Human Language Technology Research Institute

# Chinese Event Coreference Resolution: Understanding the State of the Art

**Chen Chen** and **Vincent Ng**

Human Language Technology Research Institute

The University of Texas at Dallas

# Event Coreference

- Goal: Determine which event mentions in a text refer to the same real-world event

# Event Coreference

- Goal: Determine which event mentions in a text refer to the same real-world event

Since there is few work on event coreference, our understanding of this task is fairly limited!

# Goal

- Understand how a state-of-the-art end-to-end event coreference resolver can be improved

# Goal

- Understand how a state-of-the-art end-to-end event coreference resolver can be improved
  - Question 1: To what extent is the noise inherent in the output of each of its upstream components limiting the performance?

# Goal

- Understand how a state-of-the-art end-to-end event coreference resolver can be improved
  - Question 1: To what extent is the noise inherent in the output of each of its upstream components limiting the performance?
  - Question 2: What are the major types of errors that are attributable to the resolution algorithm?

# Why Chinese Event Coreference Resolution

- Lack of publicly available results on Chinese event coreference resolution
  - Most of recent work on event coreference are for English
    - Humphreys at el.(1997), Chen et al.(2009), Bejan and Harabagiu(2010), Chen et al.(2011), Lee et al.(2012)

# Plan for the Talk

- ACE Event Coreference
- Six Upstream Components in the Pipeline
- Results and Analysis – Answer to Question 1
- Conclusion

# Plan for the Talk

- ACE Event Coreference
- Six Upstream Components in the Pipeline
- Results and Analysis – Answer to Question 1
- Conclusion

# ACE Event Coreference

(Zhang Jiarong) was cycling on (the road) (yesterday evening) and was [injured] when (two men) [stabbed] (him) with (a knife). (The thugs)' [criminal] motivation may have something to do with (Zhang Jiarong)'s testimony in a criminal case.

# ACE Event Coreference

(Zhang Jiarong) was cycling on (the road) (yesterday evening) and was [injured] when (two men) [stabbed] (him) with (a knife). (The thugs)' [criminal] motivation may have something to do with (Zhang Jiarong)'s testimony in a criminal case.

- This example contains three event mentions, each of which has a type and subtype and is associated with arguments

# ACE Event Coreference

(Zhang Jiarong) was cycling on (the road) (yesterday evening) and was [injured] when (two men) [stabbed] (him) with (a knife). (The thugs)' [criminal] motivation may have something to do with (Zhang Jiarong)'s testimony in a criminal case.

- This example contains three event mentions, each of which has a type and subtype and is associated with arguments
  - [injured]: Type LIFE; SubType INJURE; arguments: (Zhang Jiarong) , (the road) and (yesterday evening)

# ACE Event Coreference

(Zhang Jiarong) was cycling on (the road) (yesterday evening) and was [injured] when (two men) [stabbed] (him) with (a knife). (The thugs)' [criminal] motivation may have something to do with (Zhang Jiarong)'s testimony in a criminal case.

- This example contains three event mentions, each of which has a type and subtype and is associated with arguments
  - [injured]: Type LIFE; SubType INJURE; arguments: (Zhang Jiarong) , (the road) and (yesterday evening)
  - [stabbed]: Type CONFLICT; SubType: ATTACK; arguments: (two men), (him) and (a knife)

# ACE Event Coreference

(Zhang Jiarong) was cycling on (the road) (yesterday evening) and was [injured] when (two men) [stabbed] (him) with (a knife). (The thugs)' [criminal] motivation may have something to do with (Zhang Jiarong)'s testimony in a criminal case.

- This example contains three event mentions, each of which has a type and subtype and is associated with arguments
  - [injured]: Type LIFE; SubType INJURE; arguments: (Zhang Jiarong) , (the road) and (yesterday evening)
  - [stabbed]: Type CONFLICT; SubType: ATTACK; arguments: (two men), (him) and (a knife)
  - [criminal]: Type CONFLICT; SubType: ATTACK; arguments: (The thugs) and (Zhang Jiarong)

# ACE Event Coreference

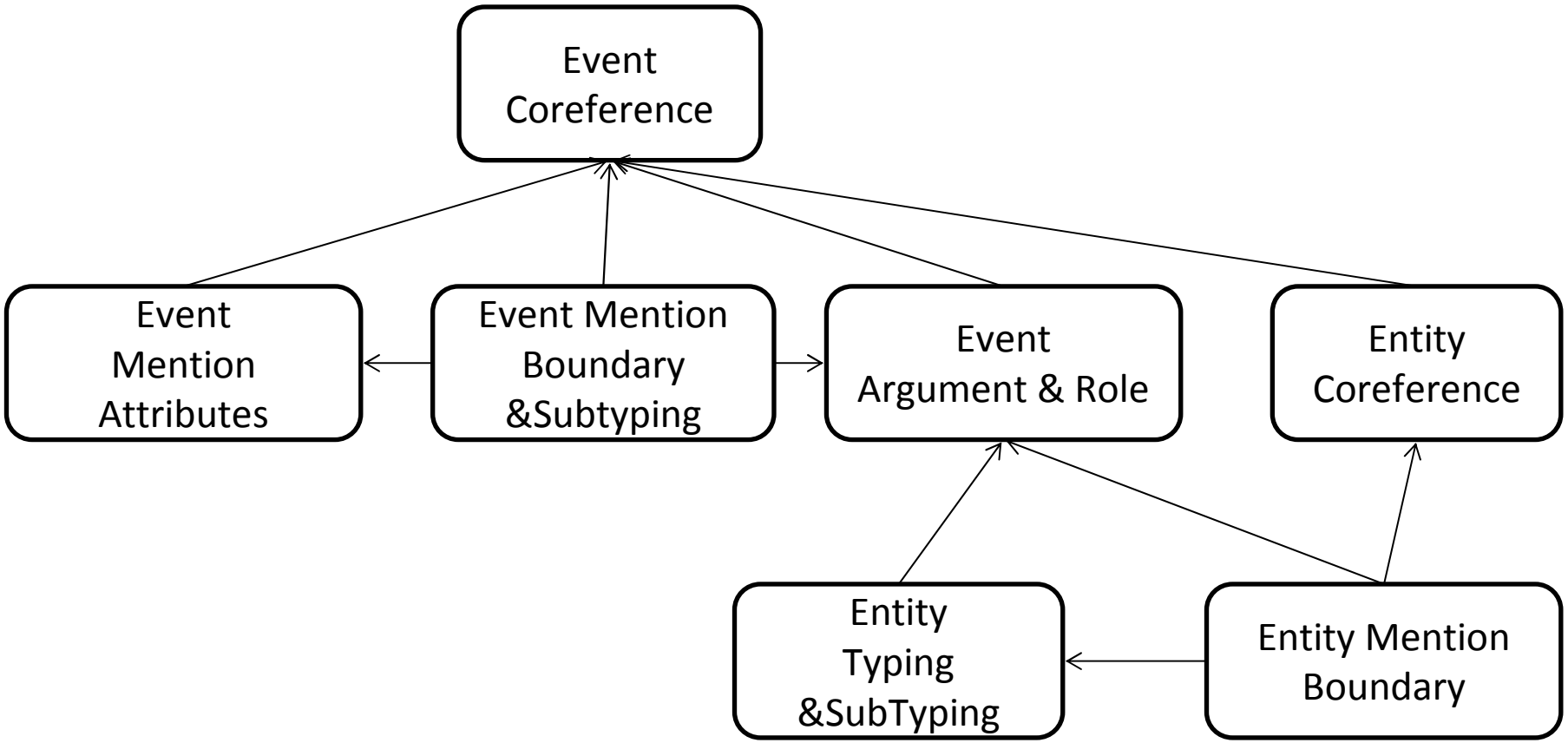
(Zhang Jiarong) was cycling on (the road) (yesterday evening) and was [injured] when (two men) [stabbed] (him) with (a knife). (The thugs)' [criminal] motivation may have something to do with (Zhang Jiarong)'s testimony in a criminal case.

- This example contains three event mentions, each of which has a type and subtype and is associated with arguments
- [stabbed] and [criminal] are coreferent because they refer to the same real-world event

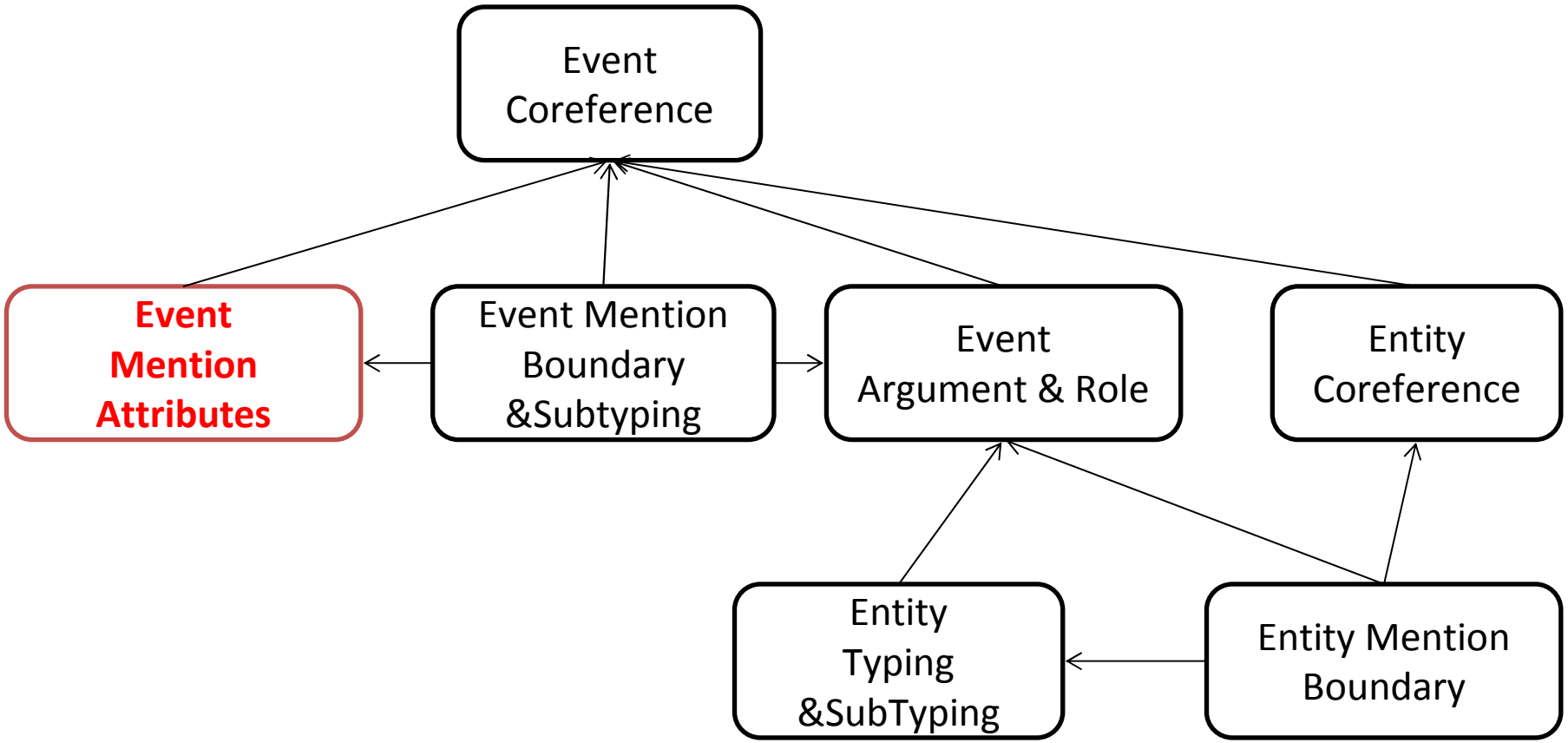
# Plan for the Talk

- ACE Event Coreference
- Six Upstream Components in the Pipeline
- Results and Analysis – Answer to Question 1
- Conclusion

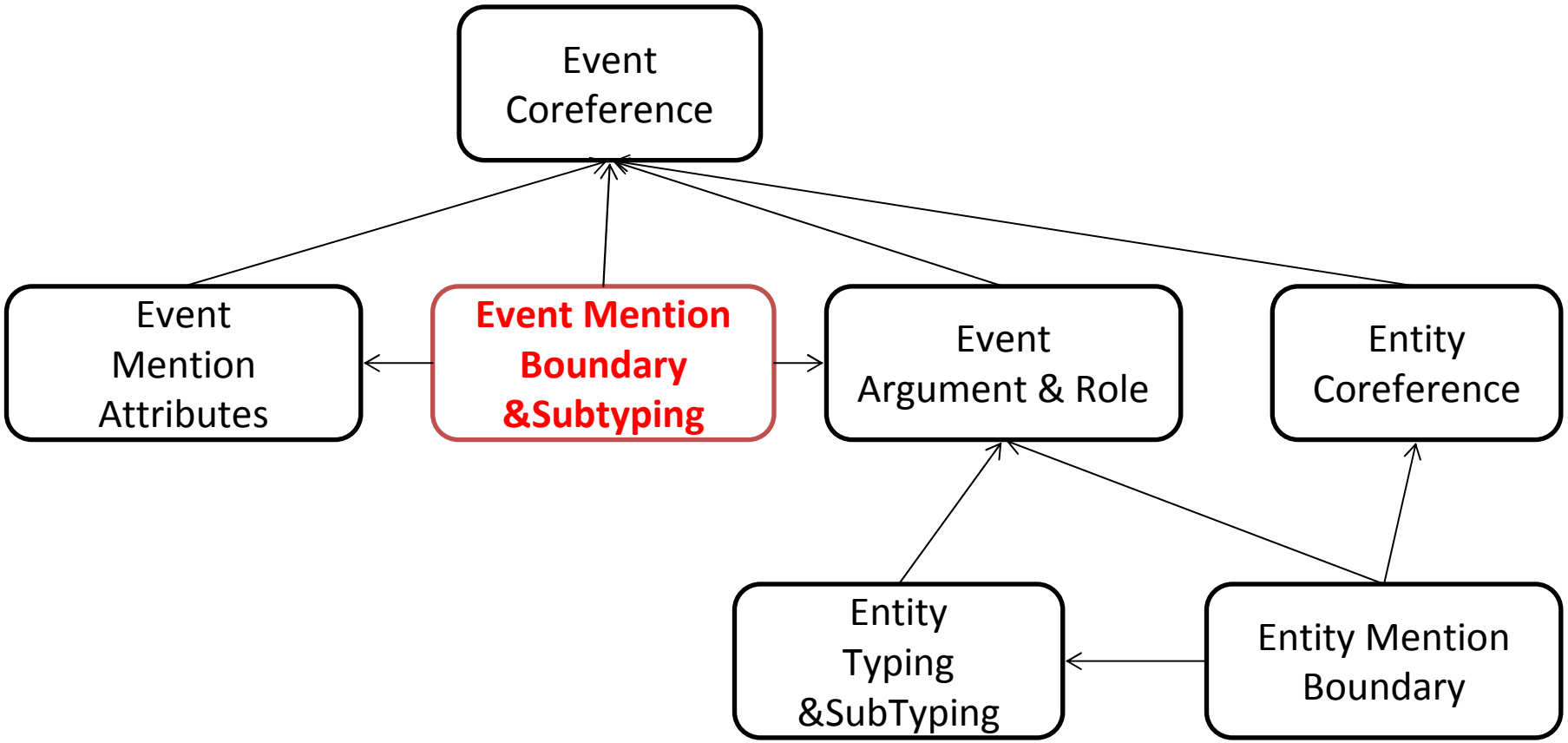
# Six Upstream Components for Event Coreference Resolution



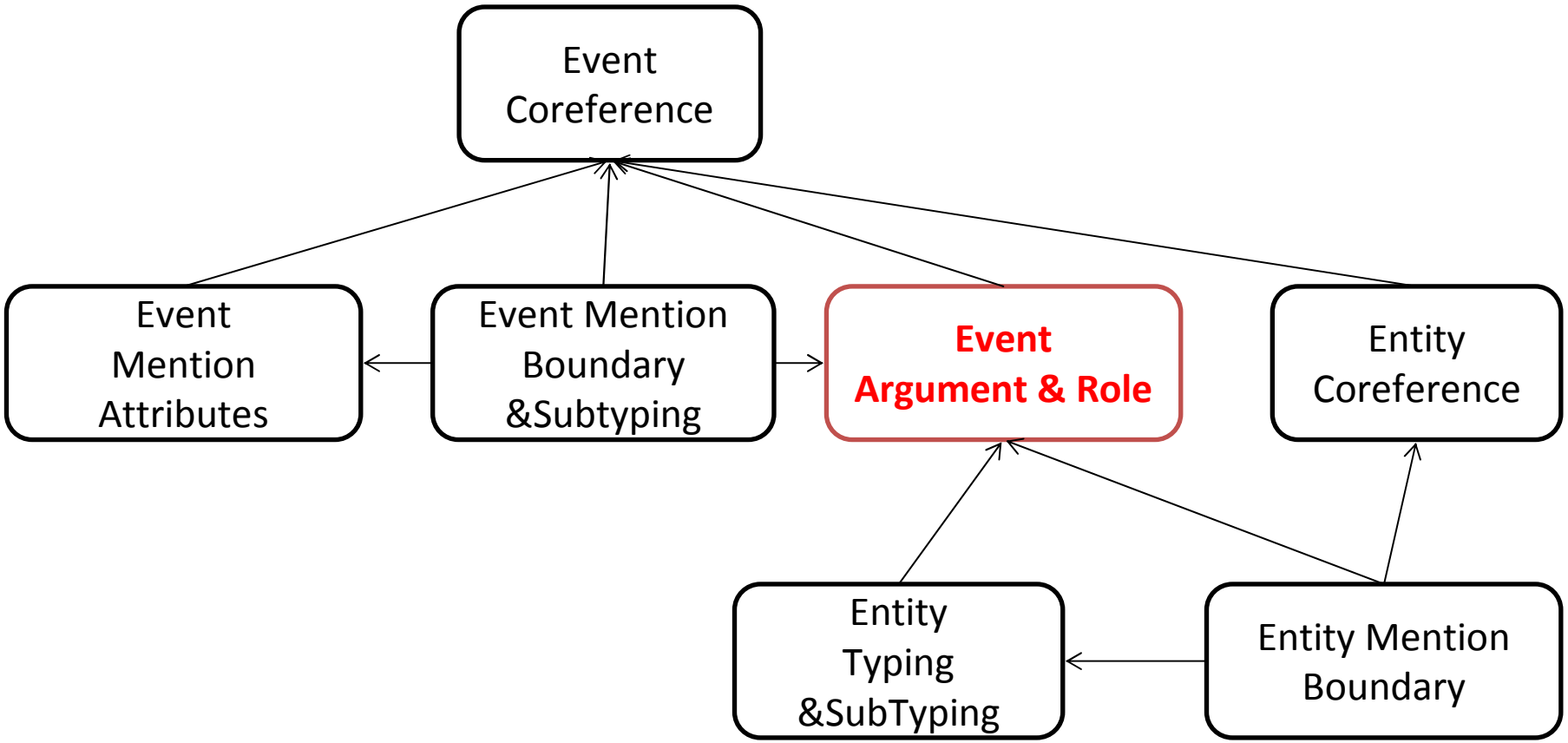
# Six Upstream Components for Event Coreference Resolution



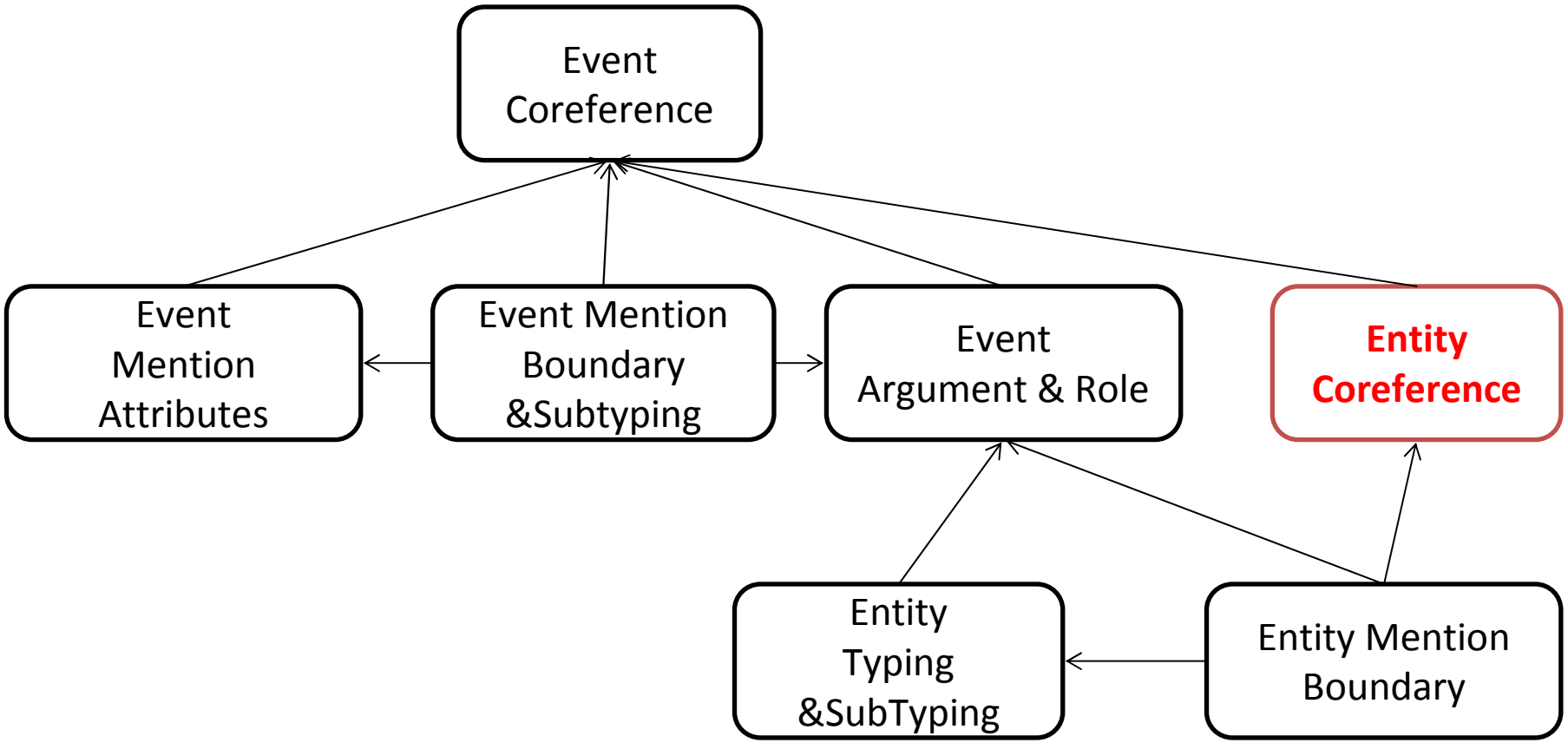
# Six Upstream Components for Event Coreference Resolution



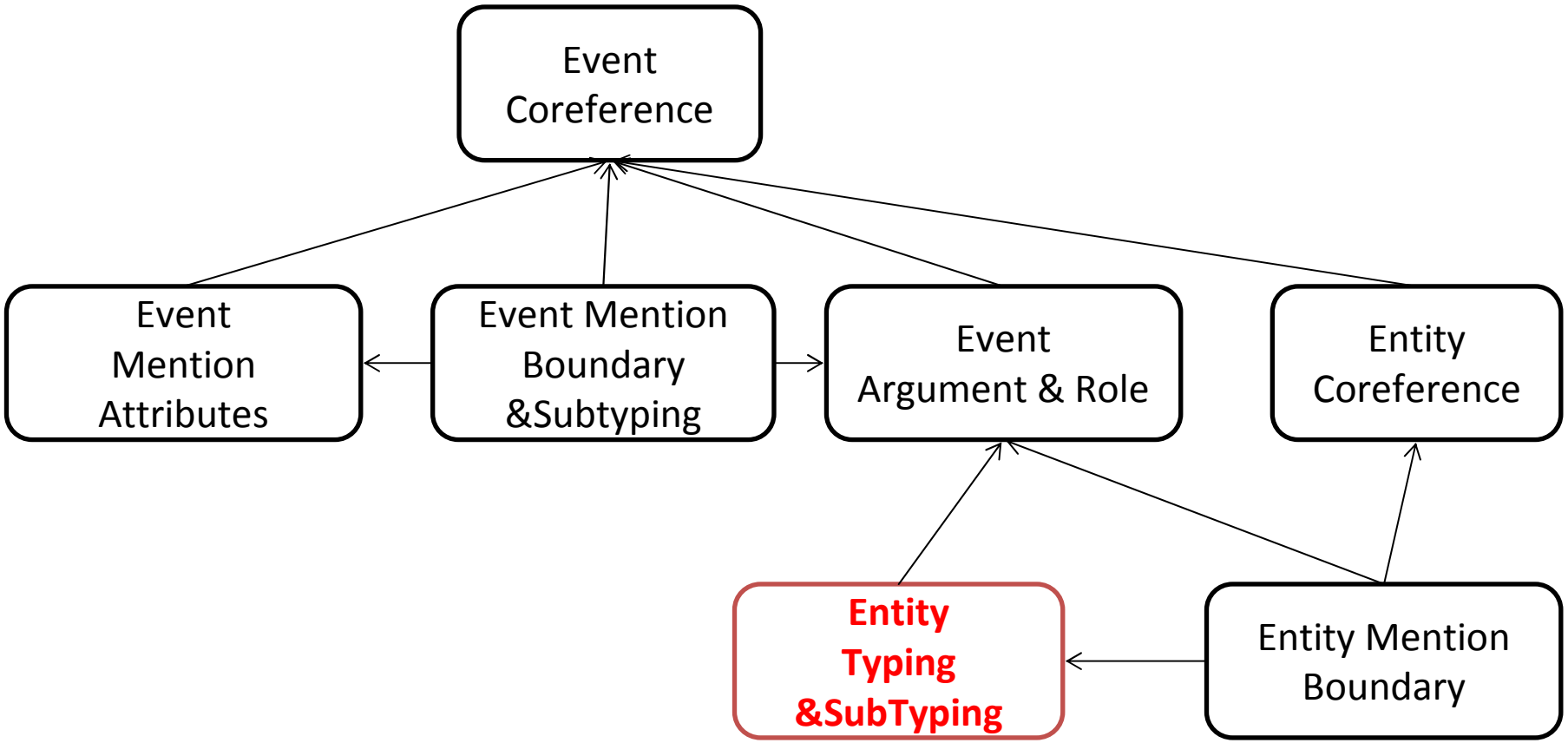
# Six Upstream Components for Event Coreference Resolution



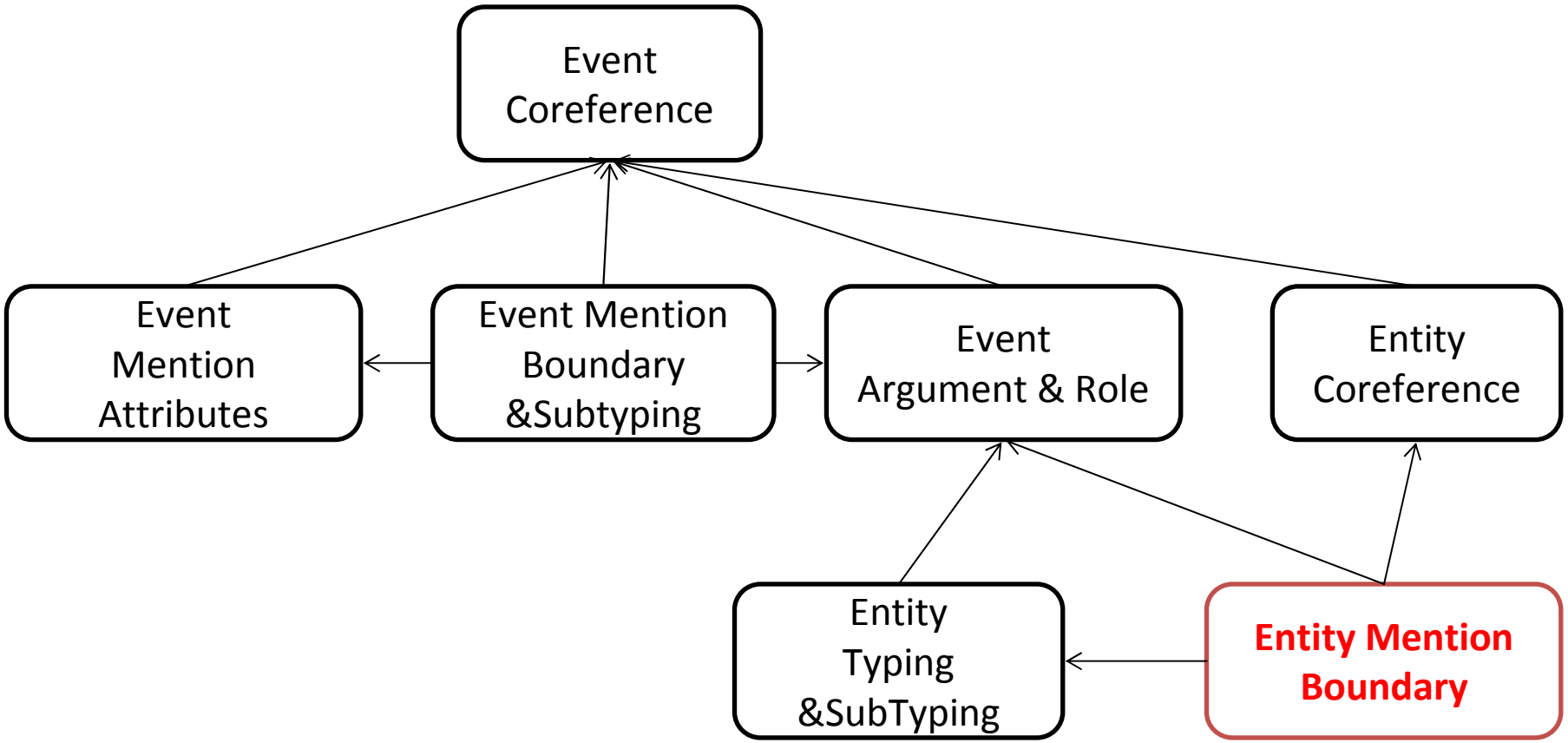
# Six Upstream Components for Event Coreference Resolution



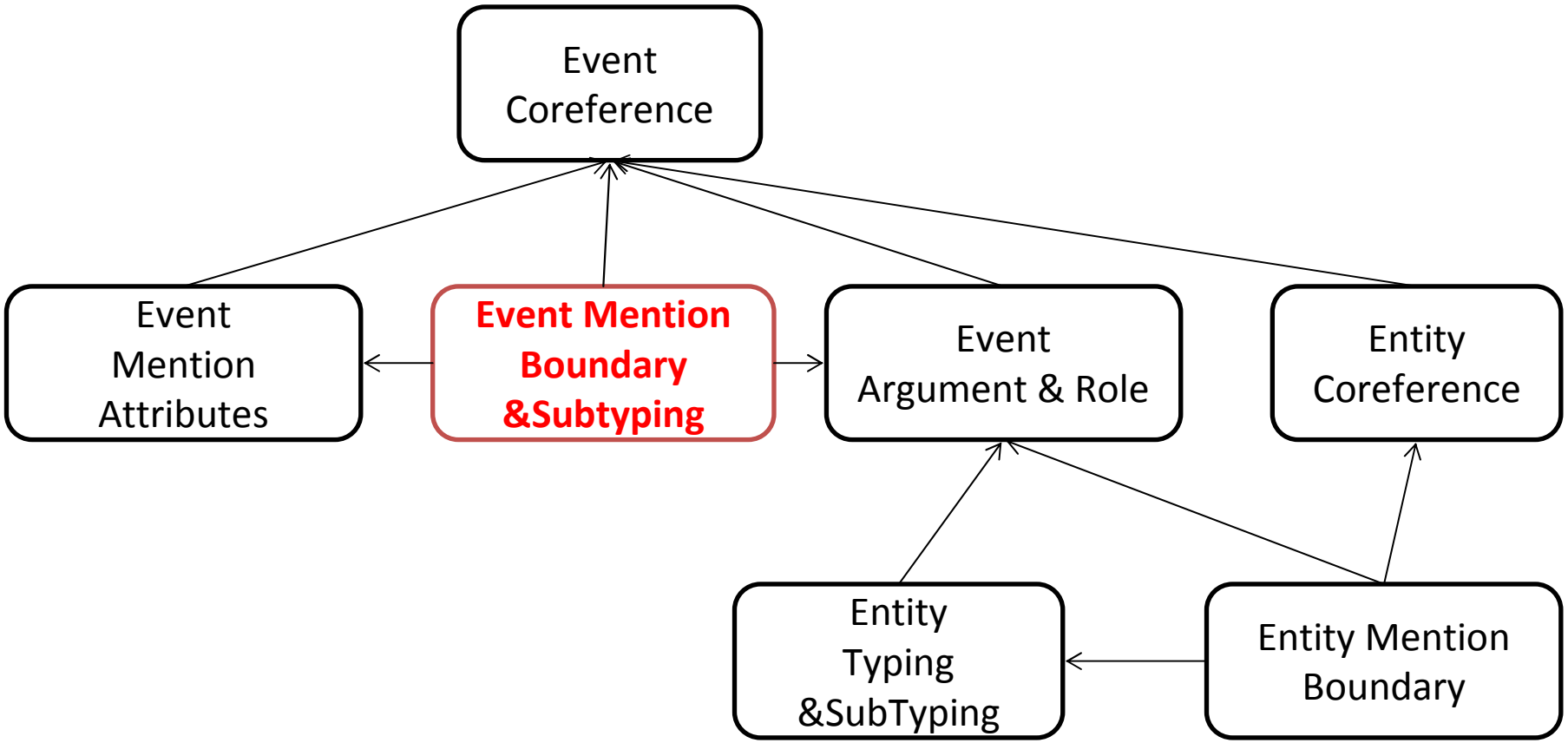
# Six Upstream Components for Event Coreference Resolution



# Six Upstream Components for Event Coreference Resolution



# Six Upstream Components for Event Coreference Resolution



# Event Mention Boundary Identification & SubTyping

- Goals
  - Provide the event mentions for event coreference

# Event Mention Boundary Identification & SubTyping

- Goals
  - Provide the event mentions for event coreference
  - Label each event mention with its event subtype

# Event Mention Boundary Identification & SubTyping

- Goals
  - Provide the event mentions for event coreference
  - Label each event mention with its event subtype,

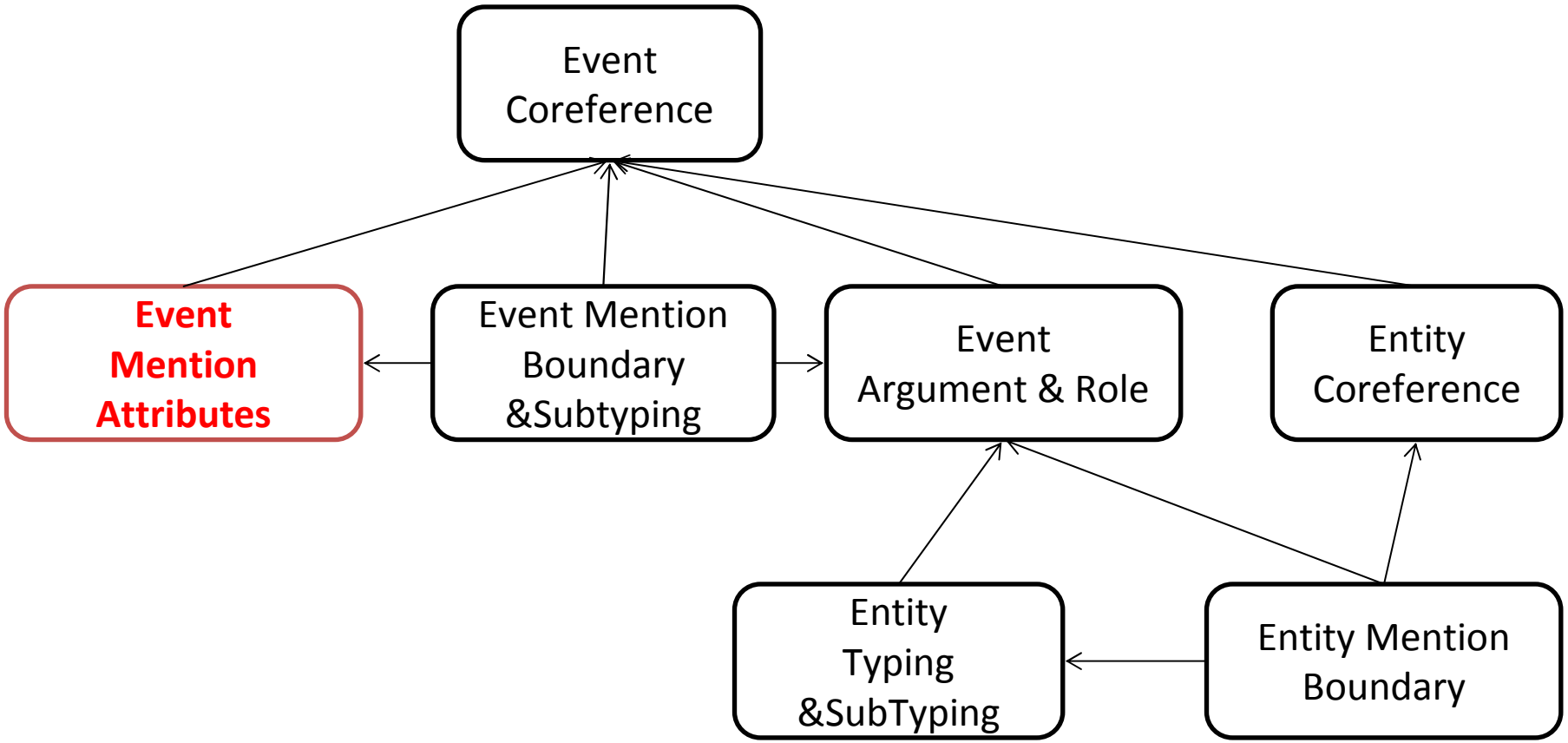
- Why is this component useful for event coreference?

# Event Mention Boundary Identification & SubTyping

- Goals
  - Provide the event mentions for event coreference
  - Label each event mention with its event subtype,

- Why is this component useful for event coreference?
  - Two event mentions with different type or subtype cannot be coreferent (useful feature for event coreference)

# Six Upstream Components for Event Coreference Resolution



# Event Mention Attribute Value Computation

- Goal
  - Assign each event mention with 4 attributes: POLARITY, MODALITY, GENERICITY and TENSE

# Event Mention Attribute Value Computation

- Goal
  - Assign each event mention with 4 attributes: POLARITY, MODALITY, GENERICITY and TENSE

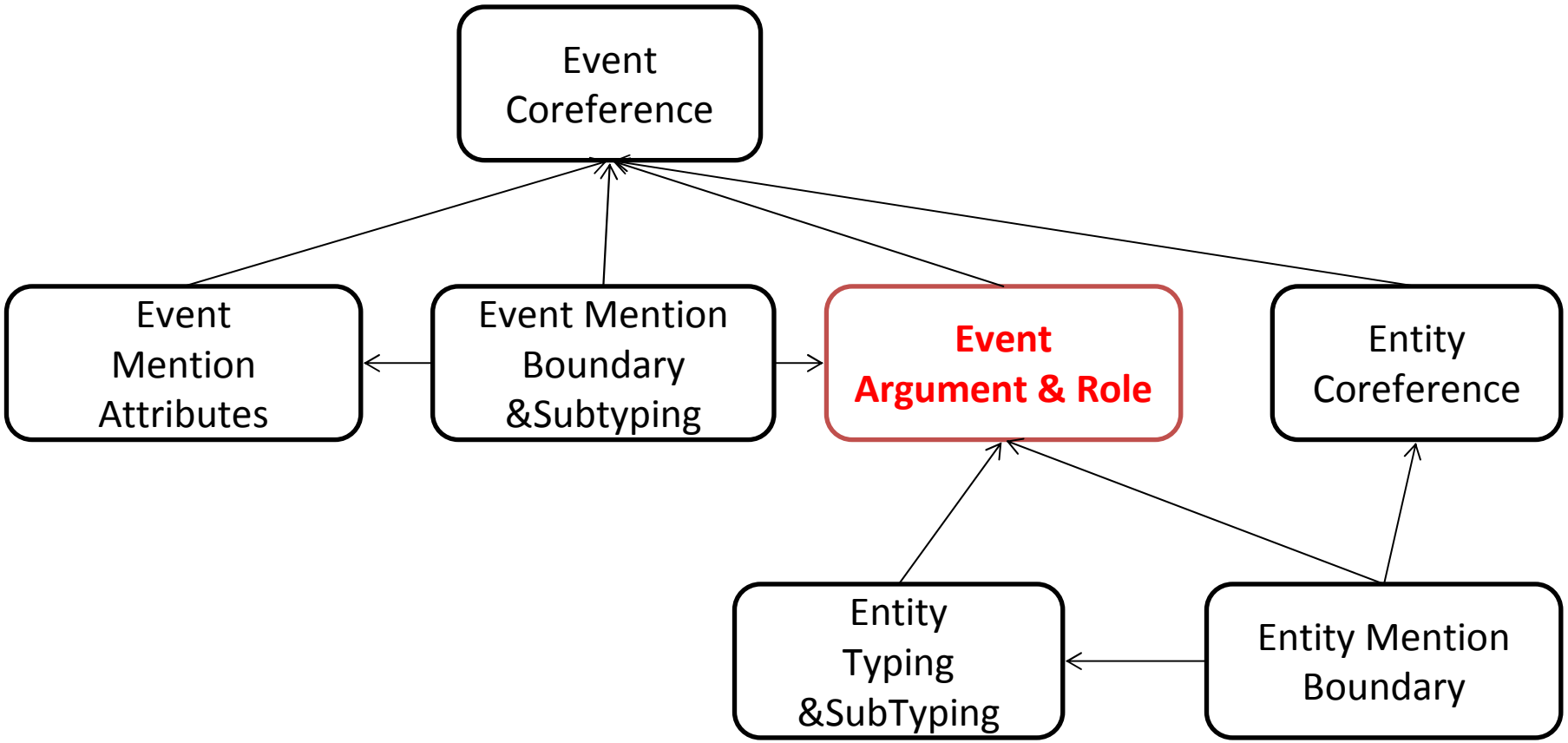
- Why is this component useful for event coreference?

# Event Mention Attribute Value Computation

- Goal
  - Assign each event mention with 4 attributes: POLARITY, MODALITY, GENERICITY and TENSE

- Why is this component useful for event coreference?
  - Two events differ in any of four attributes cannot be coreferent (useful feature for event coreference)

# Six Upstream Components for Event Coreference Resolution



# Event Argument & Role Classification

- Goals
  - Identify arguments for an event mention (e.g., the participants, time, place)

# Event Argument & Role Classification

- Goals
  - Identify arguments for an event mention (e.g., the participants, time, place)
  - Assign a role (e.g. VICTIM, PLACE, TIME-WITHIN) to each argument

# Event Argument & Role Classification

- Goals
  - Identify arguments for an event mention (e.g., the participants, time, place)
  - Assign a role (e.g. VICTIM, PLACE, TIME-WITHIN) to each argument

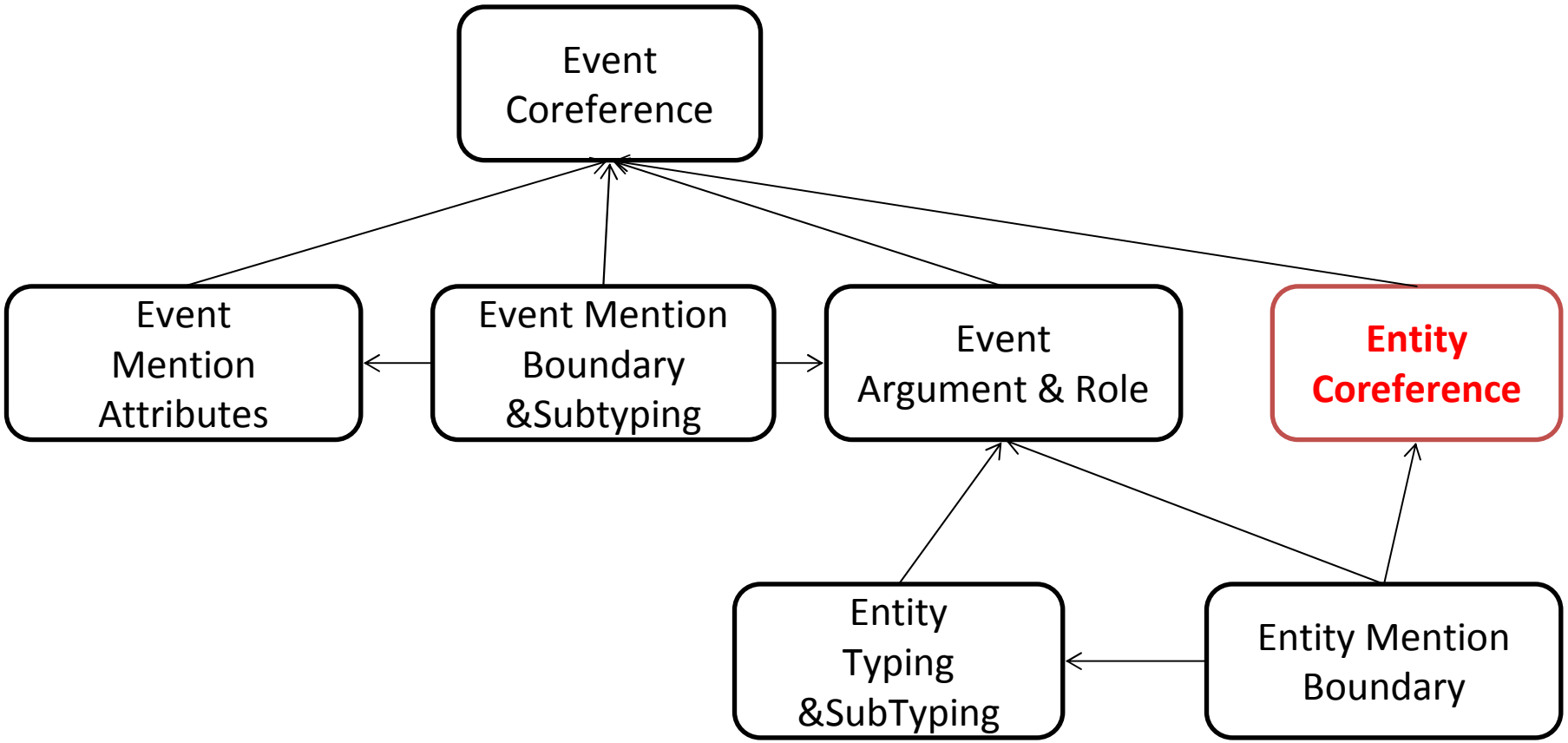
- Why is this component useful for event coreference?

# Event Argument & Role Classification

- Goals
  - Identify arguments for an event mention (e.g., the participants, time, place)
  - Assign a role (e.g. VICTIM, PLACE, TIME-WITHIN) to each argument

- Why is this component useful for event coreference?
  - Useful features for event coreference

# Six Upstream Components for Event Coreference Resolution



# Entity Coreference Resolution

- Goal
  - Create entity coreference clusters

# Entity Coreference Resolution

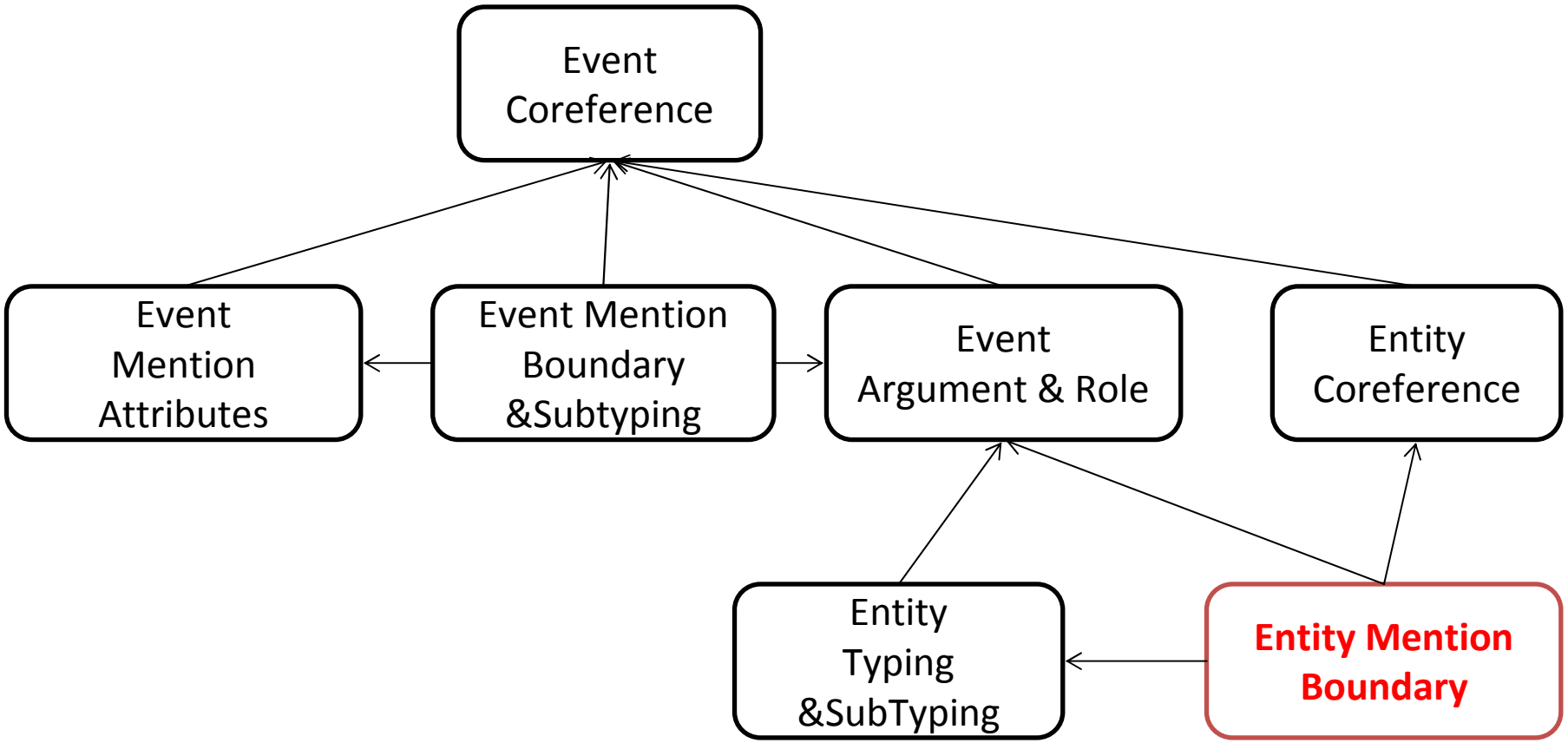
- Goal
  - Create entity coreference clusters

- Why is this component useful for event coreference?

# Entity Coreference Resolution

- Goal
  - Create entity coreference clusters
- Why is this component useful for event coreference?
  - Two event mentions having coreferent arguments are likely to be coreferent (useful feature for event coreference)

# Six Upstream Components for Event Coreference Resolution



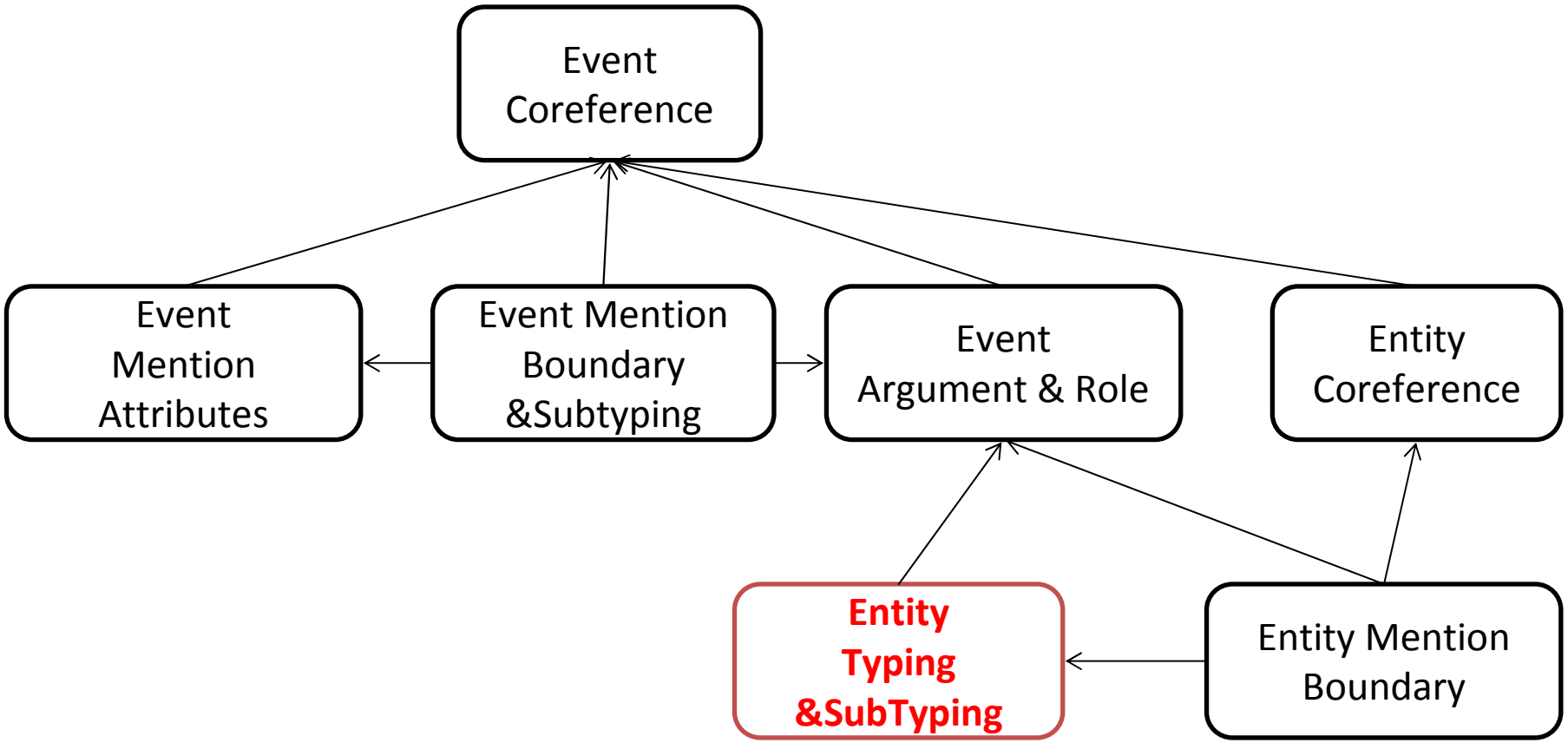
# Entity Mention Boundary Identification

- Goal
  - Provide candidate arguments and entity mentions needed by the aforementioned components

# Entity Mention Boundary Identification

- Goal
  - Provide candidate arguments and entity mentions needed by the aforementioned components
    - Indirect influence on event coreference

# Six Upstream Components for Event Coreference Resolution



# Entity Typing & SubTyping

- Goal
  - Determine the type and subtype of entity mention

# Entity Typing & SubTyping

- Goal
  - Determine the type and subtype of entity mention

    (Indirect influence on event coreference)

# Entity Typing & SubTyping

- Goal
  - Determine the type and subtype of entity mention

    (Indirect influence on event coreference)

- Why is this component useful for event coreference?

# Entity Typing & SubTyping

- Goal
  - Determine the type and subtype of entity mention
    (Indirect influence on event coreference)

- Why is this component useful for event coreference?
  - Features for classifying the role of event arguments

# Plan for the Talk

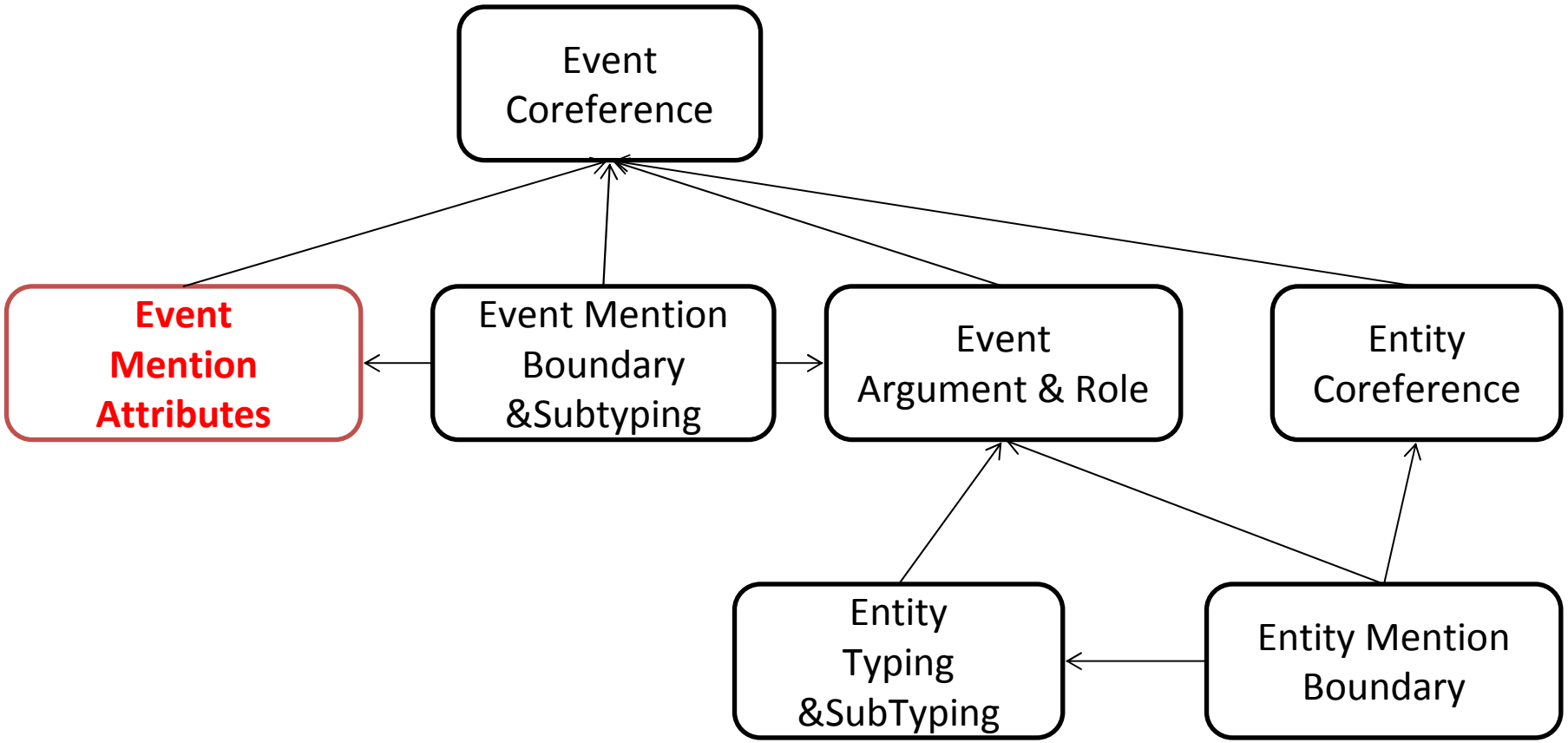
- ACE Event Coreference
- Six Upstream Components in the Pipeline
- Results and Analysis – Answer to Question 1
- Conclusion

# Evaluation Methodology

- Start with an event coreference resolver that assumes all six upstream components are error free
- Replace each oracle component with its system (i.e., automatically computed) counterpart one by one

# Replacement 1:

## Using System Event Mention Attribute Values for Train & Test



# Replacement 1:
## Using System Event Mention Attribute Values for Train & Test

- How to implement this component?
  - Following Chen et al.(2009), we train 4 classifiers to compute these attributes, with one classifier per attribute

# Replacement 1:
## Using System Event Mention Attribute Values for Train & Test

- How to implement this component?
  - Following Chen et al.(2009), we train 4 classifiers to compute these attributes, with one classifier per attribute

- Each of four attribute classifiers is only marginally better than a simple majority baseline

# Replacement 1:

## Using System Event Mention Attribute Values for Train & Test

| | | MUC | | | $B^3$ | | | $CEAF_e$ | | AvgF |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | P | F | R | P | F | R | P | F | F |
| Before | 80.4 | 70.0 | 74.8 | 88.4 | 79.7 | 83.8 | 57.3 | 66.8 | 1.7 | 73.4 |
| After | 72.5 | 64.5 | 68.3 | 83.8 | 77.4 | 80.5 | 53.1 | 59.9 | 56.3 | 68.3 |

- The average F decreases by 5.1%

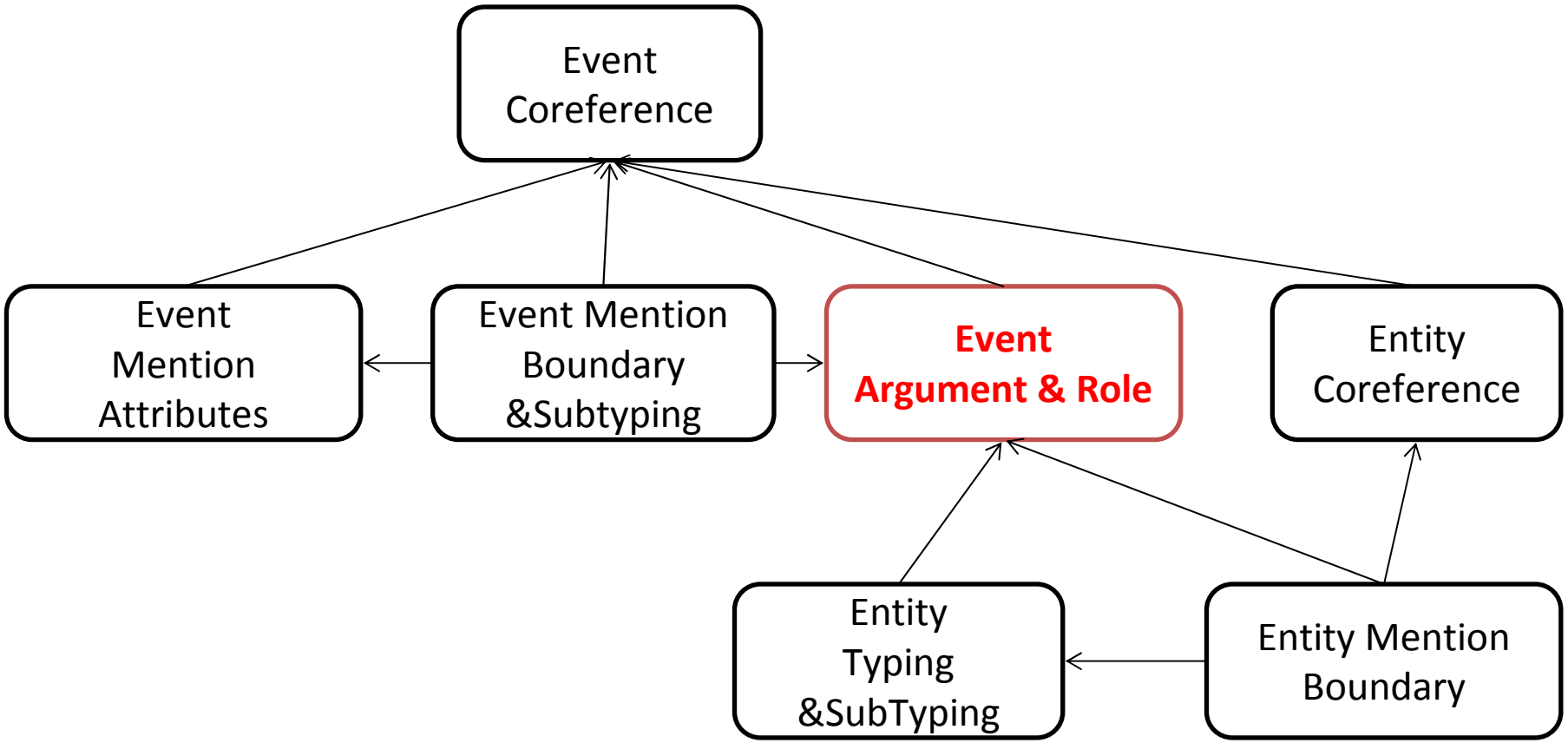
# Replacement 1:

## Using System Event Mention Attribute Values for Train & Test

| | | MUC | | | $B^3$ | | | $CEAF_e$ | | AvgF |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | P | F | R | P | F | R | P | F | F |
| Before | 80.4 | 70.0 | 74.8 | 88.4 | 79.7 | 83.8 | 57.3 | 66.8 | 1.7 | 73.4 |
| After | 72.5 | 64.5 | 68.3 | 83.8 | 77.4 | 80.5 | 53.1 | 59.9 | 56.3 | 68.3 |

- Conclusion 1:
  - Improving the four event attribute classifiers could significantly improve event coreference

# Replacement 2:
## Using System Event Argument & Role



# Replacement 2:
## Using System Event Argument & Role

- How to implement this component?
  - Implemented as part of our Chinese event extraction system (Chen and Ng, 2012c)

# Replacement 2:
## Using System Event Argument & Role

- How to implement this component?
  - Implemented as part of our Chinese event extraction system (Chen and Ng, 2012c)

- Given gold event mention boundary and subtyping, the F-score of event argument and role classification are 76.9% and 68.2%

# Replacement 2:
## Using System Event Argument & Role

| | | MUC | | | $B^3$ | | | $CEAF_e$ | | AvgF |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | P | F | R | P | F | R | P | F | F |
| Before | 72.5 | 64.5 | 68.3 | 83.8 | 77.4 | 80.5 | 53.1 | 59.9 | 56.3 | 68.3 |
| After | 71.2 | 61.2 | 65.8 | 83.9 | 74.9 | 79.1 | 49.9 | 58.0 | 53.6 | 66.2 |

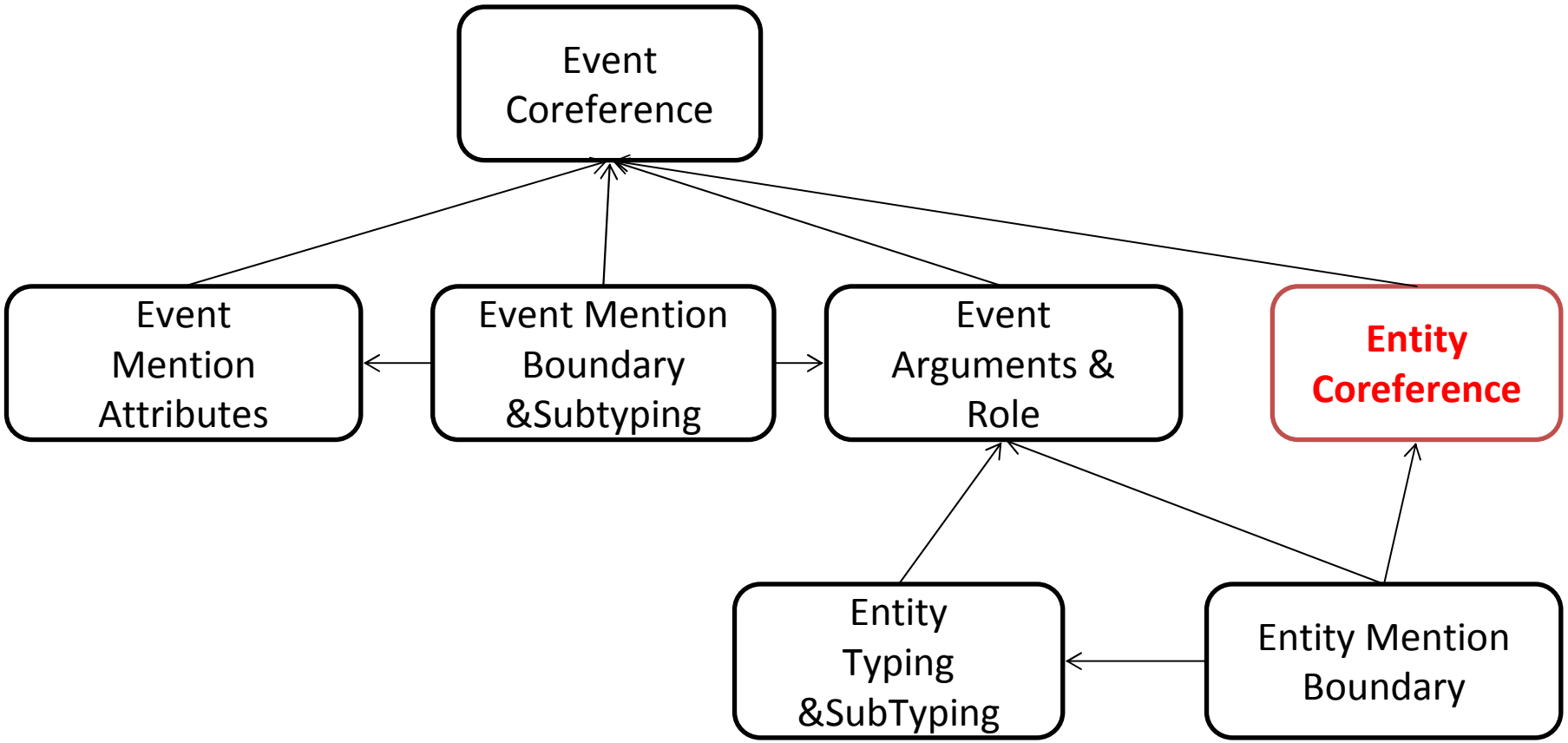
- After replacing gold with system event argument and role, average F-score drops slightly, though significantly, by 2.1%

# Replacement 2:
## Using System Event Argument & Role

| | | MUC | | | $B^3$ | | | $CEAF_e$ | | AvgF |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | P | F | R | P | F | R | P | F | F |
| Before | 72.5 | 64.5 | 68.3 | 83.8 | 77.4 | 80.5 | 53.1 | 59.9 | 56.3 | 68.3 |
| After | 71.2 | 61.2 | 65.8 | 83.9 | 74.9 | 79.1 | 49.9 | 58.0 | 53.6 | 66.2 |

- Conclusion 2:
  - Event argument and role classification have a small, but significant, impact on event coreference performance

# Replacement 3:
# Using System Entity Coreference



# Replacement 3: Using System Entity Coreference

- How to implement this component?
  - Provided by our Chinese entity coreference resolver (Chen and Ng, 2012b)

# Replacement 3: Using System Entity Coreference

- How to implement this component?
  - Provided by our Chinese entity coreference resolver (Chen and Ng, 2012b)

- Given gold entity mentions, our system entity coreference resolver achieves a MUC F-score of 78.0%

# Replacement 3: Using System Entity Coreference

| | | MUC | | | $B^3$ | | | $CEAF_e$ | | AvgF |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | P | F | R | P | F | R | P | F | F |
| Before | 71.2 | 61.2 | 65.8 | 83.9 | 74.9 | 79.1 | 49.9 | 58.0 | 53.6 | 66.2 |
| After | 61.6 | 58.5 | 60.0 | 79.0 | 75.7 | 77.3 | 49.1 | 51.5 | 50.3 | 62.5 |

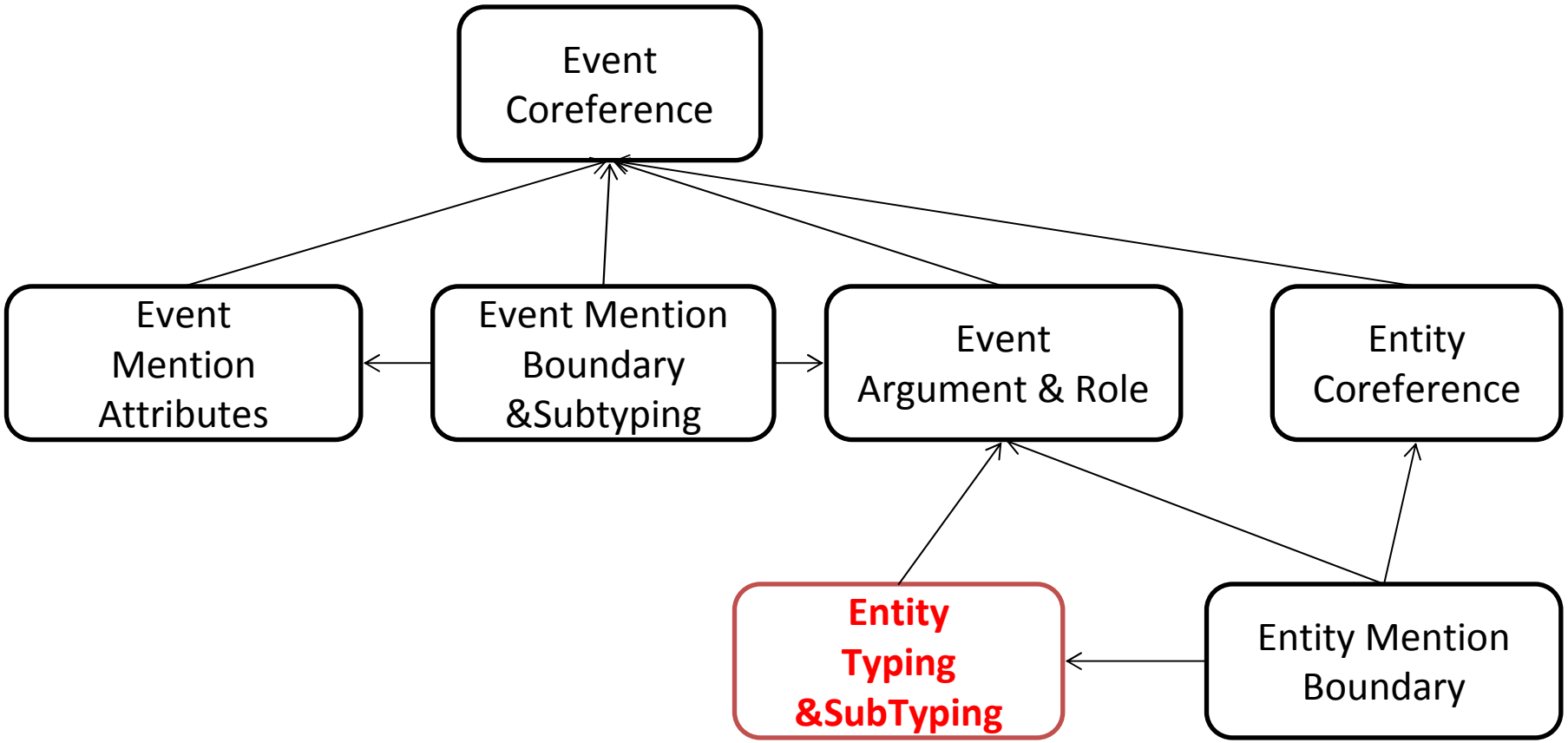
- Replacing gold with system entity coreference incurs a 3.7% drop

# Replacement 3: Using System Entity Coreference

| | | MUC | | | $B^3$ | | | $CEAF_e$ | | AvgF |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | P | F | R | P | F | R | P | F | F |
| Before | 71.2 | 61.2 | 65.8 | 83.9 | 74.9 | 79.1 | 49.9 | 58.0 | 53.6 | 66.2 |
| After | 61.6 | 58.5 | 60.0 | 79.0 | 75.7 | 77.3 | 49.1 | 51.5 | 50.3 | 62.5 |

- Conclusion 3:
  - Improve entity coreference could significantly improve event coreference

# Replacement 4:
# Using System Entity Typing & SubTyping



# Replacement 4:
## Using System Entity Typing & SubTyping

- How to implement this component?
  - We determine entity types and subtypes by training two SVM classifiers

# Replacement 4:
## Using System Entity Typing & SubTyping

- How to implement this component?
  - We determine entity types and subtypes by training two SVM classifiers

- Given gold entity mentions, system entity type and subtype classifiers achieve F-scores of 90.1% and 81.6%

# Replacement 4:
## Using System Entity Typing & SubTyping

| | | MUC | | | $B^3$ | | | $CEAF_e$ | | AvgF |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | P | F | R | P | F | R | P | F | F |
| Before | 61.6 | 58.5 | 60.0 | 79.0 | 75.7 | 77.3 | 49.1 | 51.5 | 50.3 | 62.5 |
| After | 62.2 | 57.9 | 60.0 | 79.4 | 75.2 | 77.2 | 49.0 | 52.3 | 50.6 | 62.6 |

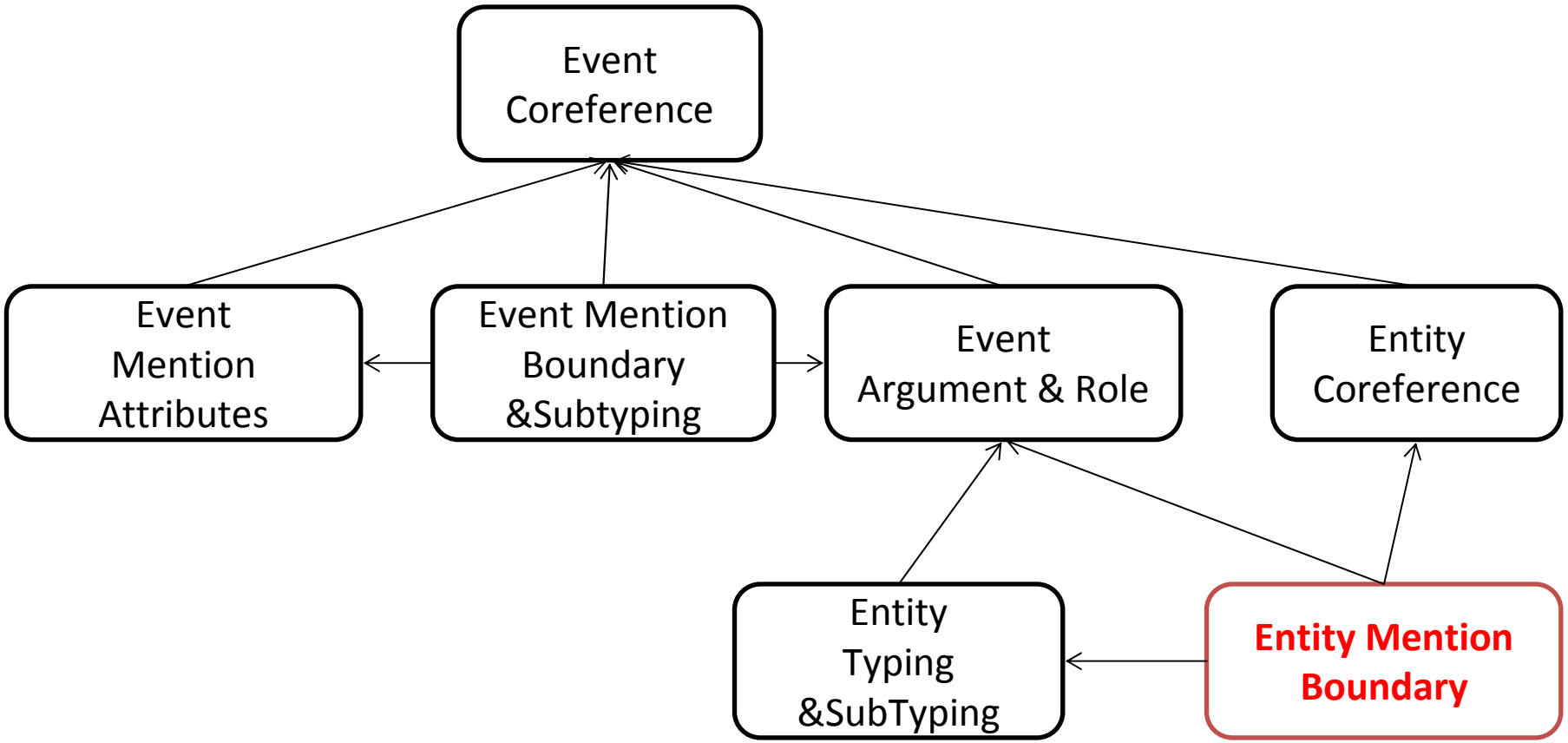
- After replacing gold with system entity types and subtypes, event coreference performance does not drop

# Replacement 4:
## Using System Entity Typing & SubTyping

| | | MUC | | | $B^3$ | | | $CEAF_e$ | | AvgF |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | P | F | R | P | F | R | P | F | F |
| Before | 61.6 | 58.5 | 60.0 | 79.0 | 75.7 | 77.3 | 49.1 | 51.5 | 50.3 | 62.5 |
| After | 62.2 | 57.9 | 60.0 | 79.4 | 75.2 | 77.2 | 49.0 | 52.3 | 50.6 | 62.6 |

- Conclusion 4:
  - Improving entity typing & subTyping classification is unlikely to improve event coreference

# Replacement 5:
## Using System Entity Mention Boundary detection



# Replacement 5:
## Using System Entity Mention Boundary detection

- How to implement this component?
  - We train CRF classifiers to extract entity mentions, time expressions and value expressions

# Replacement 5:
## Using System Entity Mention Boundary detection

- How to implement this component?
  - We train CRF classifiers to extract entity mentions, time expressions and value expressions

- System entity mention boundary detection component achieves an F-score of 84.7%

# Replacement 5:
## Using System Entity Mention Boundary detection

| | | MUC | | | $B^3$ | | | $CEAF_e$ | | AvgF |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | P | F | R | P | F | R | P | F | F |
| Before | 62.2 | 57.9 | 60.0 | 79.4 | 75.2 | 77.2 | 49.0 | 52.3 | 50.6 | 62.6 |
| After | 63.3 | 57.4 | 60.2 | 80.2 | 74.4 | 77.2 | 48.2 | 52.8 | 50.4 | 62.6 |

- Replacing gold with system entity mention boundary detection does not alter event coreference performance

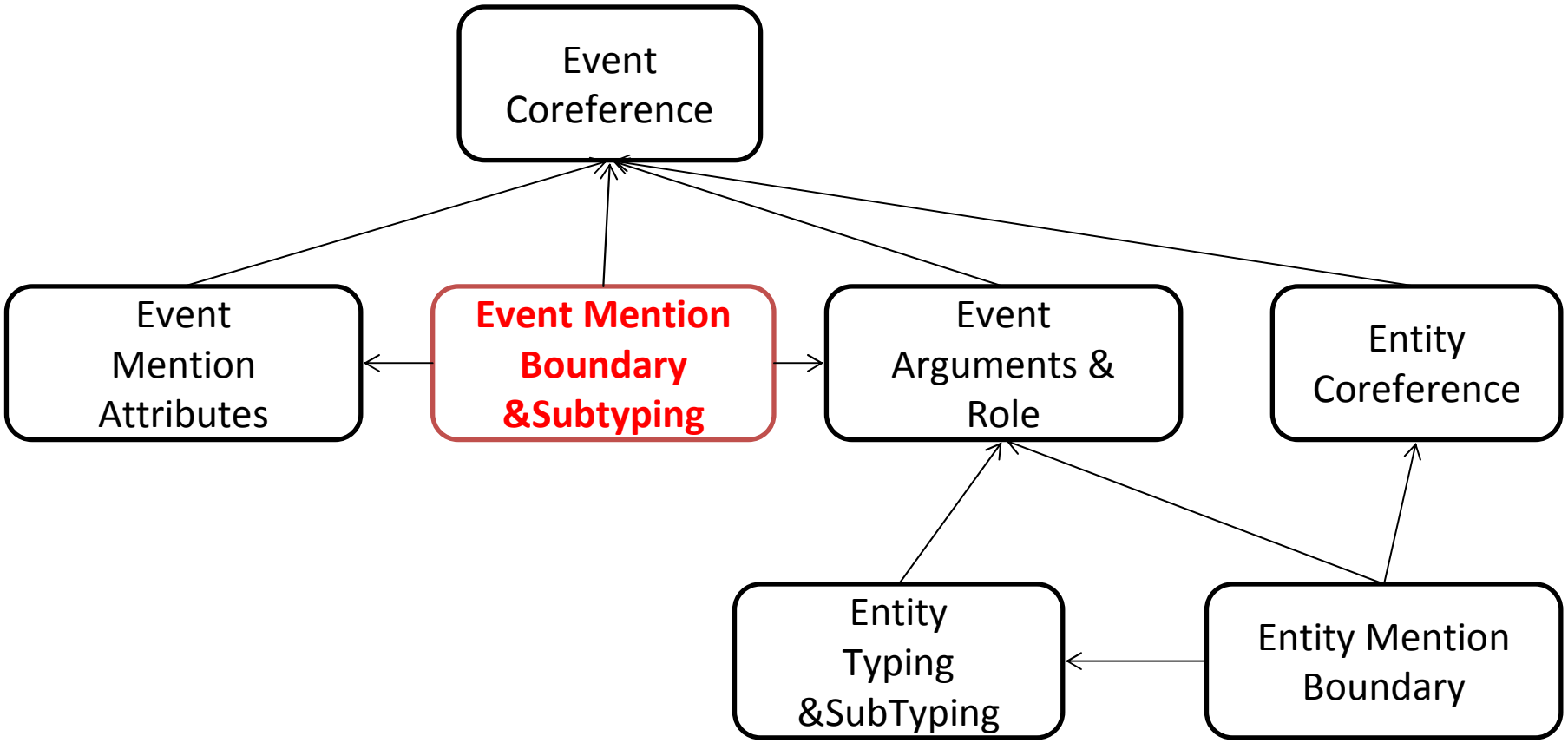
# Replacement 5:

## Using System Entity Mention Boundary detection

| | | MUC | | | $B^3$ | | | $CEAF_e$ | | AvgF |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | P | F | R | P | F | R | P | F | F |
| Before | 62.2 | 57.9 | 60.0 | 79.4 | 75.2 | 77.2 | 49.0 | 52.3 | 50.6 | 62.6 |
| After | 63.3 | 57.4 | 60.2 | 80.2 | 74.4 | 77.2 | 48.2 | 52.8 | 50.4 | 62.6 |

- Conclusion 5:
  - Improving entity mention boundary detection may not improve event coreference

# Replacement 6:
## Using System Event Mention Boundary Identification & subtyping



# Replacement 6:
## Using System Event Mention Boundary Identification & subtyping

- How to implement this component?
  - Implemented as part of our Chinese event extraction system (Chen and Ng, 2012c)

# Replacement 6:
## Using System Event Mention Boundary Identification & subtyping

- How to implement this component?
  - Implemented as part of our Chinese event extraction system (Chen and Ng, 2012c)

- System event mention boundary identifier achieves an F-score of 65.1%
- System event subtype classifier achieves an F-score of 61.3%

# Replacement 6:
## Using System Event Mention Boundary Identification & subtyping

| | | MUC | | | $B^3$ | | | $CEAF_e$ | | AvgF |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | P | F | R | P | F | R | P | F | F |
| Before | 63.3 | 57.4 | 60.2 | 80.2 | 74.4 | 77.2 | 48.2 | 52.8 | 50.4 | 62.6 |
| After | 37.4 | 36.7 | 37.1 | 72.8 | 71.1 | 71.9 | 40.6 | 41.1 | 40.8 | 49.9 |

- Replacing gold with system event mention boundary and subtyping causes average F-score to drop by 12.7%

# Replacement 6:
## Using System Event Mention Boundary Identification & subtyping

| | | MUC | | | $B^3$ | | | $CEAF_e$ | | AvgF |
|---|---|---|---|---|---|---|---|---|---|---|
| | R | P | F | R | P | F | R | P | F | F |
| Before | 63.3 | 57.4 | 60.2 | 80.2 | 74.4 | 77.2 | 48.2 | 52.8 | 50.4 | 62.6 |
| After | 37.4 | 36.7 | 37.1 | 72.8 | 71.1 | 71.9 | 40.6 | 41.1 | 40.8 | 49.9 |

- Conclusion 6:
  - Event mention boundary identification and subtyping is the upstream component that has the largest impact

# Answer to Question 1

- Components whose noise have an impact on event coreference performance (in decreasing order of impact):
  - Event Mention Boundary Identification & SubTyping
  - Event Mention Attribute Value Computation
  - Entity Coreference Resolution
  - Event Argument & Role Classification
- Components whose noise do not have an impact:
  - Entity Mention Boundary Identification
  - Entity Typing & SubTyping

# Plan for the Talk

- ACE Event Coreference
- Six Upstream Components
- Results and Analysis – Answer to Question 1
- Error Analysis – Answer to Question 2
- Conclusion

# Precision Errors

- Lack of event timestamping
  - Only events occurring exactly at the same time can be coreferent
  - TENSE is just a very rough approximation

# Lack of event timestamping

E1: In last March, Yang Guangnan was [arrested] in Shanghai for the first time.

E2: Yang Guangnan was [arrested] again in Shanghai.

# Precision Errors

- Incompatible triggers
  - Two events containing coreferent arguments, but triggers are semantically incompatible

# Incompatible triggers

E1: On the 28th, Sam Nujoma arrived in Pyongyang by plane for an official goodwill [visit] to the DPRK.

E2: Namibian President Sam Nujoma [arrived] in Pyongyang by plane on the 28th.

# Precision Errors

- Incompatible important arguments
  - Two events containing strong hint to be coreferent, while some of their important arguments are incompatible

# Incompatible important arguments

E1: The delegation [visited] Sweden.

E2: During their [visit] in Denmark, the Chinese Christian delegation held a press conference.

# Recall Errors

- Coreferent mentions with synonymous triggers
  - Event mentions that have synonymous but lexically different trigger words

# Coreferent mentions with synonymous triggers

E1: Jewish [violence] against the Arabs.

E2: Two parties of [conflict].

# Recall Errors

- Coreferent mentions with compatible arguments
  - Though two argument are not coreferent, while they are compatible

# Coreferent mentions with compatible arguments

E1: Yugoslavia's head of state [visited] Bosnia-Herzegovina for the first time.

E2: Kostunica [visited] Sarajevo, the capital of Bosnia-Herzegovina.

# Plan for the Talk

- ACE Event Coreference
- Six Upstream Components in the Pipeline
- Results and Analysis – Answer to Question 1
- Conclusion

# Conclusion

- We analyzed an ACE-style Chinese event coreference system by investigating:
  - The extent to which its performance is affected by the errors made by its upstream components
  - Types of errors made by the resolution algorithm