Cross-Document Coreference Resolution for Entities and Events

Thesis Proposal

Chris Tanner
Advisor: Eugene Charniak
Committee: Michael Littman, Stefanie Tellex, Ellie Pavlik

Problem Introduction

Related Work

Completed Work

Proposed Work

Real-World Example

The New York Times

Tuesday, May 15, 2018 | Today's Paper | Video | 74°F | Nasdaq -0.81%

North Korea Postpones Talks With South, Hinting Kim-Trump Summit Is in Peril

[…]

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

Real-World Example

The New York Times

Tuesday, May 15, 2018 | Today's Paper | Video | 74°F | Nasdaq -0.81% ↓

North Korea Postpones Talks With South, Hinting Kim-Trump Summit Is in Peril

? […]

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

Real-World Example

The New York Times

Tuesday, May 15, 2018 | Today's Paper | Video | 74°F | Nasdaq -0.81%

North Korea Postpones Talks With South, Hinting Kim-Trump Summit Is in Peril

[…]

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

Real-World Example

The New York Times

Tuesday, May 15, 2018 | Today's Paper | Video | 74°F | Nasdaq -0.81% ↓

North Korea Postpones Talks With South, Hinting Kim-Trump Summit Is in Peril

[…]

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

Real-World Example

The New York Times

Tuesday, May 15, 2018 | Today's Paper | Video | 74°F | Nasdaq -0.81% ↓

North Korea Postpones Talks With South, Hinting Kim-Trump Summit Is in Peril

[…]

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

Real-World Example

The New York Times

Tuesday, May 15, 2018 | Today's Paper | Video | 74°F | Nasdaq -0.81%

North Korea Postpones Talks With South, Hinting Kim-Trump Summit Is in Peril

[…]

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

Real-World Example

The New York Times

Tuesday, May 15, 2018 | Today's Paper | Video | 74°F | Nasdaq -0.81% ↓

North Korea Postpones Talks With South, Hinting Kim-Trump Summit Is in Peril

[…]

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

Real-World Example

The New York Times

Tuesday, May 15, 2018 | Today's Paper | Video | 74°F | Nasdaq -0.81%

North Korea Postpones Talks With South, Hinting Kim-Trump Summit Is in Peril

[…]

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

Real-World Example

The New York Times

Tuesday, May 15, 2018 | Today's Paper | Video | 74°F | Nasdaq -0.81%

North Korea Postpones Talks With South, Hinting Kim-Trump Summit Is in Peril

[…]

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

Easy for humans!

Real-World Example

The New York Times

Tuesday, May 15, 2018 | Today's Paper | Video | 74°F | Nasdaq -0.81%

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

State-of-the-art Computer System?

Real-World Example

The New York Times

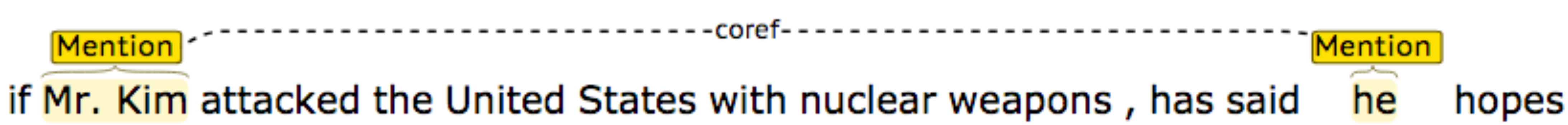
Tuesday, May 15, 2018 | Today's Paper | Video | 74°F | Nasdaq -0.81% ↓

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

↓

State-of-the-art Computer System?

↓



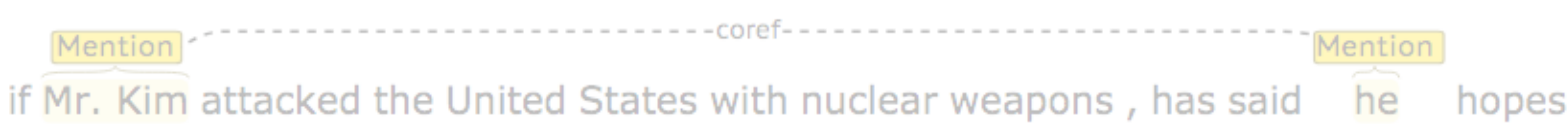
Real-World Example


The New York Times

Tuesday, May 15, 2018 | Today's Paper | Video | 74°F | Nasdaq -0.81%

Mr. Trump, who threatened North Korea less than a year ago with “fire and fury” if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades

Hard for computers!



Motivation

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

Motivation

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

Coreference Resolution

definition: grouping together all words which refer to the same underlying thing

Motivation

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

Coreference Resolution

definition: grouping together all words which refer to the same underlying thing

benefits:

- Question Answering (Narayanan and Harabagui, 2004)
- Information Extraction (Humphreys, et. al., 1997)
- Document Summarization (Daniel, et. al., 2003)

Multiple Documents

Good models should be able to group words from across multiple documents

Multiple Documents

Good models should be able to group words from across multiple documents

The Washington Post

Democracy Dies in Darkness

May 15, 2018

North Korea threatens to cancel summit with Trump over U.S.-South Korean military drills

[…]

North Korea suggested that the drills were putting the proposed summit between Trump and Kim, scheduled for June 12, in jeopardy.

Trump and Kim are due to meet in Singapore, which would be the first time a North Korean leader had met with a sitting U.S. president.

Multiple Documents

Good models should be able to group words from across multiple documents

The Washington Post

Democracy Dies in Darkness

May 15, 2018

North Korea threatens to cancel summit with Trump over U.S.-South Korean military drills

[…]

North Korea suggested that the drills were putting the proposed summit between Trump and Kim, scheduled for June 12, in jeopardy.

Trump and Kim are due to meet in Singapore, which would be the first time a North Korean leader had met with a sitting U.S. president.

?

Multiple Documents

Good models should be able to group words from across multiple documents

The Washington Post

Democracy Dies in Darkness

May 15, 2018

North Korea threatens to cancel summit with Trump over U.S.-South Korean military drills

[…]

North Korea suggested that the drills were putting the proposed summit between Trump and Kim, scheduled for June 12, in jeopardy.

Trump and Kim are due to meet in Singapore, which would be the first time a North Korean leader had met with a sitting U.S. president.

Multiple Documents

Good models should be able to group words from across multiple documents

The Washington Post

Democracy Dies in Darkness

May 15, 2018

North Korea threatens to cancel summit with Trump over U.S.-South Korean military drills

[…]

North Korea suggested that the drills were putting the proposed summit between Trump and Kim, scheduled for June 12, in jeopardy.

Trump and Kim are due to meet in Singapore, which would be the first time a North Korean leader had met with a sitting U.S. president.

?

Multiple Documents

Good models should be able to group words from across multiple documents

The Washington Post

Democracy Dies in Darkness

May 15, 2018

North Korea threatens to cancel summit with Trump over U.S.-South Korean military drills

[…]

North Korea suggested that the drills were putting the proposed summit between Trump and Kim, scheduled for June 12, in jeopardy.

Trump and Kim are due to meet in Singapore, which would be the first time a North Korean leader had met with a sitting U.S. president.

Motivation

Good models should be able to group words from across multiple documents

The New York Times

Tuesday, May 15, 2018 | Today's Paper | Video | 74°F | Nasdaq -0.81%

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

The Washington Post

Democracy Dies in Darkness

May 15, 2018

North Korea suggested that the drills were putting the proposed summit between Trump and Kim, scheduled for June 12, in jeopardy.

Trump and Kim are due to meet in Singapore, which would be the first time a North Korean leader had met with a sitting U.S. president.

Definitions

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility. The New York Times

Definitions

mention = word(s) which refer to an underlying entity or event

Mr. Trump who he Mr. Kim

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility. The New York Times

Definitions

mention = word(s) which refer to an underlying entity or event

Mr. Trump who he Mr. Kim

entity = a person, location, organization, or time

Mr. Trump, who threatened North Korea less than a year ago with “fire and fury” if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility. The New York Times

Definitions

mention = word(s) which refer to an underlying entity or event

Mr. Trump who he Mr. Kim

entity = a person, location, organization, or time

event = a specific action

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility. The New York Times

Definitions

mention = word(s) which refer to an underlying entity or event

Mr. Trump who he Mr. Kim threatened attacked

entity = a person, location, organization, or time

event = a specific action

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility. The New York Times

Definitions

I research **Event Coreference** because it helps complete the picture.

Different axis of info from Entity Coreference

Event Coref is hard:

Wide-reading:

The attack took place yesterday.
The bombing killed four people.

Paraphrase:

She gave him the book.
He was given the book by her.

Two events co-refer if they share the same spatiotemporal properties and participants (Quine, 1985)

Common Systems

	entities	events
within-doc		
cross-doc		

Common Systems

Most systems do:

	entities	events
within-doc	✓	
cross-doc		

Common Systems

Few systems do:

	entities	events
within-doc		✓
cross-doc		✓

Common Systems

Few systems do:

(My system currently does this)

	entities	events
within-doc		✓
cross-doc		✓

Common Systems

Nobody* does:

	entities	events
within-doc	✓	✓
cross-doc	✓	✓

Common Systems

~~Nobody* does:~~

I propose doing:

	entities	events
within-doc	✓	✓
cross-doc	✓	✓

Definitions

Most systems do:

	entities	events
within-doc	✔	
cross-doc		

The New York Times

Tuesday, May 15, 2018 | Today's Paper | Video | 74°F | Nasdaq -0.81%

Mr. Trump, who threatened North Korea less than a year ago with “fire and fury” if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

The Washington Post

Democracy Dies in Darkness

May 15, 2018

North Korea suggested that the drills were putting the proposed summit between Trump and Kim, scheduled for June 12, in jeopardy.

Trump and Kim are due to meet in Singapore, which would be the first time a North Korean leader had met with a sitting U.S. president.

Definitions

Some systems do:

(**My system does**)

	entities	events
within-doc		✓
cross-doc		✓

The New York Times

Tuesday, May 15, 2018 | Today's Paper | Video | 74°F | Nasdaq -0.81%

Mr. Trump, who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

The Washington Post

Democracy Dies in Darkness

May 15, 2018

North Korea suggested that the drills were putting the proposed summit between Trump and Kim, scheduled for June 12, in jeopardy.

Trump and Kim are due to meet in Singapore, which would be the first time a North Korean leader had met with a sitting U.S. president.

Definitions

~~Nobody* does:~~
I propose doing:

	entities	events
within-doc	✓	✓
cross-doc	✓	✓

The New York Times

Tuesday, May 15, 2018 | Today's Paper | Video | 74°F | Nasdaq -0.81% ↓

Mr. Trump who threatened North Korea less than a year ago with "fire and fury" if Mr. Kim attacked the United States with nuclear weapons, has said he hopes the planned June 12 meeting in Singapore will lead to improved relations with North Korea after decades of hostility.

The Washington Post

Democracy Dies in Darkness

May 15, 2018

North Korea suggested that the drills were putting the proposed summit between Trump and Kim, scheduled for June 12, in jeopardy.

Trump and Kim are due to meet in Singapore, which would be the first time a North Korean leader had met with a sitting U.S. president.

Thesis Statement

This research will test the hypothesis that a cross-document coreference resolution system which *jointly* models entities and events can improve state-of-the-art results from having modelled them individually.

Related Work

Related Work

Entity Coreference

- Long history (Hobbs, 1978)
- Pronouns are hard
- Corpus: CoNLL-2012:
- within-doc annotations only
  - train set: 2,802 docs
  - dev set: 343 docs
  - test set: 345 docs

Related Work

Entity Coreference

- Long history (Hobbs, 1978)
- Pronouns are hard
- Corpus: CoNLL-2012:
- within-doc annotations only
  - train set: 2,802 docs
  - dev set: 343 docs
  - test set: 345 docs

Event Coreference

- Shorter history (MUC/ACE ,1997)
- Events not as unique as Entities
  - Sony announced today
  - Friday, Obama announced
- Varying lexical representations:
  - The casting of Smith
  - Smith stepped into the role
  - Smith was handed the keys to play

Related Work

We

- performed Event coreference
- used Deep Learning

Related Work

We

- performed Event coreference
- used Deep Learning

Event Coreference Research

	Model	Corpus
Yang, et. al. (2015)	HDDCRP	ECB+
Choubey and Huang (2017)	FFNN	ECB+
our work	CCNN	ECB+

Entity Coreference Research

	Model	Corpus
Wiseman, et. al. (2015)	FFNN	CoNLL-2012
Wiseman, et. al. (2016)	RNN	CoNLL-2012
Clark and Manning (2016)	FFNN	CoNLL-2012
Clark and Manning (2016)	RL	CoNLL-2012
Lee, et. al. (2017)	LSTM + FFNN	CoNLL-2012

Related Work

Takeaway #1: all **Event** research uses ECB+ corpus

Yang, et. al. (2015)	HDDCRP	ECB+
Choubey and Huang (2017)	FFNN	ECB+
our work	CCNN	ECB+

Wiseman, et. al. (2015)	FFNN	CoNLL-2012
Wiseman, et. al. (2016)	RNN	CoNLL-2012
Clark and Manning (2016)	FFNN	CoNLL-2012
Clark and Manning (2016)	RL	CoNLL-2012
Lee, et. al. (2017)	LSTM + FFNN	CoNLL-2012

Related Work

Takeaway #1: all **Event** research uses ECB+ corpus

Takeaway #2: Only ~4 successful Deep Learning papers

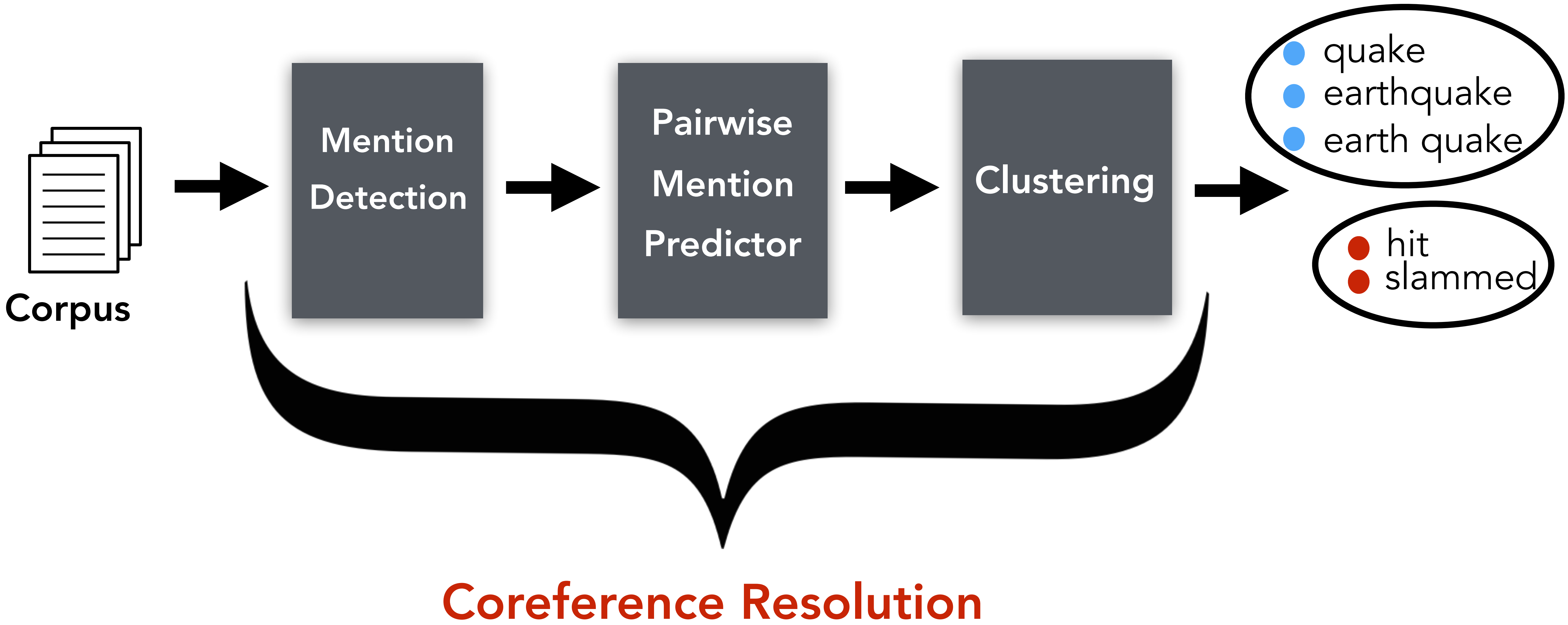
	Model	Corpus
Yang, et. al. (2015)	HDDCRP	ECB+
Choubey and Huang (2017)	FFNN	ECB+
our work	CCNN	ECB+

Wiseman, et. al. (2015)	FFNN	CoNLL-2012
Wiseman, et. al. (2016)	RNN	CoNLL-2012
Clark and Manning (2016)	FFNN	CoNLL-2012
Clark and Manning (2016)	RL	CoNLL-2012
Lee, et. al. (2017)	LSTM + FFNN	CoNLL-2012

The Big Picture for Any Coreference System (entity or event)

The Big Picture for Any Coreference System (entity or event)

The Big Picture for Any Coreference System (entity or event)



The Big Picture for Any Coreference System (entity or event)

Mention Detection: determines which words comprise a Mention

Coreference Resolution

The Big Picture for Any Coreference System (entity or event)

Mention Detection: determines which words comprise a Mention

Document 1

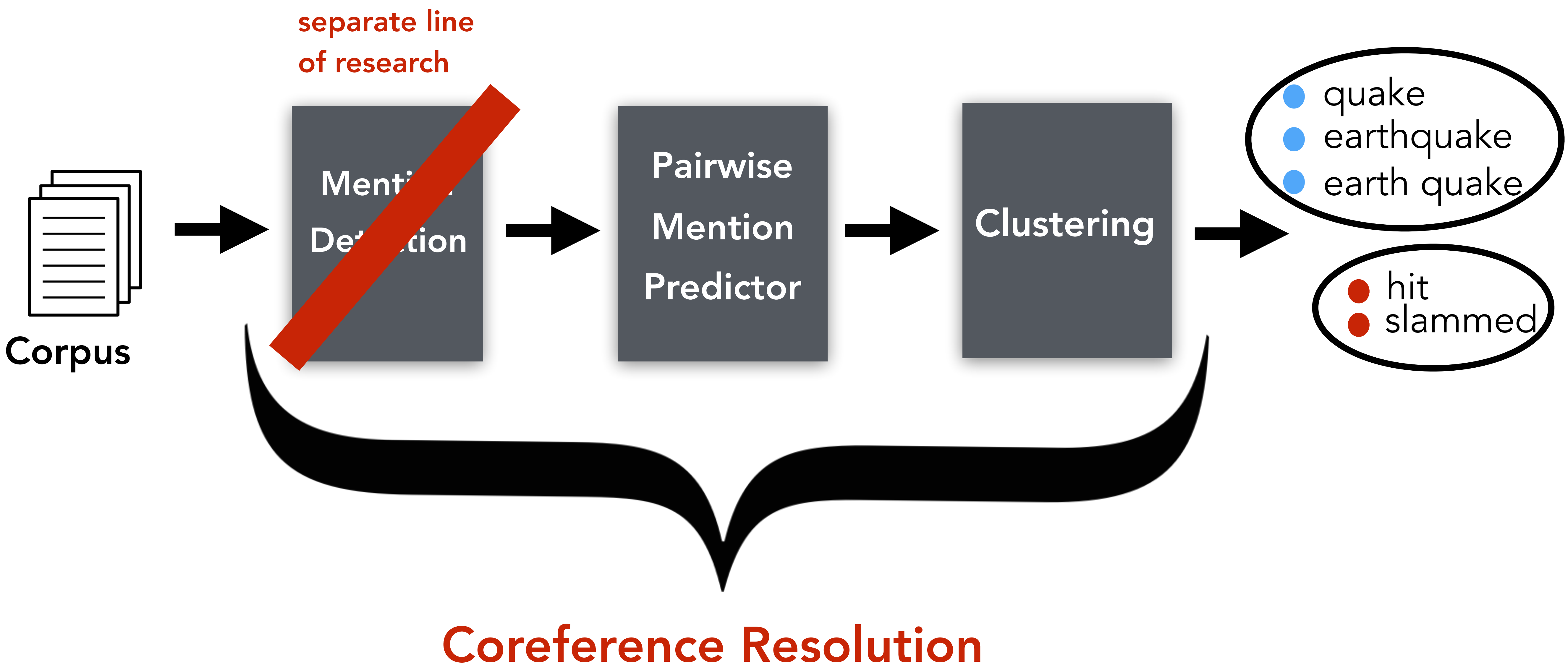
Sentence #1 $\boxed{\text{Lindsay Lohan}}$ is $\boxed{\text{MIA}}$ after $\boxed{\text{pretending}}$ to $\boxed{\text{check into}}$ $\boxed{\text{unlicensed Calif. rehab center}}$

Sentence #2 $\boxed{\text{May 2, 2013 | 5:03pm}}$

Sentence #3 ``The ‘Mean Girls’ $\boxed{\text{star}}$ $\boxed{\text{fooled}}$ an L.A Superior Court $\boxed{\text{judge}}$, a Santa Monica $\boxed{\text{prosecutor}}$ and apparently $\boxed{\text{her}}$ own $\boxed{\text{lawyer}}$, by $\boxed{\text{pretending}}$ to $\boxed{\text{check into}}$ $\boxed{\text{a rehabilitation facitility}}$, but $\boxed{\text{chickening out}}$ when $\boxed{\text{she}}$ got $\boxed{\text{there}}$.”

Coreference Resolution

The Big Picture for Any Coreference System (entity or event)

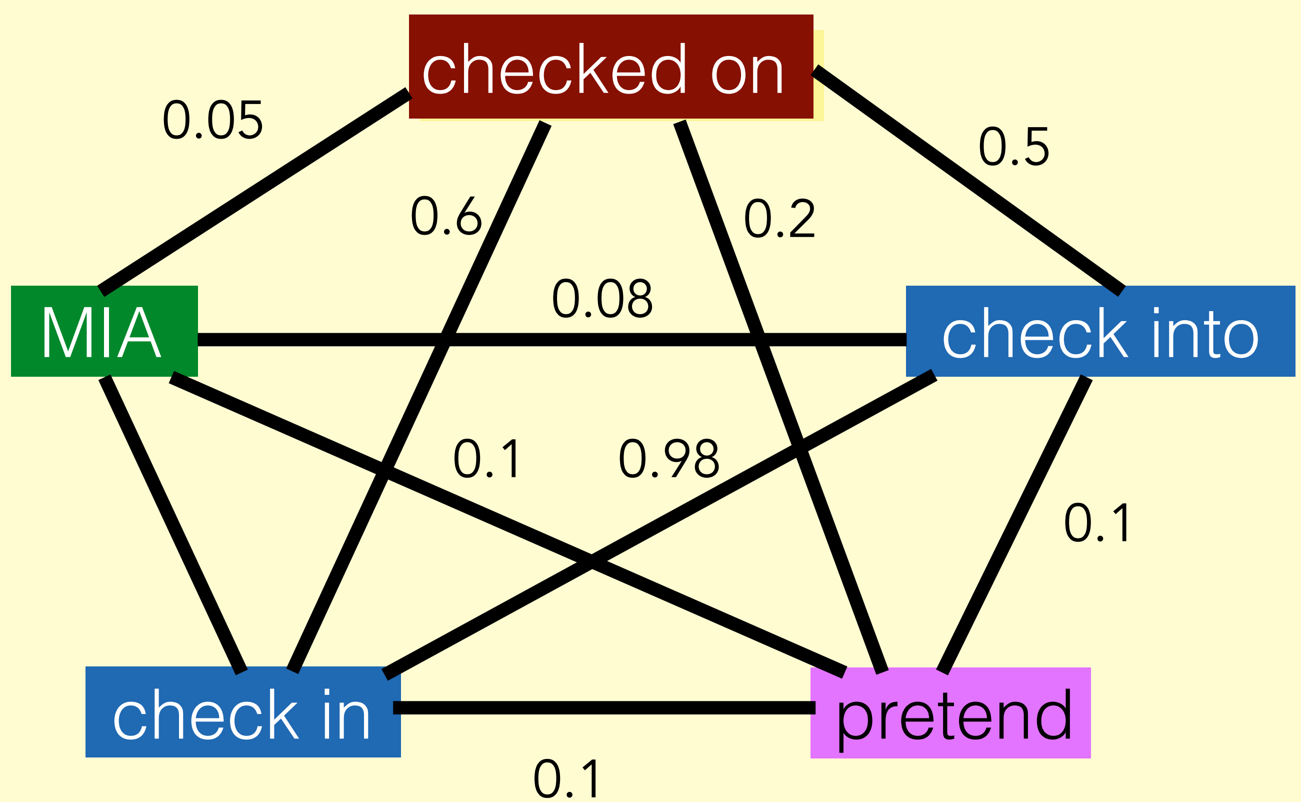


The Big Picture for Any Coreference System (entity or event)

Mention Pair Model: scores every pair of Mention in terms of likelihood of being coreferent

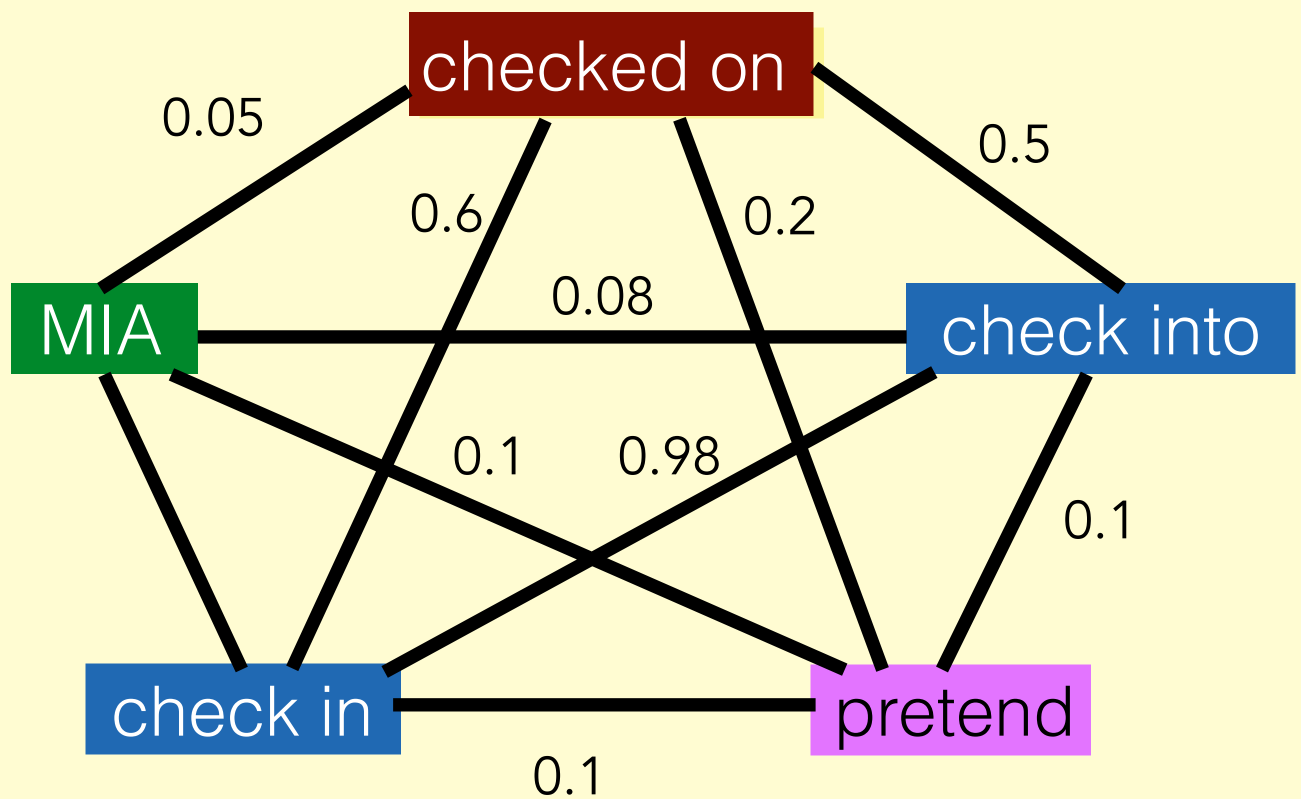
The Big Picture for Any Coreference System (entity or event)

Mention Pair Model: scores every pair of Mention in terms of likelihood of being coreferent



The Big Picture for Any Coreference System (entity or event)

Mention Pair Model: scores every pair of Mention in terms of likelihood of being coreferent



Related Work

Coreference Resolution Paradigms:

- Mention-Ranking Models
- Mention-Pair Models
- Entity-Based

Related Work

Coreference Resolution Paradigms:

- **Mention-Ranking Models**
- Mention-Pair Models
- Entity-Based

Objective: for each mention, determine its antecedent

Related Work

Coreference Resolution Paradigms:

- **Mention-Ranking Models**
- Mention-Pair Models
- Entity-Based

For each mention m_i, rank each mention m_j, where

$m_j \in \{m_1, m_2, ..., m_{i-1}, \epsilon\}$

Objective: for each mention, determine its antecedent

Related Work

Coreference Resolution Paradigms:

- **Mention-Ranking Models**
- Mention-Pair Models
- Entity-Based

?

ϵ

Lindsay Lohan is MIA after pretending to check into …

For each mention m_i, rank each mention m_j, where

$m_j \in \{m_1, m_2, ..., m_{i-1}, \epsilon\}$

Objective: for each mention, determine its antecedent

Related Work

Coreference Resolution Paradigms:

- **Mention-Ranking Models**
- Mention-Pair Models
- Entity-Based

ϵ

? ?

Lindsay Lohan is MIA after pretending to check into …

For each mention m_i, rank each mention m_j, where

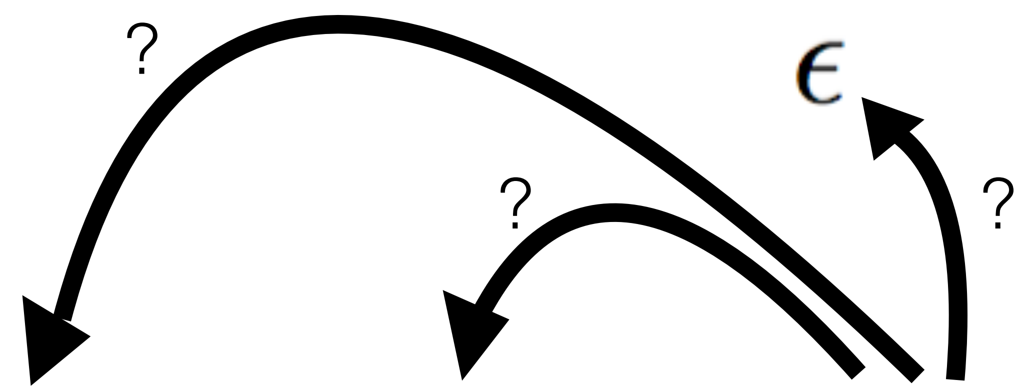
$m_j \in \{m_1, m_2, ..., m_{i-1}, \epsilon\}$

Objective: for each mention, determine its antecedent

Related Work

Coreference Resolution Paradigms:

- **Mention-Ranking Models**
- Mention-Pair Models
- Entity-Based



Lindsay Lohan is MIA after pretending to check into …

For each mention m_i, rank each mention m_j, where

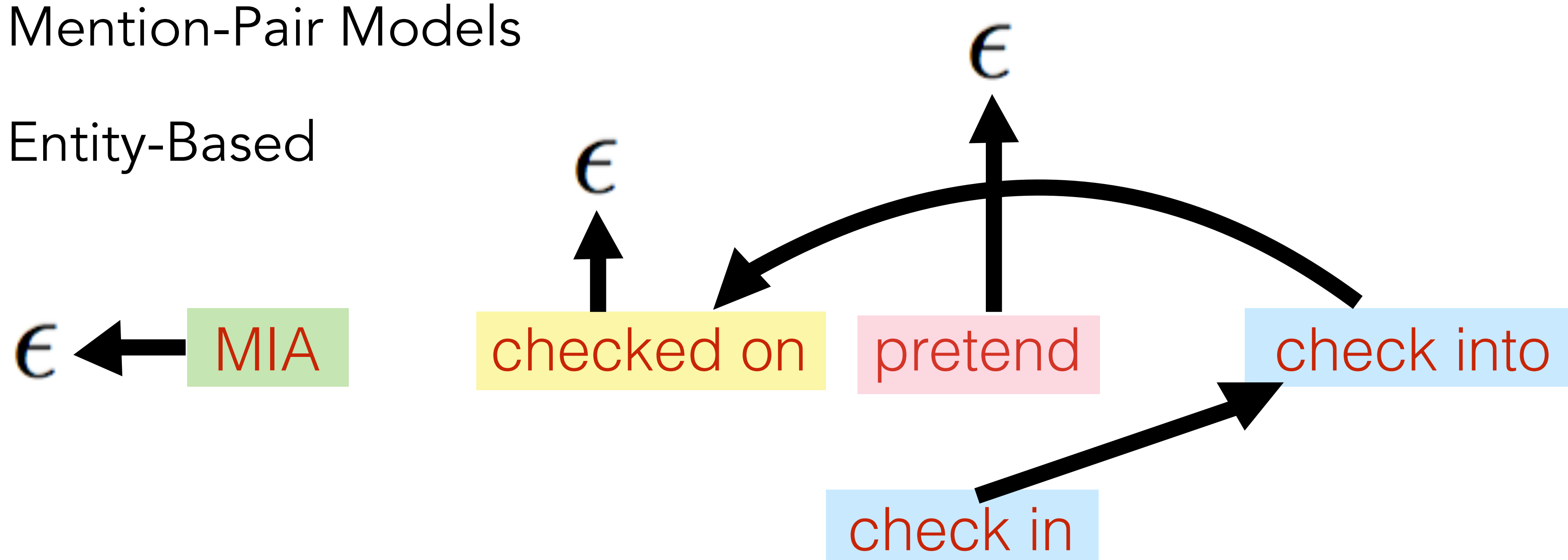
$m_j \in \{m_1, m_2, ..., m_{i-1}, \epsilon\}$

Objective: for each mention, determine its antecedent

Related Work

Coreference Resolution Paradigms:

- **Mention-Ranking Models**
- Mention-Pair Models
- Entity-Based



For each mention m_i, rank each mention m_j, where

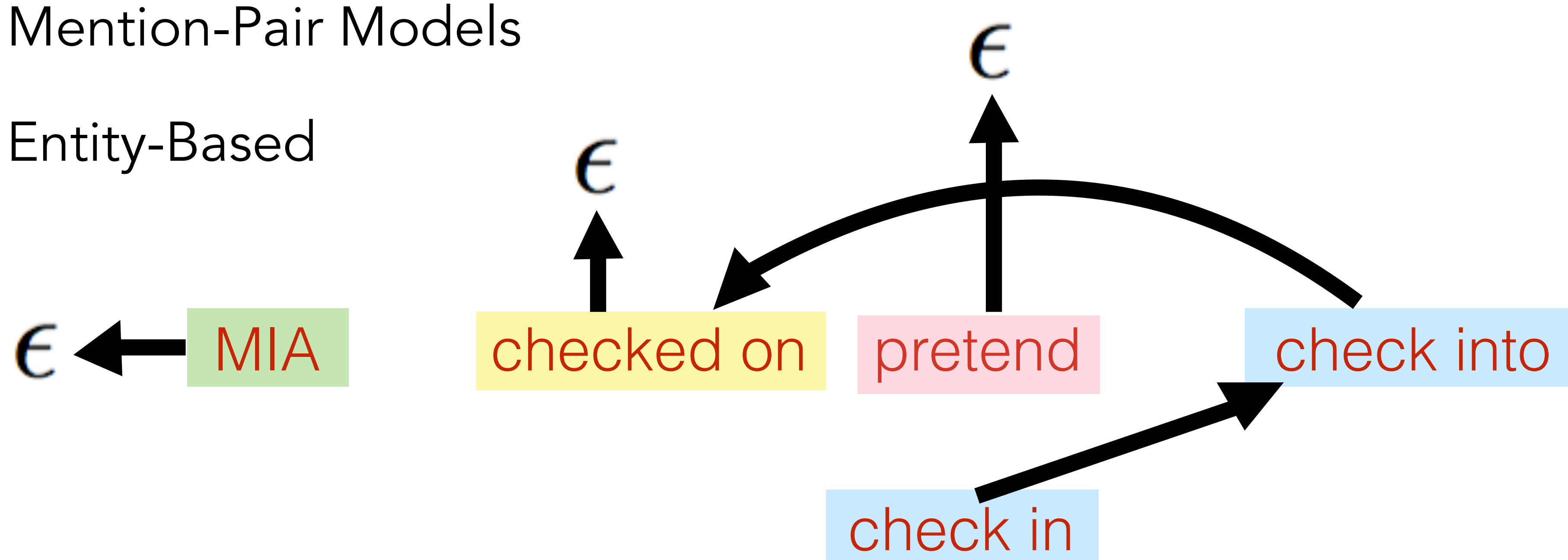
$m_j \in \{m_1, m_2, ..., m_{i-1}, \epsilon\}$

Objective: for each mention, determine its antecedent

Related Work

Coreference Resolution Paradigms:

- **Mention-Ranking Models**
- Mention-Pair Models
- Entity-Based



Weakness #1: clustering is mention-to-mention based

Related Work

Coreference Resolution Paradigms:

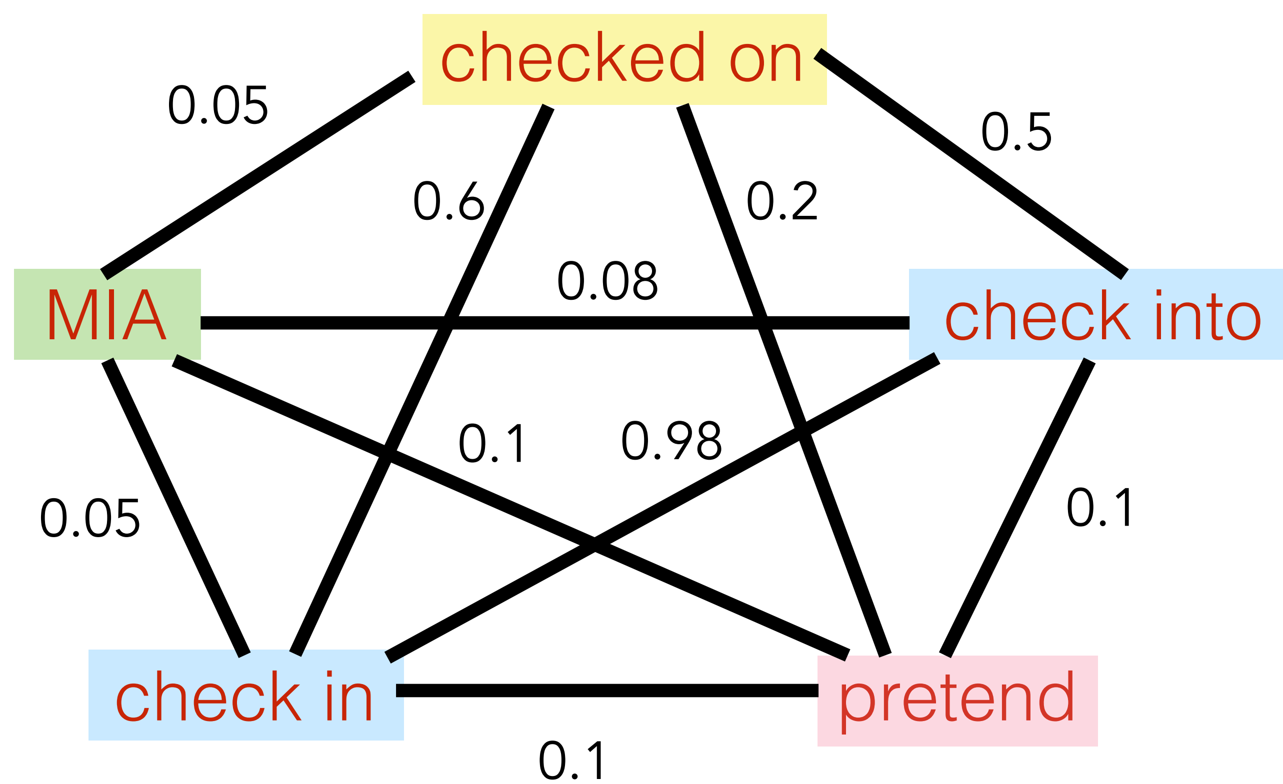
- Mention-Ranking Models
- **Mention-Pair Models**
- Entity-Based

Objective: score every pair of mentions

Related Work

Coreference Resolution Paradigms:

- Mention-Ranking Models
- **Mention-Pair Models**
- Entity-Based



Objective: score every pair of mentions

Related Work

Coreference Resolution Paradigms:

- Mention-Ranking Models
- **Mention-Pair Models**
- Entity-Based

  - Similar to mention-ranking
  - undirected
  - conducive for cross-document

Objective: score every pair of mentions

Related Work

Coreference Resolution Paradigms:

- Mention-Ranking Models
- Mention-Pair Models
- **Entity-Based**

Objective: build 1 representation for each underlying entity

Related Work

Coreference Resolution Paradigms:

- Mention-Ranking Models
- Mention-Pair Models
- **Entity-Based**

?

Objective: build 1 representation for each underlying entity

Related Work

Coreference Resolution Paradigms:

- Mention-Ranking Models
- Mention-Pair Models
- **Entity-Based**

features (coarse or sparse):

- all, most, none (e.g., 'all-female=true')
- common-noun + pronoun + pronoun

Objective: build 1 representation for each underlying entity

Related Work

Coreference Resolution Paradigms:

- Mention-Ranking Models
- Mention-Pair Models
- **Entity-Based**
    - Difficult to represent
    - How many entities to represent?
    - Doesn't make sense for events

Objective: build 1 representation for each underlying entity

Related Work

Weakness #2: Many features. Which are most useful?

Related Work

ENTITY COREF

Wiseman, et. al. (2015)

	Mention Features (ϕ_a)	
	Feature	Value Set
	Mention Head	$\mathcal{V}$
	Mention First Word	$\mathcal{V}$
	Mention Last Word	$\mathcal{V}$
	Word Preceding Mention	$\mathcal{V}$
	Word Following Mention	$\mathcal{V}$
	# Words in Mention	$\{1, 2, \ldots\}$
	Mention Synt. Ancestry	see BCS (2013)
	Mention Type	$\mathcal{T}$
+	Mention Governor	$\mathcal{V}$
+	Mention Sentence Index	$\{1, 2, \ldots\}$
+	Mention Entity Type	NER tags
+	Mention Number	{sing.,plur.,unk}
+	Mention Animacy	{an.,inan.,unk}
+	Mention Gender	{m,f,neut.,unk}
+	Mention Person	{1,2,3,unk}

	Pairwise Features (ϕ_p)	
	Feature	Value Set
	BASIC features on Mention	see above
	BASIC features on Antecedent	see above
	Mentions between Ment., Ante.	$\{0 \ldots 10\}$
	Sentences between Ment., Ante.	$\{0 \ldots 10\}$
	i-within-i	{T,F}
	Same Speaker	{T,F}
	Document Type	{Conv.,Art.}
	Ante., Ment. String Match	{T,F}
	Ante. contains Ment.	{T,F}
	Ment. contains Ante.	{T,F}
	Ante. contains Ment. Head	{T,F}
	Mention contains Ante. Head	{T,F}
	Ante., Ment. Head Match	{T,F}
	Ante., Ment. Synt. Ancestries	see above
+	BASIC+ features on Ment.	see above
+	BASIC+ features on Ante.	see above
+	Ante., Ment. Numbers	see above
+	Ante., Ment. Genders	see above
+	Ante., Ment. Persons	see above
+	Ante., Ment., Entity Types	see above
+	Ante., Ment. Heads	see above
+	Ante., Ment. Types	see above

Related Work

ENTITY COREF

Wiseman, et. al. (2016)

Features We use the raw BASIC+ feature sets described by Wiseman et al. (2015), with the following modifications:

- We remove all features from ϕ_{p} that concatenate a feature of the antecedent with a feature of the current mention, such as bi-head features.
- We add true-cased head features, a current speaker indicator feature, and a 2-character genre (out of {bc,bn,mz,nw,pt,tc,wb}) indicator to ϕ_{p} and ϕ_{a}.
- We add features indicating if a mention has a substring overlap with the current speaker (ϕ_{p} and ϕ_{a}), and if an antecedent has a substring overlap with a speaker distinct from the current mention's speaker (ϕ_{p}).
- We add a single centered, rescaled document position feature to each mention when learning $\boldsymbol{h}_{\mathrm{c}}$. We calculate a mention x_n's rescaled document position as $\frac{2n-N-1}{N-1}$.

Related Work

ENTITY COREF

Clark and Manning (2016)

Embedding Features: Word embeddings of the head word, dependency parent, first word, last word, two preceding words, and two following words of the mention. Averaged word embeddings of the five preceding words, five following words, all words in the mention, all words in the mention's sentence, and all words in the mention's document.

Additional Mention Features: The type of the mention (pronoun, nominal, proper, or list), the mention's position (index of the mention divided by the number of mentions in the document), whether the mentions is contained in another mention, and the length of the mention in words.

Document Genre: The genre of the mention's document (broadcast news, newswire, web data, etc.).

Distance Features: The distance between the mentions in sentences, the distance between the mentions in intervening mentions, and whether the mentions overlap.

Speaker Features: Whether the mentions have the same speaker and whether one mention is the other mention's speaker as determined by string matching rules from Raghunathan et al. (2010).

String Matching Features: Head match, exact string match, and partial string match.

The vectors for all of these features are concatenated to produce an I-dimensional vector $\boldsymbol{h}_0$, the input to the neural network. If $a = \text{NA}$, the features defined over mention pairs are not included. For this case, we train a separate network with an identical architecture to the pair network except for the input layer to produce anaphoricity scores.

Related Work

EVENT COREF

Yang, et. al. (2015)

where ψ is a feature vector, containing a rich set of features based on event mentions i and j: (1) head word string match, (2) head POS pair, (3) cosine similarity between the head word embeddings (we use the pre-trained 300-dimensional word embeddings from word2vec[1]), (4) similarity between the words in the event mentions (based on term frequency (TF) vectors), (5) the Jaccard coefficient between the WordNet synonyms of the head words, and (6) similarity between the context words (a window of three words before and after each event mention). If both event mentions involve participants, we consider the similarity between the words in the participant mentions based on the TF vectors, similarly for the time mentions and the location mentions. If the SRL role information is available, we also consider the similarity between words in each SRL role, i.e. Arg0, Arg1, Arg2.

Related Work

EVENT COREF

Yang, et. al. (2015)

Weakness #3: Manually-defining relational features

where ψ is a feature vector, containing a rich set of features based on event mentions i and j: (1) head word string match, (2) head POS pair, (3) cosine similarity between the head word embeddings (we use the pre-trained 300-dimensional word embeddings from word2vec[1]), (4) similarity between the words in the event mentions (based on term frequency (TF) vectors), (5) the Jaccard coefficient between the WordNet synonyms of the head words, and (6) similarity between the context words (a window of three words before and after each event mention). If both event mentions involve participants, we consider the similarity between the words in the participant mentions based on the TF vectors, similarly for the time mentions and the location mentions. If the SRL role information is available, we also consider the similarity between words in each SRL role, i.e. Arg0, Arg1, Arg2.

Current Work

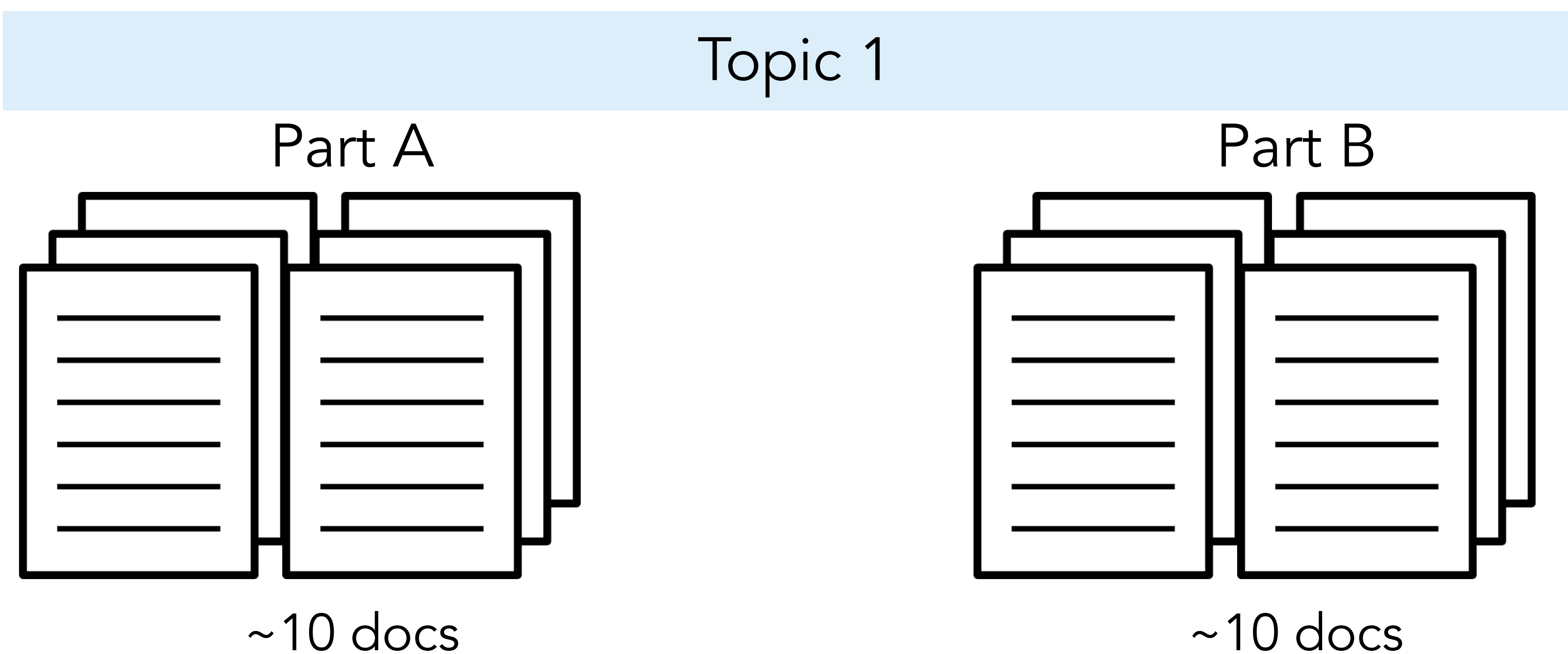
Event Coreference Corpora:

- ECB (482 docs)
- EECB (added entities to ECB)
- ECB+ (added entities to ECB plus doubled size)

Current Work

Event Coreference Corpora:

- ECB (482 docs)
- EECB (same but added entities)
- **ECB+ (982 docs; ECB plus another half)**
  - 43 distinct "topics"



Current Work

ECB+ (just the events)

- # docs: 982
- 1,840 annotated sentences (from 15,812)
- 1,826 unique entities/events (referents)
    - 399 singletons (22%) and 1,427 non-singletons (78%)
- # mentions: 11,957
    - train: 2,848 entities and 2,117 events
    - dev: 480 entities and 327 events
    - test: 3,571 entities and 2,614 events

Current Work

ECB+ (just the events)

4.6 Magnitude Quake Recorded in Sonoma County

~~Thursday, March 14, 2013~~

An earthquake with a preliminary magnitude of 4.6 was recorded in the North Bay this morning, according to the U.S. Geological Survey. The quake occurred at 2:09 a.m. about 14 miles north-northeast of Healdsburg and had a depth of 1.2 miles. ~~It was followed by a 2.9 aftershock at 2:12 a.m. and a 2.2 at 2:15 a.m.… there are no reports of injuries or major damage.~~

Doc 1

4.6 Magnitude Quake Rattles Sonoma County Early Thursday

~~Posted: 03/14/2013 06:37:46 AM PDT~~

~~Updated: 03/14/2013 07:51:21 AM PDT~~

An earthquake measuring 4.6 rattled Sonoma and Lake counties early Thursday, according to the U.S. Geological Survey. ~~The quake occurred at 2:09 a.m., about 14 miles northeast of Healdsburg, on the Maacama Fault with a depth of 12 miles. A Sonoma County Sheriff's dispatcher said around 7 a.m. that there had been no reports of damage or injuries.~~

Doc 2

Current Work

Features

- Word Embeddings (300 dimensions)
- Lemma Embeddings (300 dimensions)
- Character Embeddings (20 dimensions * 20 max length)
- POS Tag Embeddings (100 dimensions)
- Dependency Parents'/Children's Lemma Embeddings (300 dim)
- WordNet Similarity
- Bag-of-Words (summed Lemma Embeddings)

Current Work

Features

m_1

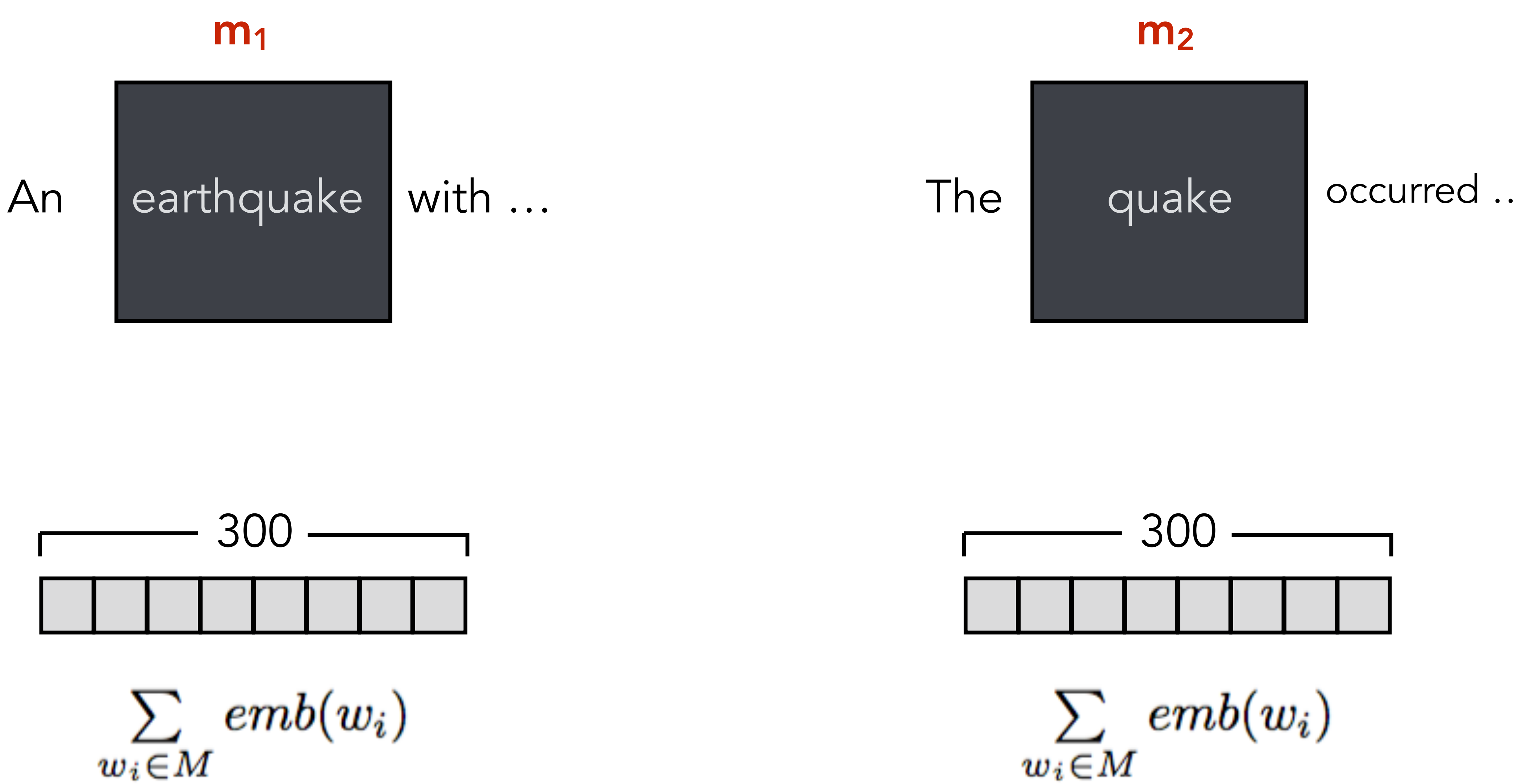
An earthquake with …

m_2

The quake occurred …

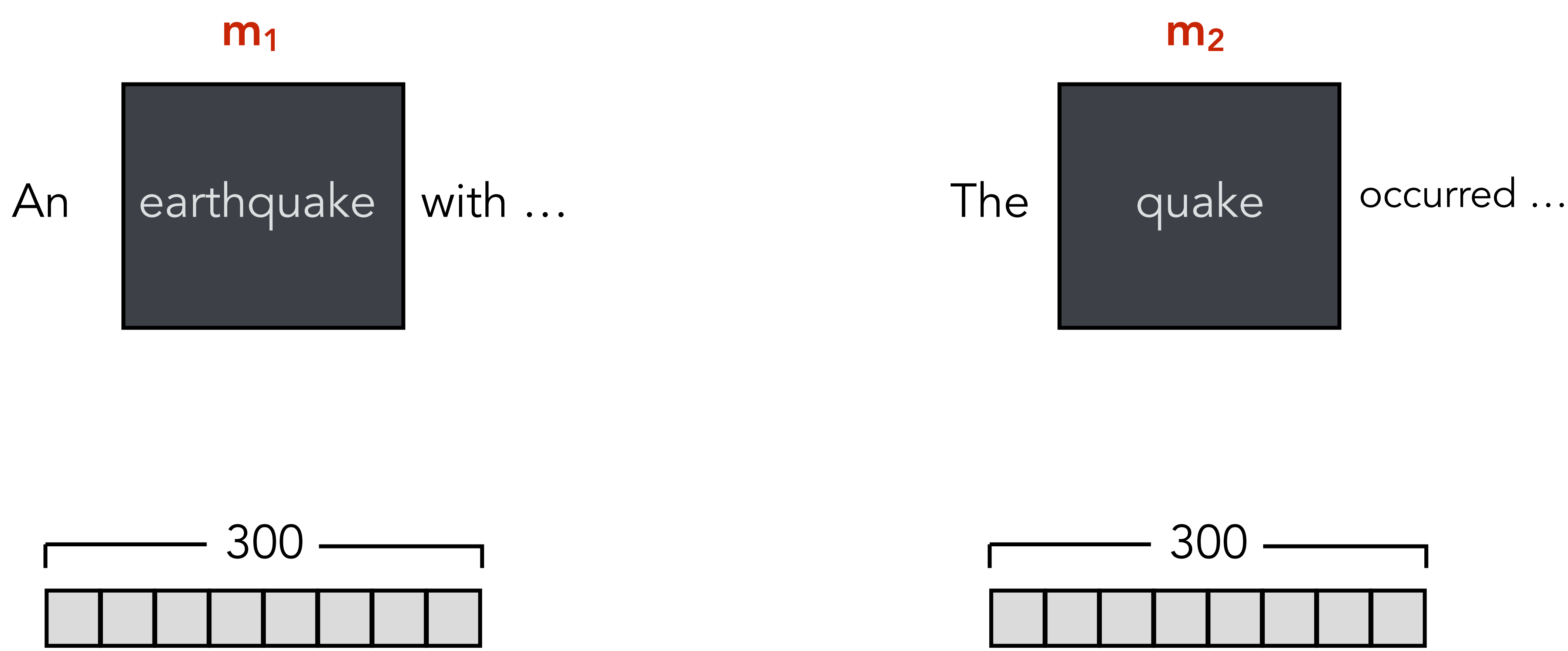
Current Work

Feature: Word Embeddings



Current Work

Feature: Lemma Embeddings

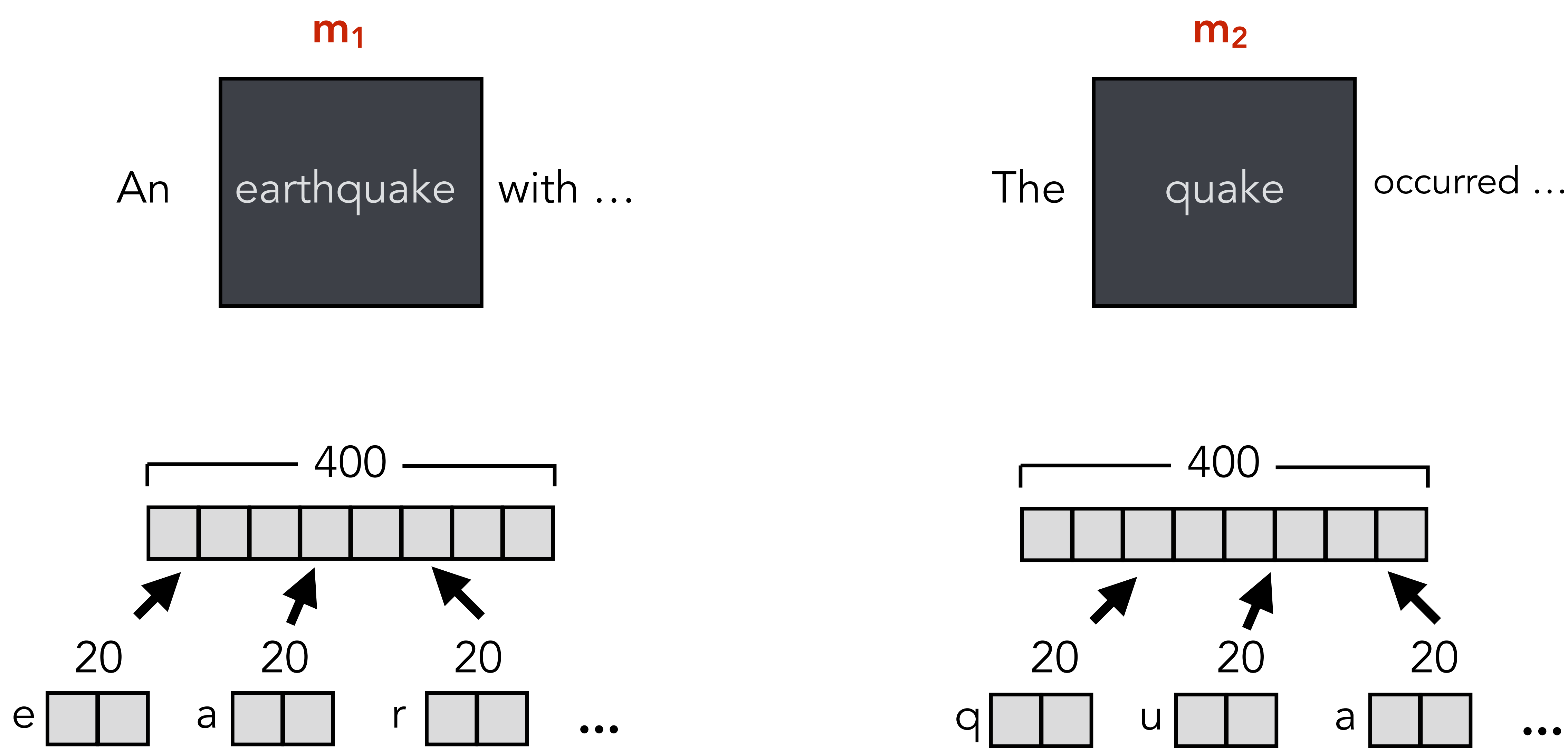


$$\sum_{w_i \in M} emb(lemma(w_i))$$

$$\sum_{w_i \in M} emb(lemma(w_i))$$

Current Work

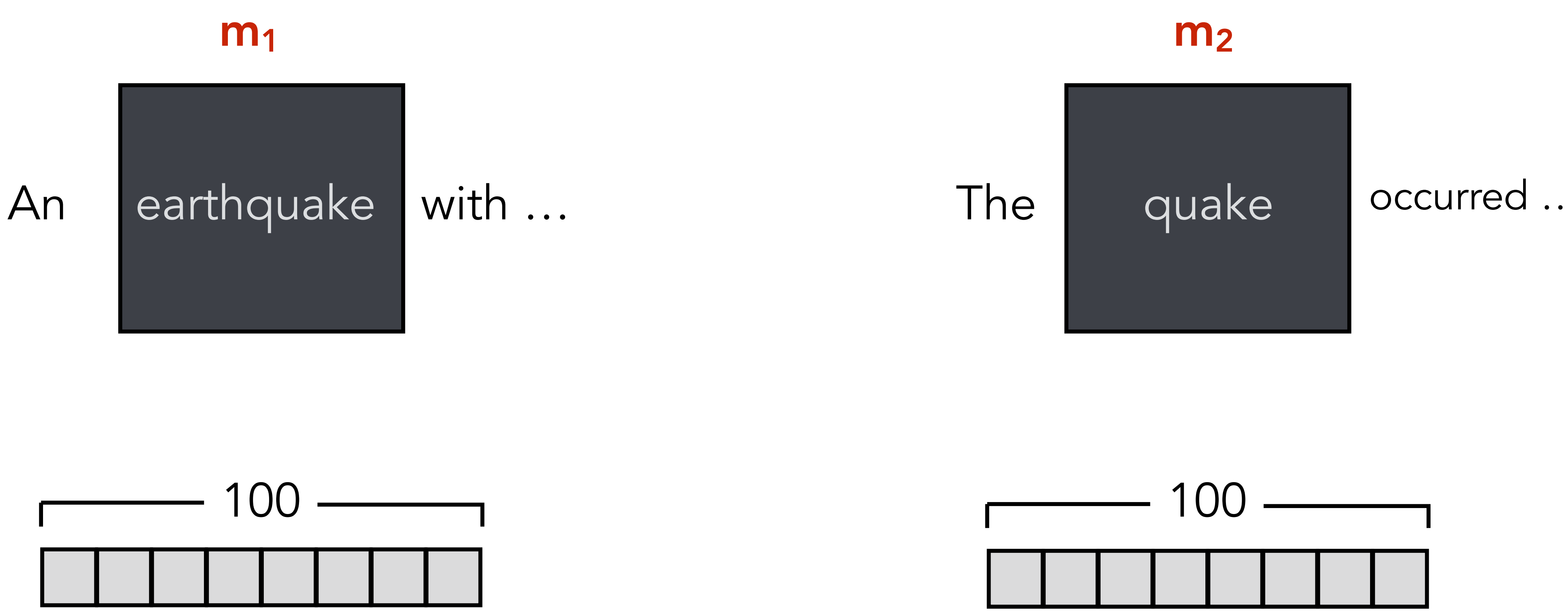
Feature: Character Embeddings



$emb(m_{c_1}) \oplus emb(m_{c_2}) \oplus \cdots \oplus emb(m_{c_n})$ $emb(m_{c_1}) \oplus emb(m_{c_2}) \oplus \cdots \oplus emb(m_{c_n})$

Current Work

Feature: POS Tag Embeddings

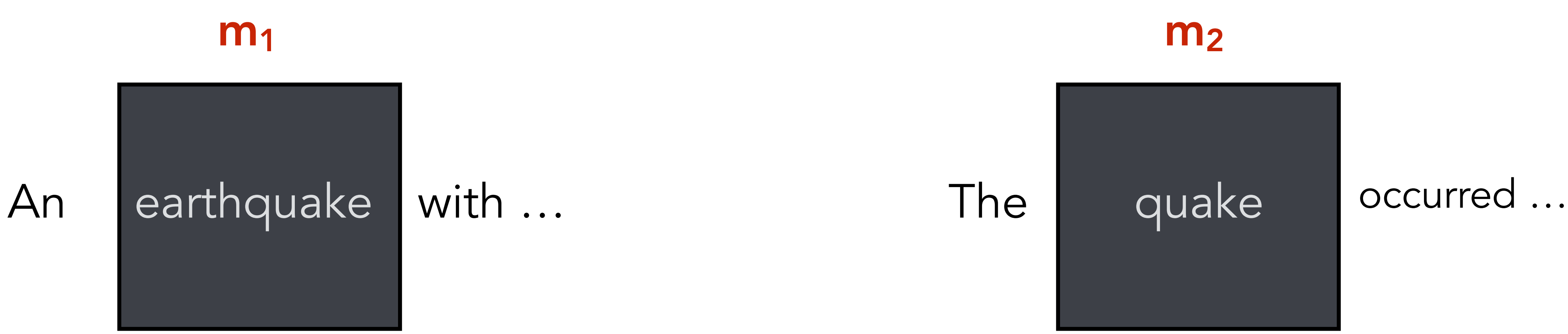


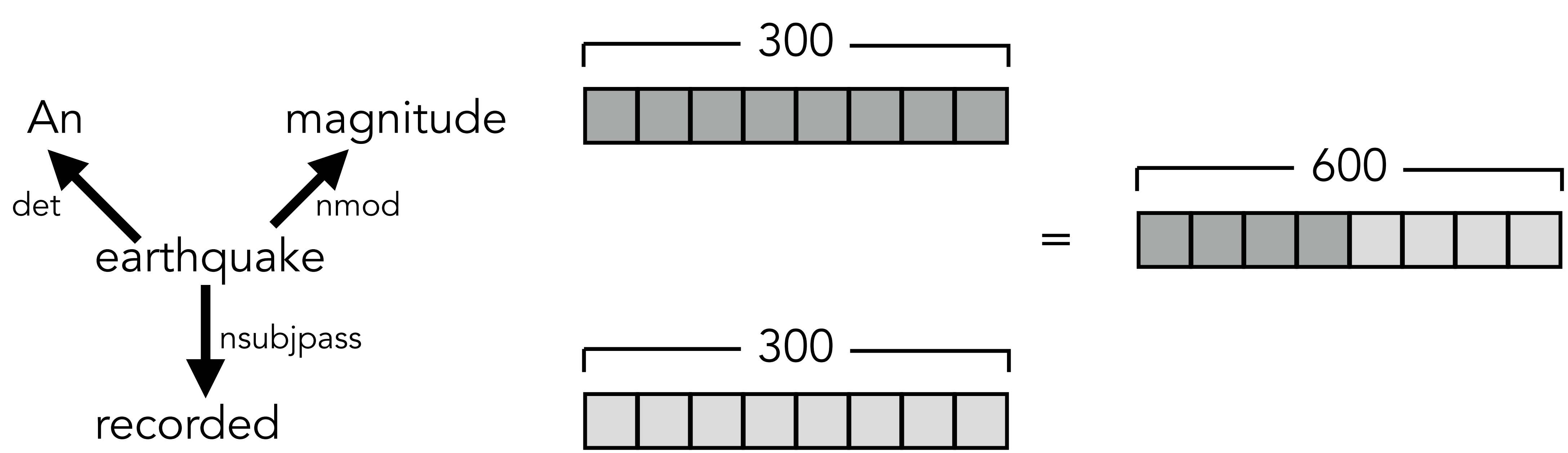
$$\sum_{w_i \in M} emb(POS(w_i))$$

$$\sum_{w_i \in M} emb(POS(w_i))$$

Current Work

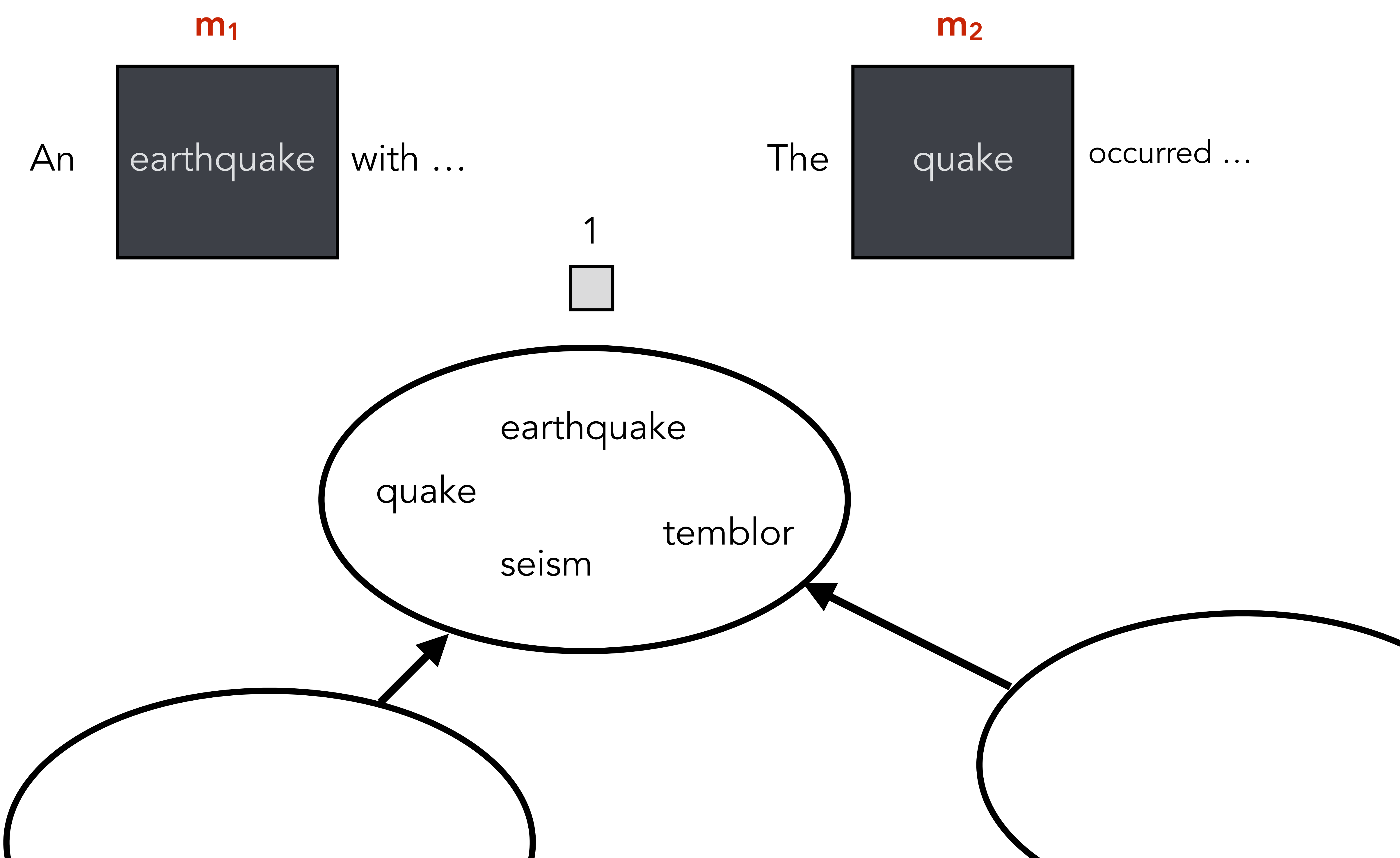
Feature: Dependency Parents'/Children's Lemma Embeddings





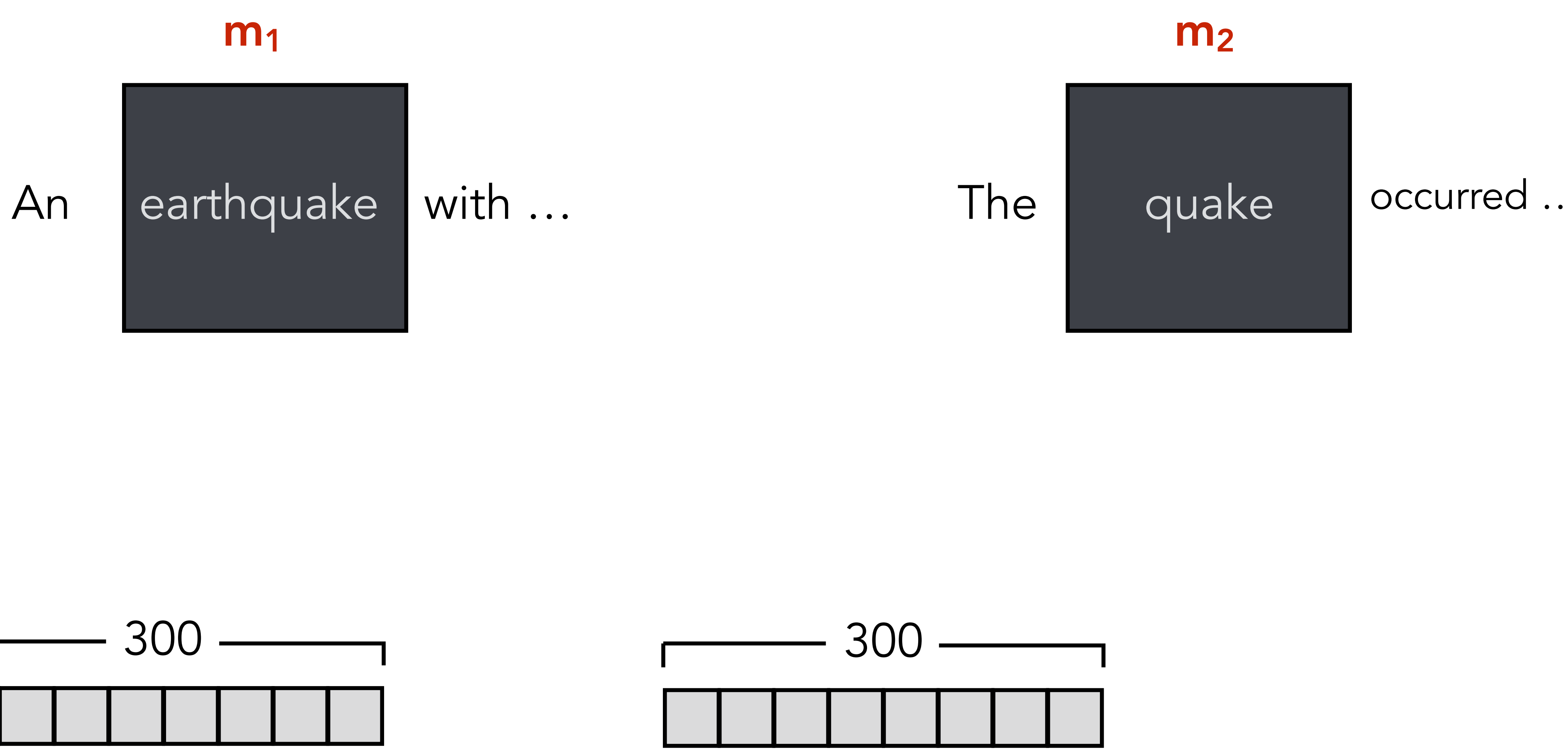
Current Work

Feature: WordNet Similarity



Current Work

Feature: Bag-of-Words Embeddings



$$\sum_{w_i \in M_{pre}} emb(lemma(w_i)) \oplus \sum_{w_i \in M_{post}} emb(lemma(w_i))$$

Current Work

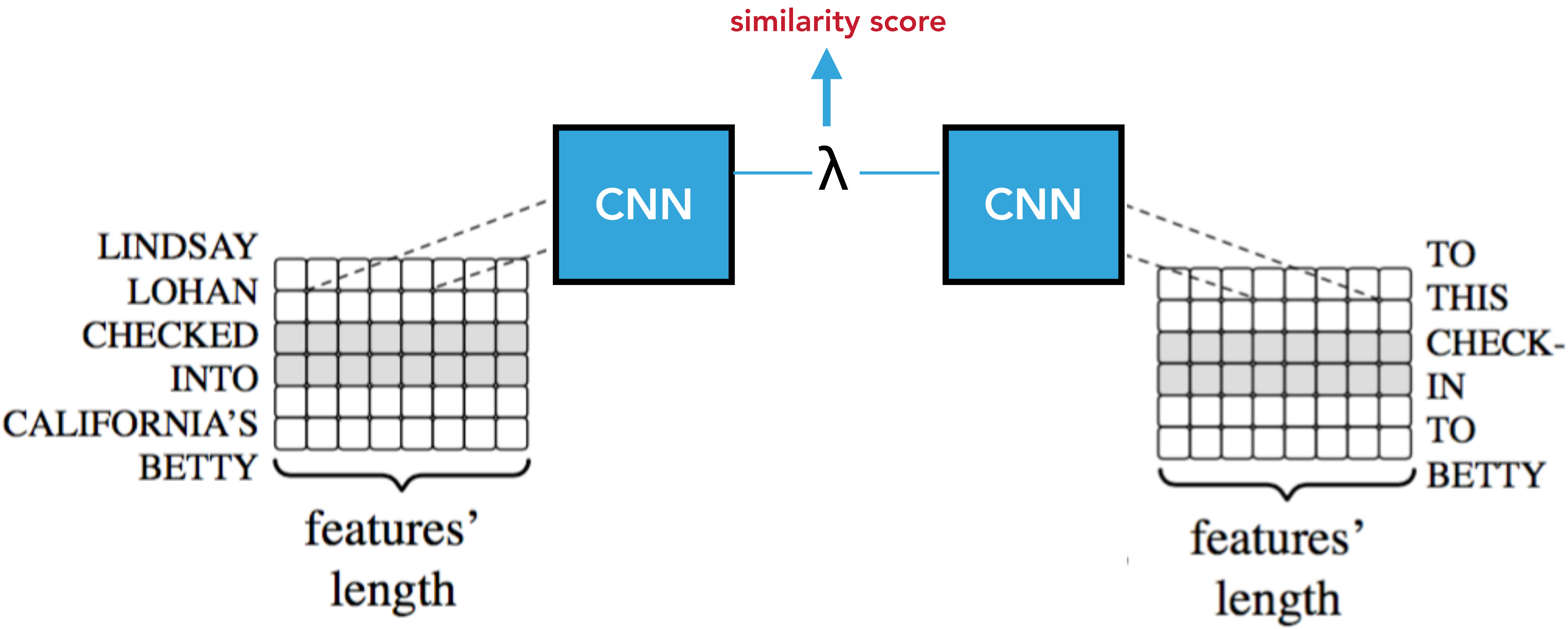
Model

We want

- mention-pair model
- automatically learns relational information
- is symmetric/agnostic w.r.t mention order
- fewest features as possible

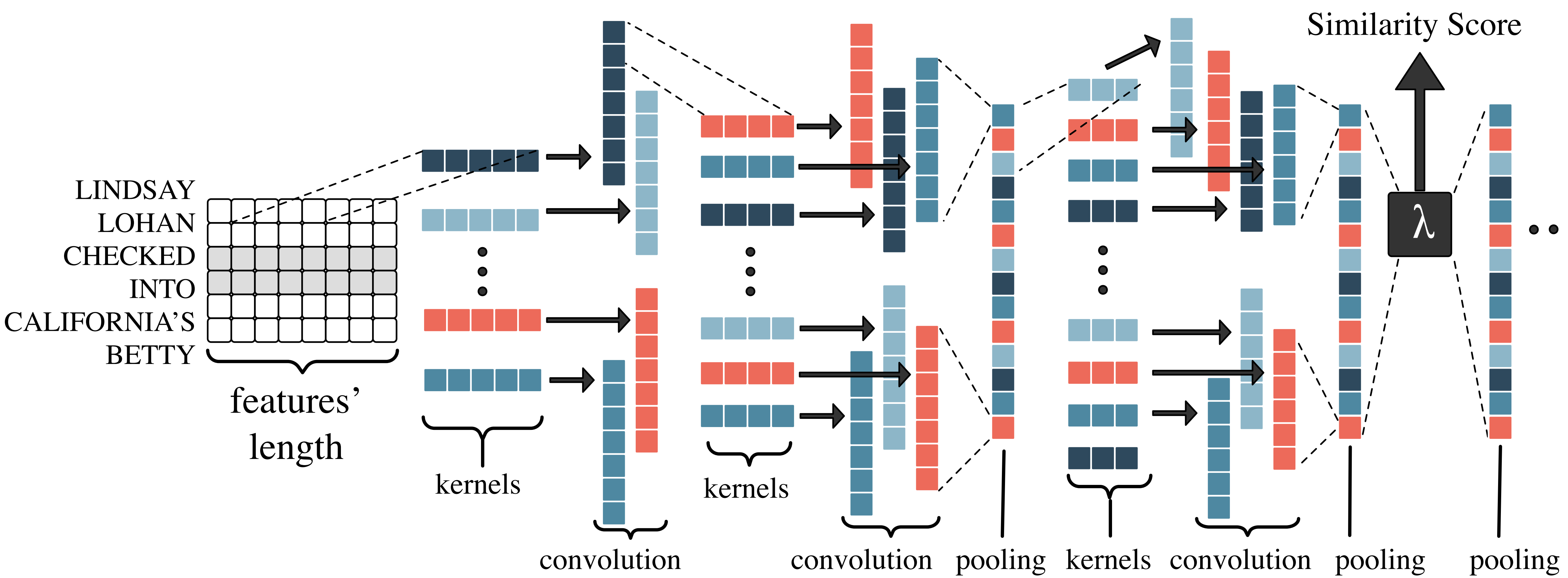
Current Work

Model: Conjoined Convolutional Neural Network



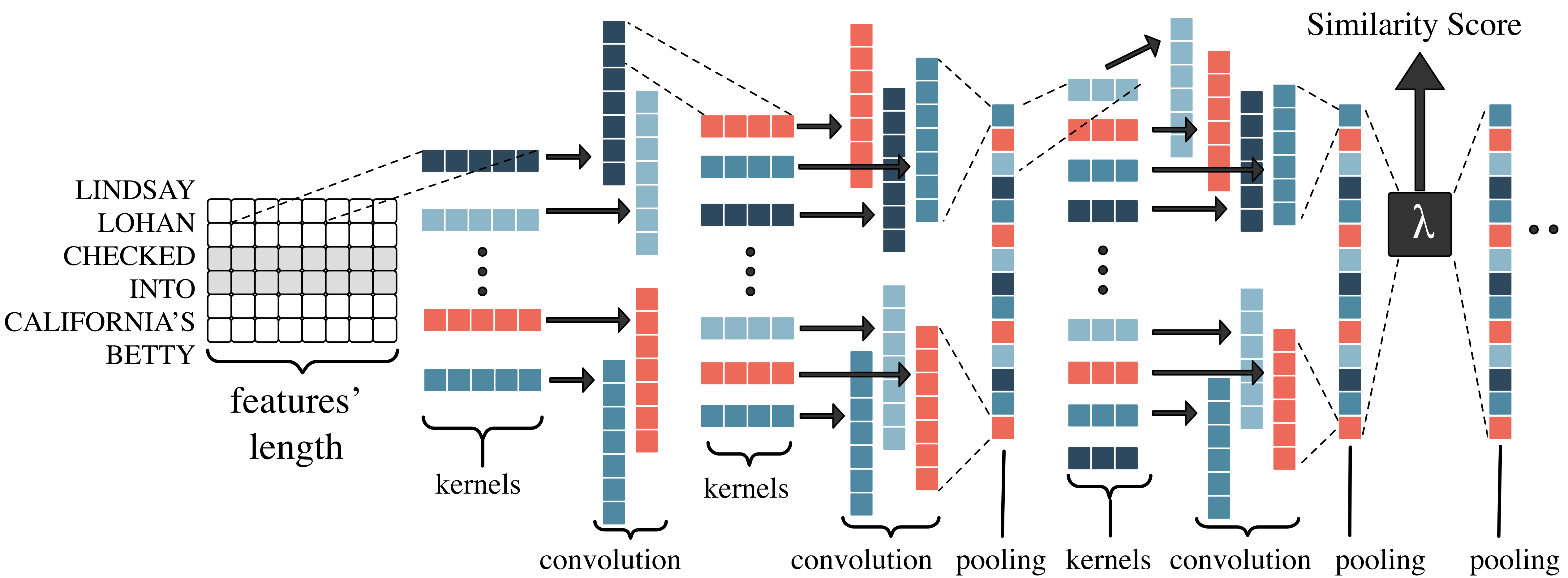
Current Work

Model: Conjoined Convolutional Neural Network



Current Work

Model: Conjoined Convolutional Neural Network



Distance Function: L^2 norm

Loss Function: Contrastive Loss

$$(1-Y)\frac{1}{2}(D_W)^2 + (Y)\frac{1}{2}\{max(0, m - D_W)\}^2$$

Current Work

Pairwise Evaluation (Within-doc)

m_{17}, m_2	1.0	erupted	erupted
m_{17}, m_4	1.0	erupted	erupted
m_5, m_{923}	0.97	announced	announce
m_{78}, m_{57}	0.95	erupt	erupted
⋮		⋮	⋮
m_{801}, m_{39}	0.03	revealed	broke into
m_{26}, m_{48}	0.02	handed down	confirmed

accuracy: **92.4**

precision: **64.0**

recall: **67.9**

f1: **65.9**

Current Work

Pairwise Evaluation

	Within-Doc	Cross-Doc
SameLemma	-	-
LibSVM	52.1	53.2
FFNN	62.1	64.8
CCNN	65.9	66.3

LibSVM and FFNN received same features as CCNN,

plus relational features (e.g., cosine sim., dot-product, WordNet)

Current Work

Pairwise Evaluation — Understanding CCNN Results vs Gold Truth

True-Positives:

False-Positives:

True-Negatives:

False-Negat [illegible] es):

Current Work

Pairwise Evaluation — Understanding CCNN Results vs FFNN's

CCNN (true-positive), FFNN (false-negative) — count:

CCNN (true-negative), FFNN (false-positive) — count:

Current Work

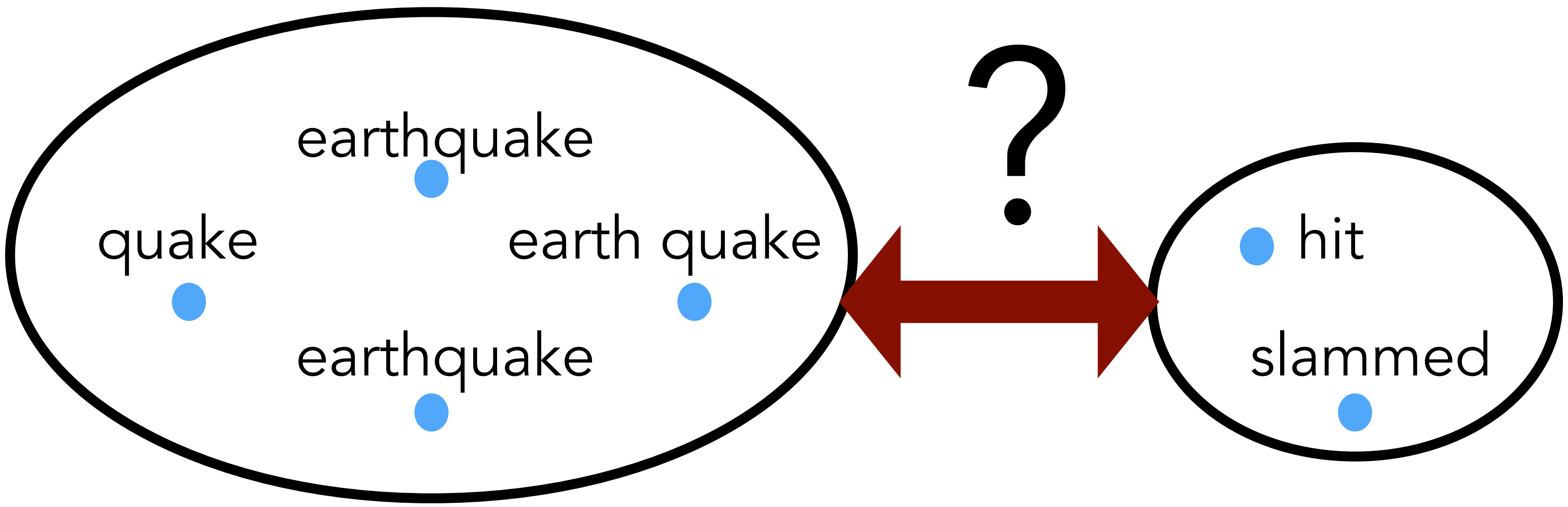
Clustering

We want

- not mention-to-mention based
- less susceptible to non-uniform predictions across topics
- no additional stopping parameter
- prevention against an all-subsuming cluster

Current Work

Neural Clustering



Simple Features:

- min-pair distance: $\min_{m_i, m_j} d(m_i, m_j)$
- avg-pair distance: $\frac{\sum_{m_i, m_j} d(m_i, m_j)}{\|C_x\| \|C_y\|}$
- max-pair distance: $\max_{m_i, m_j} d(m_i, m_j)$
- size of candidate cluster: $\frac{\|C_x\| + \|C_y\|}{\sum_z \|C_z\|}$

Construct Training Data:

- randomly sample a gold cluster
- positive: 2 random-sized subsets
- negative: 1 subset, 1 set from a *different* cluster

Current Work

Neural Clustering — Evaluation Metrics

- **MUC** (Vilain, et. al. 1995)

  minimum # of link modifications required to make the predicted cluster equal gold cluster

- **B^3** (Bagga and Baldwin, 1998)

  accuracy of all individual mentions, w.r.t. mapping them to an event

- **CEAF** (Luo, 2005)

  unlike B^3, doesn't use any event more than once; uses best possible alignment b/w events (computationally expensive)

- **CoNLL F1**

  the average F1 score of {MUC F1, B^3 F1, CEAF F1}

Current Work

Neural Clustering

ECB+ Development Performance

POS	■		■	■	■	■	■	■	■	■	■	■					■	■	■	■							■				
Lemma	■	■		■	■	■	■	■	■				■	■	■		■				■	■	■					■			
Dependency	■	■	■		■	■	■			■	■		■	■		■		■			■			■	■				■		
Character	■	■	■	■		■		■		■		■	■		■	■			■			■		■		■				■	
Word	■	■	■	■	■				■		■	■		■	■	■				■			■		■	■					■
Within-Doc	79.9	79.6	78.2	79.9	79.1	78.9	78.6	**83.2**	81.0	76.6	77.1	78.5	80.7	80.6	**82.6**	77.8	78.5	79.3	81.0	78.1	82.3	**84.9**	82.8	79.6	77.8	**83.7**	77.2	**84.5**	80.6	**83.8**	81.9
Cross-Doc	72.0	73.8	73.5	77.5	73.2	73.7	73.2	**78.0**	76.6	72.0	73.2	76.5	72.6	75.5	**78.5**	71.5	75.7	73.4	78.4	76.6	74.2	**79.5**	75.0	74.1	71.4	**78.4**	74.7	**79.0**	63.2	**75.6**	74.9

■ = Feature Used

= Best Performances

Lemma + Character Embeddings yield the best performance

Current Work

Neural Clustering — Evaluation on Gold Truth Mentions

	Within-Document				Cross-Document			
	MUC	B^3	CEAF	CoNLL F1	MUC	B^3	CEAF	CoNLL F1
Test Set: ECB+ Gold Mentions								
SameLemma	58.3	83.0	75.9	72.4	84.2	68.2	48.0	66.8
FFNN+AGG	59.9	85.6	78.4	74.6	77.7	69.9	50.1	65.9
FFNN+NC	60.7	86.7	79.4	75.6	74.9	67.8	56.3	67.0
CCNN+AGG	70.5	89.1	83.5	81.0	84.1	70.7	55.5	70.1
CCNN+NC	70.9	88.9	83.6	**81.2**	86.4	71.7	59.1	**72.4**

Current Work

Neural Clustering — Evaluation against state-of-the-art

Gold Test Mentions:

… as Peter Capaldi stepped into Matt Smith's soon to be vacant …

HDDCRP's Predicted Test Mentions:

… as Peter Capaldi stepped into Matt Smith's soon to be vacant …

token precision = 1.0
token recall = 0.5

Current Work

Neural Clustering — Evaluation against state-of-the-art

HDDCRP's Predicted Test Mentions:

	prec	recall	f1
events	**0.98**	**0.8**	**0.88**
entities	0	0	0
events + entities	0.99	0.24	0.39

measured against the gold test set:

3,571 entity mentions

2,614 event mentions

Current Work

Neural Clustering — Evaluation against state-of-the-art

	Within-Document				Cross-Document			
	MUC	B^3	CEAF	CoNLL F1	MUC	B^3	CEAF	CoNLL F1
Test Set: HDDCRP's Predicted Mentions								
SameLemma	40.4	66.4	66.2	57.7	66.7	51.4	46.2	54.8
HDDCRP	53.4	75.4	71.7	66.8	73.1	53.5	49.5	58.7
CCNN+NC	54.0	75.5	72.2	**67.2**	71.3	57.0	49.6	**59.3**

Current Work

Neural Clustering — Evaluation against state-of-the-art

```
  So, we did an additional post-processing (i believe even Bishan did this,
  though I am not sure) in which we also removed those predicted clusters
  which don't have any event mention in annotated coreference clusters. I


On Sun, Oct 22, 2017 at 9:21 PM, [redacted]> wrote=
:

> No, I am not sure about what she did. Are you able to run her code?
> Another guy was assisting me in running her code and he couldn't run the
> code after spending a month. Are you using the Perl scorer or mine one? I
> haven't tested my scorer on ECB+ corpus and I am not sure if that's corre=
ct
> :P.
```

Current Work

Neural Clustering — Evaluation against state-of-the-art

	Within-Document				Cross-Document			
	MUC	B^3	CEAF	CoNLL F1	MUC	B^3	CEAF	CoNLL F1
Test Set: HDDCRP's Predicted Mentions								
SameLemma	40.4	66.4	66.2	57.7	66.7	51.4	46.2	54.8
HDDCRP	53.4	75.4	71.7	66.8	73.1	53.5	49.5	58.7
CCNN+NC	54.0	75.5	72.2	**67.2**	71.3	57.0	49.6	**59.3**
Test Set: Choubey's et. al. Mentions								
SameLemma	48.8	66.7	65.1	60.2	68.1	53.3	47.2	56.2
Choubey	62.6	72.4	71.8	68.9	73.4	80.4	56.5	63.6
CCNN+NC	67.3	73.0	69.5	**69.9**	77.0	56.3	60.2	**64.5**

Current Work

Neural Clustering — Understanding Results

Current Work

Event Coreference — Lessons Learned

- event corpora are challenging
  - size
  - comparisons to other systems
- vital to do WD then CD
- worth exploring smarter models for including semantics and relations:
  - entities
  - dependencies
  - paraphrases?

Proposed Work

Short Term (~2 weeks)

Possible Improvements

- EMNLP deadline in 6 days
- Ensemble Approach (FFNN average scores)
- Self-training

Proposed Work

Let's add entities!

1. How well do we do on Entities?

Mention Detection

- Stanford Core NLP
- NLTK
- Bi-LSTM trained on ECB+

Entity Coreference (with and without pronouns)

- Stanford Core NLP
- Our CCNN (try other features)

Corpora

- ECB+
- EECB
- CoNLL-2012
- NIST's TAC — Nov 13, 2018 (*Entity Discovery and Linking Workshop*)

Proposed Work

Joint Entities and Events

One of the key suspected Mafia bosses arrested yesterday has hanged himself.

Police said **Lo Presti** had hanged himself.

His suicide appeared to be related to clan feuds.

Entities can help resolve Events

Proposed Work

Joint Entities and Events

The New Orleans Saints placed **Reggie Bush** on the injured list on Wednesday.

Saints put **Bush** on I.R.

Events can help resolve Entities

Proposed Work

Past Work on Joint Entities and Events

- **Haghighi and Klein (2010)** include a feature which concerns the governor of the head of the nominal mentions (which could be an event)
- **Rahman and Ng (2011)** use semantic roles of entity mentions, plus the actual verbs, as features
- **Choubey and Huang (2017)** uses hand-made relational rule to determine if two event clusters should be merged
- **Lee, et. al. (2012)** performs Entity Coref via StanfordCoreNLP then builds up Event clusters based on semantic role features.

entity coreference

event coreference

Proposed Work

Past Work o

- **Haghighi a**

  governor o

  event)

- **Rahman ar**

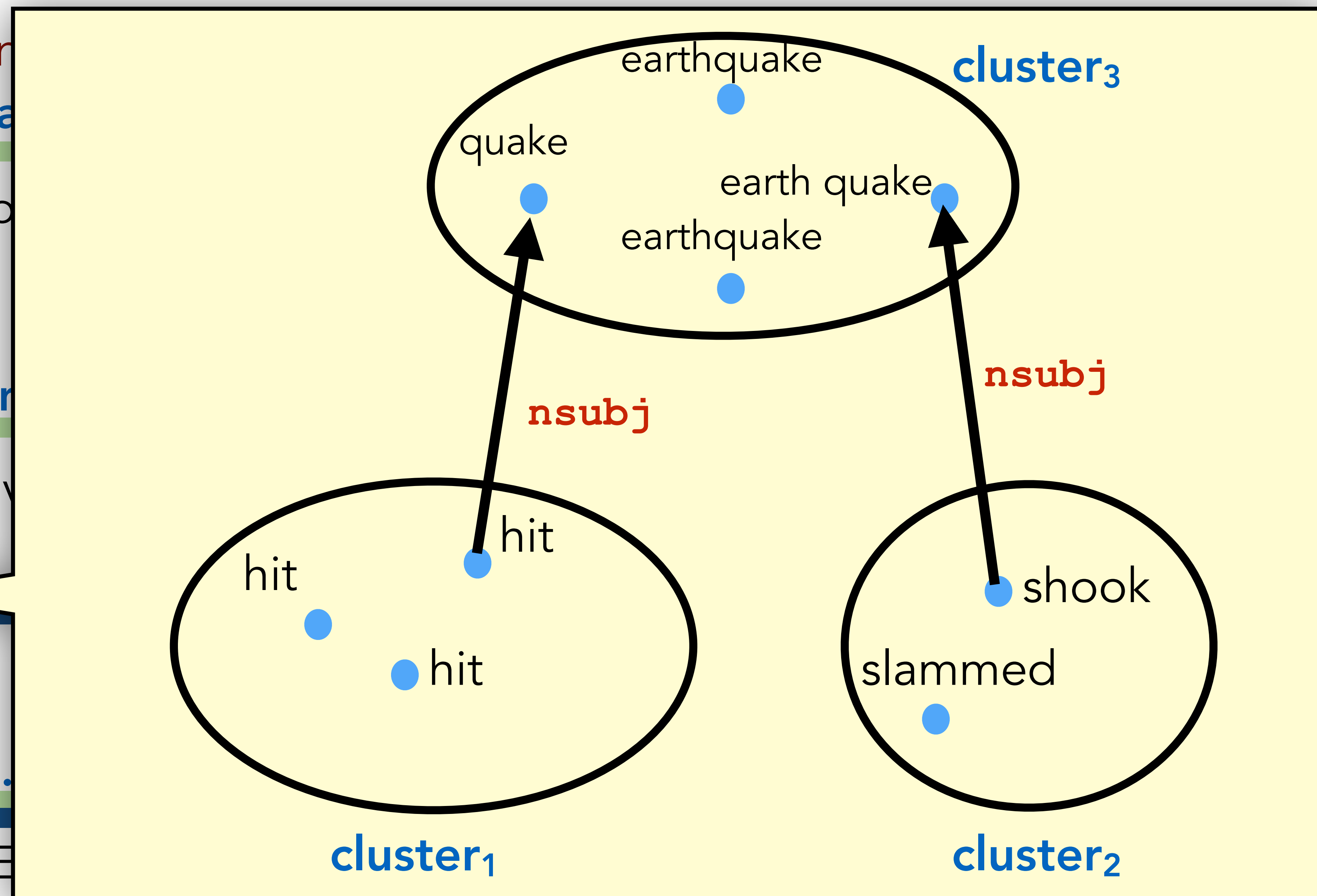
  the actual v

- **Choubey**

  determine

- **Lee, et. al.**

  builds up E



entity coreference

event coreference

Proposed Work

Past Work o

- **Haghighi a**

  governor o

  event)

- **Rahman a**

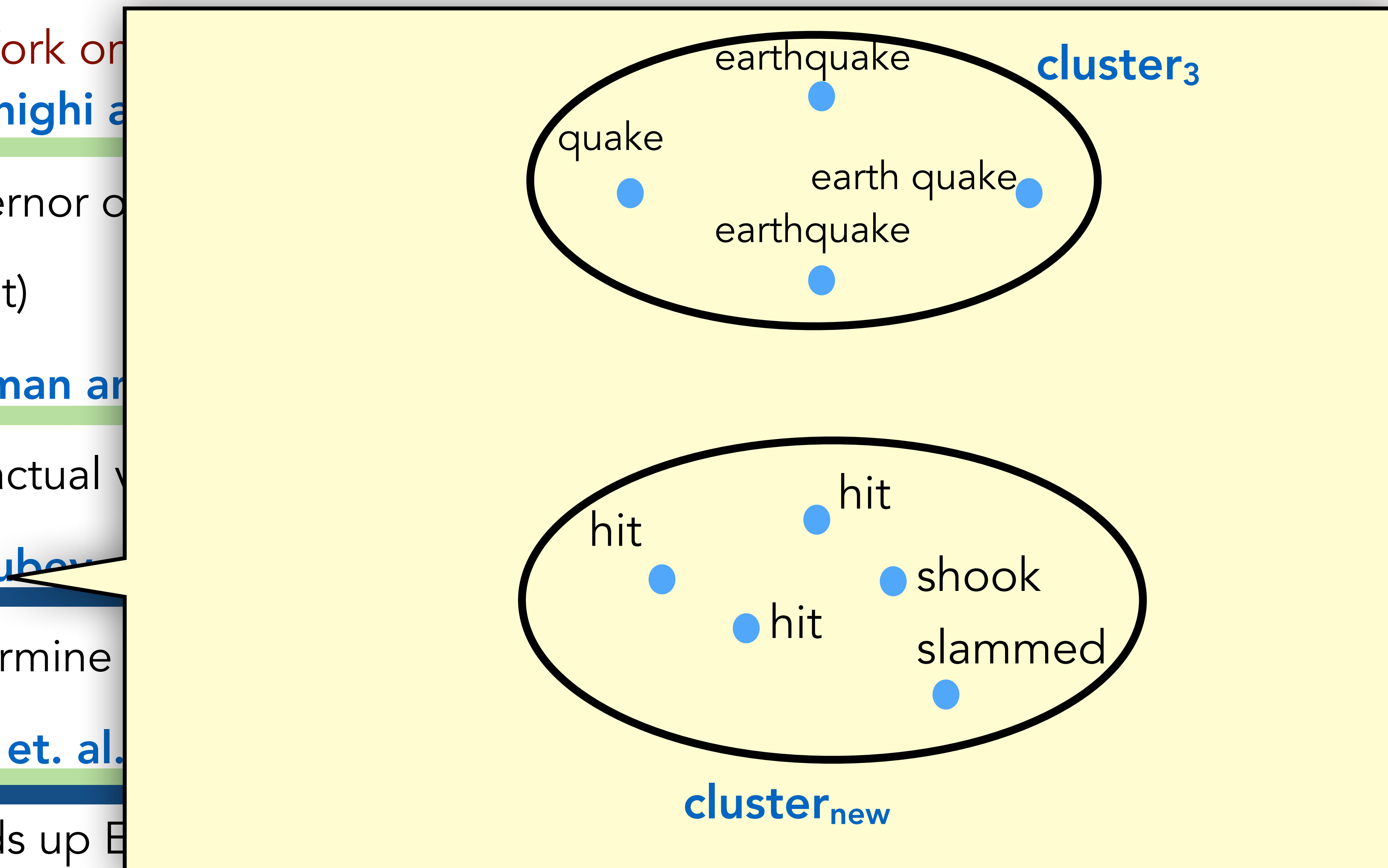
  the actual

- **Choub**

  determine

- **Lee, et. al.**

  builds up E



entity coreference

event coreference

Proposed Work

Past Work or

- **Haghighi a**

  governor o

  event)

- **Rahman ar**

  the actual v

- **Choubey**

  determine

- **Lee, et. al.**

  builds up E

Weaknesses:

- hand-defined, hard-threshold rule
- most events don't directly depend on other events
- doesn't use entities

entity coreference

event coreference

Proposed Work

Past Work on Joint Entities and Events

- **Haghighi and Klein (2010)** include a feature which concerns the governor of the head of the nominal mentions (which could be an event)
- **Rahman and Ng (2011)** use semantic roles of entity mentions, plus the actual verbs, as features
- **Choubey and Huang (2017)** uses hand-made relational rule to determine if two event clusters should be merged
- **Lee, et. al. (2012)** performs Entity Coref via StanfordCoreNLP then builds up Event clusters based on semantic role features.

entity coreference

event coreference

Proposed Work

2. Jointly Model Entities and Events

Confidence of Event Coreference influence Entity Coreference and vice versa

One of the key suspected Mafia bosses arrested yesterday has hanged himself.

Police said **Lo Presti** had hanged himself.

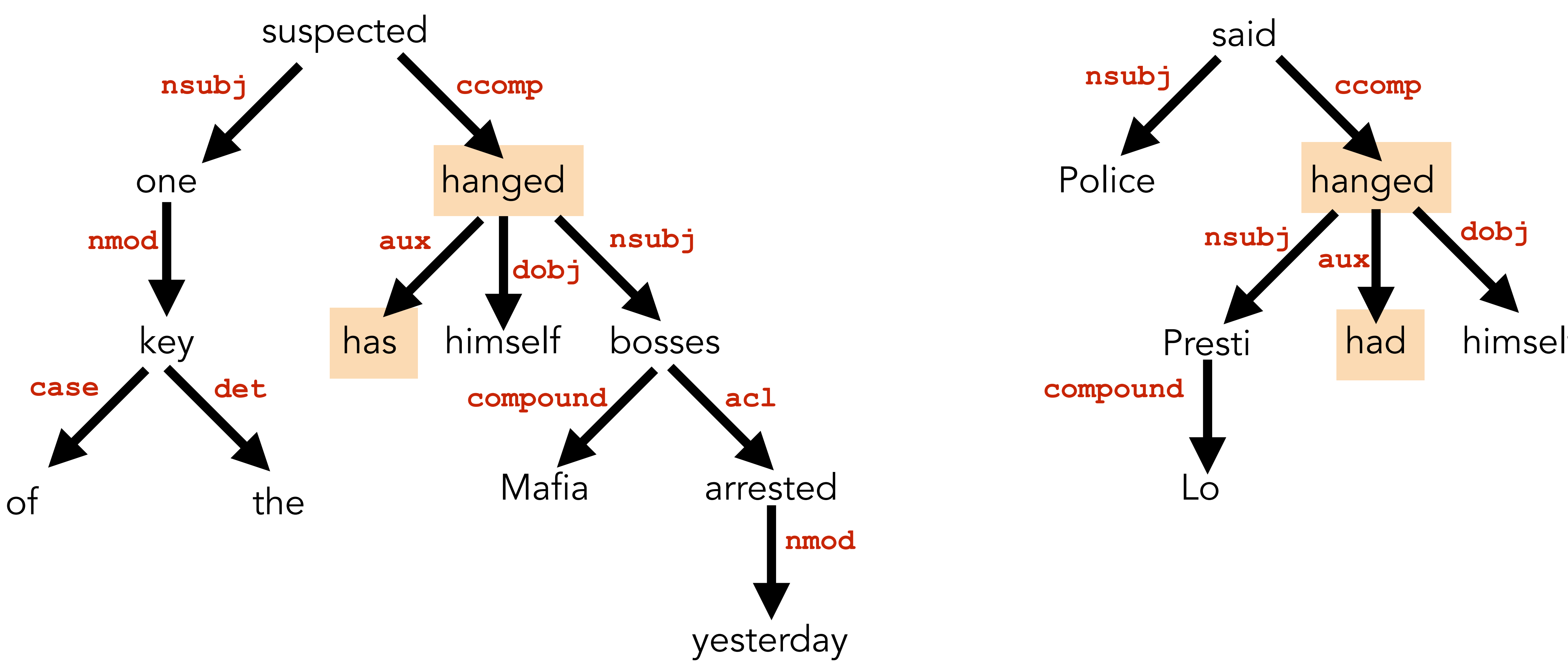
Proposed Work

2. Jointly Model Entities and Events

Model dependencies organically

One of the key suspected Mafia bosses arrested yesterday has hanged himself.

Police said **Lo Presti** had hanged himself.



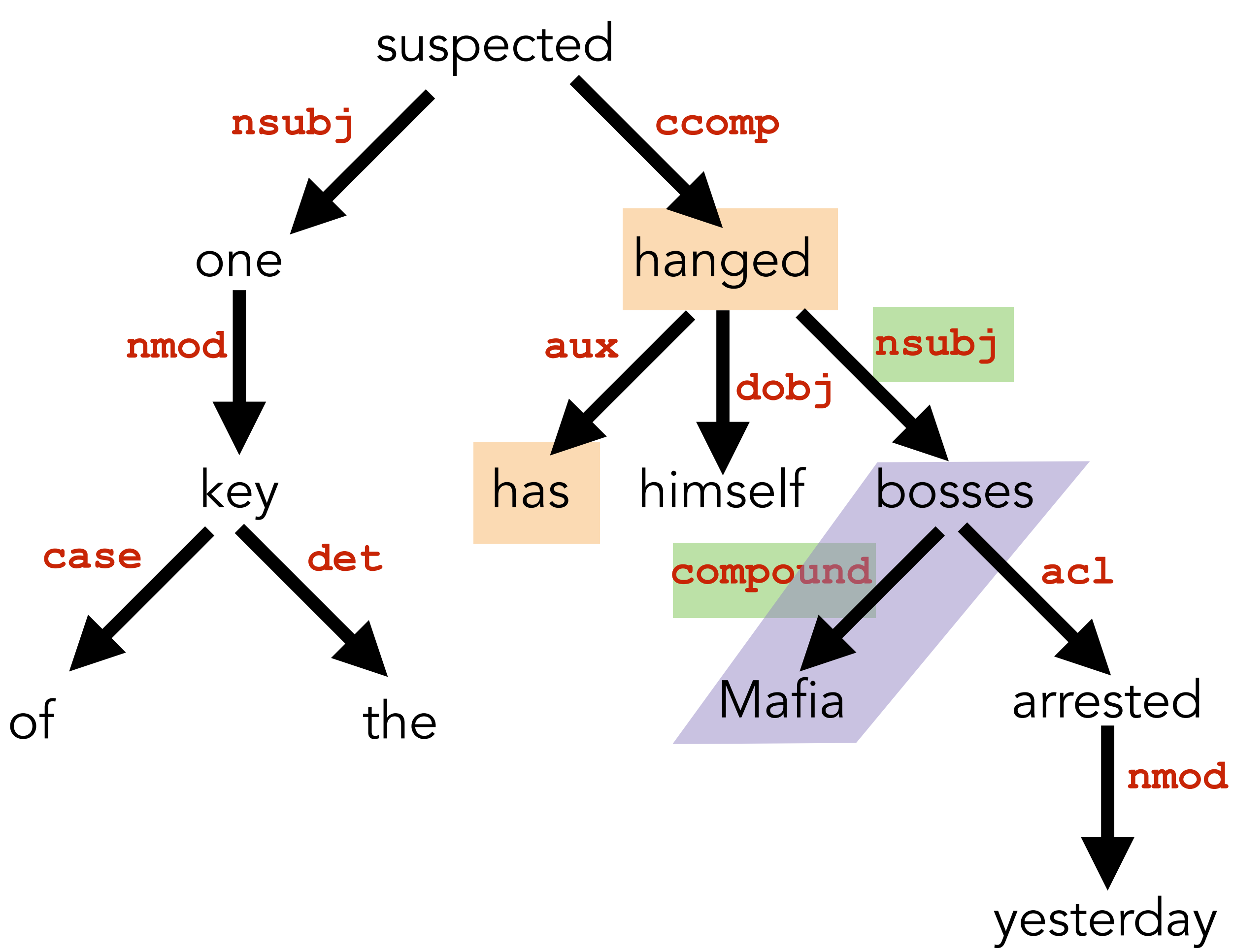
Proposed Work

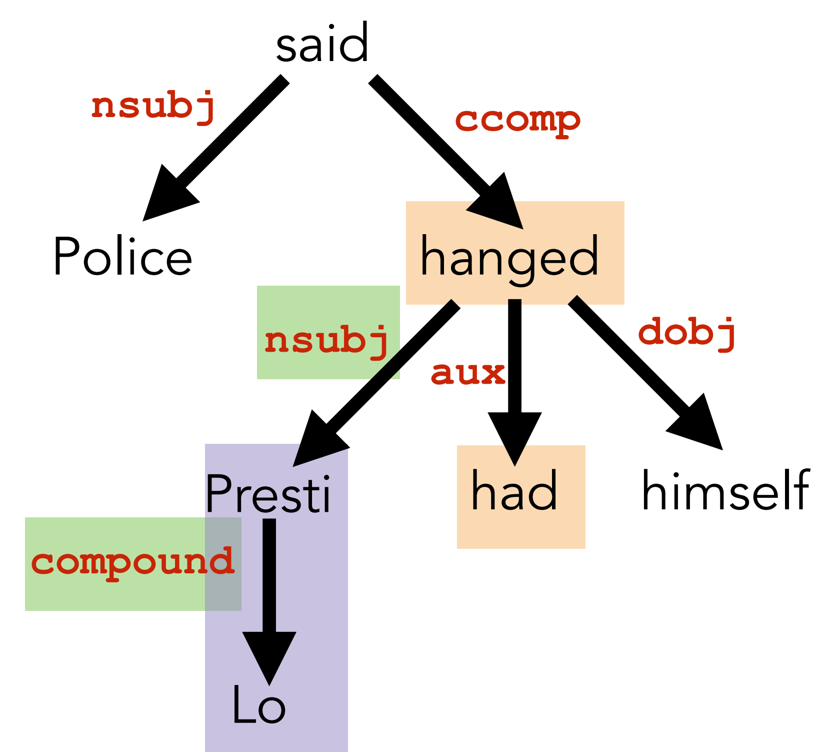
2. Jointly Model Entities and Events

Model dependencies organically

One of the key suspected Mafia bosses arrested yesterday has hanged himself.

Police said **Lo Presti** had hanged himself.





Proposed Work

3. Jointly Detect Mentions and Perform Coreference

My Stretch Goal:

- 1 system performs both mention detection and coreference
- few manually-defined features

  (e.g. not BoW windows, cosine sim. b/w X and Y, governor?)
- little lexical mark-up

  (e.g., dependency relations, POS, word embeddings, lemmas)

Proposed Work

3. Jointly Detect Mentions and Perform Coreference

My Stretch Goal:

- 1 system performs both mention detection and coreference
- few manually-defined features

  (e.g. not BoW windows, cosine sim. b/w X and Y, governor?)
- little lexical mark-up

  (e.g., dependency relations, POS, word embeddings, lemmas)

End-to-End Neural Coreference Resolution **(Lee, et. al., 2017)**

CoNLL (entities)

Within-Doc

Proposed Work

3. Jointly Detect Mentions and Perform Coreference

Areas for Improvement:

- Include Events
- Cross-document (or subsuming meta-document)
- Joint Mention Detection and Coreference

Rough Ideas:

- Bi-LSTM input
- Sample pairs of mention spans to evaluate
    - target output serves as a weighted vote for each mention existing and a vote for the pair to co-occur
    - final clusters gleaned by highest democratic votes

Completed Work

baseline Entity Coref experiments

Joint Entity + Event Coref

continue above or if finished, mention detection

write dissertation

defend

Proposed Work

Estimated Timeline

May 22	submit to EMNLP
June	improve Event Coref via ensemble + self-training
July - August	baseline Entity Coref experiments
August - Nov	Joint Entity + Event Coref
Nov - March	continue above or if finished, mention detection
March - May	write dissertation
May	defend

Proposed Work

Thesis Statement

This research aims to develop the first *comprehensive* cross-document coreference resolution system, by improving event coreference and showing the benefits of jointly modelling both events and entities.

Proposed Work

Thesis Statement

This research aims to develop the first *comprehensive* cross-document coreference resolution system, by improving event coreference and showing the benefits of jointly modelling both events and entities.

Thanks!

Questions?