Carnegie Mellon University

Language Technologies Institute

Ph.D. Thesis Defense

Extraction of Event Structures from Text

May 29, 2018

Jun Araki

Carnegie Mellon University

Thesis Committee:
Teruko Mitamura (Chair), Eduard Hovy,
Graham Neubig, and Luke Zettlemoyer

Events are Everywhere

Earthquakes

Olympic games

Payment

Picnics

Why Events? — Practical Reasons

- An overwhelming amount of text about events
- Event-oriented text analysis is crucial for stakeholders to make sensible decisions from a holistic view

Text

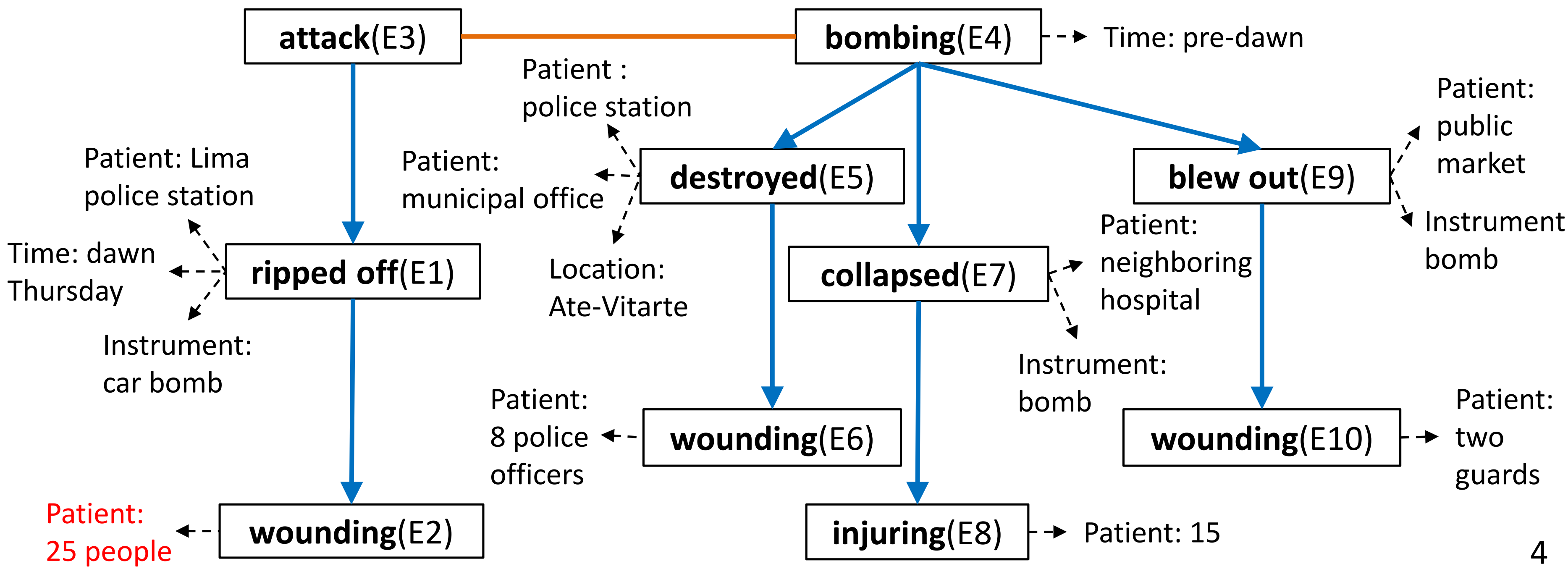
Knowledge bases & visualization

Stakeholders

Why Events? — Theoretical Reasons

- Events are a core component for natural language understanding

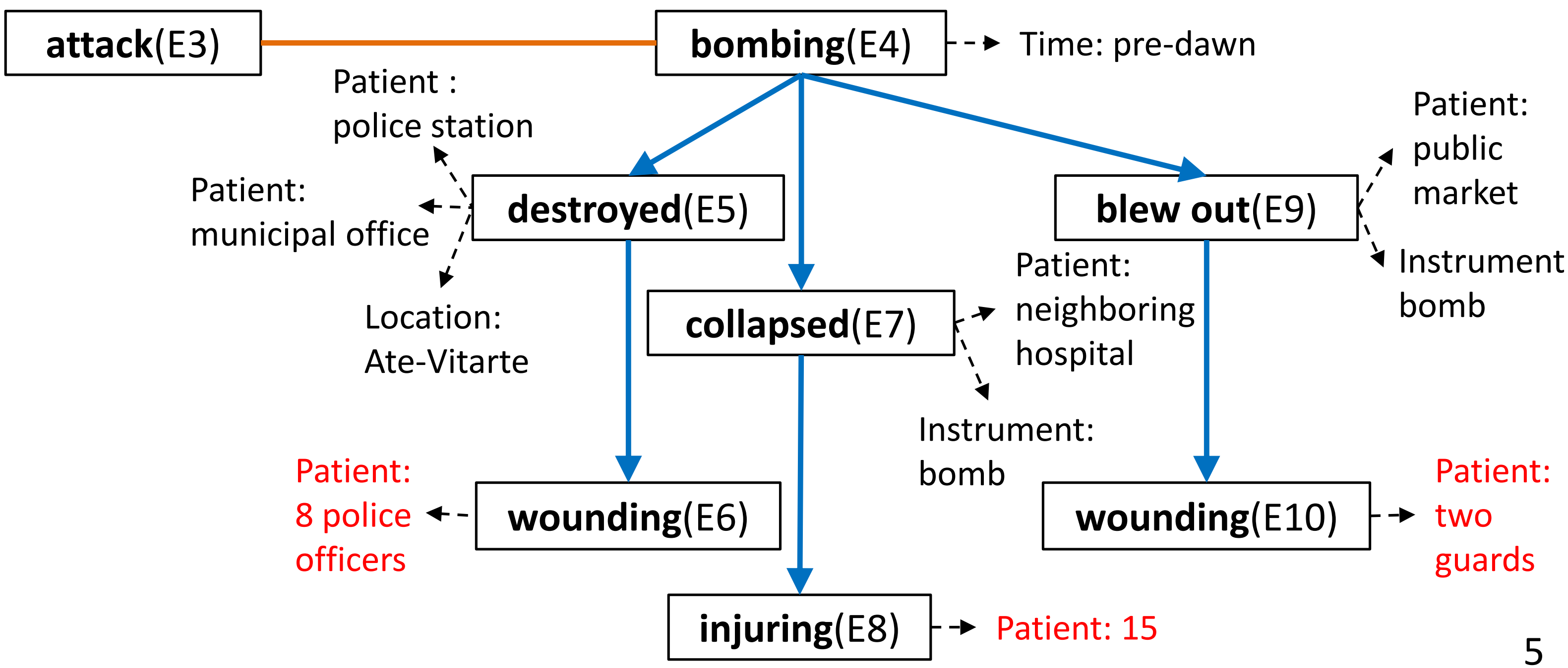
A car bomb that police said was set by Shining Path guerrillas **ripped off**(E1) the front of a Lima police station before dawn Thursday, **wounding**(E2) 25 people. The **attack**(E3) marked the return to the spotlight of the feared Maoist group, recently overshadowed by a smaller rival band of rebels. The pre-dawn **bombing**(E4) **destroyed**(E5) part of the police station and a municipal office in Lima's industrial suburb of Ate-Vitarte, **wounding**(E6) 8 police officers, one seriously, Interior Minister Cesar Saucedo told reporters. The bomb **collapsed**(E7) the roof of a neighboring hospital, **injuring**(E8) 15, and **blew out**(E9) windows and doors in a public market, **wounding**(E10) two guards.



Why Events? — Theoretical Reasons

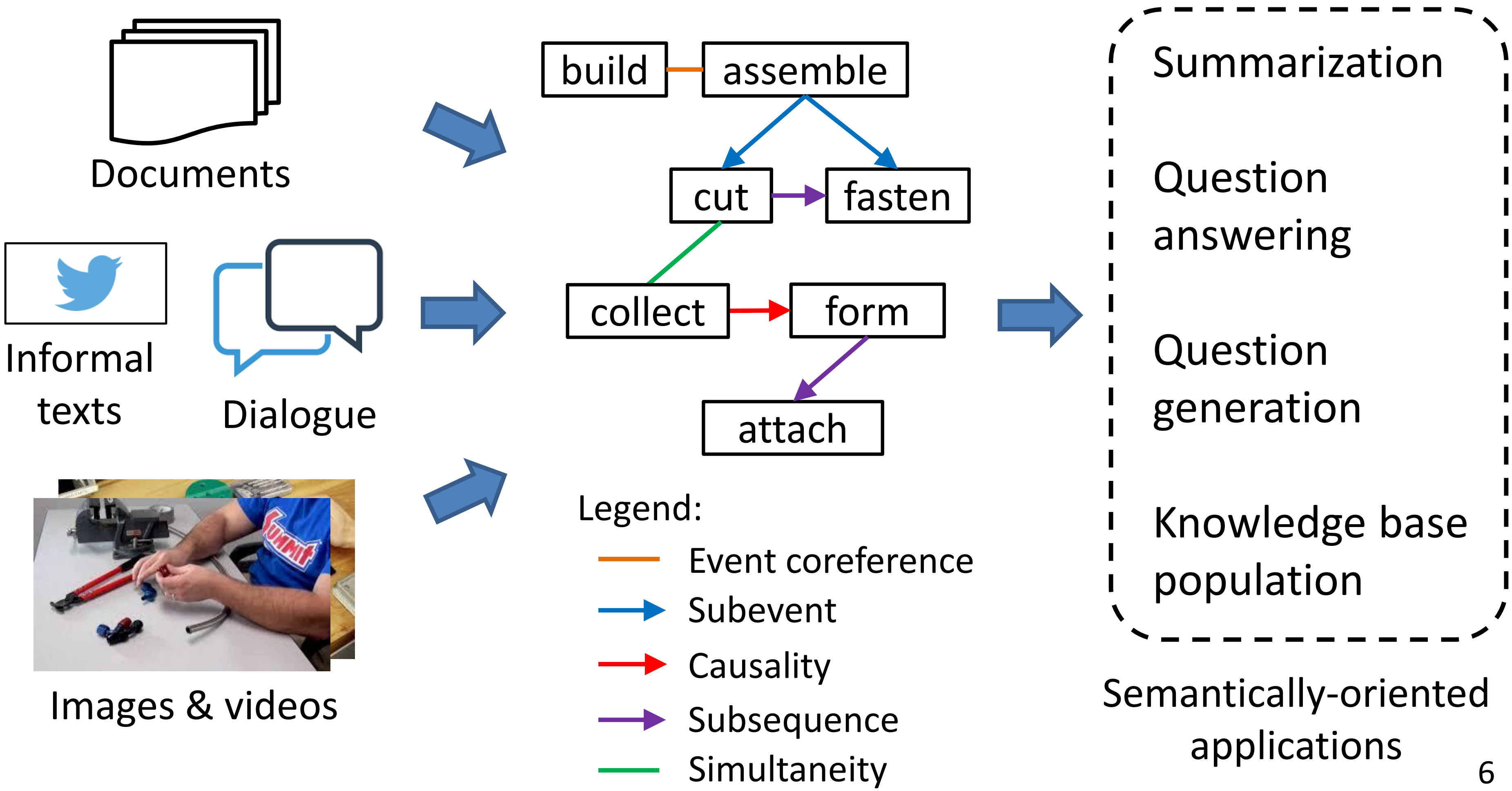
- Events are a core component for natural language understanding

The **attack**(E3) marked the return to the spotlight of the feared Maoist group, recently overshadowed by a smaller rival band of rebels. The pre-dawn **bombing**(E4) **destroyed**(E5) part of the police station and a municipal office in Lima's industrial suburb of Ate-Vitarte, **wounding**(E6) 8 police officers, one seriously, Interior Minister Cesar Saucedo told reporters. The bomb **collapsed**(E7) the roof of a neighboring hospital, **injuring**(E8) 15, and **blew out**(E9) windows and doors in a public market, **wounding**(E10) two guards.



Research Vision

- Event structures represent **core semantic backbones**
  - A meaningful representation to go beyond sentence-level NLP



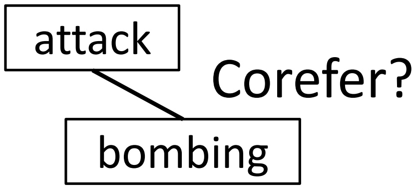
Thesis Goal

- The central goal of this thesis is:

> To devise a computational method that models the structural property of events in a principled framework for event detection and event coreference resolution

Overview: Thesis Contributions

- <u>Before</u> this thesis

Task	Problem	
Event detection	P1: Restricted annotation	Closed domains (e.g., 33 types in ACE) "turn the TV on"?
Event detection	P2: Data sparsity	Human annotation is expensive
Event detection / Event coreference resolution	P3: Event interdependencies	Pipeline models propagate errors
Event coreference resolution	P4: Lack of subevent detection	
Event coreference resolution	P5: Limited applications	Applications for NLU by humans?

Overview: Thesis Contributions

- After this thesis

Task	Problem	Approach	Theory
Event detection	P1: Restricted annotation	Open-domain event detection	Eventualities
	P2: Data sparsity	Distant supervision	Realis
	P3: Event interdependencies	Joint modeling	Event identity
Event coreference resolution	P4: Lack of subevent detection	Subevent structure detection	Educational theory
	P5: Limited applications	Question generation	

Outline

- Introduction
- Event detection

  P1: Restricted annotation — Open-domain event detection [Araki+ COLING 2018]

  P2: Data sparsity — Distant supervision

- Event coreference resolution

  P3: Event interdependencies — Joint modeling [Araki+ EMNLP 2015]

  P4: Lack of subevent detection — Subevent structure detection [Araki+ LREC 2014]

  P5: Limited applications — Question generation [Araki+ COLING 2016]

- Conclusion & future work

Problems with Closed-Domain Event Detection

- Limited coverage of events
  - Prior work focuses on limited event types
    - MUC, ACE, TAC KBP, GENIA, BioNLP, and ProcessBank
- Lack of training data
  - Human annotation of events is expensive
    - Supervised models overfit to small data

Task: TAC KBP 2017 Detection of event spans and types

	Model	Precision	Recall	F1
Prior work (Official results)	Top 5	57.02	42.29	48.56
	Top 4	47.10	50.18	48.60
	Top 3	54.27	46.59	50.14
	Top 2	52.16	48.71	50.37
	Top 1	56.83	**55.57**	56.19
Our models	BLSTM	69.79	41.31	51.90
	BLSTM-CRF	**70.15**	41.06	51.80
	BLSTM-MLC	68.03	48.53	**56.65**

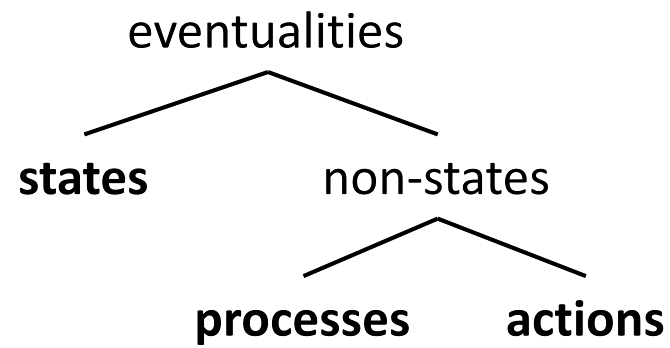
Problems with Open-Domain Event Detection

- Limited coverage of events
  - Some prior work has conceptually different focuses
    - PropBank, NomBank, and FrameNet
  - Other prior work focuses on limited syntactic types
    - OntoNotes, TimeML, ECB+, and RED
- Lack of training data
  - Human annotation of events in the open domain is further expensive

- We propose **a new paradigm of open-domain event detection**:
  - Detect all kinds of events without any specific event types
  - Generate high-quality training data automatically

Definition of Events

- Eventualities [Bach 1986]
  - A broader notion of events
  - Consist of 3 components:



Component	Definition	Examples
states	a class of notions that are durative and changeless	want, own, love, resemble
processes	a class of notions that are durative and do not have any explicit goals	walking, sleeping, raining
actions	a class of notions that have explicit goals or are momentaneous happenings	build, walk to Pittsburgh, recognize, arrive, clap

Bach, E. The algebra of events. Linguistics and Philosophy, 9:5–16. 1986.

Definition of Events

- Event nuggets [Mitamura+ 2015]
  - A semantically meaningful unit that expresses an event
- Syntactic scope:
  - Verbs
    - Single-word verbs
    - Verb phrases
      - Continuous
      - Discontinuous
  - Nouns
    - Single-word nouns
    - Noun phrases
    - Proper nouns
  - Adjectives
  - Adverbs

Examples:

The child **broke** a window …

She **picked up** a letter.
He **turned** the TV **on** … / She **sent** me an **email**.

The **discussion** was …
… maintained by **quality control** of …
Hurricane Katrina was …

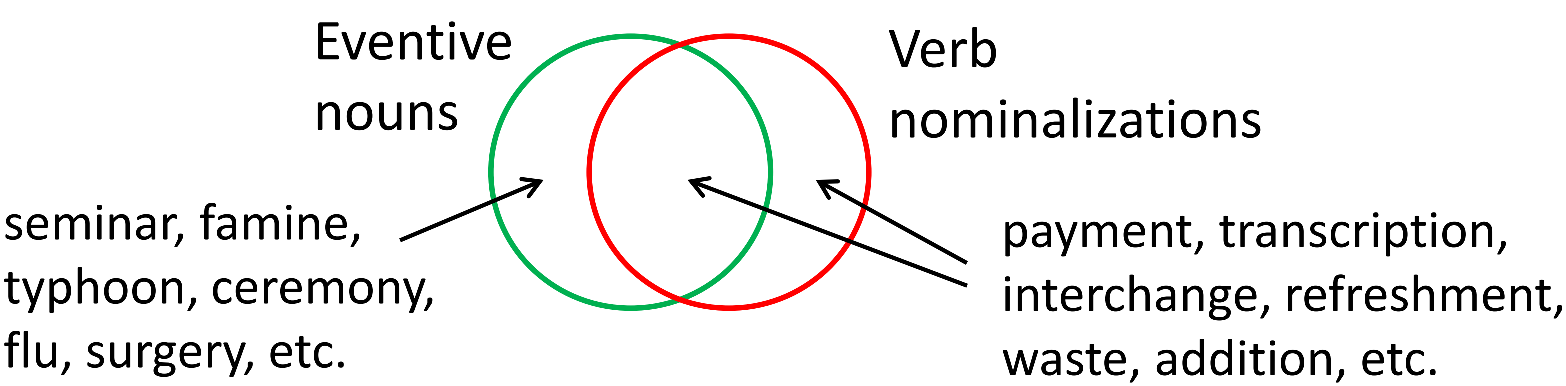
She was **talkative** at the party.

She **replied dismissively** to …

Mitamura, T., Yamakawa, Y., Holm, S., Song, Z., Bies, A., Kulick, S., and Strassel, S. Event nugget annotation: Processes and issues. NAACL-HLT 2015 Workshop on Events: Definition, Detection, Coreference, and Representation.

Difficult Cases

- Ambiguities on eventiveness (**events** vs. *non-events*):
  - That is what I **meant**.
  - ‘Enormous’ *means* ‘very big.’
  - His **payment** was late.
  - His *payment* was $10.
  - **Force** equals mass times acceleration.
  - Mary was **talkative** at the party.
  - Mary is a *talkative* person.
- Eventive nouns
  - Cannot be simply approximated by verb nominalizations



Distant Supervision from WordNet

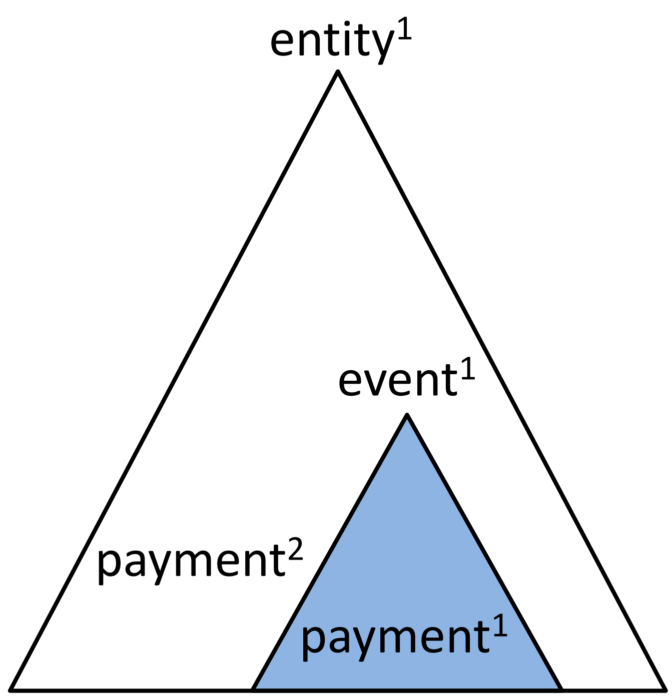
- Assumption:
  - There is a semantically adequate correspondence between components of eventualities and WordNet senses

Eventualities (by Bach)		WordNet	
Component	Definition	Sense	Gloss (Brief Definition)
states	a class of notions that are durative and changeless	**state**^{2}	the way something is with respect to its main attributes
processes	a class of notions that are durative and do not have any explicit goals	**process**^{6}	a sustained phenomenon or one marked by gradual changes through a series of states
actions	a class of notions that have explicit goals or are momentaneous happenings	**event**^{1}	something that happens at a given place and time

Distant Supervision from WordNet

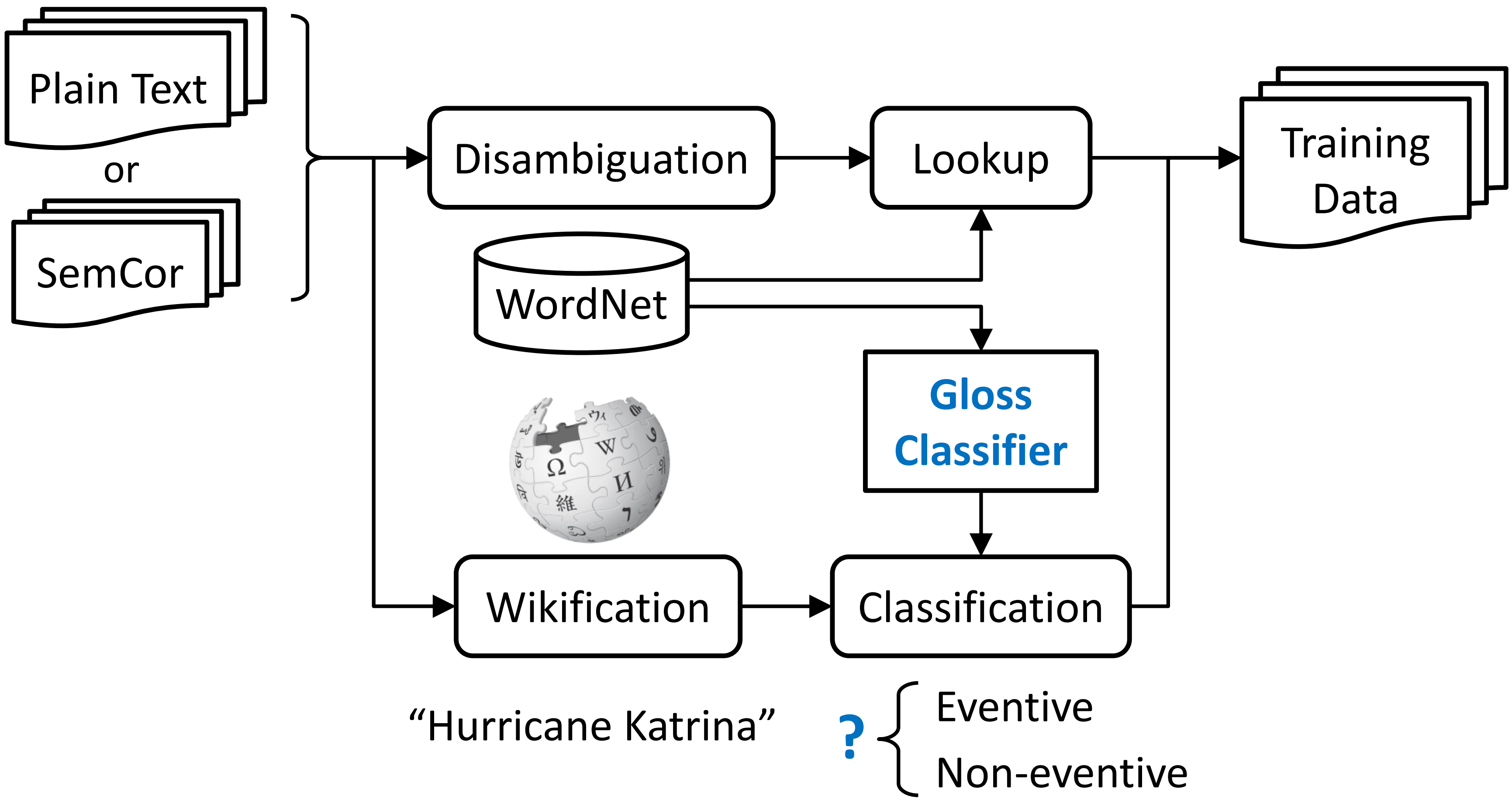
- Assumption:
  - WordNet's hyponym taxonomy provides a reasonable approximation of eventive nouns

Label	Sense	Gloss
Eventive	payment^1	the act of paying money
Non-eventive	payment^2	a sum of money paid or a claim discharged



Training Data Generation: Overview

- Baseline: Disambiguation + WordNet lookup
- Capture proper nouns using Wikipedia knowledge
  - WordNet coverage is limited



Gloss Classification — Heuristics-based

- Assumptions:
  - The first sentence of a Wikipedia article provides a high-quality gloss
  - The **syntactic head** of the gloss represents a high-level concept to decide eventiveness
- Example:

Entry	The first sentence of the Wikipedia article
Hurricane Katrina	Hurricane Katrina was an extremely destructive and deadly tropical **cyclone** that is tied with Hurricane Harvey of 2017 as the costliest hurricane on record.

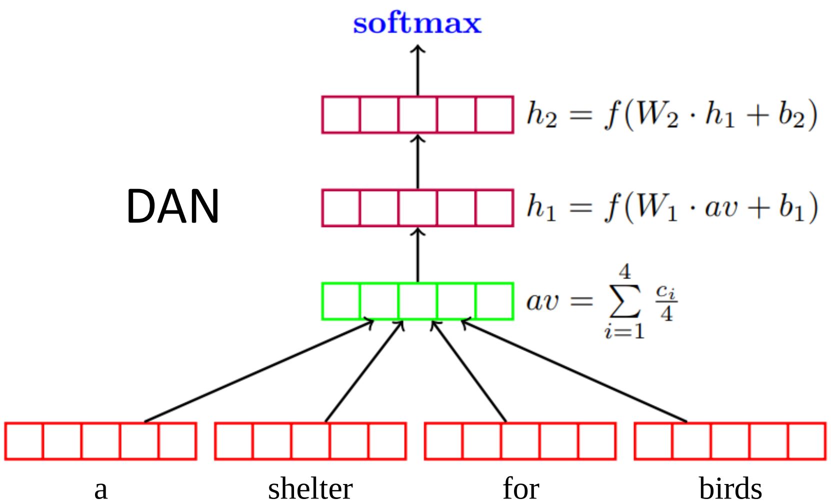
Wikipedia gloss

- Heuristics-based algorithm: HeadLookup
  - (1) Get the head and disambiguate it
  - (2) Look up the head's sense in WordNet

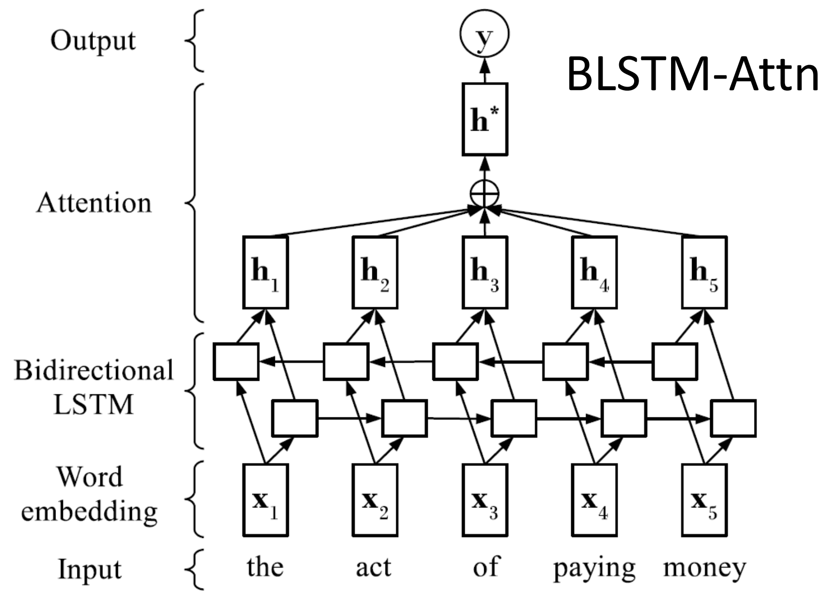
Gloss Classification — Learning-based

- Collect gloss dataset $D = D_p \cup D_n$ from WordNet automatically
  - D_p = {gloss whose sense is under state[2], process[6], or event[1]} $|D_p| = 13{,}415$
  - D_n = {all the other glosses of WordNet nouns} $|D_n| = 68{,}700$
- Train classifiers to minimize binary cross-entropy loss
  - Bag-of-words model with logistic regression
  - Deep average network (DAN) [Iyyer+ 2015]
  - BLSTM with self-attention [Lin+ 2017]

DAN



BLSTM-Attn

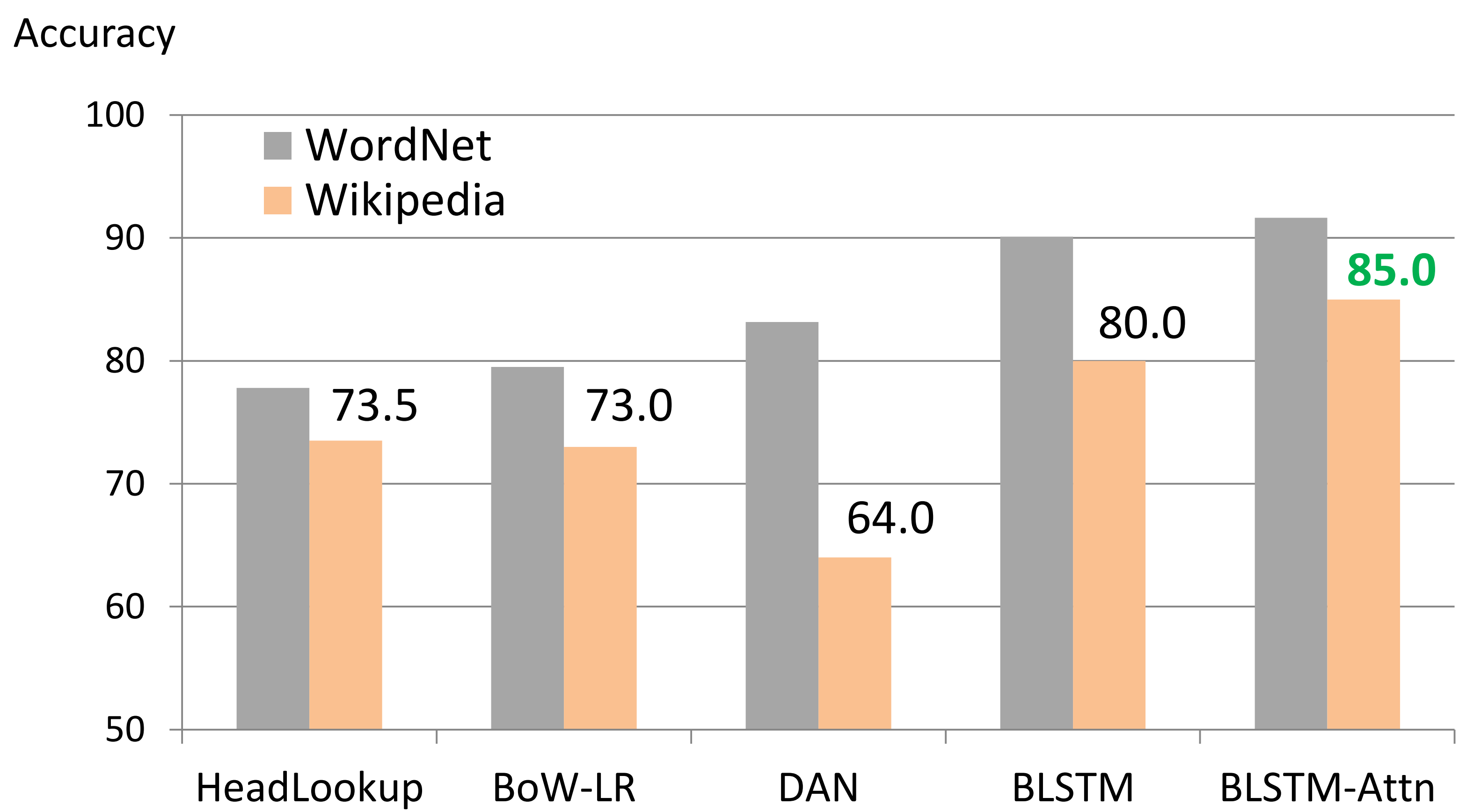


Iyyer, M., Manjunatha, V., Boyd-Graber J., and Daume III, H. Deep unordered composition rivals syntactic methods for text classification. ACL 2015.

Lin, Z., Feng, M., Santos, C., Yu, M., Xiang, B., Zhou, B., and Bengio, Y. A structured self-attentive sentence embedding. ICLR 2017.

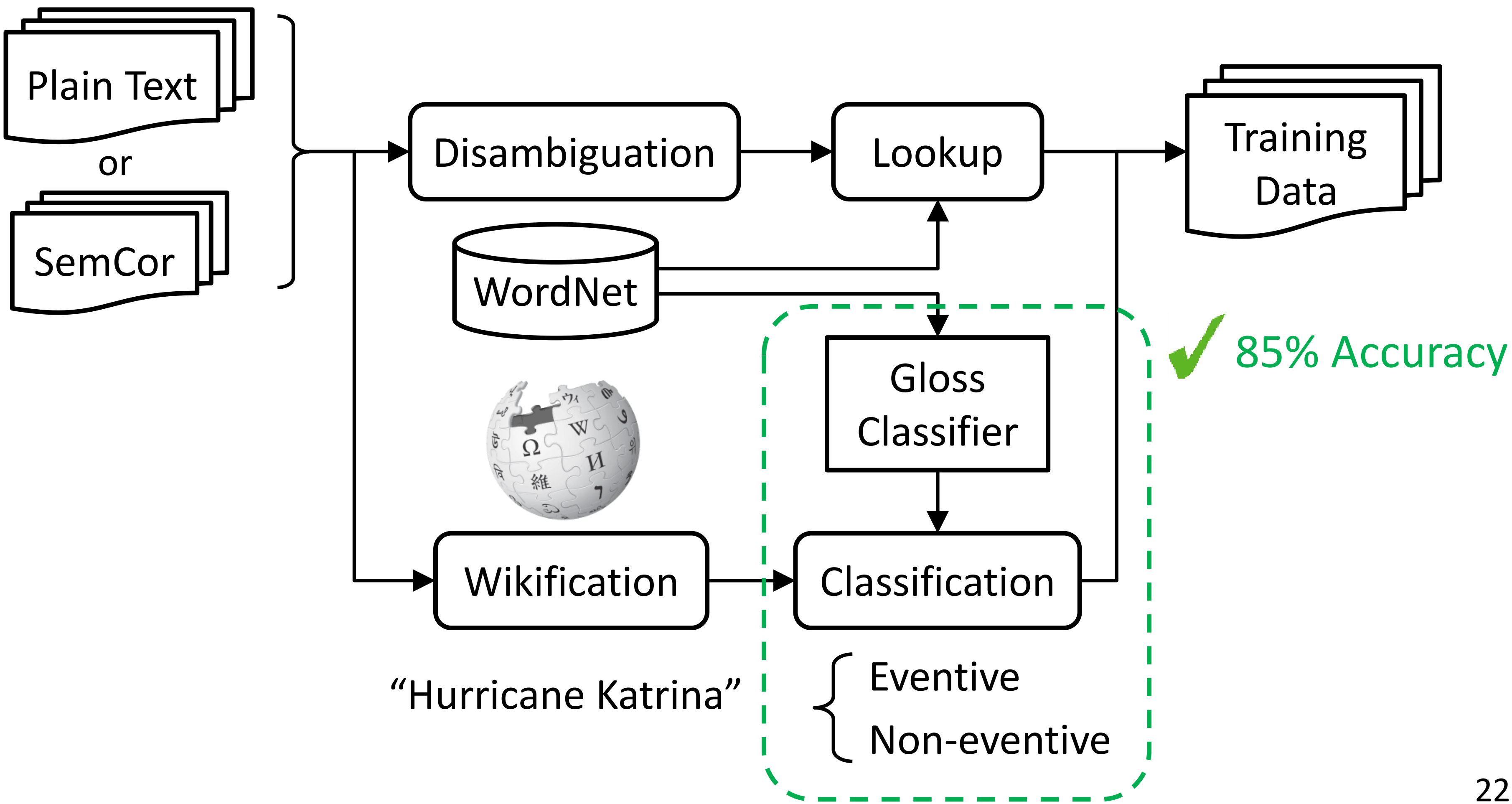
Results: Gloss Classification

- Test data
  - WordNet: 2,000 examples randomly sampled from D_p and D_n
  - Wikipedia: 200 examples manually created in 10 domains



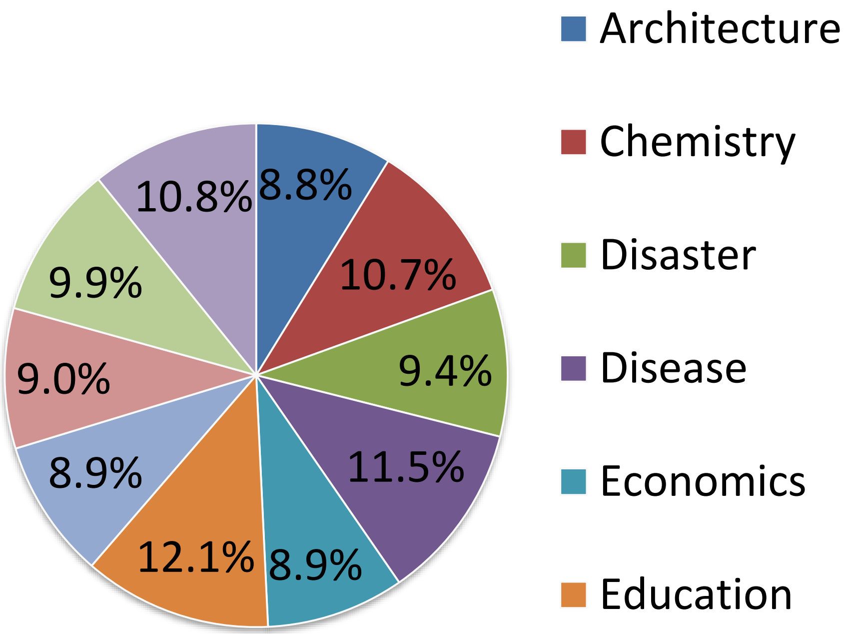
Training Data Generation: Overview

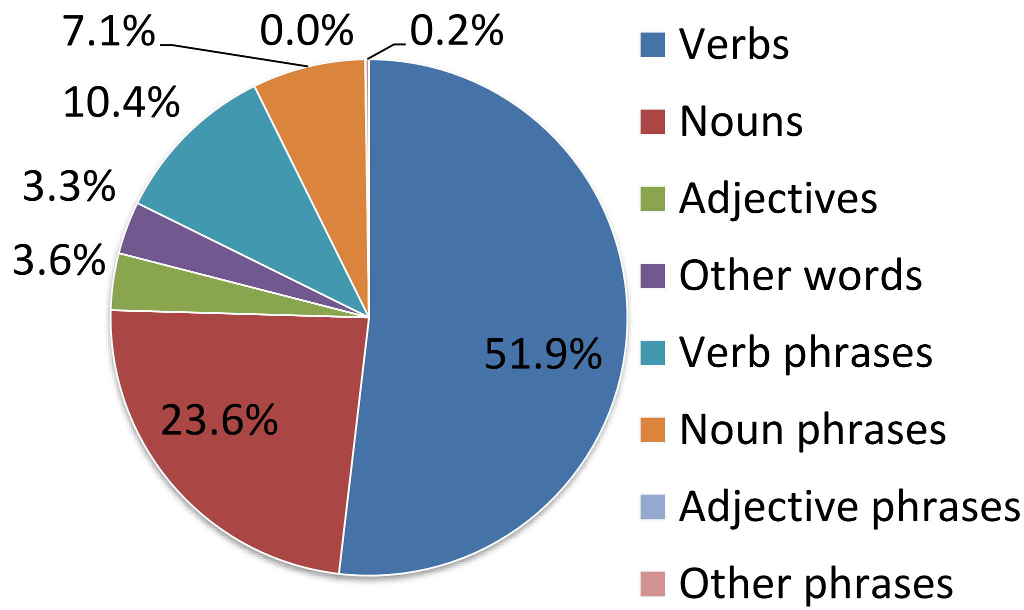
- Training data needs to be as accurate as possible
  - How well does this rule-based event detector perform?



Open-Domain Event Corpus

- Manually annotated 100 articles in Simple Wikipedia
  - 5,397 event nuggets in 10 different domains
  - Inter-annotator agreement (average of pairwise F1 scores):
    - 80.7% (strict match) and 90.3% (partial match)





Results: Training Data Generation

- Dataset: Simple Wikipedia corpus
- Observations:
  - Our WordNet-based heuristics work well
  - The neural gloss classifier gives the best performance

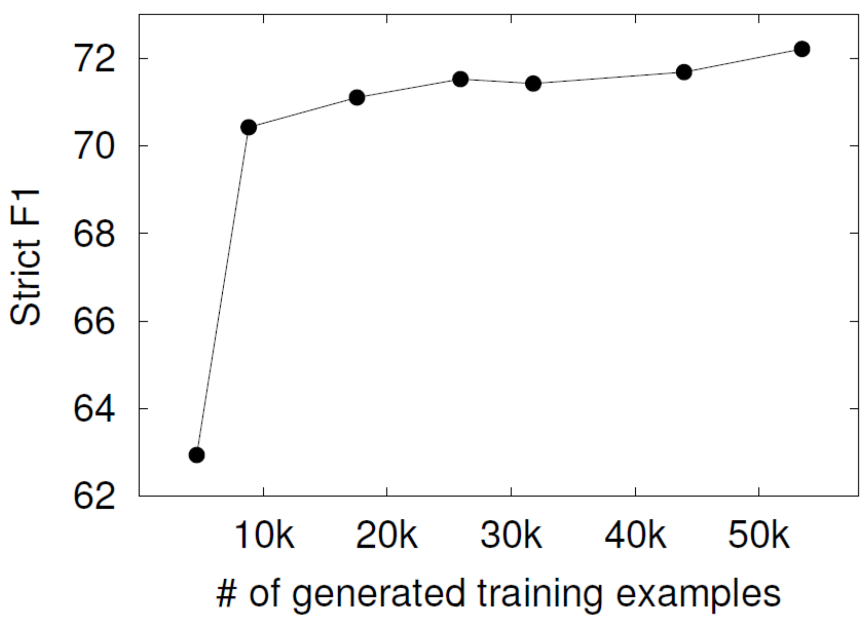
Model	Strict match			Partial match		
	Precision	Recall	F1	Precision	Recall	F1
VERB (Baseline)	79.5	51.7	62.7	95.4	62.0	75.2
RULE	80.1	77.0	78.5	89.0	85.5	87.2
RULE-WP-HL	80.5	77.5	79.0	88.6	85.3	86.9
RULE-WP-GC	**80.8**	**77.7**	**79.2**	**89.1**	**85.7**	**87.3**

Use HeadLookup for Wikipedia proper nouns

Use BLSTM-Attn for Wikipedia proper nouns

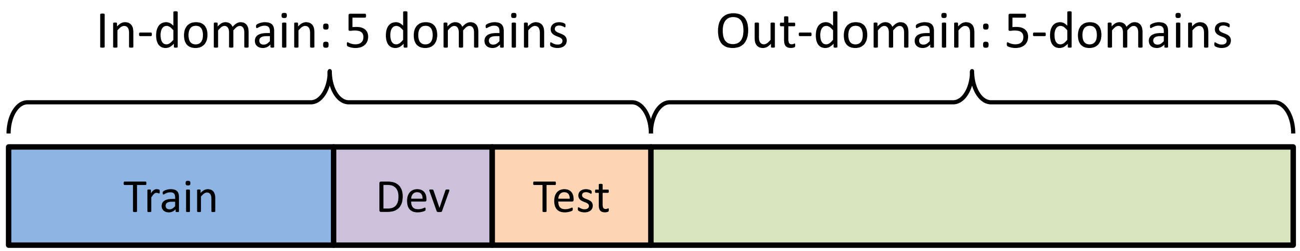
Results: Training Data Generation

- We use SemCor as input to eliminate disambiguation error
  - Generates ~60k event nuggets in total
- Train BLSTM models on the data
  - Use POS embeddings with pre-trained word embeddings
  - Sequence labeling with {B, I, DB, DI, O}
  - Minimize cross-entropy loss
- The model performs better with larger training data



Comparison with Supervised Models

- In-domain and out-domain settings



- The distantly supervised model **performs robustly**
  - Better than supervised models in both settings
  - Averages of F1 scores in 3 runs:

Setting	Model	Strict F1	Partial F1
In-domain	BLSTM	73.8	85.9
	DS-BLSTM	**76.1**	**88.0**
Out-domain	BLSTM	67.9	82.8
	DS-BLSTM	**71.3**	**86.6**

Outline

- Introduction
- Event detection
  - P1: Restricted annotation — Open-domain event detection [Araki+ COLING 2018]
  - P2: Data sparsity — Distant supervision
- Event coreference resolution
  - P3: Event interdependencies — Joint modeling [Araki+ EMNLP 2015]
  - P4: Lack of subevent detection — Subevent structure detection [Araki+ LREC 2014]
  - P5: Limited applications — Question generation [Araki+ COLING 2016]
- Conclusion & future work

Definition of Event Coreference

- Event coreference is a linguistic phenomenon that two event mentions refer to the same event
- 5 types of full identity of events [Hovy+ 2013]:

Type	Example
Lexical identity	"move" and "movement"
Pronouns	"an earthquake" and "it"
Synonyms	"wound" and "injure"
Paraphrases	"Mary **gave** John the book" and "John was **given** the book by Mary"
Wide-reading	"The **attack** took place yesterday. The **bombing** killed four people."

Hovy, E., Mitamura, T., Verdejo, F., Araki, J., and Philpot, A. Events are Not Simple: Identity, Non-Identity, and quasi-identity. NAACL-HLT 2013 Workshop on Events: Definition, Detection, Coreference, and Representation.

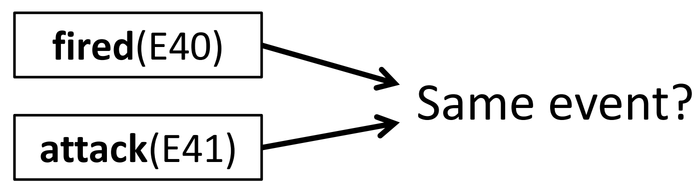
Subevents as Partial Event Coreference

- Definition of subevents: **Partial identity** of events [Hovy+ 2013]

> Mention 1 is a **subevent** of mention 2 if:
> - mention 2 represents a stereotypical sequence of events, or a script, and
> - mention 1 is one of events executed as part of that script

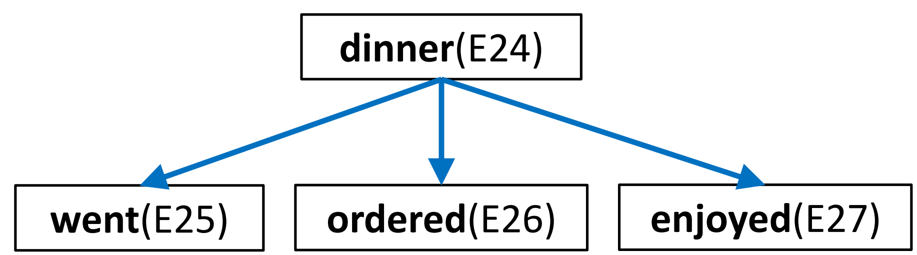
- Subevents can be helpful for full event coreference resolution

> In the town of Ercis, suspected rebels **fired**(E40) rockets at a police station. No one was injured in the **attack**(E41).



- Subevents can provide domain knowledge backbones

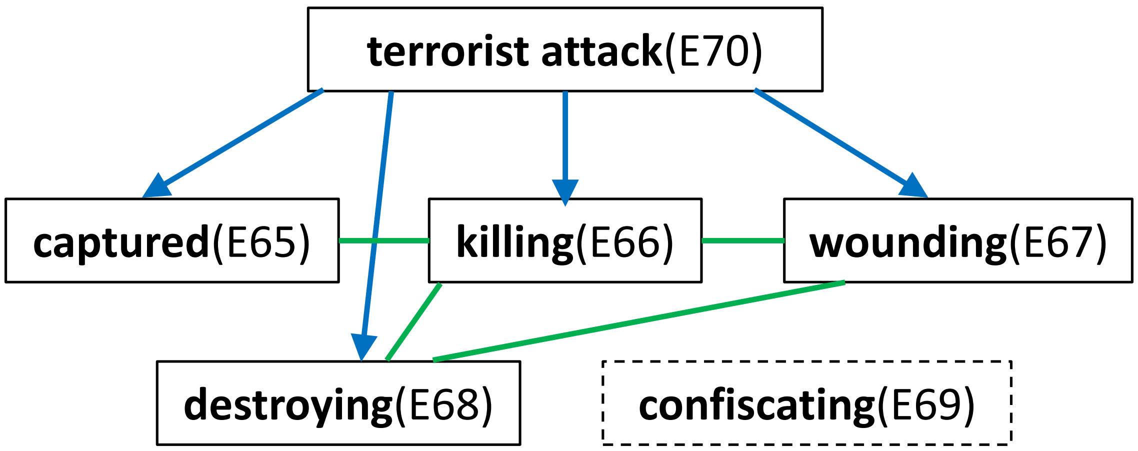
> He had a good **dinner**(E24) last night. He **went**(E25) to a famous restaurant, and **ordered**(E26) a recommended menu. He **enjoyed**(E27) beef steak with a glass of red wine.



Hovy, E., Mitamura, T., Verdejo, F., Araki, J., and Philpot, A. Events are Not Simple: Identity, Non-Identity, and quasi-identity. NAACL-HLT 2013 Workshop on Events: Definition, Detection, Coreference, and Representation.

Subevent Structure Detection

- We proposed a two-stage approach for subevent detection [Araki+ 2014]
  - Stage 1: Find event coreference and subevent parent-child and sibling relations using multinomial logistic regression
  - Stage 2: Find the most likely parents for subevents using voting algorithms



Task: Detection of subevent parent-child relations

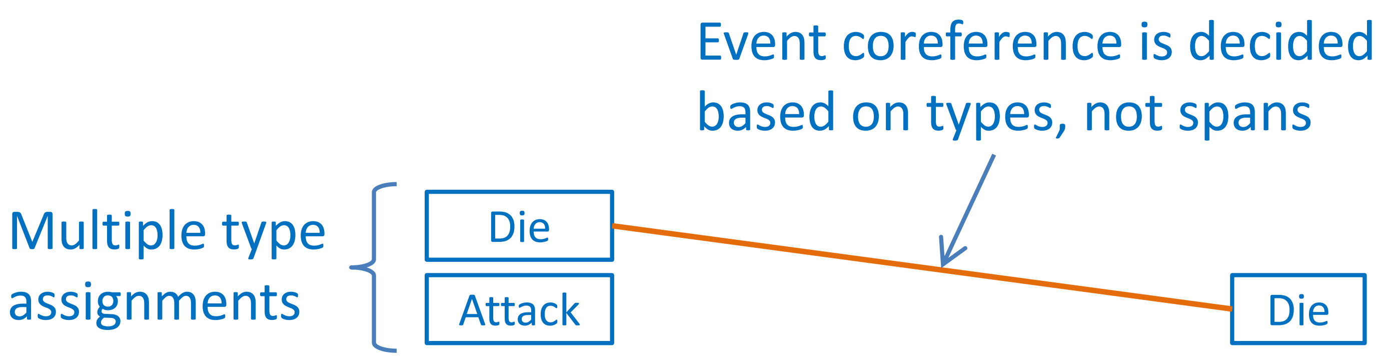
Test data: IC corpus

Model	Avg F1
Stage 1	56.19
Stage 2	**59.45**

Araki, J., Liu, Z., Hovy, E., and Mitamura, T. Detecting subevent structure for event coreference resolution. LREC 2014.

End-to-End Event Coreference Resolution

- TAC KBP Event Nugget and Coreference task [Mitamura+ 2017]
  - Closed-domain (event ontology: 18 event types)
  - Input: Plain text
  - Output:
    - Spans, types, and realis values of event nuggets
    - Event coreference



The city was **attacked** last week. Ten people were **killed**.

Mitamura, T., Liu, Z., and Hovy, E. Events detection, coreference and sequencing: What’s next? Overview of the TAC KBP 2017 Event track. TAC 2017.

Realis

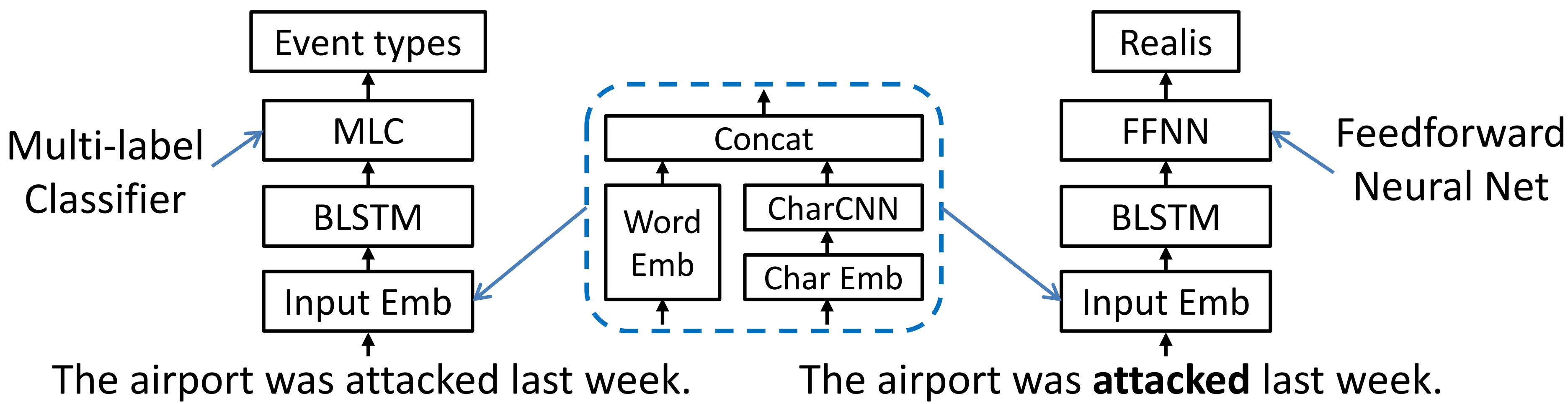
- **Realis** is the epistemic status of events about whether they occurred or not
- Definition of realis used in TAC KBP:
  - ACTUAL := events that actually happened
  - GENERIC := general events (e.g., “Children **grow**.”)
  - OTHER := events that are neither ACTUAL or GENERIC (e.g., negated, hypothetical, or future events)
- Statistics of the TAC KBP datasets
  - Most (>88%) of coreferential events have the same realis value

	Train	Test
# documents	737	167
# non-singleton event clusters	2588	605
A only or G only or O only	**2280 (88.1%)**	**558 (92.2%)**
A only	1331 (51.4%)	322 (53.2%)
G only	380 (14.7%)	81 (13.4%)
O only	569 (22.0%)	155 (25.6%)

Legend
A: ACTUAL
G: GENERIC
O: OTHER

Supervised Neural Models

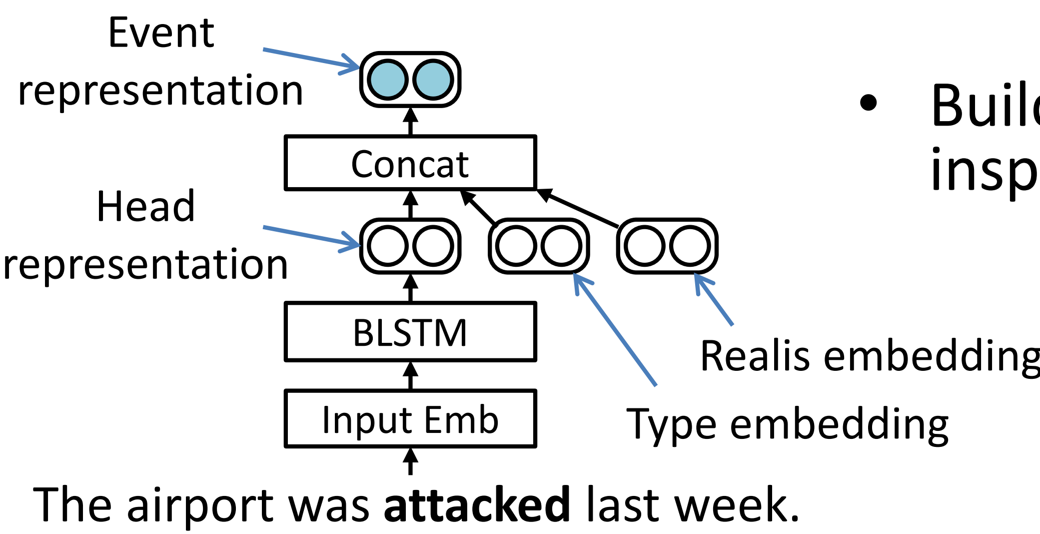
- BLSTM-based models: **(1) → (2)**
  - (1) Event detection
    - Minimize multi-label one-versus-all loss (maximum entropy)
    - Tune a probability threshold to cut off type predictions
  - (2) Realis prediction
    - Minimize cross-entropy loss



(1) Event detection model **(2) Realis model**

Supervised Neural Models

- Build a mention-ranking model inspired by [Lee+ 2017]

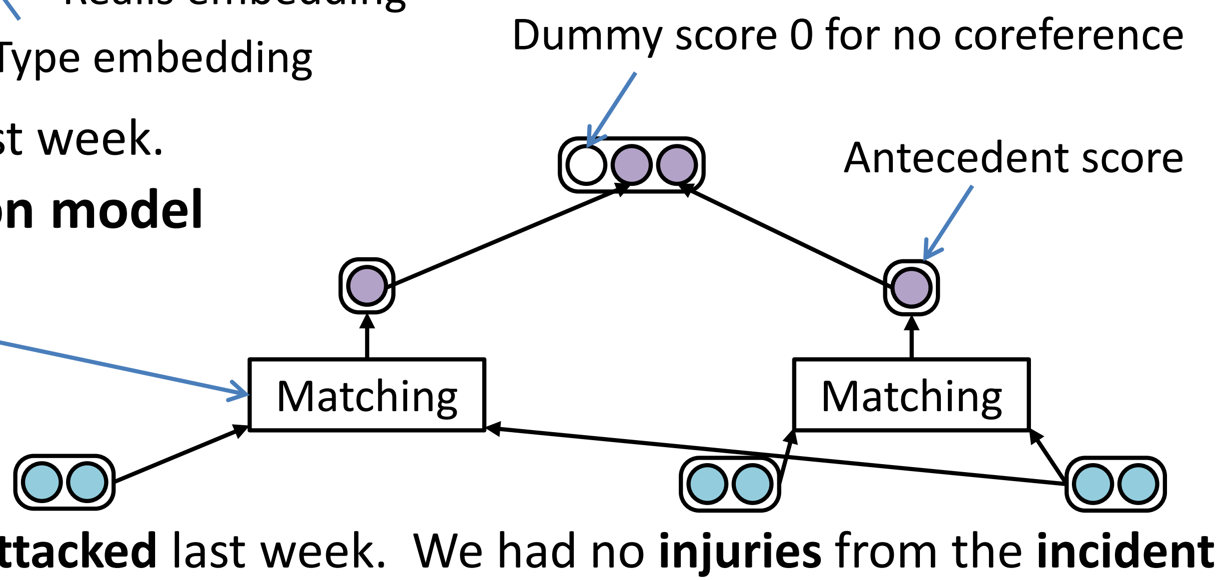


(3a) Event representation model

Heuristic matching technique inspired by [Mou+ 2017]:

$$\mathbf{m} = [\mathbf{e}_i; \mathbf{e}_j; |\mathbf{e}_i - \mathbf{e}_j|; \mathbf{e}_i \circ \mathbf{e}_j]$$

$$s_a(i,j) = \mathbf{w}_a \cdot \mathbf{m} + b_a$$



(3b) Event coreference model

Lee, K., He, L., Lewis, M., and Zettlemoyer, L. End-to-end neural coreference resolution. EMNLP 2017.

Mou, L., Men, R., Li, G., Xu, Y., Zhang L., Yan, R., and Jin, Z. Natural language inference by tree-based convolution and heuristic matching. ACL 2016.

Results: Event Detection

- Our neural models **outperform the state-of-the-art**

Task: TAC KBP 2017
Detection of span+type

Model	P	R	F1
Top 3	54.27	46.59	50.14
Top 2	52.16	48.71	50.37
Top 1	56.83	**55.57**	56.19
BLSTM	69.79	41.31	51.90
BLSTM-CRF	**70.15**	41.06	51.80
BLSTM-MLC	68.03	48.53	**56.65**

Task: TAC KBP 2017
Detection of span+type+realis (overall)

Model	P	R	F1
Top 3	39.69	38.81	39.24
Top 2	42.52	36.50	39.28
Top 1	38.51	**41.03**	39.73
BLSTM	55.09	32.61	40.97
BLSTM-CRF	**55.20**	32.31	40.76
BLSTM-MLC	52.84	37.69	**44.00**

Results: Event Coreference Resolution

- Our neural models **outperform the state-of-the-art**

Task: TAC KBP 2017
Event coreference resolution

Model	MUC	B^3	$CEAF_e$	BLANC	Avg
Top 3	22.90	34.34	33.63	17.94	27.20
Top 2	33.79	39.88	35.73	26.06	33.87
Top 1	30.63	43.84	39.86	26.97	35.33
LTR (Baseline)	29.94	43.92	41.60	25.64	35.28
NEC-TR	30.19	44.38	42.88	26.17	35.91
NEC	**33.95**	**44.88**	**43.02**	**28.06**	**37.48**

Event Interdependencies

- Individual event mentions interact with each other via event coreference

Trebian was **born**(E11) on November 4th. We were praying that his
Be-Born
father would get here on time, but unfortunately he missed **it**(E12).
?

In a village near the West Bank town of Qalqiliya, an 11-year-old Palestinian boy was **killed**(E13) during an exchange of **gunfire**(E14).
Die Attack
Also Monday, Israeli soldiers **fired**(E15) on four diplomatic vehicles in
Attack
the northern Gaza town of Beit Hanoun, diplomats said. There were no **injuries**(E16) from the **incident**(E17).
Injure **?**

Event Interdependencies

- Individual event mentions interact with each other via event coreference

Trebian was **born**(E11) on November 4th. We were praying that his
Be-Born
father would get here on time, but unfortunately he missed **it**(E12).
?

In a village near the West Bank town of Qalqiliya, an 11-year-old Palestinian boy was **killed**(E13) during an exchange of **gunfire**(E14).
Die Attack
Also Monday, Israeli soldiers **fired**(E15) on four diplomatic vehicles in
Attack
the northern Gaza town of Beit Hanoun, diplomats said. There were no **injuries**(E16) from the **incident**(E17).
Injure ?

Event Interdependencies

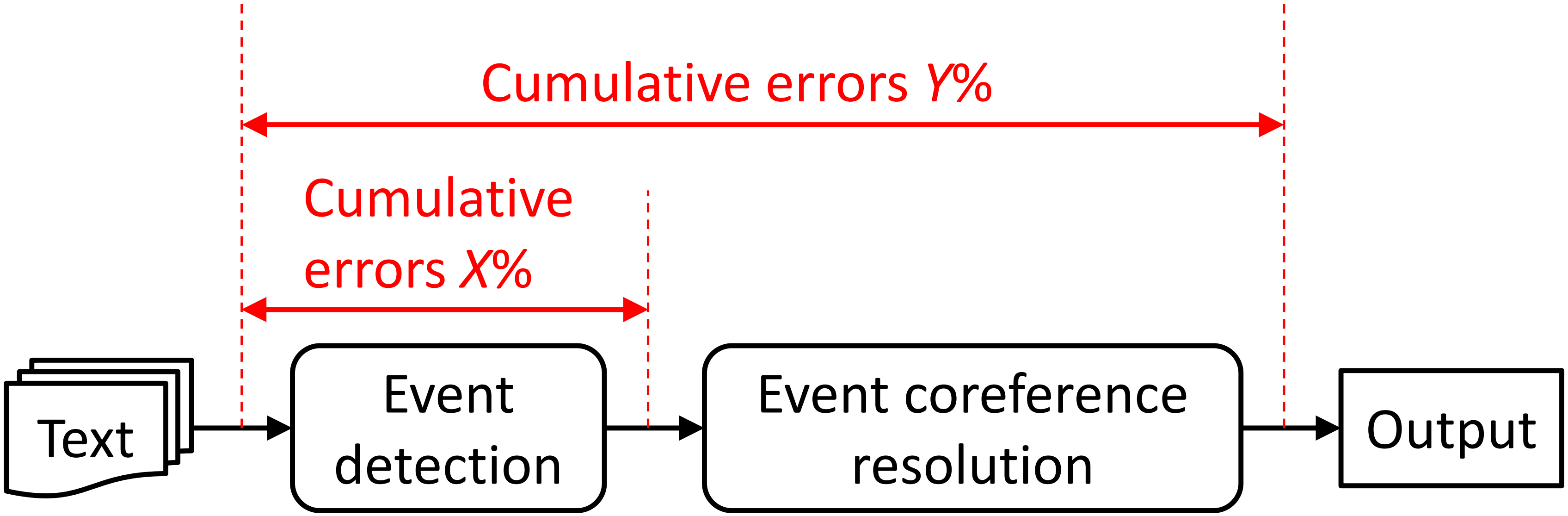
- Individual event mentions interact with each other via event coreference

Trebian was **born**(E11) on November 4th. We were praying that his
Be-Born
father would get here on time, but unfortunately he missed **it**(E12).
Be-Born

In a village near the West Bank town of Qalqiliya, an 11-year-old Palestinian boy was **killed**(E13) during an exchange of **gunfire**(E14).
Die
Attack
Also Monday, Israeli soldiers **fired**(E15) on four diplomatic vehicles in
Attack
the northern Gaza town of Beit Hanoun, diplomats said. There were no **injuries**(E16) from the **incident**(E17).
Injure
Attack

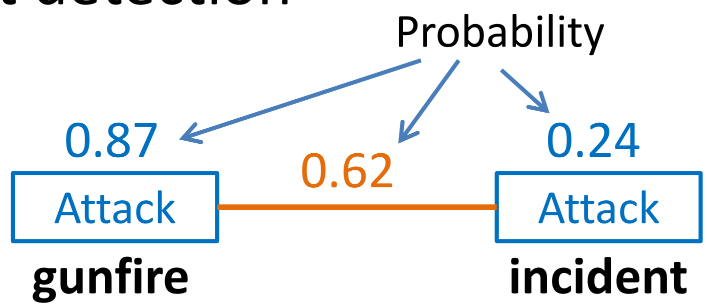
Problems with Pipeline Models

- Prior work has addressed event detection and event coreference resolution separately
- Pipeline models propagate errors ➡ normally $Y > X$

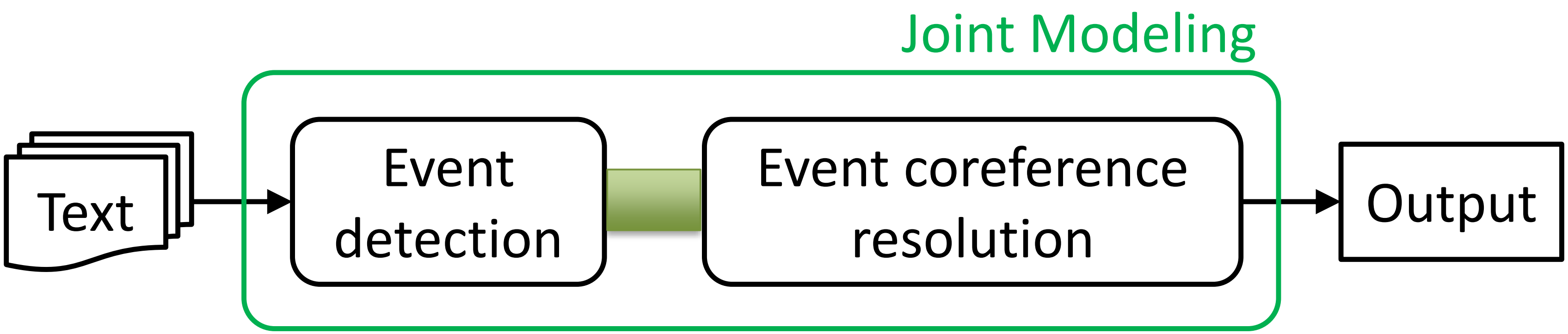


Joint Modeling

- Explore more possibilities while not committing to single output of event detection

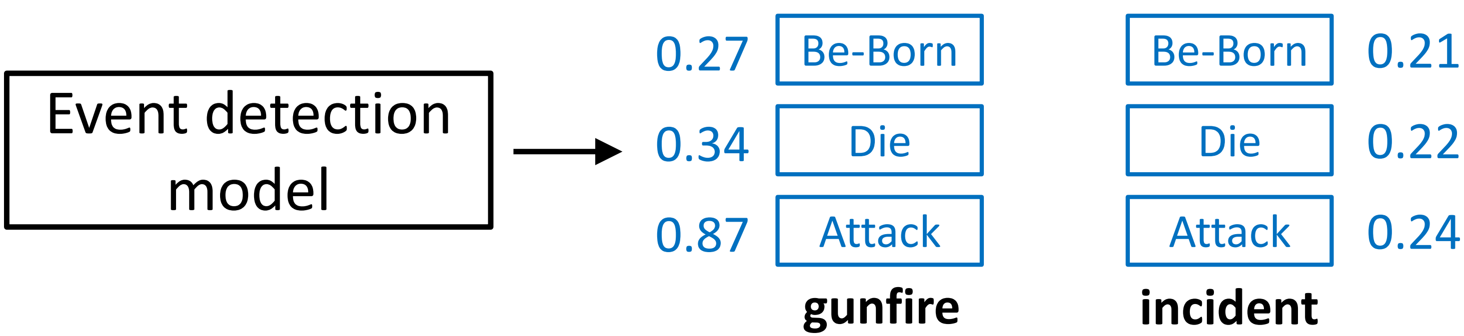


- Assumption:
  - Improve recall in both event detection and event coreference resolution



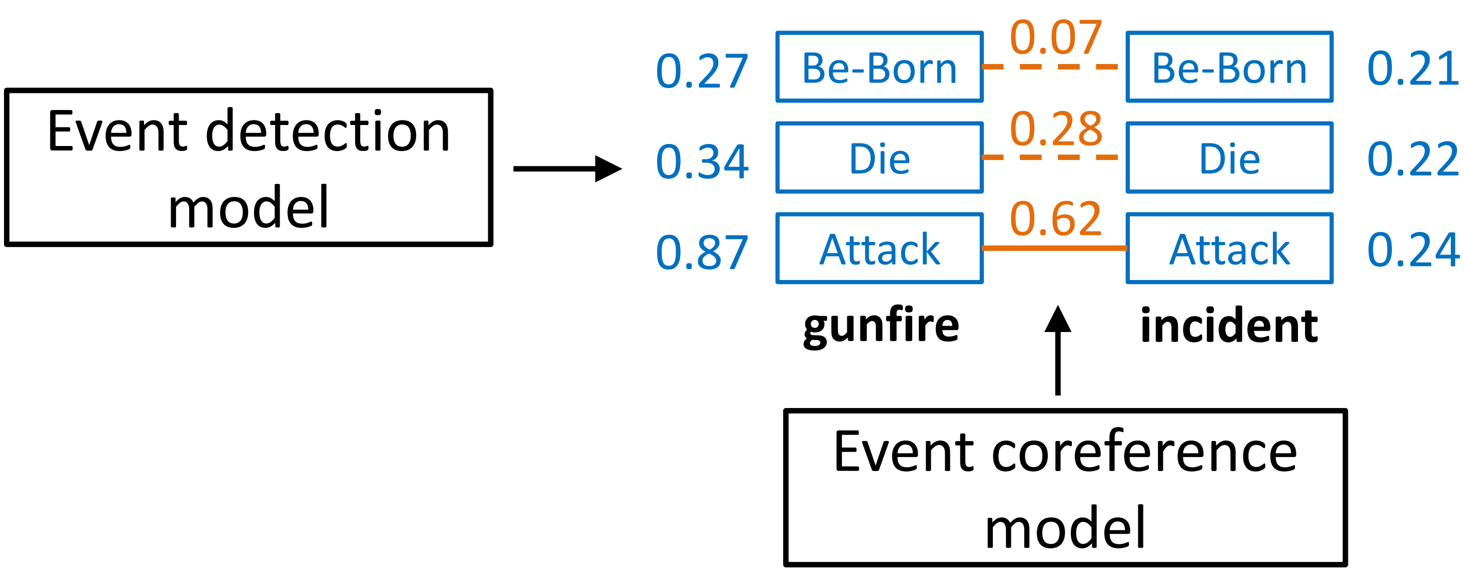
Joint Modeling (1): Joint Decoding

- Use individually pre-trained event detection and event coreference models
- **Leave low-scoring type predictions** for further consideration of event coreference
  - If event coreference is found, we keep the type predictions
  - If not (ending up with singletons), we prune them



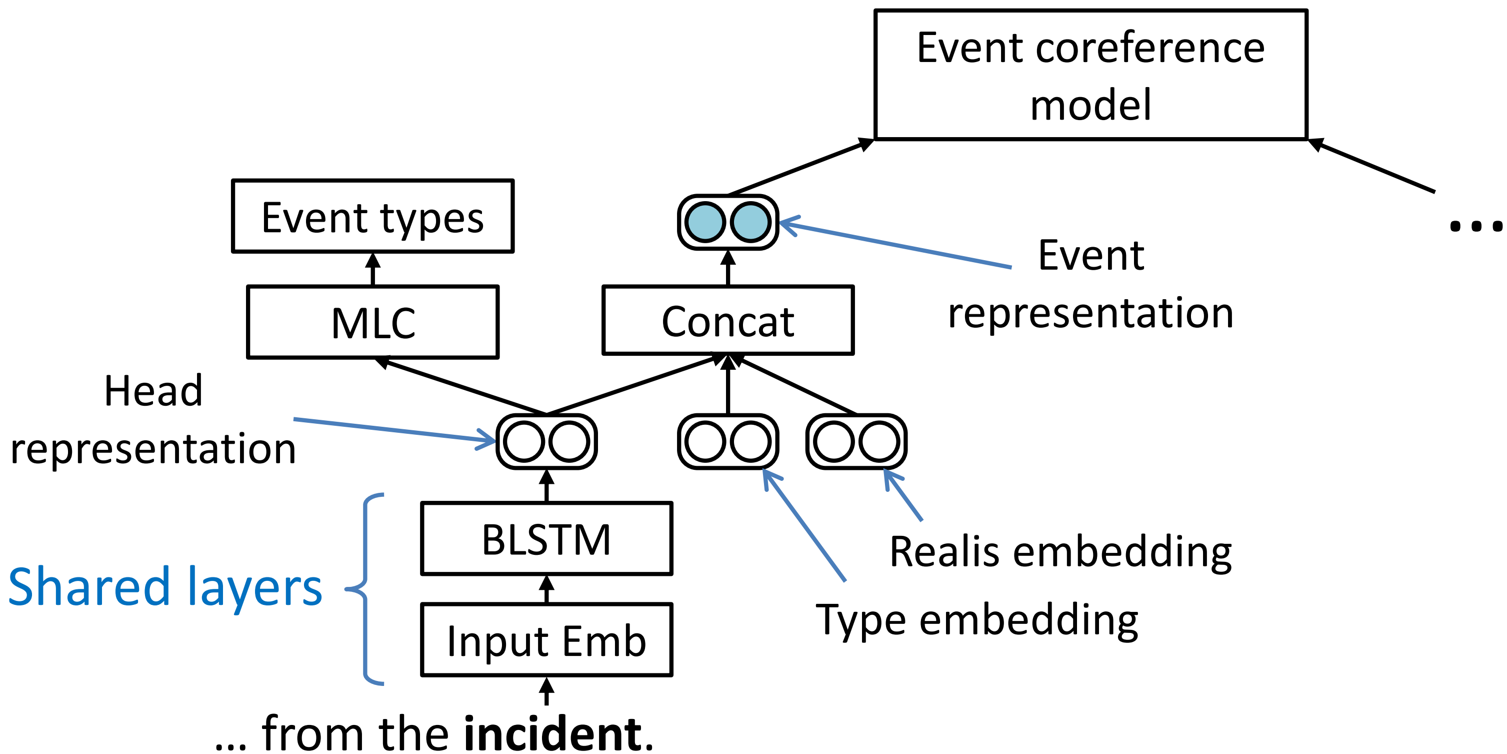
Joint Modeling (1): Joint Decoding

- Use individually pre-trained event detection and event coreference models
- **Leave low-scoring type predictions** for further consideration of event coreference
  - If event coreference is found, we keep the type predictions
  - If not (ending up with singletons), we prune them



Joint Modeling (2): Joint Training

- Jointly train event detection and event coreference models
  - Share input embedding and BLSTM layers
  - Assumption: **Multi-task learning effect**
    - Training signals from related tasks provide superior regularization
- Use joint decoding in the inference phase



Results: Event Detection

- Our joint models **further makes an improvement**

Task: TAC KBP 2017
Detection of span+type

Model	P	R	F1
Top 3	54.27	46.59	50.14
Top 2	52.16	48.71	50.37
Top 1	56.83	**55.57**	56.19
BLSTM	69.79	41.31	51.90
BLSTM-CRF	**70.15**	41.06	51.80
BLSTM-MLC	68.03	48.53	56.65
JD	67.61	48.97	56.90
JT+JD	65.44	50.53	**57.03**

Task: TAC KBP 2017
Detection of span+type+realis (overall)

Model	P	R	F1
Top 3	39.69	38.81	39.24
Top 2	42.52	36.50	39.28
Top 1	38.51	**41.03**	39.73
BLSTM	55.09	32.61	40.97
BLSTM-CRF	**55.20**	32.31	40.76
BLSTM-MLC	52.84	37.69	44.00
JD	52.56	38.07	44.16
JT+JD	50.72	39.16	**44.20**

Results: Event Coreference Resolution

- Our joint models **further makes an improvement**

Task: TAC KBP 2017
Event coreference resolution

Model	MUC	B^3	$CEAF_e$	BLANC	Avg
Top 3	22.90	34.34	33.63	17.94	27.20
Top 2	33.79	39.88	35.73	26.06	33.87
Top 1	30.63	43.84	39.86	26.97	35.33
LTR (Baseline)	29.94	43.92	41.60	25.64	35.28
NEC-TR	30.19	44.38	42.88	26.17	35.91
NEC	33.95	44.88	43.02	28.06	37.48
JD	34.04	**45.02**	**43.15**	28.15	37.59
JT+JD	**35.81**	44.87	41.98	**29.47**	**38.03**

Applications of Event Coreference

- Most applications let **systems** use event coreference for a downstream task
  - e.g., textual entailment

Text: Amazon was found by Jeff Bezos.
Hypothesis: Bezos established a company.

found
established

→  → "T entails H"

- Problem: Limited applications of event coreference
  - Hypothesis: Event coreference can be useful for natural language understanding by **humans**

found
established

→  → 

Event Coreference for Question Generation

- Goal:
  - Generate more sophisticated questions from multiple sentences for English-as-a-second-language (ESL) students
    - Enhance language learning tools, e.g., SmartReader [Azab+ 2013]
- Background: Educational theory
  - **Higher-level questions** have more educational benefits for reading comprehension [Anderson+ 1975; Andre, 1979]
- Problems
  - Prior work generates questions from single sentences
    - Generated questions tend to be too specific and low-level
    - They just assess the ability to compare sentences

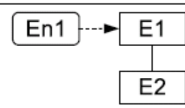
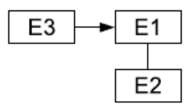
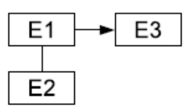
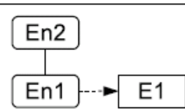
Azab, M., Salama, A., Oflazer, K., Shima, H., Araki, J., and Mitamura, T. An English reading tool as an NLP showcase. In Proceedings of IJCNLP 2013: System Demonstrations.

Anderson, R. and Biddle, B. On asking people questions about what they are reading. Psychology of Learning and Motivation, 9:90–132. 1975.

Andre, T. Does answering higher level questions while reading facilitate productive learning? Review of Educational Research, 49(2):280–318. 1979.

Our Approach: Template-based QG

- **Inference step**: resolution of event or entity coreference, or detection of a paraphrase
- Generate questions based on templates:

Semantic relation	Question patterns	Answer	Question templates
Event coreference	P1. 	En1	T1. What [*verbal trigger* + *subsequent arguments*]?
	P2. 	E3	T2. What causes [*nominal trigger* + *subsequent arguments*]? T3. What makes it happen to [*verbal trigger* + *subsequent arguments*]? T4. What makes it happen that [*event clause*]?
	P3. 	E3	T5. What is a result of [*nominal trigger* + *subsequent arguments*]? T6. What happens when [*event clause*]?
Entity coreference	P4. 	En2	T1. What [*verbal trigger* + *subsequent arguments*]?
Paraphrase	P5. 	En1	

Evaluation for Generated Questions

- Questions are evaluated by two human annotators
- Metrics:
  - **Grammatical correctness**: Whether a question is syntactically well-formed
    - 1 (best): no grammatical error, 2: 1 or 2 errors, 3 (worst): 3 or more errors
  - **Answer existence**: Whether the answer to a question can be inferred from the passage associated with the question
    - 1 (yes): the answer can be inferred from the passage, 2 (no): otherwise
  - **Inference steps**: How many semantic relations humans need to understand in order to answer a question

Results of Question Generation

- Baseline: [Heilman+ 2010]
- Data: 200 questions generated from ProcessBank

Lower is better　　　　Higher is better

System	Grammatical Correctness			Answer Existence			Inference Steps		
	Ann1	Ann2	Total	Ann1	Ann2	Total	Ann1	Ann2	Total
Ours	1.52	1.48	1.50	1.17	1.26	1.21	0.80	0.71	**0.76**
Baseline	1.42	1.25	**1.34**	1.20	1.14	**1.17**	0.13	0.19	0.16

- Observation:
  - Our system is able to generate **higher-level questions** that require a larger number of inference steps, while retaining grammatical correctness and answer existence

Heilman, M. and Smith, N. Good Question! Statistical Ranking for Question Generation. NAACL-HLT 2010.

Outline

- Introduction
- Event detection

  P1: Restricted annotation — Open-domain event detection [Araki+ COLING 2018]

  P2: Data sparsity — Distant supervision

- Event coreference resolution

  P3: Event interdependencies — Joint modeling [Araki+ EMNLP 2015]

  P4: Lack of subevent detection — Subevent structure detection [Araki+ LREC 2014]

  P5: Limited applications — Question generation [Araki+ COLING 2016]

- Conclusion & future work

Conclusion (1/2)

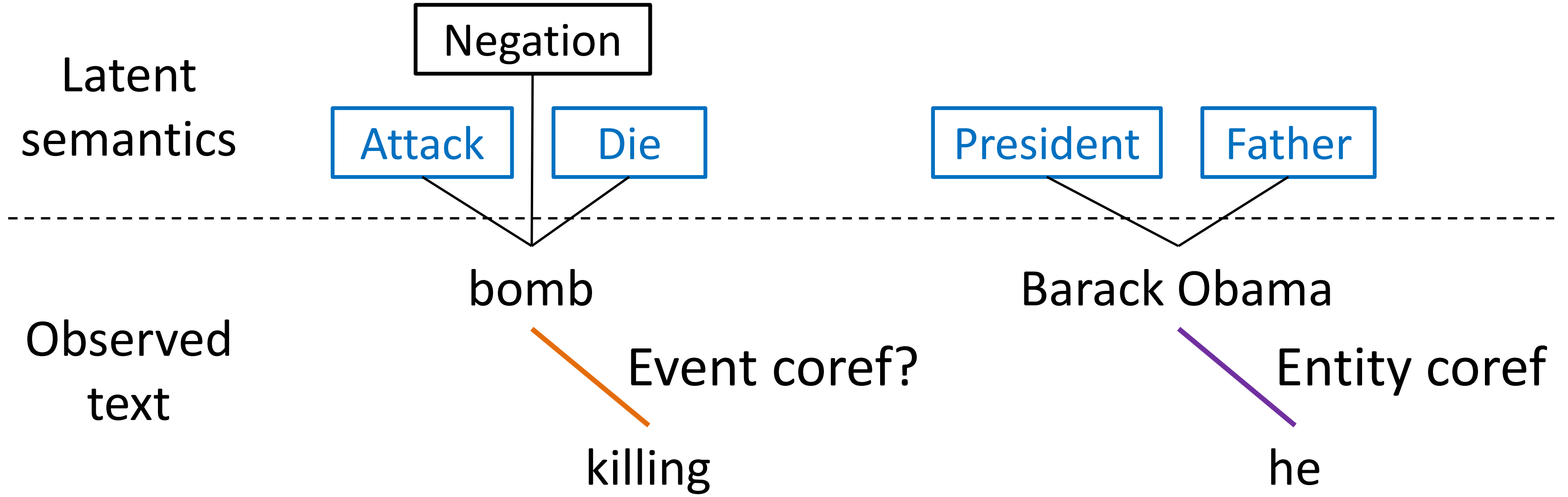
- Event detection
  – We introduced **a new paradigm of open-domain event detection**
    - Despite our relatively wide and flexible annotation of events, we achieved **high inter-annotator agreement**: 80.7% F1 (strict match) and 90.3% F1 (partial match)
  – We showed that it is feasible for our distant supervision approach to **generate high-quality training data** while obviating the need for human annotation
  – **State-of-the-art performance**
    - Our neural event detection and joint models outperform the best system in TAC KBP 2017

Conclusion (2/2)

- Event coreference resolution
  - Our **joint modeling framework** can capture event interdependencies adequately, improving recall
  - **State-of-the-art performance**
    - Our neural event coreference and joint models outperform the best system in TAC KBP 2017
  - We proposed the **first work for subevent detection**
    - Our two-stage approach can improve subevent structures
  - Using event coreference, our question generation system can **generate more sophisticated questions** that require deeper semantic understanding

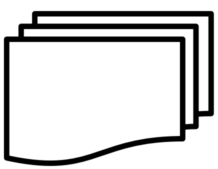
Connections to Other NLP Tasks

- **Event detection and entity detection**
  - Events tend to have more single-word expressions
  - Events can have discontinuous expressions
- **Event coreference and entity coreference**
  - Events are a structured representation involving agents, patients, times, and locations
  - Events tend to have more ambiguous multifaceted semantics
  - Events have realis (can be negated, hypothesized, etc.)



Future Work: Cross-X

- **Cross-document**
  - Event coreference resolution
- **Cross-language**
  - Events are language-independent phenomena
- **Cross-modality**
  - Events are also found in informal texts, dialogue, audios, and videos



Documents

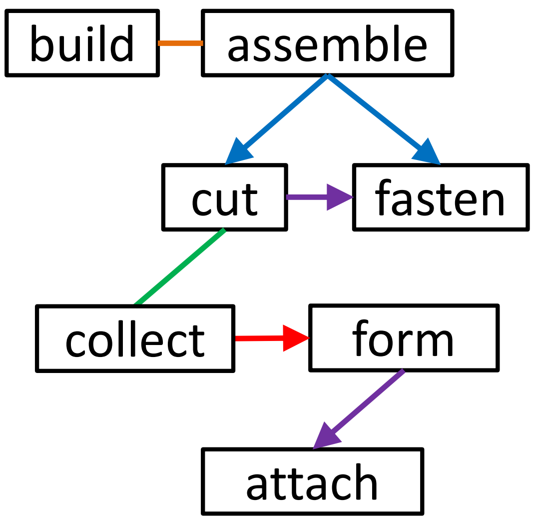
Informal texts

Dialogue

Images & videos

Future Work: Ontology & Applications

- Event-centered knowledge bases (KBs) facilitate more advanced reasoning, enabling more sophisticated applications
  - Challenge: Construction of event type taxonomies

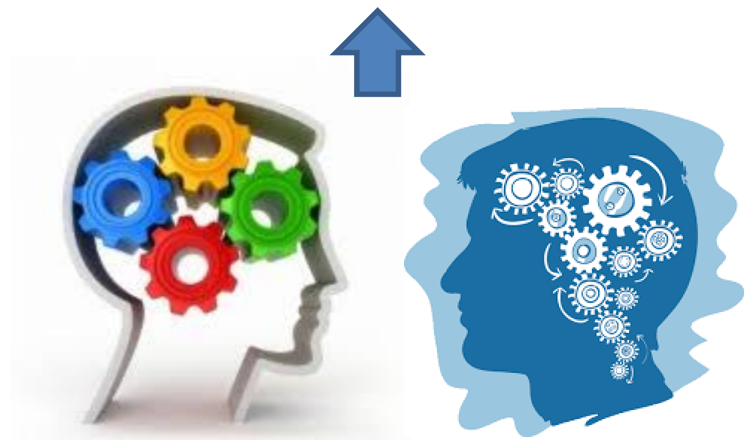


Legend:

- Event coreference
- Subevent
- Causality
- Subsequence
- Simultaneity

Summarization

Question answering



Event KBs Entity KBs

Common-sense and domain-specific knowledge

References

- Araki, J. and Mitamura, T. Open-Domain Event Detection using Distant Supervision. COLING 2018. To appear.
- Araki, J., Rajagopal, D., Sankaranarayanan, S., Holm, S., Yamakawa, Y., and Mitamura, T. Generating Questions and Multiple-Choice Answers using Semantic Analysis of Texts. COLING 2016.
- Araki, J. and Mitamura, T. Joint Event Trigger Identification and Event Coreference Resolution with Structured Perceptron. EMNLP 2015.
- Araki, J., Liu, Z., Hovy, E., and Mitamura, T. Detecting Subevent Structure for Event Coreference Resolution. LREC 2014.
- Hovy, E., Mitamura, T., Verdejo, F., Araki, J., and Philpot, A. Events are Not Simple: Identity, Non-Identity, and Quasi-Identity. NAACL-HLT 2013 Workshop on Events: Definition, Detection, Coreference, and Representation.
- Azab, M., Salama, A., Oflazer, K., Shima, H., Araki, J., and Mitamura, T. An English Reading Tool as an NLP Showcase. In Proceedings of IJCNLP 2013: System Demonstrations.