Event and Commonsense

Event-centric Natural Language Processing (Part IV)

Hongming Zhang

UPenn

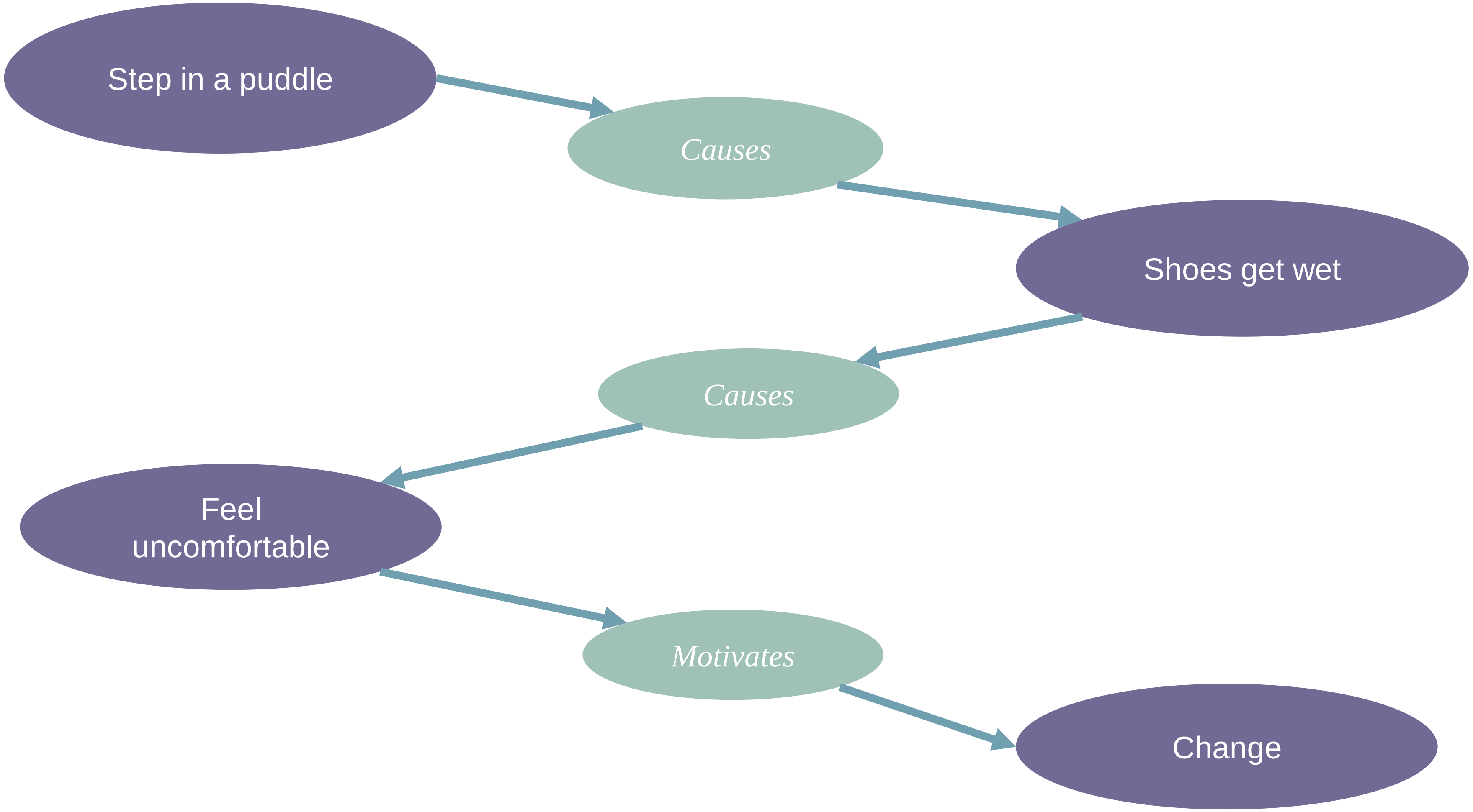
Aug 2021

ACL Tutorials

Event-centric Natural Language Understanding

Commonsense is crucial for NLU

Example: John stepped in a puddle and had to go home to change.



Outline

- Understanding Commonsense from the Angle of Events
- Instance-level Event Knowledge Acquisition
  - Human Annotation
  - Automatic Event Knowledge Extraction
  - Language Modeling
- Schema-level Event Knowledge Acquisition
- Conclusion

Commonsense Knowledge

- Modern Definition of Commonsense Knowledge (Liu & Singh, 2004)
  - “While to the average person the term ‘commonsense’ is regarded as synonymous with ‘good judgement’”
  - “the AI community it is used in a technical sense to refer to the millions of basic facts and understandings possessed by most people.”
  - “Commonsense is about preference and not always true”
    - If you forget someone’s birthday, they may be unhappy with you.
    - But if your friends understand that you are busy, he will not by angry.

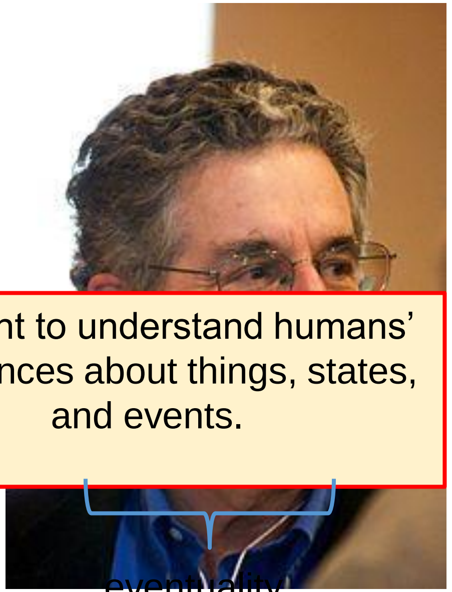
Unlike factual knowledge, they are not inevitably true.

Commonsense is about preference.

Hugo Liu and Push Singh, ConceptNet - a practical commonsense reasoning tool-kit, BTTJ, 2004

What kinds of preference?

- Semantic meaning in our language can be described as "a finite set of mental primitives and a finite set of mental combination." (Jackendoff, 1990)
- The primitive units of semantic meanings include
  - Thing (or entity)
    - cat
  - State
    - The cat is cute.
    - The cat is smiling.
  - Event
    - The cat is running.



We want to understand humans' preferences about things, states, and events.

eventuality

States describe things.

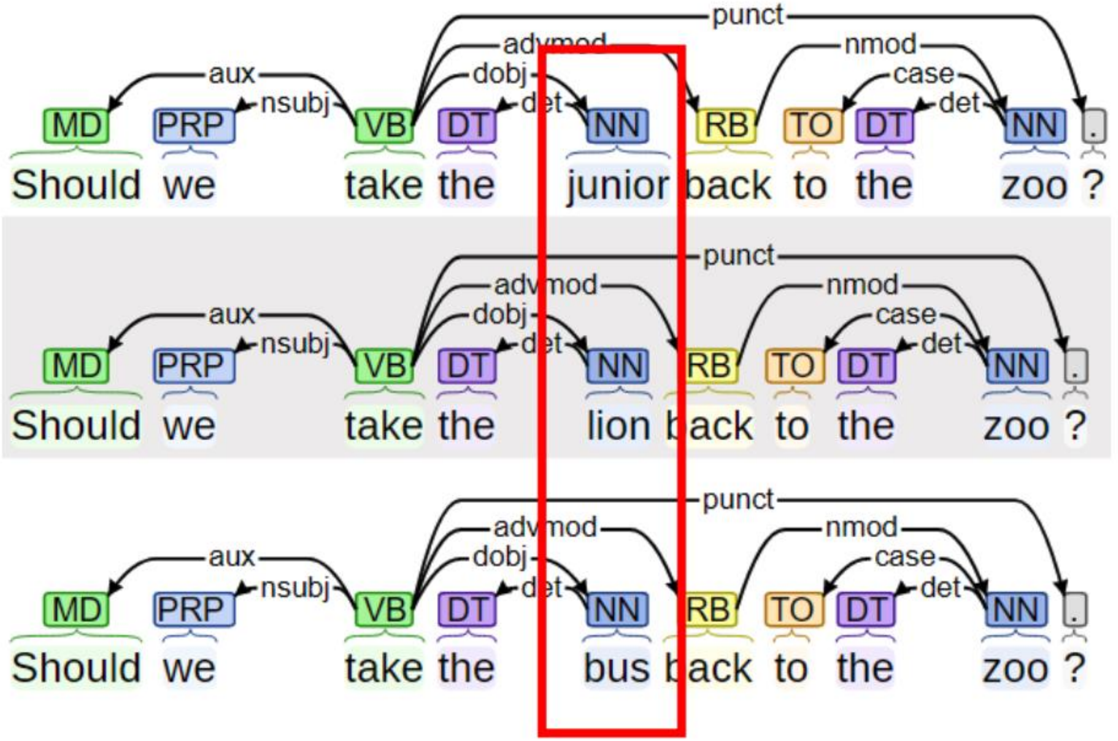
Events describe the changing of states.

Ray Jackendoff Semantic Structures. Cambridge, Massachusetts: MIT Press. 1990

How to represent the preference?

- The lower bound of a semantic theory (Katz and Fodor, 1963)
  - Linguistic description – grammar = semantics
  - Understanding language needs both "the speaker's knowledge of his language and his knowledge about world" (Katz and Fodor, 1963)

It is so dangerous!!!



When the grammar is controlled, the selection we made can reflect our understanding about the world.

Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. Language, 39(2), 170–210.

Selectional Preference

- Selectional Preference (Resnik, 1993)
  - A relaxation of selectional restrictions (Katz and Fodor, 1963) and is often used as syntactic features (Chomsky, 1965).
  - Applied to IsA hierarchy in WordNet and verb-object relations.
  - With this formulation, we can easily use the frequency/plausibility scores of different combinations to reflect humans' preference.

  - Examples:
    - ("Cat" -IsA- "Animal") > ("Cat" -IsA- "Plant")
    - ("eat" -dobj- "food") > ("eat" -dobj- "rock")

Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. Language, 39(2), 170–210.

Noam Chomsky. 1965. Aspects of the Theory of Syntax. MIT Press, Cambridge, MA.

Philip Resnik. 1993. Selection and information: A class-based approach to lexical relationships. Ph.D. thesis, University of Pennsylvania.

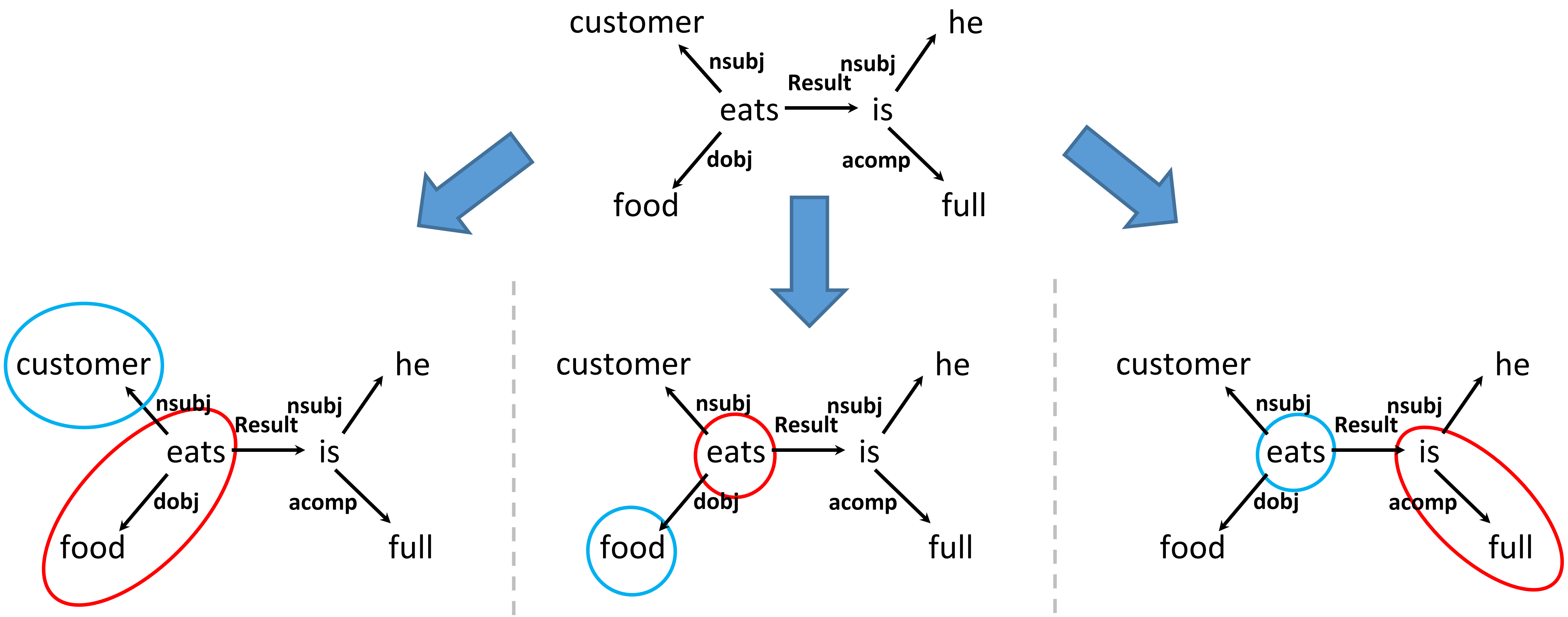
Higher-order Selectional Preference

- First-order
  - dobj: (“eat”–>dobj->“food”) > (“eat”->dobj->“house”)
  - Nsubj: (“sing”->nsubj->“singer”) > (“sing”->nsubj->“house”)
  - …
- Second-order (Zhang et al., 2019)
  - Nsubj-amod / dobj-amod
  - (“eat”->nsubj->“[SUB]”->amod->“hungry”) > (“eat”->dobj>“[OBJ]”->amod->“hungry”)
- Higher-order
  - (“I eat dinner”->Causes->“I am full”) > (“I eat dinner”->Causes->“I am hungry”)

Commonsense can be represented by the higher-order selectional preference over eventualities.

Hongming Zhang, Hantian Ding, and Yangqiu Song. SP-10K: A Large-Scale Evaluation Set for Selectional Preference Acquisition. ACL 2019.

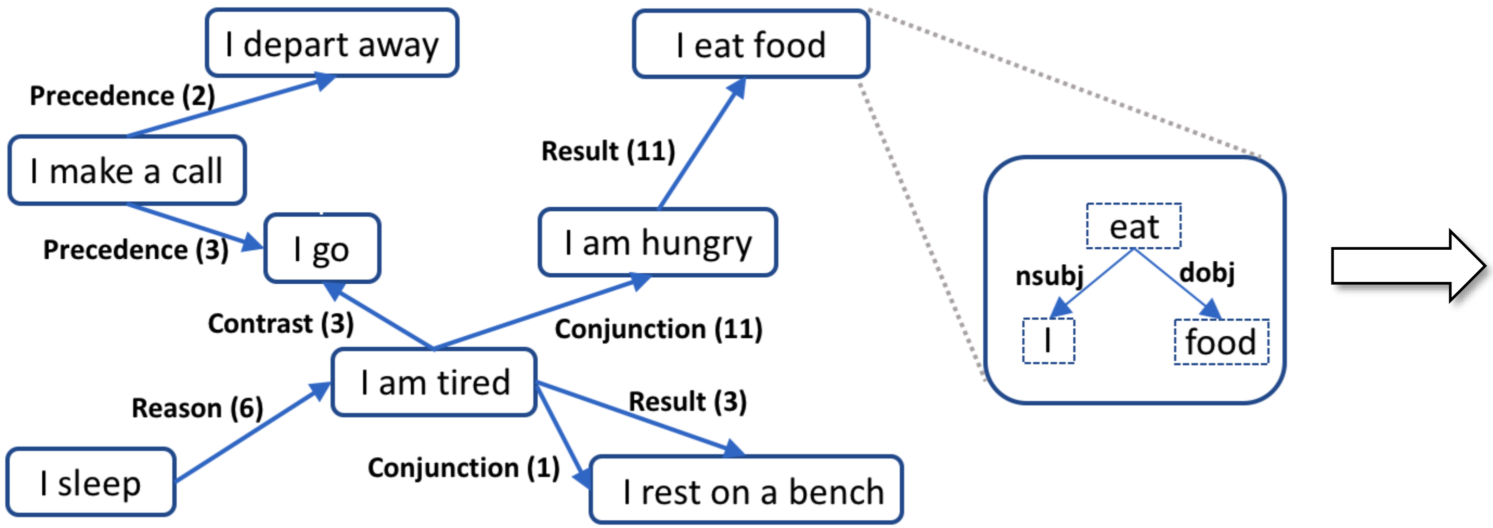
Transferability from event knowledge to Commonsense



(‘customer’-*CapableOf*-‘eat food’) | (‘food’-*ReceivesAction*-‘eat’) | (‘eat’-*Causes*-‘be full’)

Hongming Zhang, Daniel Khashabi, Yangqiu Song, and Dan Roth. TransOMCS: From Linguistic Graphs to Commonsense Knowledge. IJCAI 2020.

Transferability from event knowledge to Commonsense



Event-centric KG

“human” *CapableOf*

1. stand
2. think
3. die
4. learn
5. make mistake
6. lie
7. typically have 🤔
8. create society
9. have cell
10. create life

“love” *Causes*

1. be friendly
2. be happy
3. pain
4. marriage
5. be quaint 🤔
6. be unhappy
7. be allergic 🤔
8. be desperate
9. be apart
10. be silly

Human-defined commonsense

Outline

- Understanding Commonsense from the Angle of Events
- Instance-level Event Knowledge Acquisition
  - Human Annotation
  - Automatic Event Knowledge Extraction
  - Language Modeling
- Schema-level Event Knowledge Acquisition
- Conclusion

Event-centric KBs

	# Events	# Event relation	# Relation Types
FrameNet (Baker et al., 1998)	27,691	1,709	7
ACE (Aguilar et al., 2014)	3,290	0	0
PropBank (Palmer et al., 2005)	112,917	0	0
NomBank (Meyers et al., 2004)	114,576	0	0
TimeBank (Pustejovsky et al., 2003)	7,571	8,242	1
ConceptNet (Liu and Singh, 2004)	74,989	116,097	4
Event2Mind (Smith et al., 2018)	24,716	57,097	3
ProPora (Dalvi et al., 2018)	2,406	16,269	1
ATOMIC (Sap et al., 2019)	309,515	877,108	9
ATOMIC 2020* (Hwang et al., 2020)	-	165,164	4

Pro: High quality
Con: Expensive; Small Scale; Limited relation types

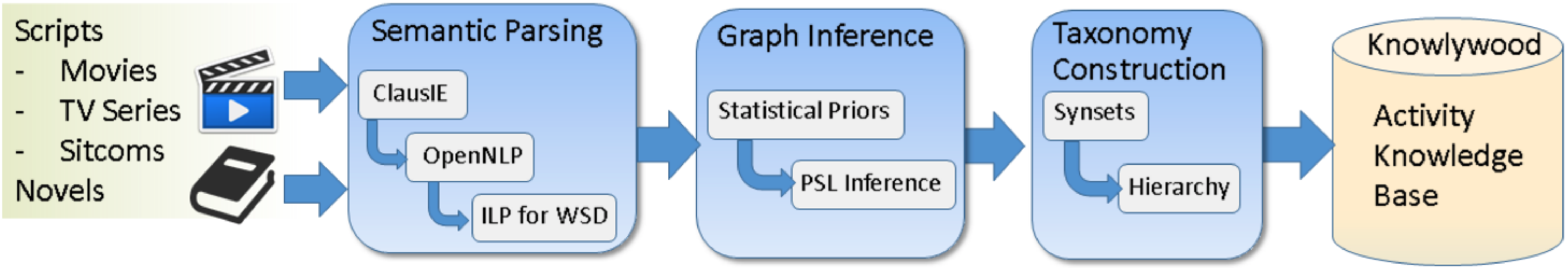
*For ATOMIC 2020, we only count the unique edges and ignore the edges it inherits from other KBs.

Outline

- Understanding Commonsense from the Angle of Events
- Instance-level Event Knowledge Acquisition
  - Human Annotation
  - Automatic Event Knowledge Extraction
  - Language Modeling
- Schema-level Event Knowledge Acquisition
- Conclusion

Knowlywood (Tandon et al., 2015)

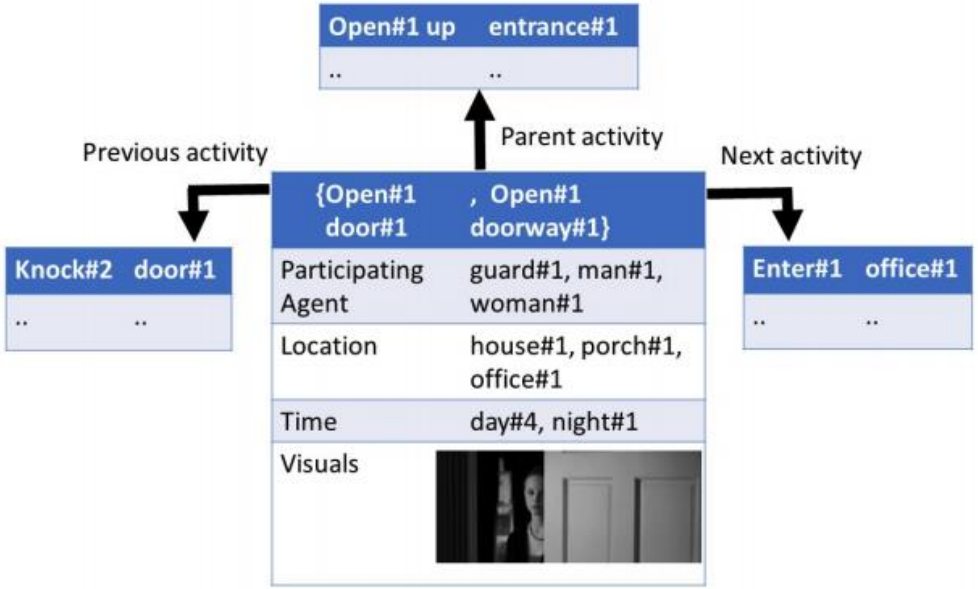
- KG Format
  - Node: Verb + Object
  - Edge: Temporal Relation
- Resource
  - 560 movie scripts
- Extraction Methodology



Niket Tandon, Gerard de Melo, Abir De, and Gerhard Weikum. 2015. Knowlywood: Mining Activity Knowledge From Hollywood Narratives. CIKM 2015.

Knowlywood

- Example



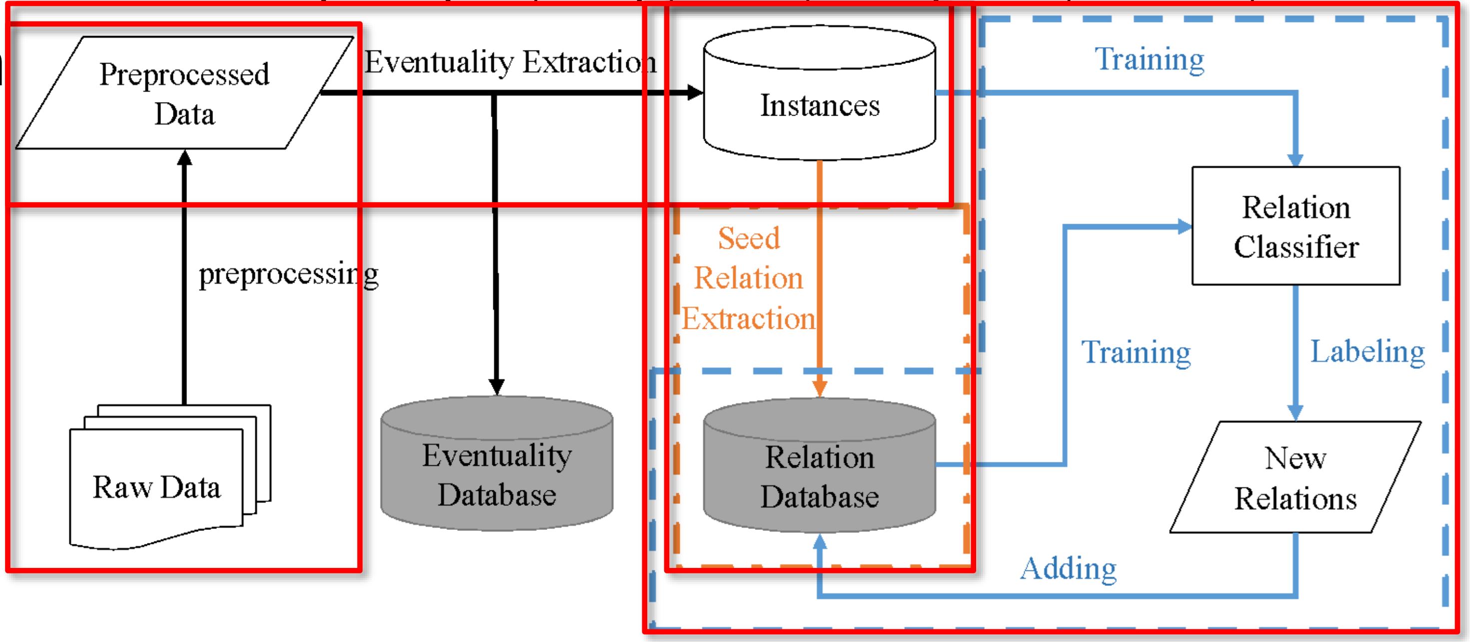
“Knock door”->“open up entrance”->“enter office”

- Quantity

Source	#Input Scripts	#Scenes	#Unique Activities	Parent	Participant	Prev	Next	Loc.	Time	Avg.
Movie scripts	560	148,296	244,789	0.87	0.86	0.78	0.85	0.79	0.79	0.84
TV series	290	886,724	565,394	0.89	0.85	0.81	0.84	0.82	0.84	0.86
Sitcoms	179	286,266	200,550	0.88	0.85	0.81	0.87	0.81	0.83	0.87
Novels	103	383,795	137,365	0.84	0.84	0.78	0.88	0.85	0.72	0.84
Crowdsrc.	25	3,701	9,575	0.82	0.91	0.91	0.87	0.74	0.40	0.86
Knowlywood	1,157	1,708,782	964,758	0.87	0.86	0.84	0.85	0.78	0.84	**0.85±0.01**
ConceptNet 5	-	-	4,757	0.15	0.81	0.92	0.91	0.33	N/A	**0.46±0.02**

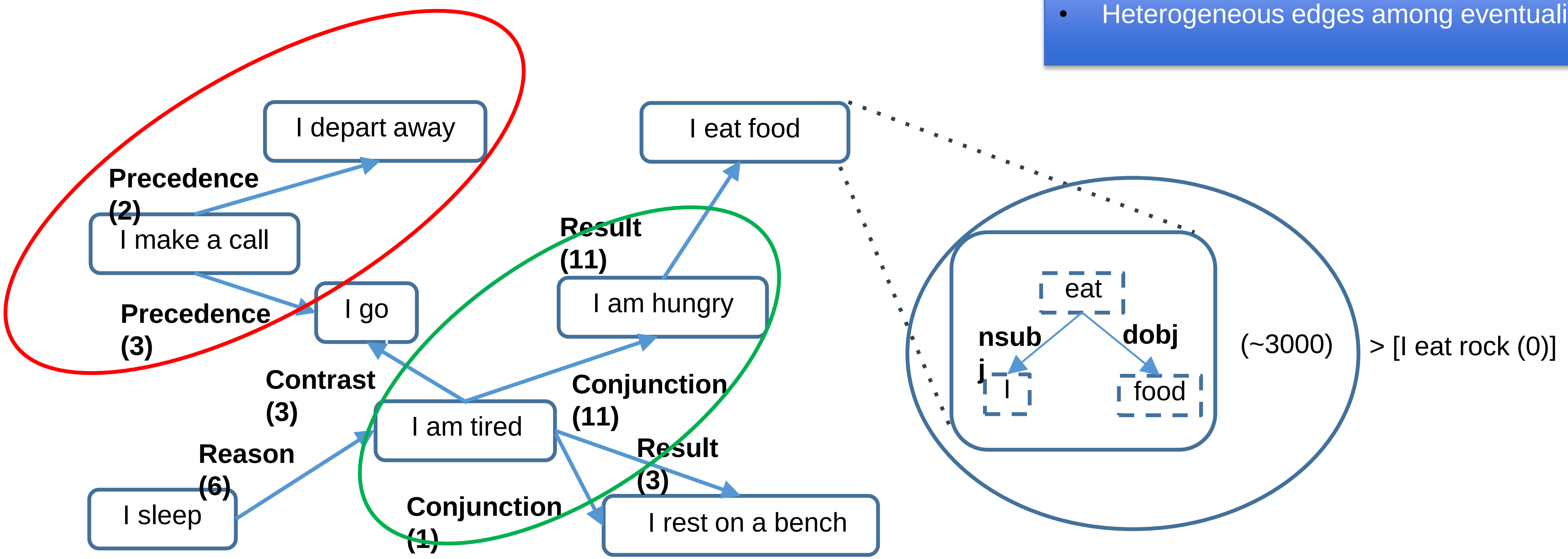
ASER (Zhang et al., 2020)

- KG Format
  - A Hybrid graph
  - Node: Eventualities in the format of dependency graphs
  - Edge: All discourse relations
- Resource
  - 11B token textual corpora (i.e., Yelp, NYT, Wikipedia, Reddit, Subtitles, E-books)
- Extraction



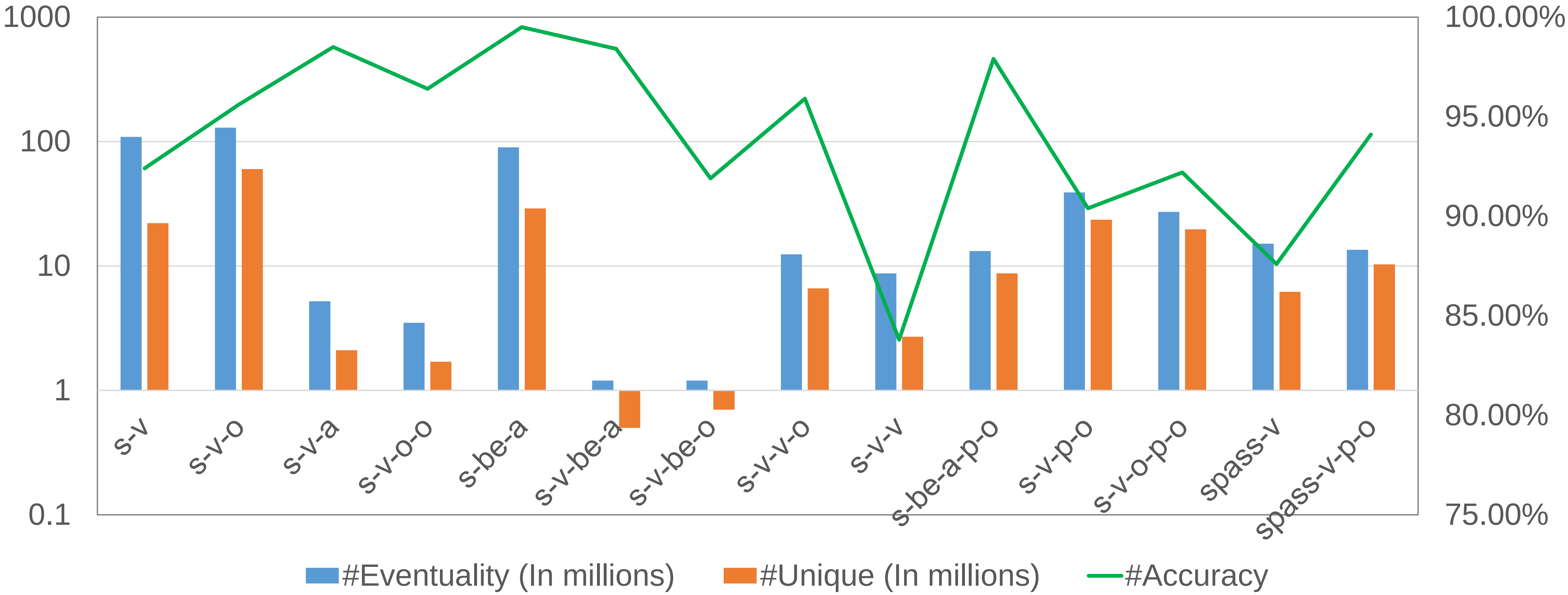
ASER Example

A hybrid graph of
- Each eventuality is a hyper-edge of words
- Heterogeneous edges among eventualities

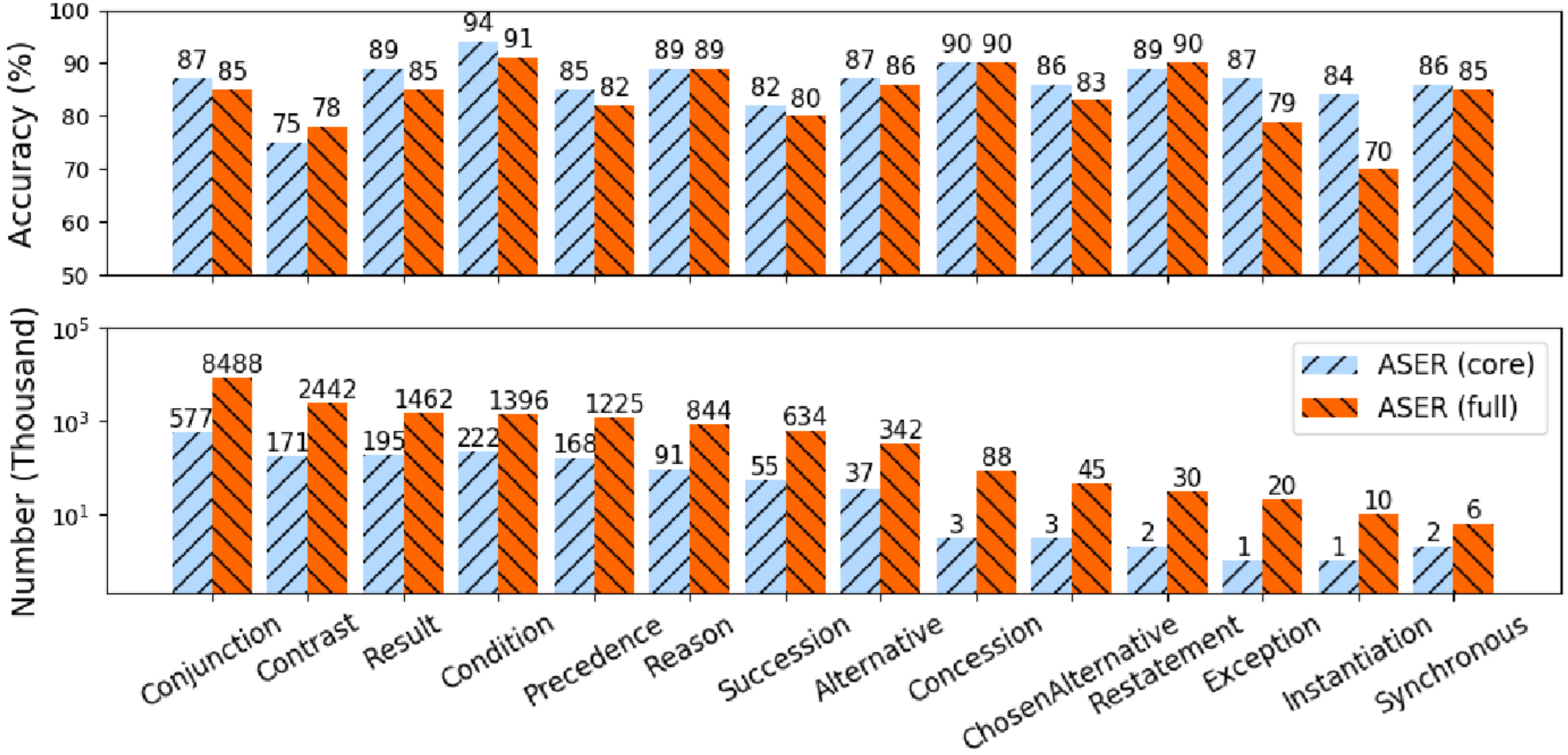


194 million eventualities, 64 million edges

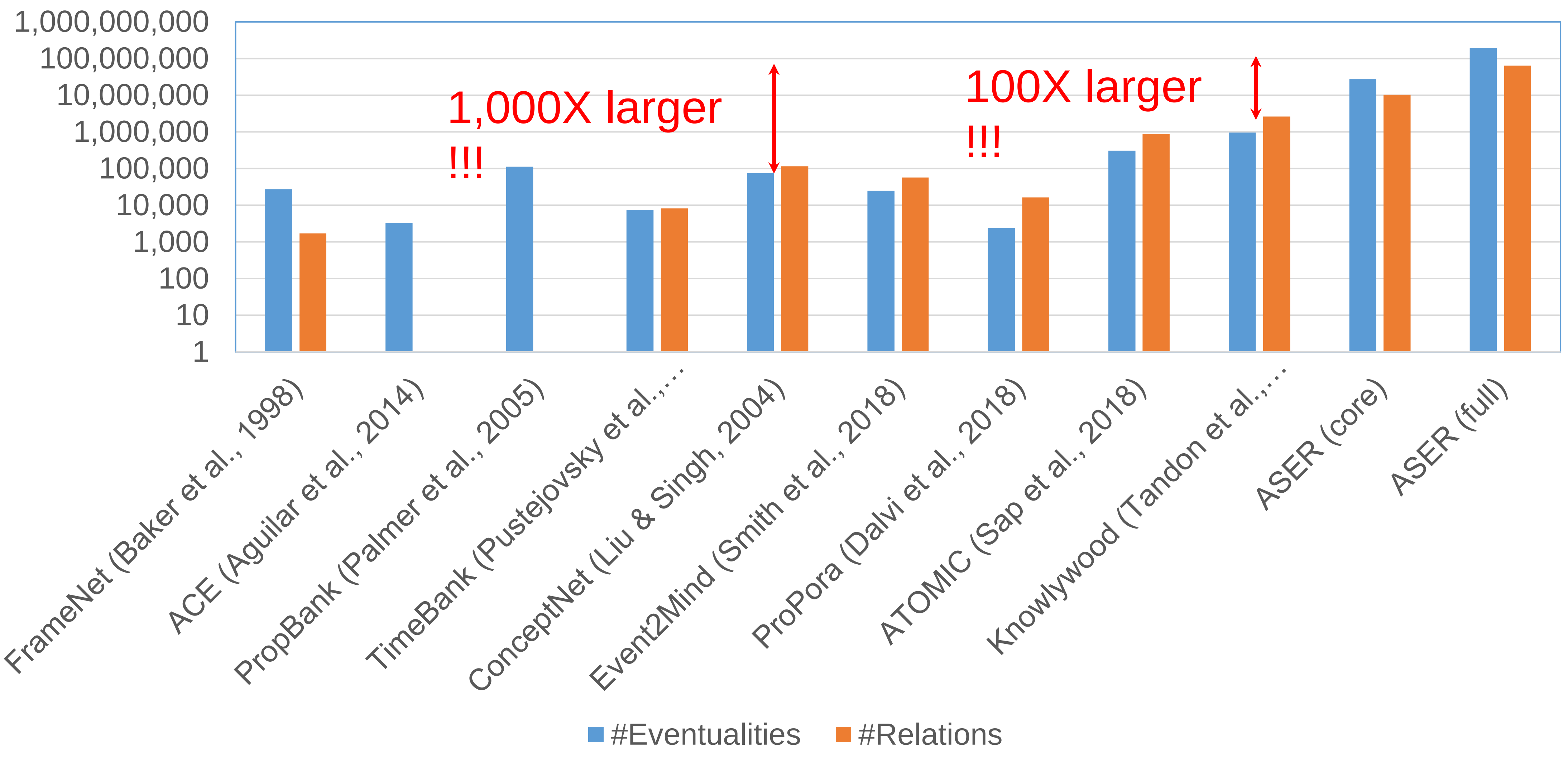
ASER Quantity and Quality (Eventuality)



ASER Quantity and Quality (Edge)



Comparison with Other event KGs

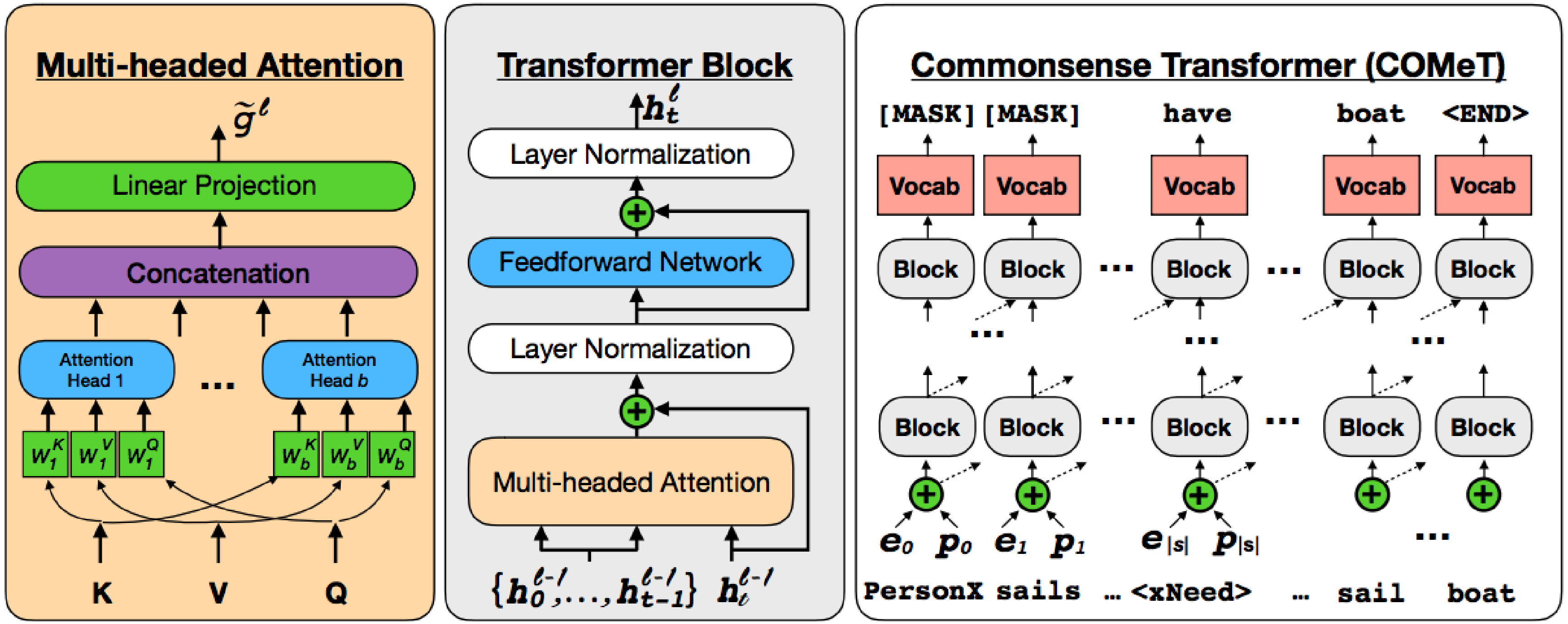


PS: In ConceptNet 5.0, more edges are added, but only the core part, which is inherited from ConceptNet 1.0 (Liu & Singh, 2004), is related to commonsense knowledge.

Outline

- Understanding Commonsense from the Angle of Events
- Instance-level Event Knowledge Acquisition
  - Human Annotation
  - Automatic Event Knowledge Extraction
  - Language Modeling
- Schema-level Event Knowledge Acquisition
- Conclusion

- Language Model
  - Examples: GPT-1/2/3
- COMET (Bosselut et al., 2019):
  - Commonsense Transformers for Automatic Knowledge Graph Construction

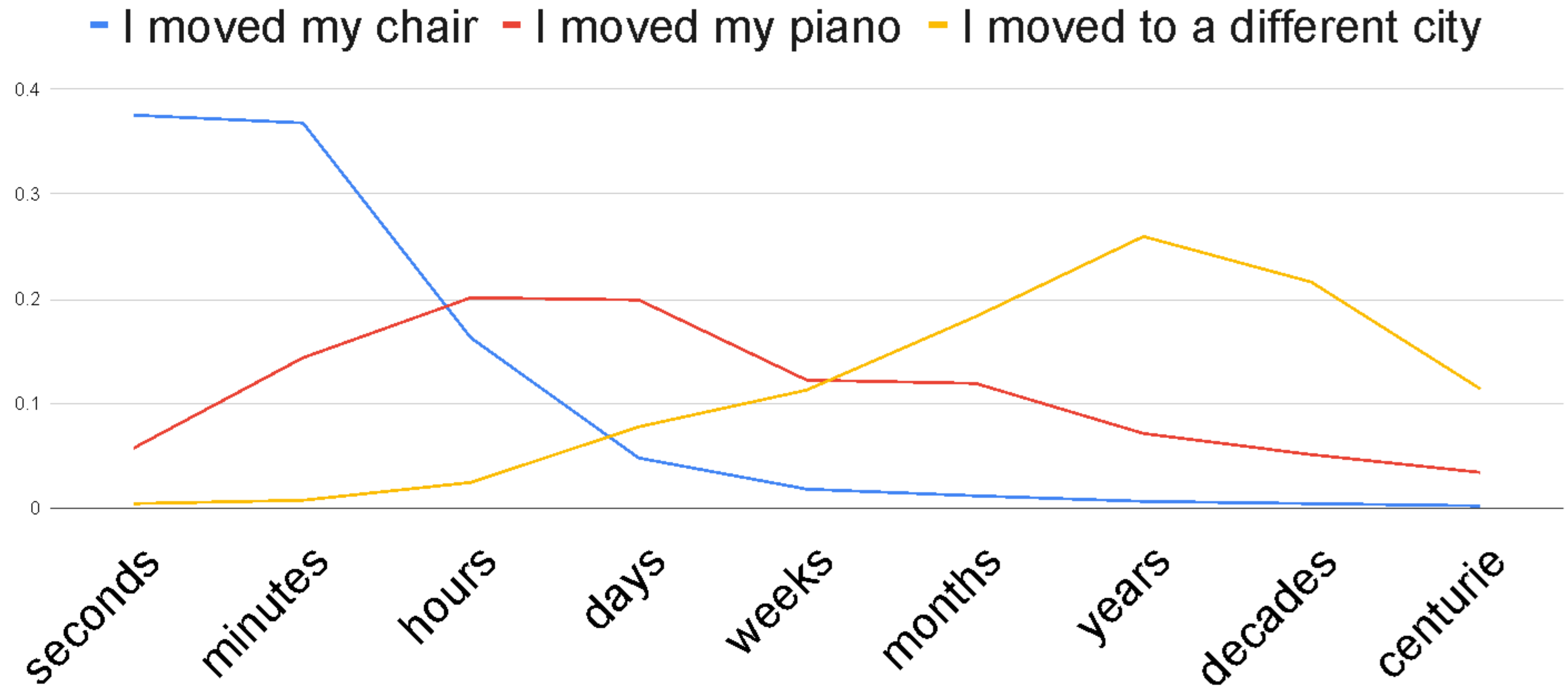


(a) (b) Event 1 (c) Event 2

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. COMET: Commonsense Transformers for Automatic Knowledge Graph Construction. ACL 2019.

Event Temporal Commonsense

- TacoLM (Zhou et al., 2020)
  - a general time-aware language model that distincts temporal properties in fine grained



Ben Zhou, Qiang Ning, Daniel Khashabi,, and Dan Roth. Temporal Common Sense Acquisition with Minimal Supervision. ACL 2020.

Event Temporal Commonsense

Goal: build a general time-aware LM with minimal supervision

Step 1: Information Extraction

- Use high-precision patterns to acquire temporal information
  - Unsupervised automatic extraction
- Overcomes reporting biases with a large amount of natural text

Step 2: Joint Language Model Pre-training

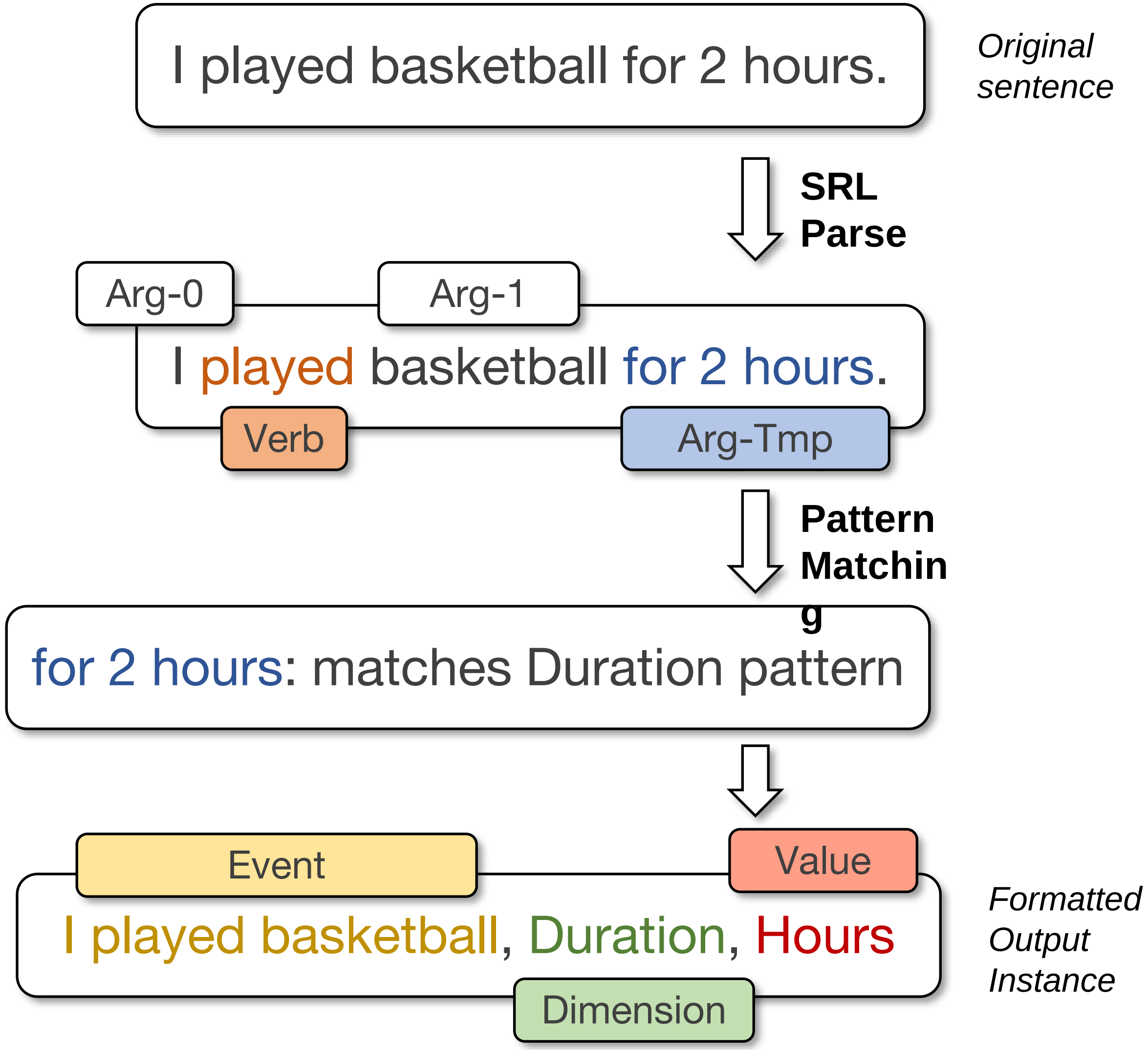
- Multiple temporal dimensions
  - Duration ~ 1 / Frequency

“I brush my teeth every morning” → Duration of “brushing teeth” < morning

  - Further generalization to combat reporting biases

Output: TacoLM- a time-aware general BERT

Event Temporal Commonsense



Information Extraction

I [M] played basketball [SEP] [M] [DUR] [HRS]

- Baseline Model: Pre-trained BERT-base
- Main objective: mask some tokens and recover them
- How we mask:
  - With some probability, mask temporal

I [M] played basketball [SEP] [M] [DUR] **[MASK]**

  - Otherwise, mask a certain portion of unchanged

I [M] **[MASK] [MASK]** [SEP] [M] [DUR] **[HRS]**

  - Max (P(Event|Dim,Val) + P(Val|Event,Dim)); Preserving original LM capability

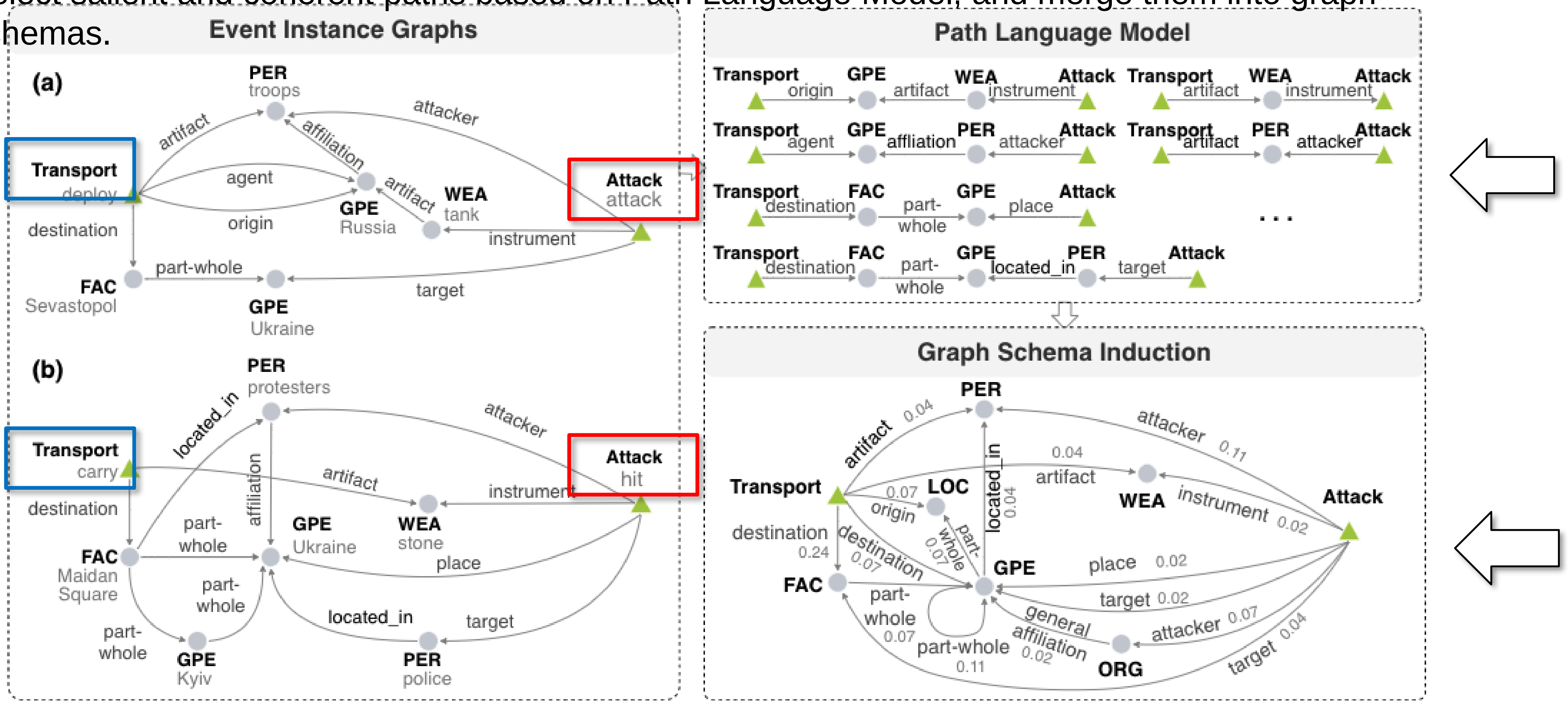
Joint training with language model

Outline

- Understanding Commonsense from the Angle of Events
- Instance-level Event Knowledge Acquisition
  - Human Annotation
  - Automatic Event Knowledge Extraction
  - Language Modeling
- Schema-level Event Knowledge Acquisition
- Conclusion

Event Graph Schema Induction (Li et al., 2020)

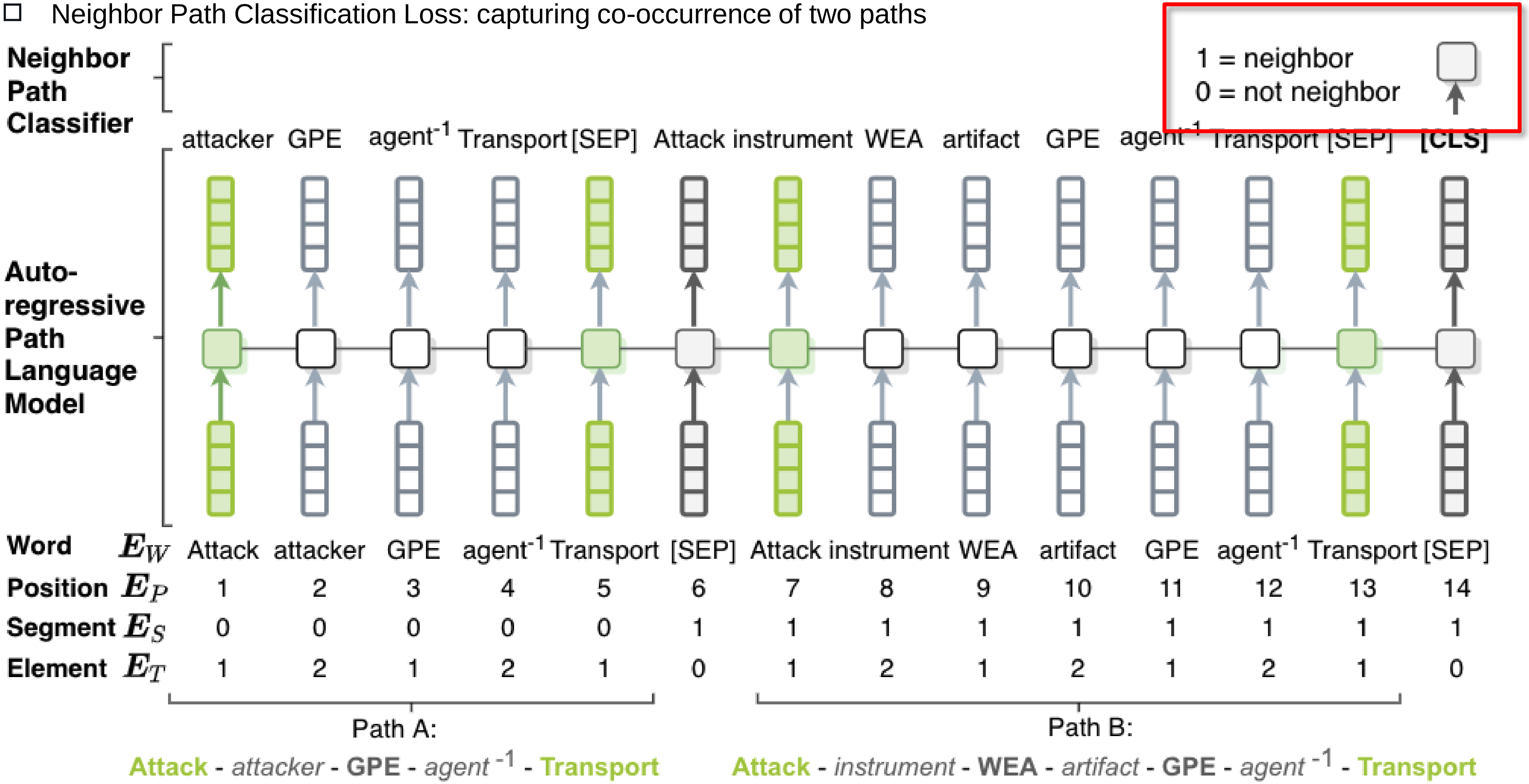
- History repeats itself: Instance graphs (a) and (b) refer to very different event instances, but they both illustrate a same scenario.
- Select salient and coherent paths based on Path Language Model, and merge them into graph schemas.



Manling Li, Qi Zeng, Kyunghyun Cho, Heng Ji, Jonathon May, Nathanael Chambers, Clare Voss. Connecting the Dots: Event Graph Schema Induction with Path Language Modeling. ACL 2020.

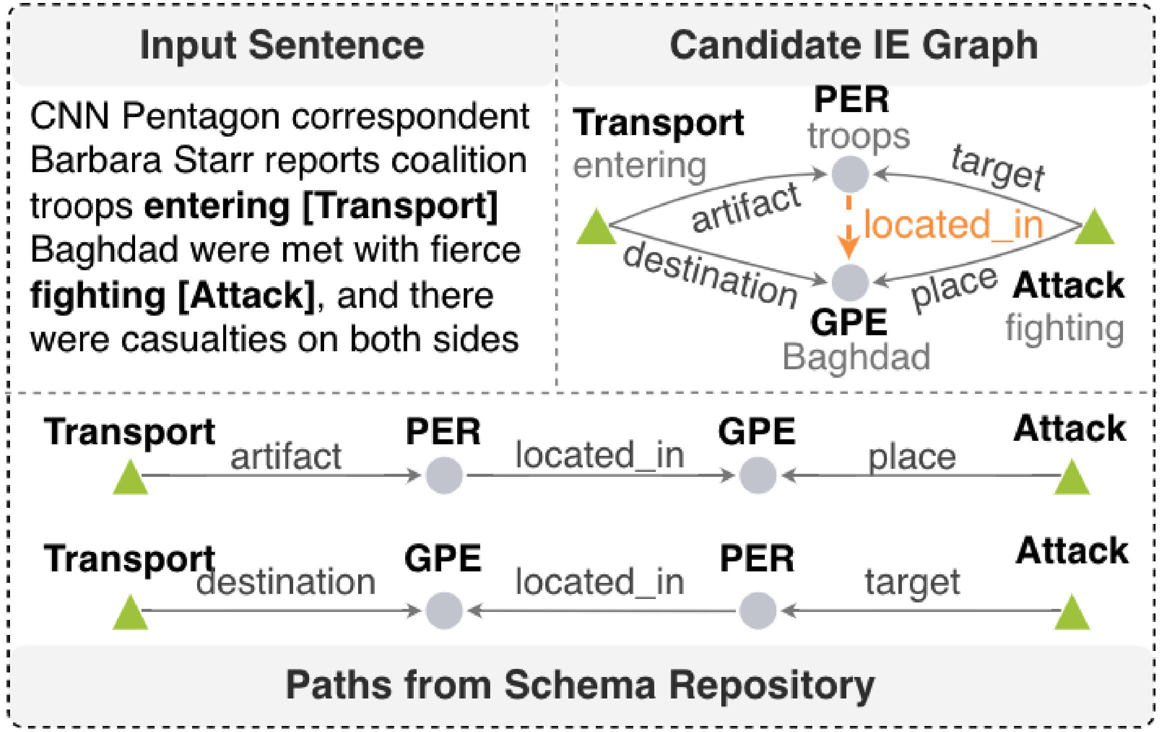
Path Language Model

- 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



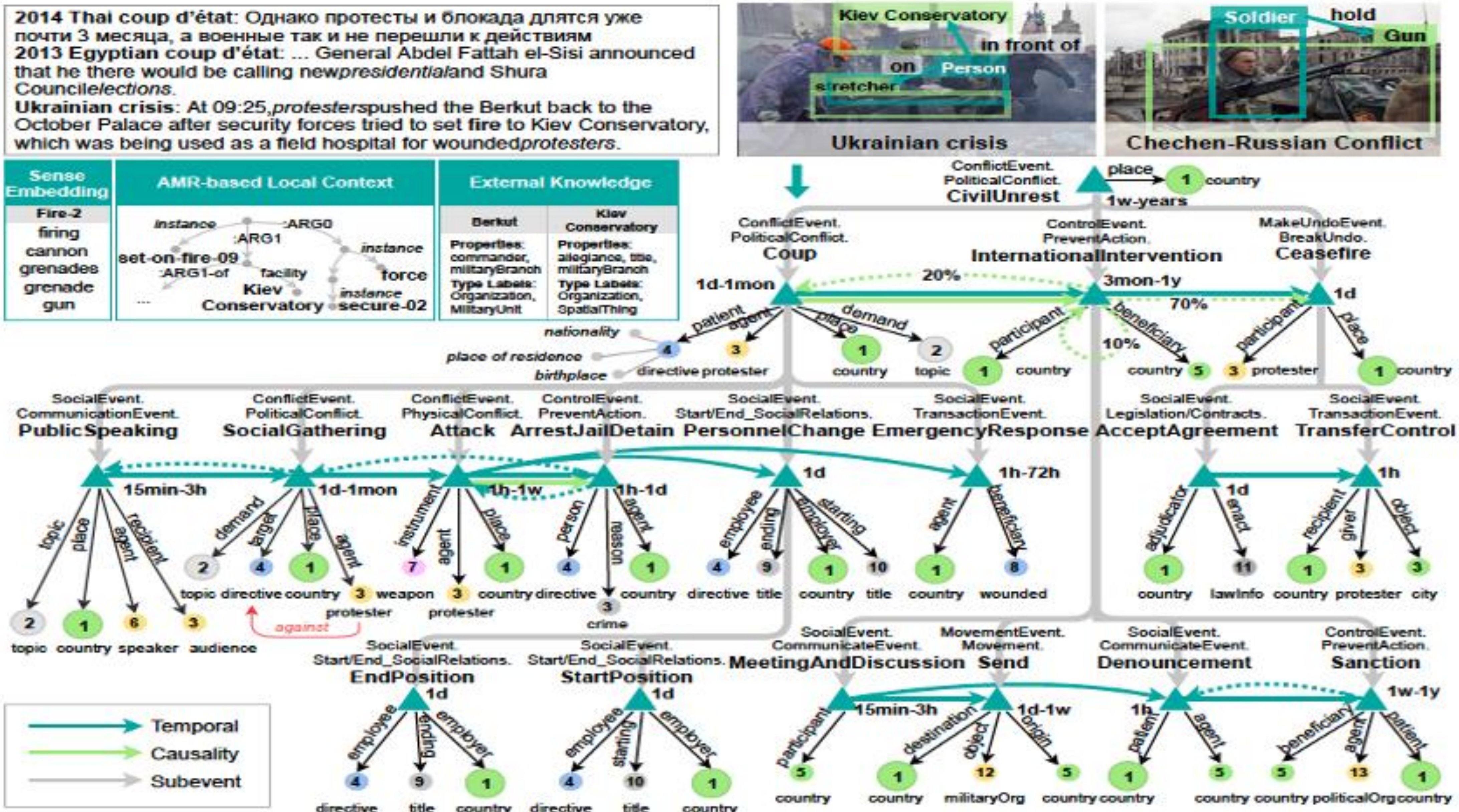
Schema-Guided Information Extraction

- 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



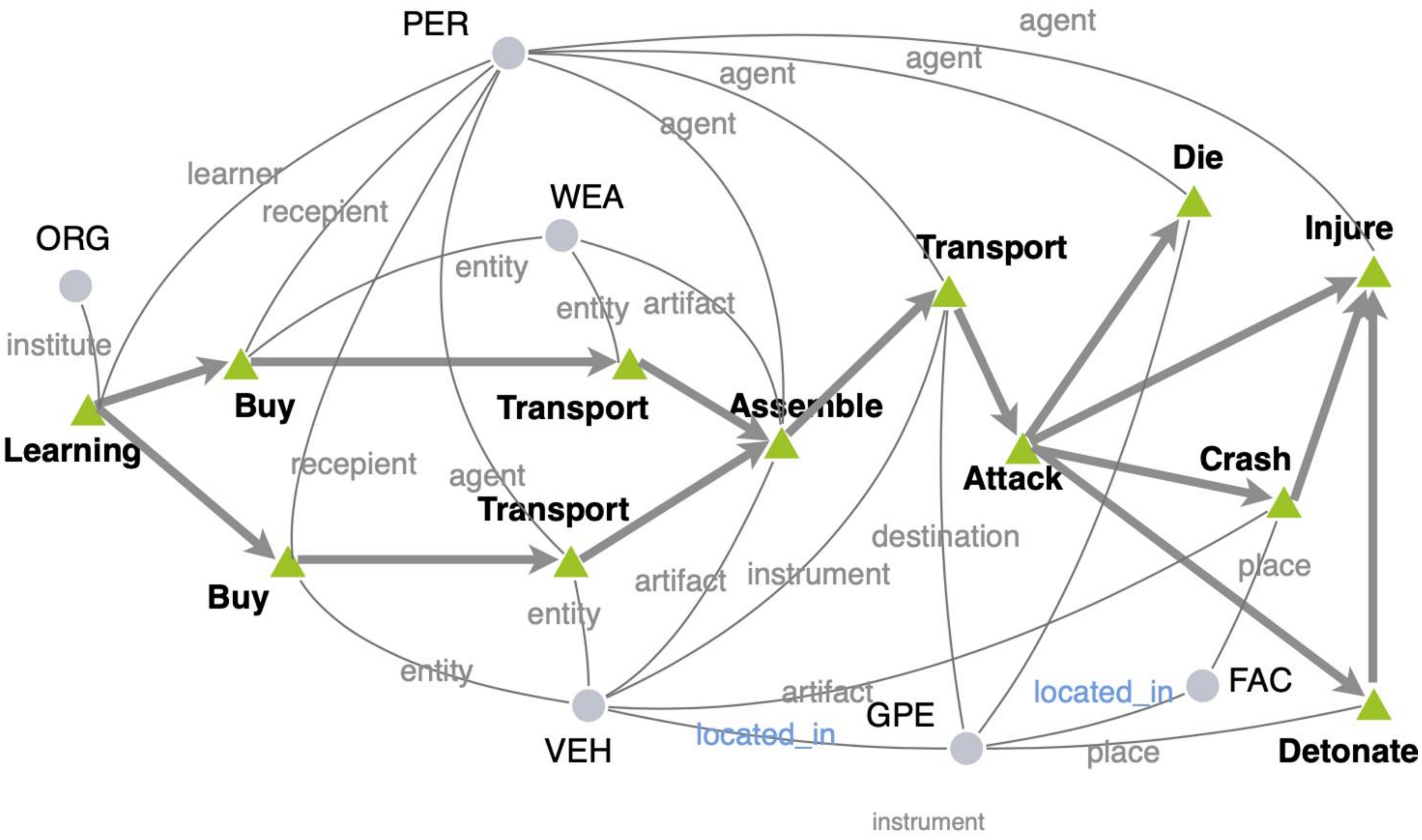
Dataset	Entity	**Event Trigger** Identification	**Event Trigger** Classification	**Event Argument** Identification	**Event Argument** Classification	Relation
Baseline	90.3	75.8	72.7	57.8	55.5	44.7
+PathLM	90.2	**76.0**	**73.4**	**59.0**	**56.6**	**60.9**

Event Schema Induction for Event Prediction



Temporal Complex Event Schema (Li et al, 2021)

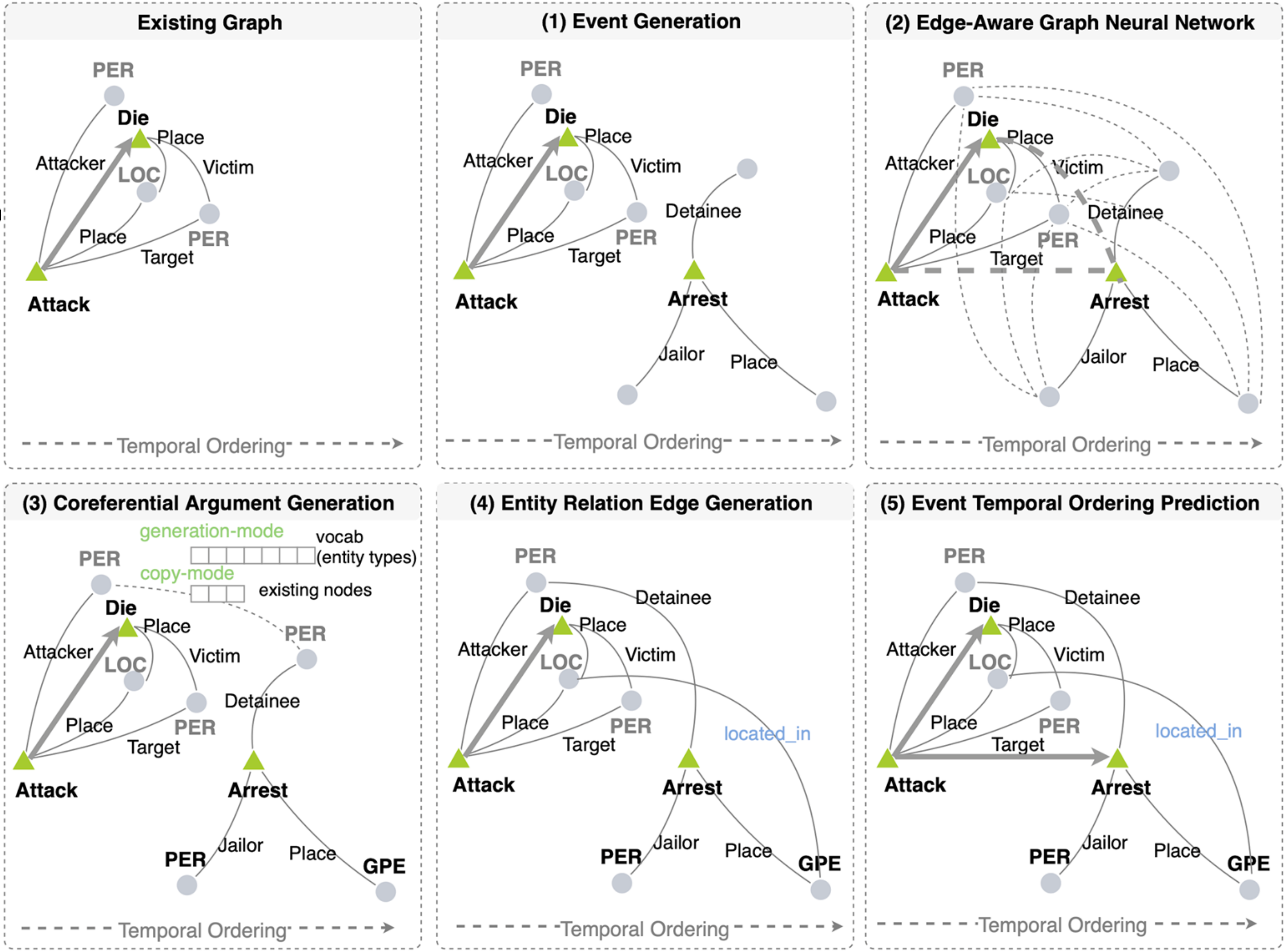
- Graph Structure Aware:
  - Encode entity coreference and entity relation
  - Capture the interdependency of events and entities (sequences can not)
- Scenario guided:
  - Train one model based on instance graphs of the same scenario
- Probabilistic:
  - Support downstream tasks, such as event



Generative Event Graph Model

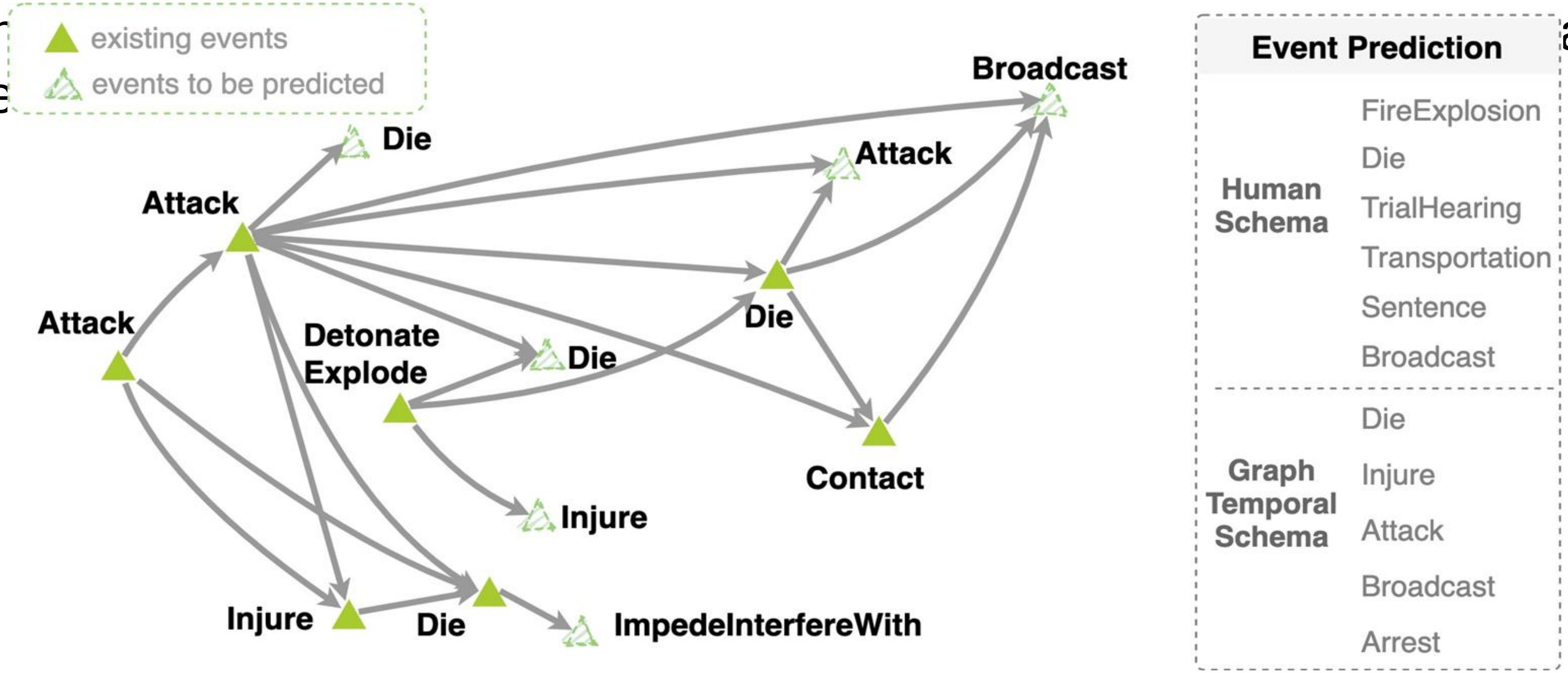
- Schemas are the hidden knowledge to control instance graph generatio

- Step 1.
  Event Node Generation
- Step 2.
  Message Passing
- Step 3.
  Argument Node Generation
- Step 4.
  Relation Edge Generation
- Step 5.
  Temporal Edge Generation



Extrinsic Evaluation

- **Schema-guided Event Prediction:** The task aims to predict ending events of each graph.
  - Considering that there can be multiple ending events in one instance graph, we rar aluation me



Dataset	Models	MRR	HITS@1
General	Human Schema	0.173	0.205
	Event Graph Model	**0.401**	**0.520**

Dataset	Models	MRR	HITS@1
IED	Human Schema	0.072	0.222
	Event Graph Model	**0.223**	**0.691**

Outline

- Understanding Commonsense from the Angle of Events
- Instance-level Event Knowledge Acquisition
  - Human Annotation
  - Automatic Event Knowledge Extraction
  - Language Modeling
- Schema-level Event Knowledge Acquisition
- Conclusion

Key takeaways

- There is a transferability from event knowledge to commonsense knowledge
- Compared with commonsense, acquiring event knowledge is cheaper and more scalable.
- All existing acquisition systems have advantages and limitations.

	Quality	Scale	Relation Coverage	Explainability	Robustness	Downstream Task
Human Annotation	**High**	Small	Middle	**High**	High	Difficult
Automatic Event Knowledge Extraction	Middle	**Large**	**High**	**High**	Middle	Difficult
Language Model	Middle	**Large**	**High**	Low	Low	**Easy**

Thanks

Key References

- WSC: Hector Levesque, Ernest Davis, and Leora Morgenstern. The winograd schema challenge. KRR 2012.
- *ConceptNet:* Hugo Liu and Push Singh, ConceptNet - a practical commonsense reasoning tool-kit, BTTJ, 2004
- ATOMIC: Maarten Sap, Ronan LeBras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A. Smith, Yejin Choi, ATOMIC: An Atlas of Machine Commonsense for If-Then Reasoning. AAAI 2019
- COMET: Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. COMET: commonsense transformers for automatic knowledge graph construction. ACL 2019.
- LAMA: Fabio Petroni, Tim Rocktaschel, Se- ¨ bastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander H. Miller. Language models as knowledge bases? EMNLP 2019.
- ASER: Hongming Zhang, Xin Liu, Haojie Pan, Yangqiu Song, and Cane Wing-Ki Leung. ASER: A Large-scale Eventuality Knowledge Graph. WWW 2020.
- TransOMCS: Hongming Zhang, Daniel Khashabi, Yangqiu Song, and Dan Roth. TransOMCS: From Linguistic Graphs to Commonsense Knowledge. International Joint Conference on Artificial Intelligence (IJCAI). 2020.
- KnowlyWood: Niket Tandon, Gerard de Melo, Abir De, and Gerhard Weikum. 2015. Knowlywood: Mining Activity Knowledge From Hollywood Narratives. CIKM 2015.
- Manling Li, Qi Zeng, Kyunghyun Cho, Heng Ji, Jonathon May, Nathanael Chambers, Clare Voss. Connecting the Dots: Event Graph Schema Induction with Path Language Modeling. ACL 2020.
- TacoLM: Ben Zhou, Qiang Ning, Daniel Khashabi, Dan Roth. Temporal Common Sense Acquisition with