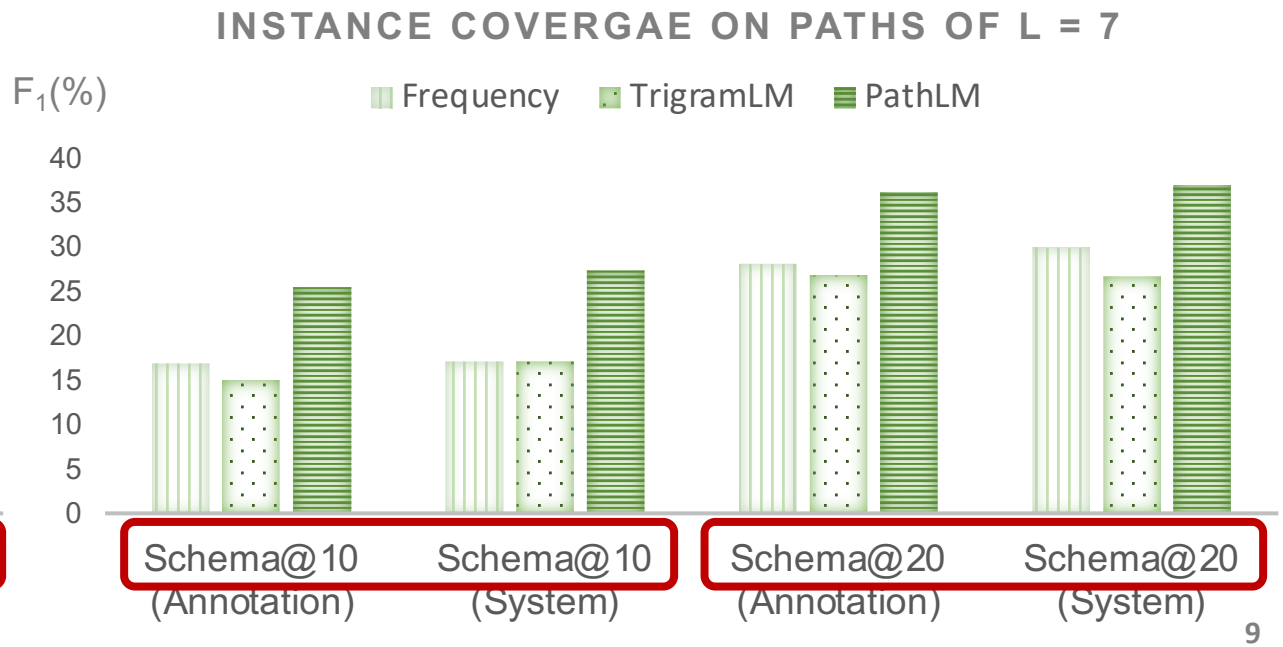




Experiment: Instance Coverage

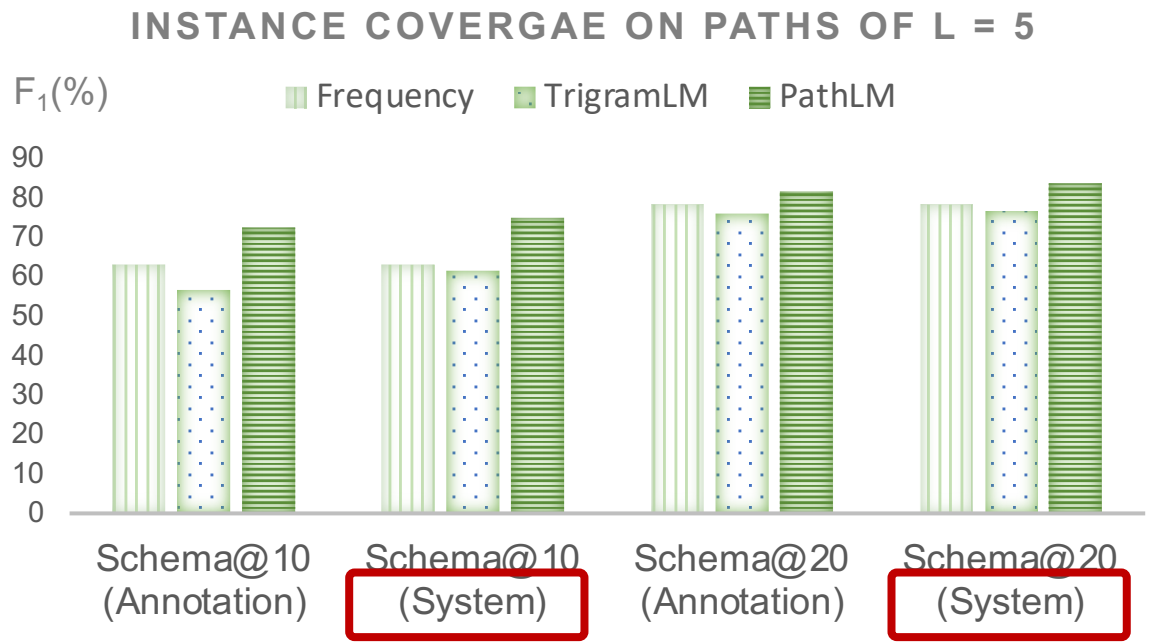
- We evaluate the intersection by substructures of the schema graph, i.e., paths of different lengths
- Proves ability of PathLM
  - Higher gains on longer path queries (e.g. l = 7) → able to capture complex graph structures involving long distance between related events.
  - Larger gains compared to baselines on Schema@10 than Schema@20, demonstrating the effectiveness of our ranking approach, especially on top ranked ones.
  - Schemas induced from automatically constructed and manually constructed event graph instances have comparable performance → robust to extraction noise

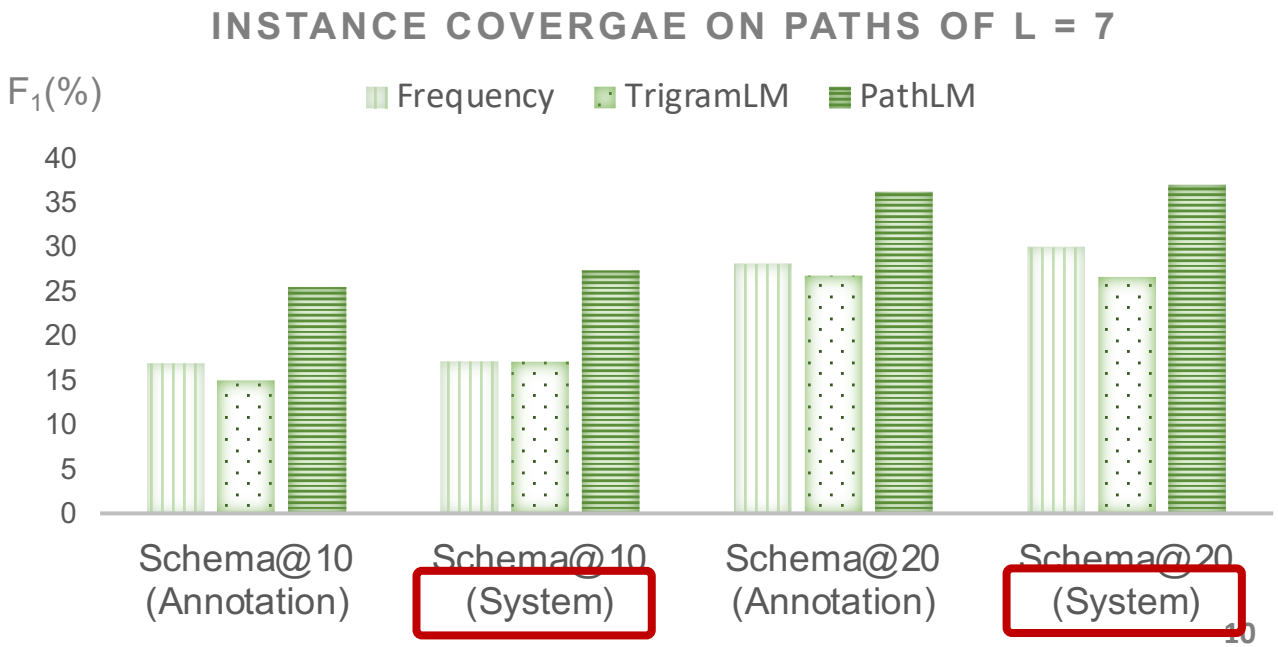




Experiment: Instance Coverage

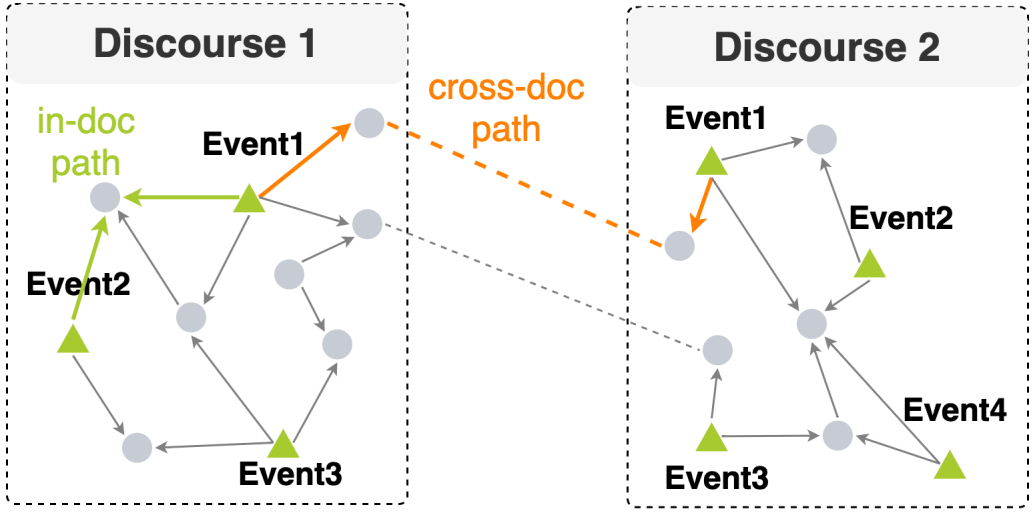
- We evaluate the intersection by substructures of the schema graph, i.e., paths of different lengths
- Proves ability of PathLM:
  - Higher gains on longer path queries (e.g. l = 7) → able to capture complex graph structures involving long distance between related events.
  - Larger gains compared to baselines on Schema@10 than Schema@20, demonstrating the effectiveness of our ranking approach, especially on top ranked ones.
  - Schemas induced from automatically constructed and manually constructed event graph instances have comparable performance → robust to extraction noise





Experiment: Instance Coherence

- we hypothesize that an instance graph between two events *v* and *v′* is coherent if *v* and *v′* are from the same discourse (e.g., a news document)
- We carefully select 24 documents with each document talking about a unique complex event such as *Iraq War* or *North Korea Nuclear Test*
- We define **Instance Coherence** as the proportion of in-doc path instances

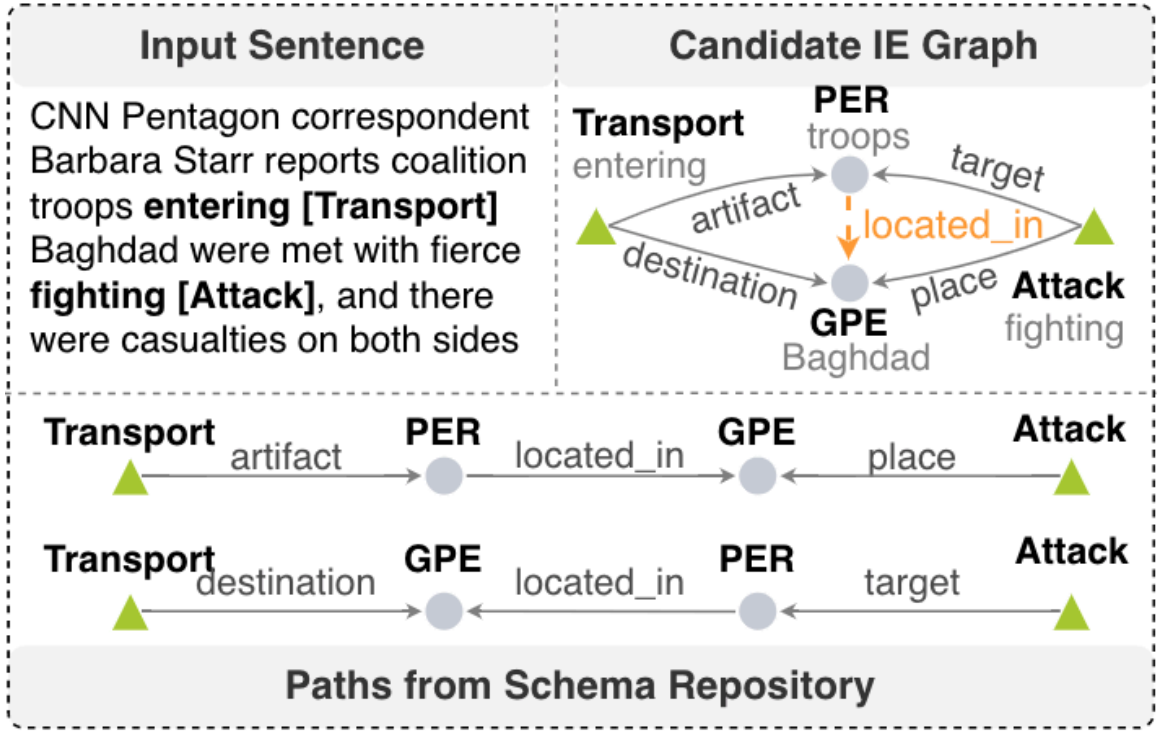


$$\text{Coherence} = \frac{\sum_{s \in \mathcal{S}} \sum_{g \in \mathcal{G}} \sum_{p \in g \cap s} f(p) \cdot \mathbb{I}_g}{\sum_{s \in \mathcal{S}} \sum_{g \in \mathcal{G}} \sum_{p \in g \cap s} f(p)};$$

Historical	Models	Schema@10 (%)	Schema@20 (%)
Annotation	Frequency	67.8	65.6
	UnigramLM	62.4	69.9
	BigramLM	59.0	67.5
	TrigramLM	56.6	64.9
	PathLM	**76.0**	**79.9**
	w/o CLS_NP	75.3	79.2
System	Frequency	60.1	65.6
	UnigramLM	61.8	70.0
	BigramLM	59.7	69.6
	TrigramLM	55.8	65.8
	PathLM	**76.4**	**78.5**
	w/o CLS_NP	73.9	77.1

Schema-Guided Information Extraction

- Use the state-of-the-art IE system OneIE (Lin et al, 2020) to decode converts each input document into an IE graph
- Each path in the graph schema is encoded as a single global feature for scoring candidate IE graphs
- OneIE promotes candidate IE graphs containing paths matching schema graphs
- http://blender.cs.illinois.edu/software/oneie
- F-scores (%) on ACE2005 data [Lin et al., ACL2020]:



Dataset	Entity	Event Trigger Identification	Event Trigger Classification	Event Argument Identification	Event Argument Classification	Relation
Baseline	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**








THANK YOU!