End-to-end Neural Coreference Resolution

Kenton Lee Luheng He Mike Lewis Luke Zettlemoyer

University of Washington

Facebook AI Research

Allen Institute for Artificial Intelligence

Coreference Resolution

Input document
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building.

Coreference Resolution

Input document
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building.

Cluster #1	A fire in a Bangladeshi garment factory	the blaze in the four-story building

Coreference Resolution

Input document
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building.

Cluster #1	A fire in a Bangladeshi garment factory	the blaze in the four-story building
Cluster #2	a Bangladeshi garment factory	the four-story building

Coreference Resolution

Input document
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building.

Cluster #1	A fire in a Bangladeshi garment factory	the blaze in the four-story building
Cluster #2	a Bangladeshi garment factory	the four-story building
Cluster #3	at least 37 people	the deceased

Two Subproblems

Input document
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building.

Mention detection

A fire in a Bangladeshi garment factory
at least 37 people
…
the four-story building

Mention clustering

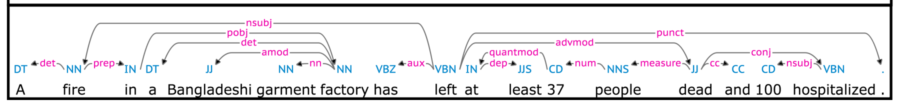
Cluster #1	A fire in a Bangladeshi garment factory	the blaze in the four-story building
Cluster #2	a Bangladeshi garment factory	the four-story building
Cluster #3	at least 37 people	the deceased

Previous Approach: Rule-based pipeline

Input document

A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized.

Syntactic parser



Hand-engineered rules

```
public static boolean inStopList(Mention m) {
  String mentionSpan = m.spanToString().toLowerCase(Locale.ENGLISH);
  if (mentionSpan.equals("u.s.") || mentionSpan.equals("u.k.")
      || mentionSpan.equals("u.s.s.r")) return true;
  if (mentionSpan.equals("there") || mentionSpan.startsWith("etc.")
      || mentionSpan.equals("ltd.")) return true;
  if (mentionSpan.startsWith("'s ")) return true;
//    if (mentionSpan.endsWith("etc.")) return true;

  return false;
}

protected void removeSpuriousMentions(Annotation doc, List<List<Mention>> predictedMentions, Dictionaries dict, boolean remov
  if(lang == Locale.ENGLISH) removeSpuriousMentionsEn(doc, predictedMentions, dict);
  else if (lang == Locale.CHINESE) removeSpuriousMentionsZh(doc, predictedMentions, dict, removeNested);
}

protected void removeSpuriousMentionsEn(Annotation doc, List<List<Mention>> predictedMentions, Dictionaries dict) {
  List<CoreMap> sentences = doc.get(CoreAnnotations.SentencesAnnotation.class);

  for(int i=0 ; i < predictedMentions.size() ; i++) {
    CoreMap s = sentences.get(i);
    List<Mention> mentions = predictedMentions.get(i);

    List<CoreLabel> sent = s.get(CoreAnnotations.TokensAnnotation.class);
    Set<Mention> remove = Generics.newHashSet();

    for(Mention m : mentions){
      String headPOS = m.headWord.get(CoreAnnotations.PartOfSpeechAnnotation.class);

      // non word such as 'hmm'
      if(dict.nonWords.contains(m.headString)) remove.add(m);
```

Candidate mentions
A fire in a Bangladeshi garment factory
garment
factory
at least 37 people dead and 100 hospitalized
…

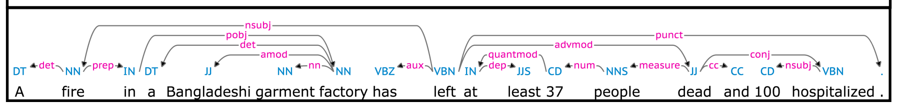
Mention #1	Mention #2	Coreferent?
A fire in a Bangladeshi garment factory	garment	✓/✗
garment	factory	✓/✗
factory	at least 37 people dead and 100 hospitalized	✓/✗
…	…	✓/✗

Previous Approach: Rule-based pipeline

Input document

A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized.

Syntactic parser



Hand-engineered rules

```
public static boolean inStopList(Mention m) {
  String mentionSpan = m.spanToString().toLowerCase(Locale.ENGLISH);
  if (mentionSpan.equals("u.s.") || mentionSpan.equals("u.k.")
      || mentionSpan.equals("u.s.s.r")) return true;
  if (mentionSpan.equals("there") || mentionSpan.startsWith("etc.")
      || mentionSpan.equals("ltd.")) return true;
  if (mentionSpan.startsWith("'s ")) return true;
//    if (mentionSpan.endsWith("etc.")) return true;

  return false;
}

protected void removeSpuriousMentions(Annotation doc, List<List<Mention>> predictedMentions, Dictionaries dict, boolean remov
  if(lang == Locale.ENGLISH) removeSpuriousMentionsEn(doc, predictedMentions, dict);
  else if (lang == Locale.CHINESE) removeSpuriousMentionsZh(doc, predictedMentions, dict, removeNested);
}

protected void removeSpuriousMentionsEn(Annotation doc, List<List<Mention>> predictedMentions, Dictionaries dict) {
  List<CoreMap> sentences = doc.get(CoreAnnotations.SentencesAnnotation.class);

  for(int i=0 ; i < predictedMentions.size() ; i++) {
    CoreMap s = sentences.get(i);
    List<Mention> mentions = predictedMentions.get(i);

    List<CoreLabel> sent = s.get(CoreAnnotations.TokensAnnotation.class);
    Set<Mention> remove = Generics.newHashSet();

    for(Mention m : mentions){
      String headPOS = m.headWord.get(CoreAnnotations.PartOfSpeechAnnotation.class);

      // non word such as 'hmm'
      if(dict.nonWords.contains(m.headString)) remove.add(m);
```

Mention clustering: main source of improvement for many years!

- Haghighi and Klein (2010)
- Raghunathan et al. (2010)
- …
- Clark & Manning (2016)

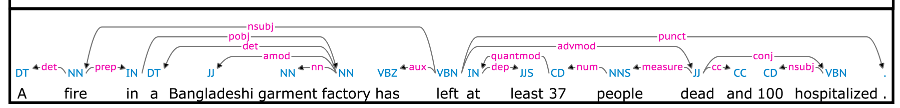
	Mention #2	Coreferent?
	garment	✓/✗
	factory	✓/✗
	least 37 people dead and 100 hospitalized	✓/✗
at	…	✓/✗

Previous Approach: Rule-based pipeline

Input document

A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized.

Syntactic parser



Hand-engineered rules

Relies on parser for:

- mention detection
- syntactic features for clustering (e.g. head words)

A fire in a Bangladeshi garment factory
garment
factory
at least 37 people dead and 100 hospitalized
…

		Coreferent?
A fire in a Bangladeshi garment factory	garment	✓/✗
garment	factory	✓/✗
factory	at least 37 people dead and 100 hospitalized	✓/✗
…	…	✓/✗

Our Contribution: End-to-end Approach

- Joint mention detection and clustering
- No preprocessing (no parser, no POS-tagger etc.)

Key Challenges

- Inference: can we do better than naive $O(N^4)$ runtime?
- Data: can we learn with partial labels?
- Model: can we induce rich features (e.g. head words)?

Inference challenge: Can we do better than $O(N^4)$?

Naive joint model is $O(N^4)$:

Input document (N words)
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building. Witnesses say the only exit door was on the ground floor, and that it was locked when the fire broke out.

Inference challenge: Can we do better than $O(N^4)$?

Naive joint model is $O(N^4)$:

Input document (N words)
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building. Witnesses say the only exit door was on the ground floor, and that it was locked when the fire broke out.

Span #1
A
A fire
A fire in
…

$O(N^2)$ spans in every document

Inference challenge: Can we do better than $O(N^4)$?

Naive joint model is $O(N^4)$:

$O(N^4)$ pairwise decisions

Input document (N words)
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building. Witnesses say the only exit door was on the ground floor, and that it was locked when the fire broke out.

Span #1	**Span #2**	**Coreferent?**
A	A fire	✓/✗
A fire	A fire in	✓/✗
A fire in	A fire in a	✓/✗
…	…	✓/✗

End-to-end Approach

- Consider all possible spans
- Learn to rank antecedent spans
- Factored model to prune search space

Span Ranking

Every span independently chooses an antecedent

Input document

A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building. Witnesses say the only exit door was on the ground floor, and that it was locked when the fire broke out.

Span Ranking

- Reason over all possible spans

- Assign an antecedent to every span

	Span	**Antecedent**
1	A	y_1
2	A fire	y_2
3	A fire in	y_3
…	…	…
M	out	y_M

Span Ranking

- Reason over all possible spans

- Assign an antecedent to every span

$$y_3 \in \{\epsilon, 1, 2\}$$

	Span	**Antecedent**
1	A	y_1
2	A fire	y_2
3	A fire in	y_3
…	…	…
M	out	y_M

Span Ranking

- Reason over all possible spans
- Assign an antecedent to every span

$$y_3 \in \{\epsilon, 1, 2\}$$

ϵ : no coreference link

	Span	Antecedent
1	A	y_1
2	A fire	y_2
3	A fire in	y_3
...	...	...
M	out	y_M

Span Ranking

- Reason over all possible spans
- Assign an antecedent to every span

	Span	Antecedent
1	A	y_1
2	A fire	y_2
3	A fire in	y_3
...	...	...
M	out	y_M

$$y_3 \in \{\epsilon, 1, 2\}$$

Coreference link from span 1 to span 3

Span Ranking

- Reason over all possible spans
- Assign an antecedent to every span

	Span	Antecedent
1	A	y_1
2	A fire	y_2
3	A fire in	y_3
…	…	…
M	out	y_M

$$y_3 \in \{\epsilon, 1, 2\}$$

Coreference link from span 2 to span 3

Example Clustering

Input document
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building. Witnesses say the only exit door was on the ground floor, and that it was locked when the fire broke out.

Span	**Antecedent (y_i)**
A	ϵ
A fire	ϵ
$\dots$	$\dots$
a Bangladeshi garment factory	ϵ
$\dots$	$\dots$
the four-story building	a Bangladeshi garment factory
$\dots$	$\dots$
out	ϵ

Example Clustering

Input document
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building. Witnesses … d floor, and that it was locked when the fire broke out.

Not a mention

Span	Antecedent (y_i)
A	ϵ
A fire	ϵ
…	…
a Bangladeshi garment factory	ϵ
…	…
the four-story building	a Bangladeshi garment factory
…	…
out	ϵ

Example Clustering

Input document
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building. Witnesses say the only exit door was on the ground floor, and that it was locked when the fire broke out.

No link with previously occurring span

Span	Antecedent (y_i)
...	...
a Bangladeshi garment factory	ϵ
...	...
the four-story building	a Bangladeshi garment factory
...	...
out	ϵ

Example Clustering

Input document
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building. Witnesses say the only exit door was on the ground floor, and that it was locked when the fire broke out.

Predicted coreference link

Span	**Antecedent (y_i)**
A	ϵ
A fire	ϵ
	$\ldots$
	ϵ
$\ldots$	$\ldots$
the four-story building	a Bangladeshi garment factory
$\ldots$	$\ldots$
out	ϵ

Span Ranking Model

$P(y_1, \ldots, y_M \mid D)$

Span Ranking Model

$$P(y_1, \ldots, y_M \mid D) = \prod_{i=1}^{M} P(y_i \mid D)$$

Independent decision for every span

Span Ranking Model

$$P(y_1, \ldots, y_M \mid D) = \prod_{i=1}^{M} P(y_i \mid D)$$

$$= \prod_{i=1}^{M} \frac{e^{s(i,y_i)}}{\sum_{y' \in \mathcal{Y}(i)} e^{s(i,y')}}$$

Pairwise coreference score $s(i, j)$ between span i and span j

Span Ranking Model

$$P(y_1, \ldots, y_M \mid D) = \prod_{i=1}^{M} P(y_i \mid D)$$

$$= \prod_{i=1}^{M} \frac{e^{s(i,y_i)}}{\sum_{y' \in \mathcal{Y}(i)} e^{s(i,y')}}$$

Factor coreference score $s(i, j)$ to enable span pruning:

$$s(i,j) = \begin{cases} s_{\mathrm{m}}(i) + s_{\mathrm{m}}(j) + s_{\mathrm{a}}(i,j) & j \neq \epsilon \\ 0 & j = \epsilon \end{cases}$$

Span Ranking Model

$$P(y_1, \ldots, y_M \mid D) = \prod_{i=1}^{M} P(y_i \mid D)$$

$$\prod_{i=1}^{M} \frac{e^{s(i,y_i)}}{\sum_{y' \in \mathcal{Y}(i)} e^{s(i,y')}}$$

Is this span a mention?

Factor coreference score $s(i,j)$ to enable span pruning:

$$s(i,j) = \begin{cases} s_{\mathrm{m}}(i) + s_{\mathrm{m}}(j) + s_{\mathrm{a}}(i,j) & j \neq \epsilon \\ 0 & j = \epsilon \end{cases}$$

Span Ranking Model

$$P(y_1, \ldots, y_M \mid D) = \prod_{i=1}^{M} P(y_i \mid D)$$

$$= \prod^{M} \frac{e^{s(i,y_i)}}{\sum \cdots}$$

Is span j an antecedent of span i?

Factor coreference score $s(i, j)$ to enable span pruning:

$$s(i,j) = \begin{cases} s_{\mathrm{m}}(i) + s_{\mathrm{m}}(j) + s_{\mathrm{a}}(i,j) & j \neq \epsilon \\ 0 & j = \epsilon \end{cases}$$

Span Ranking Model

$$P(y_1, \ldots, y_M \mid D) = \prod_{i=1}^{M} P(y_i \mid D)$$

$$= \prod_{i=1}^{M} \frac{e^{s(i,y_i)}}{\sum_{y' \in \mathcal{Y}(i)} e^{s(i,y')}}$$

Factor coreference score $s(i,j)$ to enable span pruning:

$$s(i,j) = \begin{cases} s_{\mathrm{m}}(i) + s_{\mathrm{m}}(j) + s_{\mathrm{a}}(i,j) & j \neq \epsilon \\ 0 & j = \epsilon \end{cases}$$

Dummy antecedent
has a fixed zero score

Two-stage Beam Search

Input document (N words**)**
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building. Witnesses say the only exit door was on the ground floor, and that it was locked when the fire broke out.

Two-stage Beam Search

Input document (N words)
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building. Witnesses say the only exit door was on the ground floor, and that it was locked when the fire broke out.

Span	s_{m}
A	-10
A fire	4
…	…
a Bangladeshi garment factory	6
…	…
the four-story building	10
…	…
out	-5

Two-stage Beam Search

Input document (N words)
A fire in a Bangladeshi garment factory has ... deceased were killed in the crush as worke... say the only exit door was on the ground f...

Spans with low mention scores likely to have a negative overall score

Span	s_{m}
A	-10
A fire	4
...	...
a Bangladeshi garment factory	6
...	...
the four-story building	10
...	...
out	-5

Two-stage Beam Search

Input document (N words)
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building. Witnesses say the only exit door was on the ground floor, and that it was locked when the fire broke out.

Span	s_{m}
A	-10
A fire	4
...	...
a Bangladeshi garment factory	6
...	...
the four-story building	10
...	...
out	-5

$\xrightarrow{\text{Keep top } \lambda N}$

Span	s_{m}
A fire	4
...	...
a Bangladeshi garment factory	6
...	...
the four-story building	10
...	...

Two-stage Beam Search

Input document (N words**)**
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building. Witnesses say the only exit door was on the ground floor, and that it was locked when the fire broke out.

Target Span	A fire
	…
	a Bangladeshi garment factory
	…
	the four-story building
	…

Two-stage Beam Search

Input document (N words)
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building. Witnesses say the only exit door was on the ground floor, and that it was locked when the fire broke out.

		Antecedent						
		ϵ	A fire	…	a Bangladeshi garment factory	…	the four-story building	…
Target Span	A fire	0	$-\infty$	-	$-\infty$	-	$-\infty$	-
	…	0	…	-	$-\infty$	-	$-\infty$	-
	a Bangladeshi garment factory	0	-10	-5	$-\infty$	-	$-\infty$	-
	…	0	…	…	…	-	$-\infty$	-
	the four-story building	0	2	-3	10	-5	$-\infty$	-
	…	0	…	…	…	…	…	-

Inference challenge: Can we do better than $O(N^4)$?

Naive joint model is $O(N^4)$:

Input document (N words)
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building. Witnesses say the only exit door was on the ground floor, and that it was locked when the fire broke out.

Span #1	**Span #2**	**Coreferent?**
A	A fire	✓/✗
A fire	A fire in	✓/✗
A fire in	A fire in a	✓/✗
…	…	✓/✗

Factored model enables aggressive pruning

Data Challenge: Can we learn with partial labels?

Only clusters with multiple mentions annotated:

Input document
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building…….

Data Challenge:
Can we learn with partial labels?

Only clusters with multiple mentions annotated:

Input document
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building……

Singleton mention missing from data

Learning

Marginal log-likelihood objective.

$$\log \prod_{i=1}^{M} \sum_{\hat{y} \in \mathcal{Y}(i) \cap \textsc{gold}(i)} P(\hat{y} \mid D)$$

Learning

Marginal log-likelihood objective.

$$\log \prod_{i=1}^{M} \sum_{\hat{y} \in \mathcal{Y}(i) \cap \textsc{gold}(i)} P(\hat{y} \mid D)$$

- Related to Durrett & Klein (2013)

Learning

Marginal log-likelihood objective.

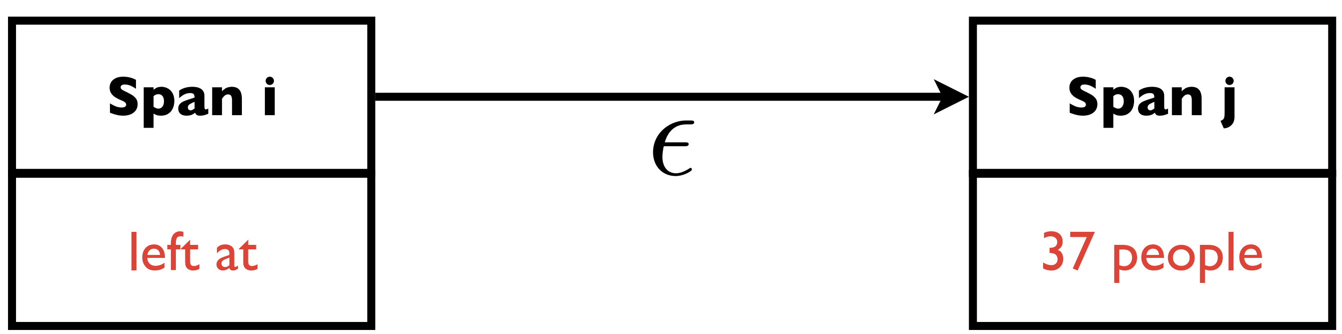
$$\log \prod_{i=1}^{M} \sum_{\hat{y} \in \mathcal{Y}(i) \cap \textsc{gold}(i)} P(\hat{y} \mid D)$$

- Related to Durrett & Klein (2013)
- Model can assign credit/blame to the mention or antecedent factors

$$s(i,j) = \begin{cases} s_{\mathrm{m}}(i) + s_{\mathrm{m}}(j) + s_{\mathrm{a}}(i,j) & j \neq \epsilon \\ 0 & j = \epsilon \end{cases}$$

Credit Assignment Example

Input document
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building.

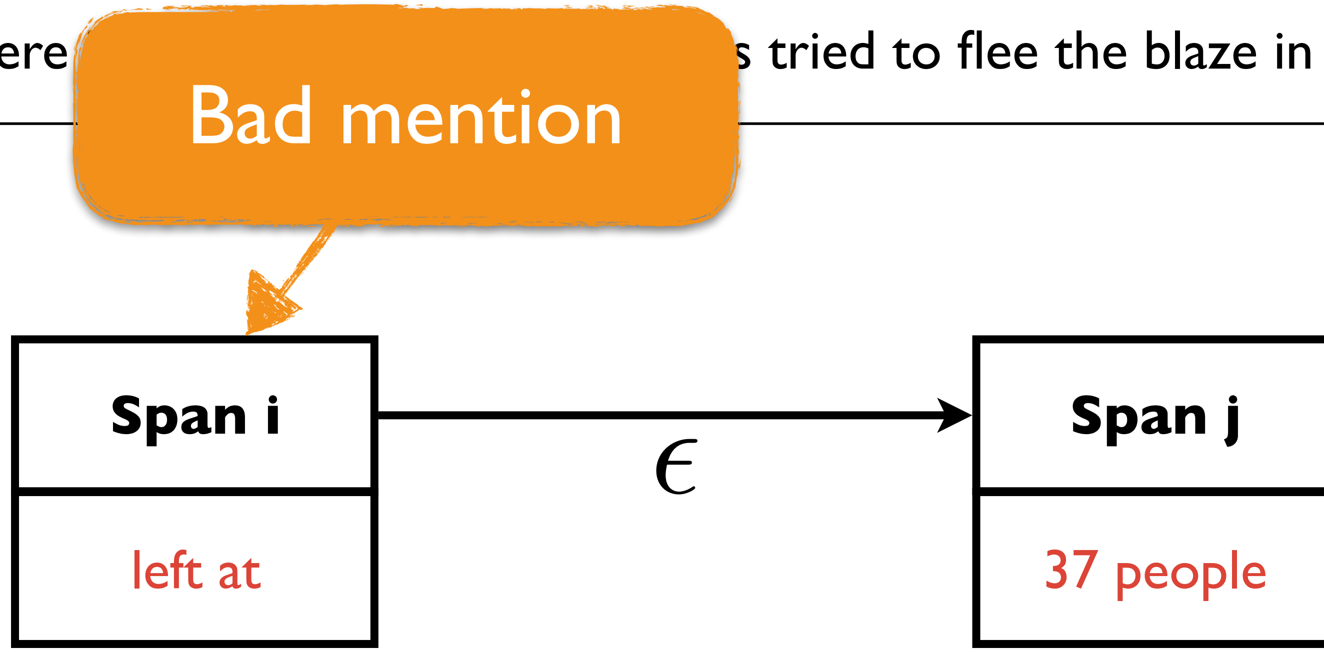


$$s(i,j) = \begin{cases} s_{\mathrm{m}}(i) + s_{\mathrm{m}}(j) + s_{\mathrm{a}}(i,j) & j \neq \epsilon \\ 0 & j = \epsilon \end{cases}$$

Credit Assignment Example

Input document

A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were … s tried to flee the blaze in the four-story building.



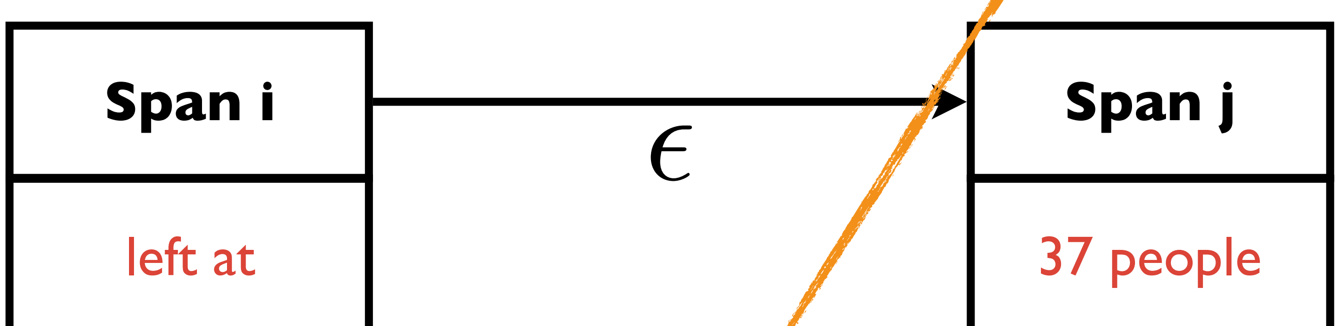
$$s(i,j) = \begin{cases} s_{\mathrm{m}}(i) + s_{\mathrm{m}}(j) + s_{\mathrm{a}}(i,j) & j \neq \epsilon \\ 0 & j = \epsilon \end{cases}$$

Credit Assignment Example

Input document

A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried tc … g.

Blame mention factor for absent link

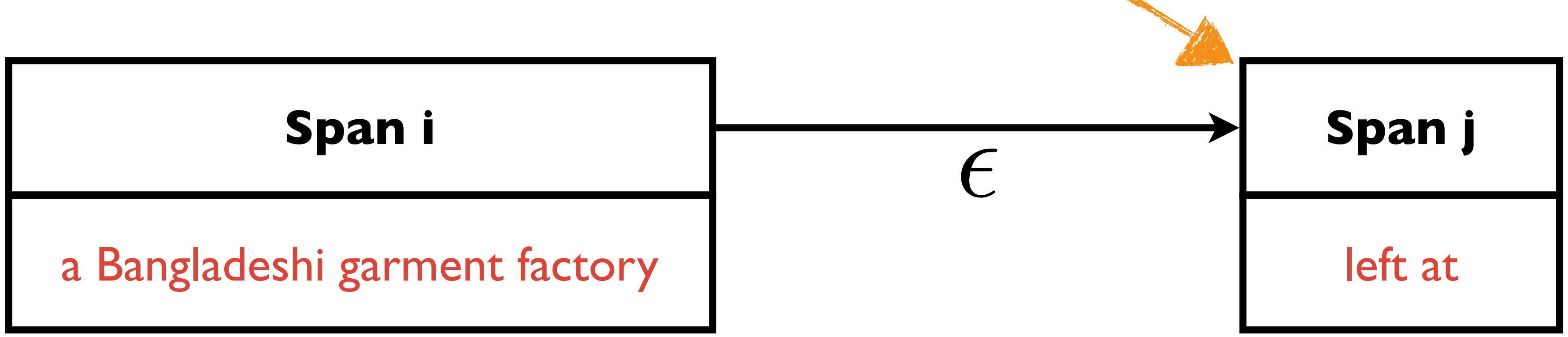


$$s(i,j) = \begin{cases} s_{\mathrm{m}}(i) + s_{\mathrm{m}}(j) + s_{\mathrm{a}}(i,j) & j \neq \epsilon \\ 0 & j = \epsilon \end{cases}$$

Credit Assignment Example

Input document
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in t[illegible] flee the blaze in the four-story building.

Bad mention

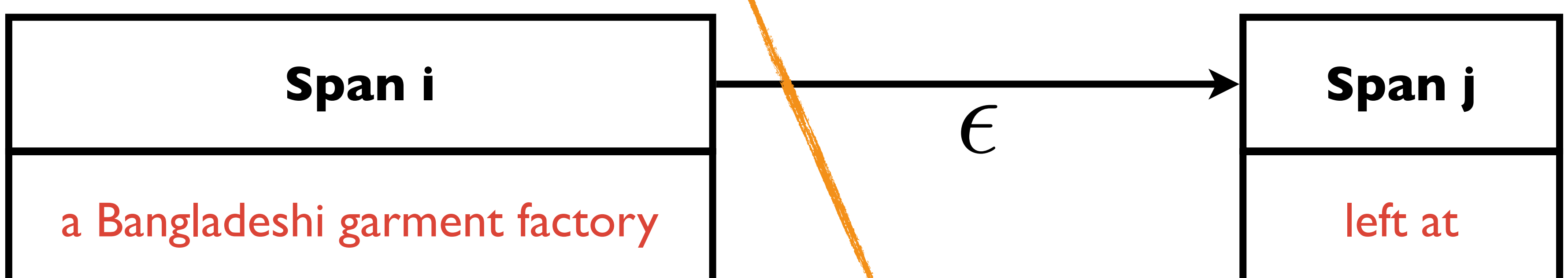


Credit Assignment Example

Input document

A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceas … tried to flee the blaze in the four-story building.

Blame mention factor for absent link



$$s(i,j) = \begin{cases} s_{\mathrm{m}}(i) + s_{\mathrm{m}}(j) + s_{\mathrm{a}}(i,j) & j \neq \epsilon \\ 0 & j = \epsilon \end{cases}$$

Credit Assignment Example

Input document
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush … -story building.

Incompatible mentions

Span i
a Bangladeshi garment factory

$\xrightarrow{\epsilon}$

Span j
37 people

$$s(i,j) = \begin{cases} s_{\mathrm{m}}(i) + s_{\mathrm{m}}(j) + s_{\mathrm{a}}(i,j) & j \neq \epsilon \\ 0 & j = \epsilon \end{cases}$$

Credit Assignment Example

Input document

A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were kill blaze in the four-story building.

Blame antecedent factor for absent link

Span i		**Span j**
a Bangladeshi garment factory	$\xrightarrow{\epsilon}$	37 people

$$s(i, j) = \begin{cases} s_{\mathrm{m}}(i) + s_{\mathrm{m}}(j) + s_{\mathrm{a}}(i, j) & j \neq \epsilon \\ 0 & j = \epsilon \end{cases}$$

Data Challenge: Can we learn with partial labels?

Only clusters with multiple mentions annotated:

Input document
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building……

Missing mentions are latent in the joint model

Model Challenge: Can we induce rich features?

Lexical and contextual cues are useful:

Input document
A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building…….

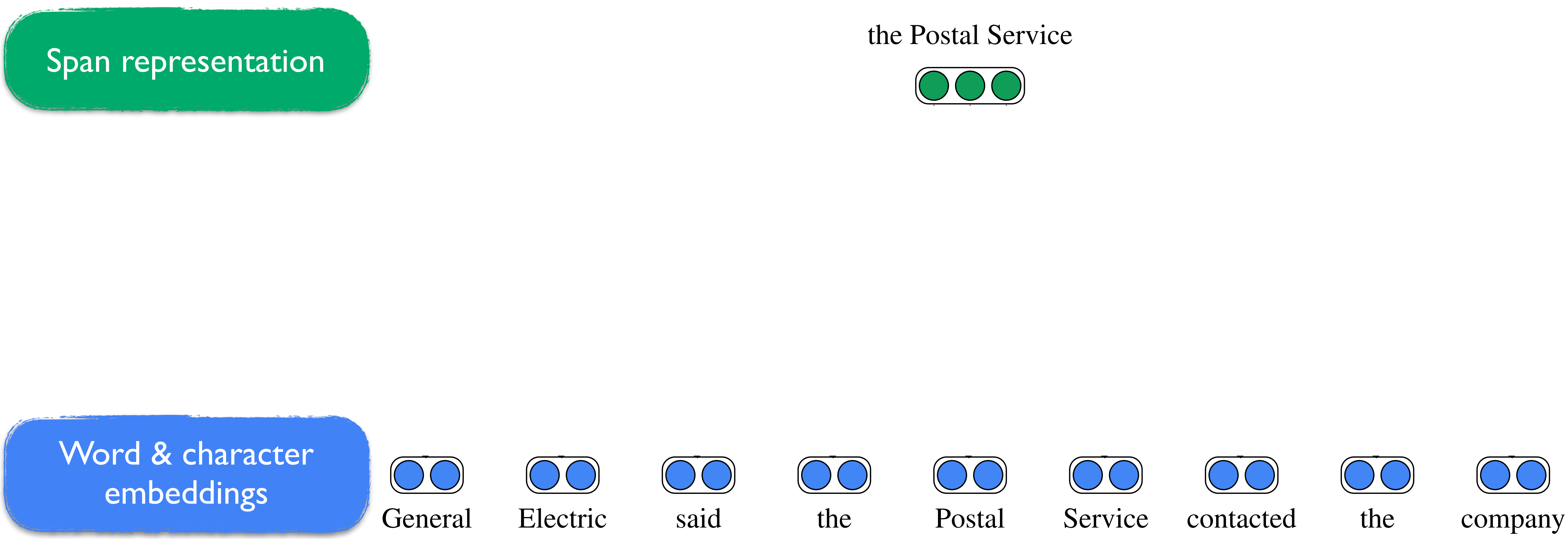
Model Challenge: Can we induce rich features?

Lexical and contextual cues are useful:

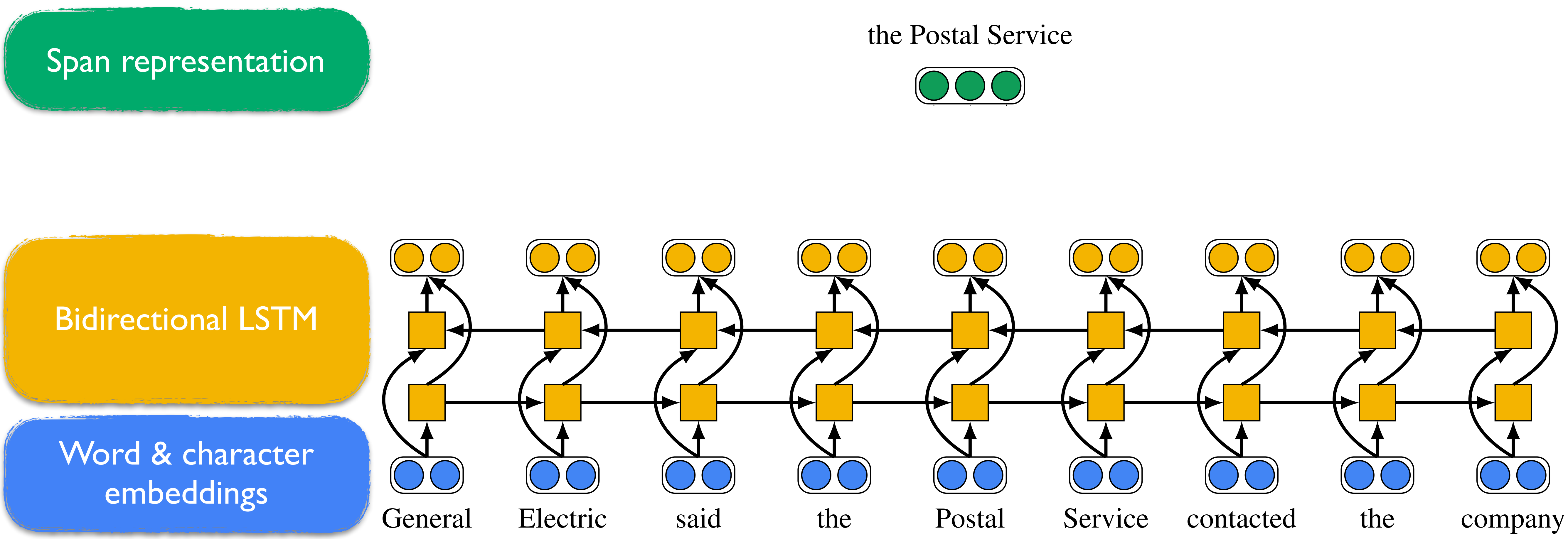
e.g. paraphrased head words

Input docume
A **fire** in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the **blaze** in the four-story building…….

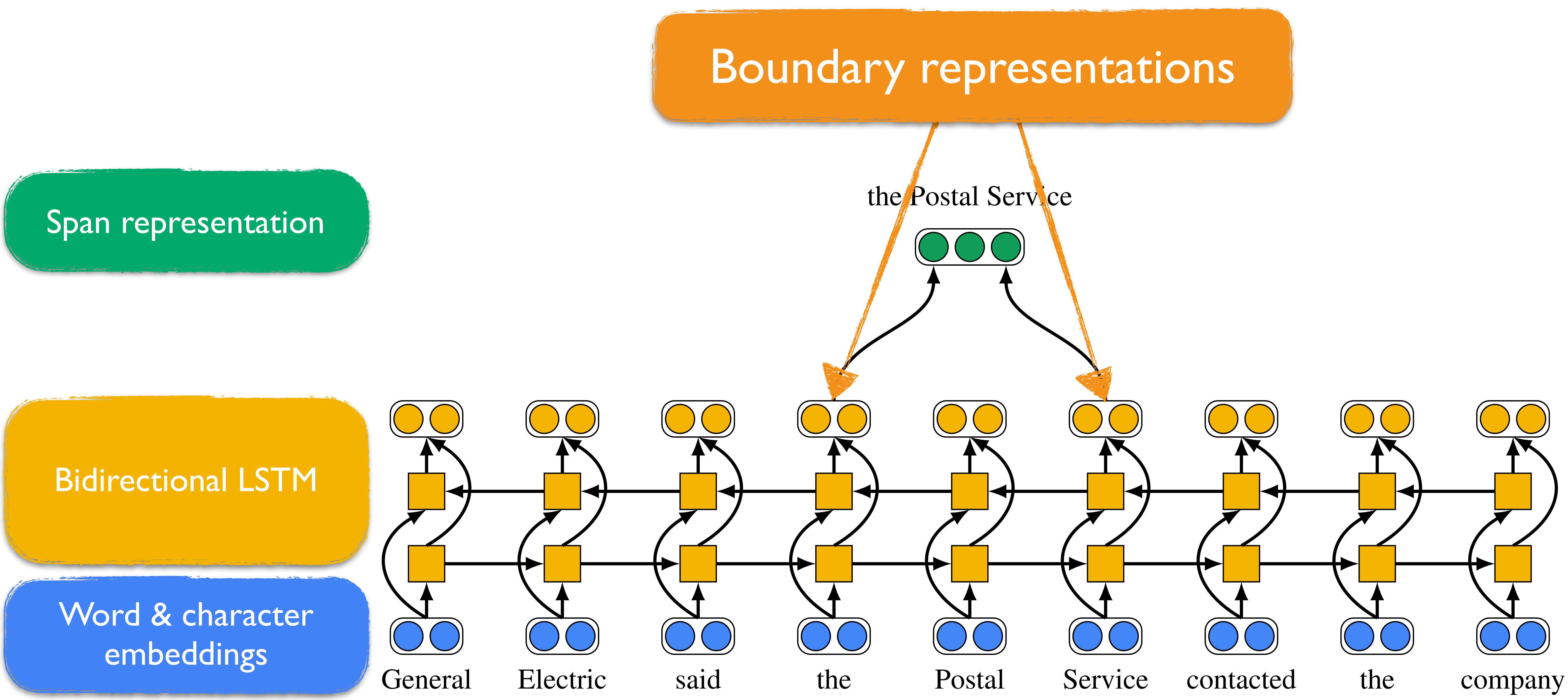
Neural Span Representations



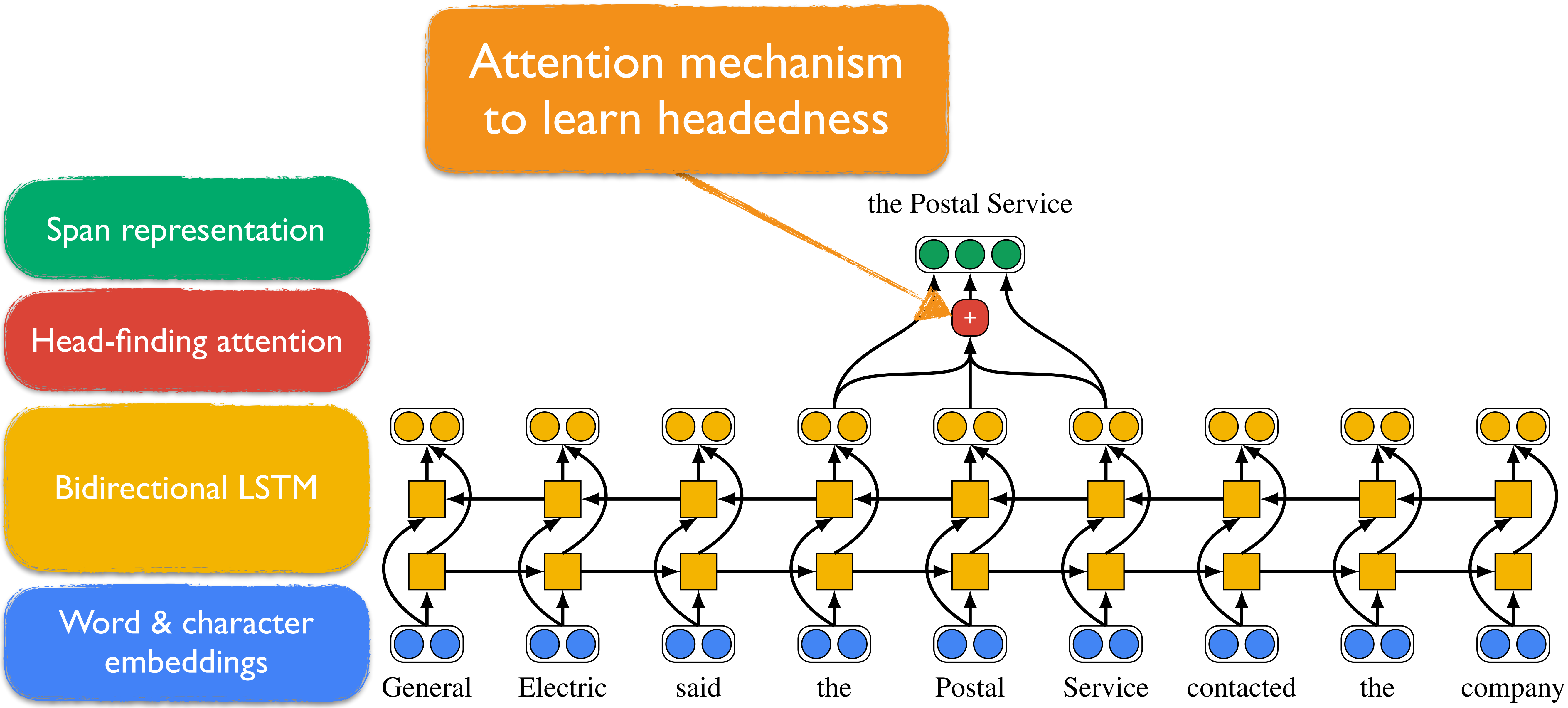
Neural Span Representations



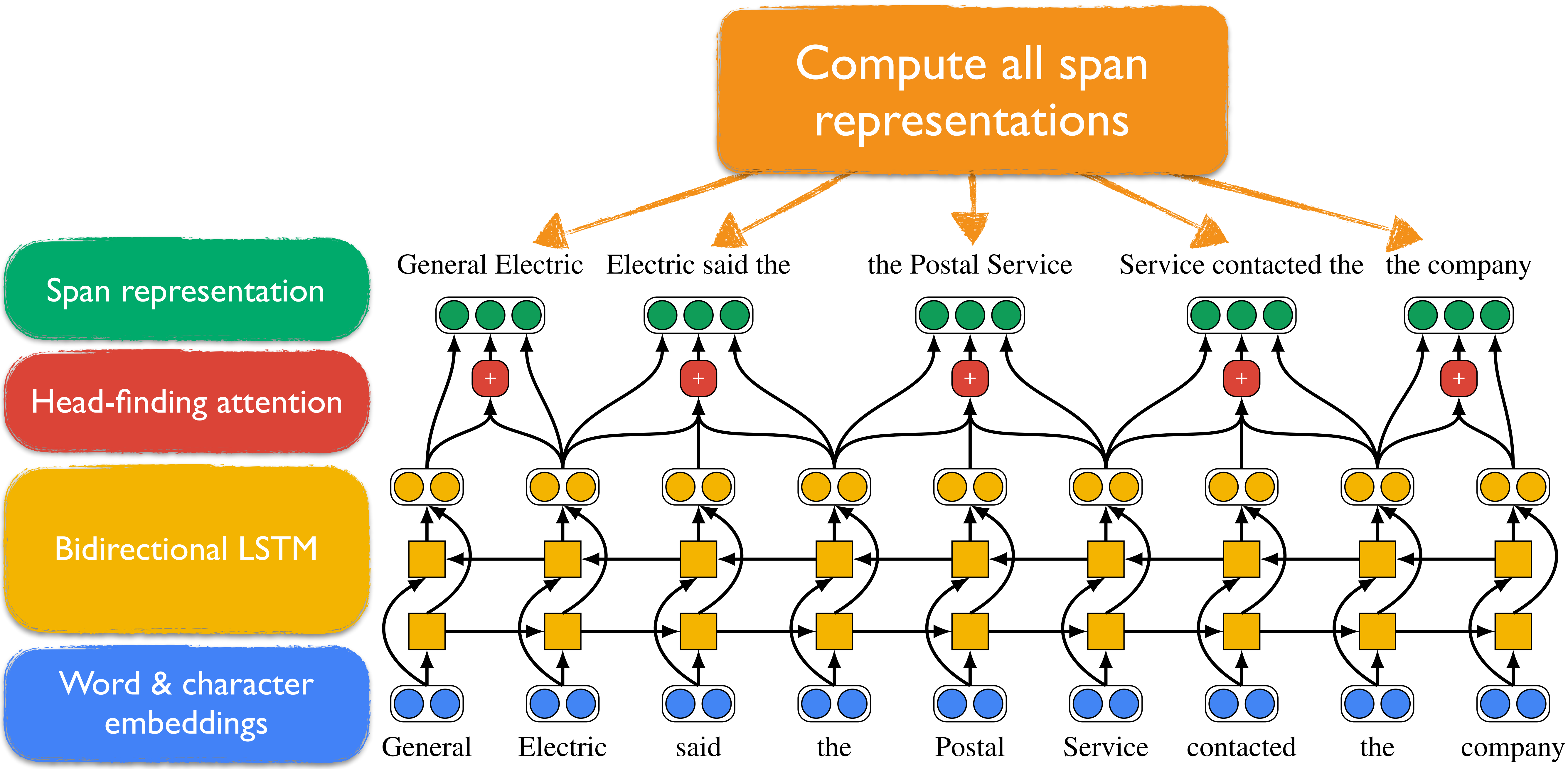
Neural Span Representations



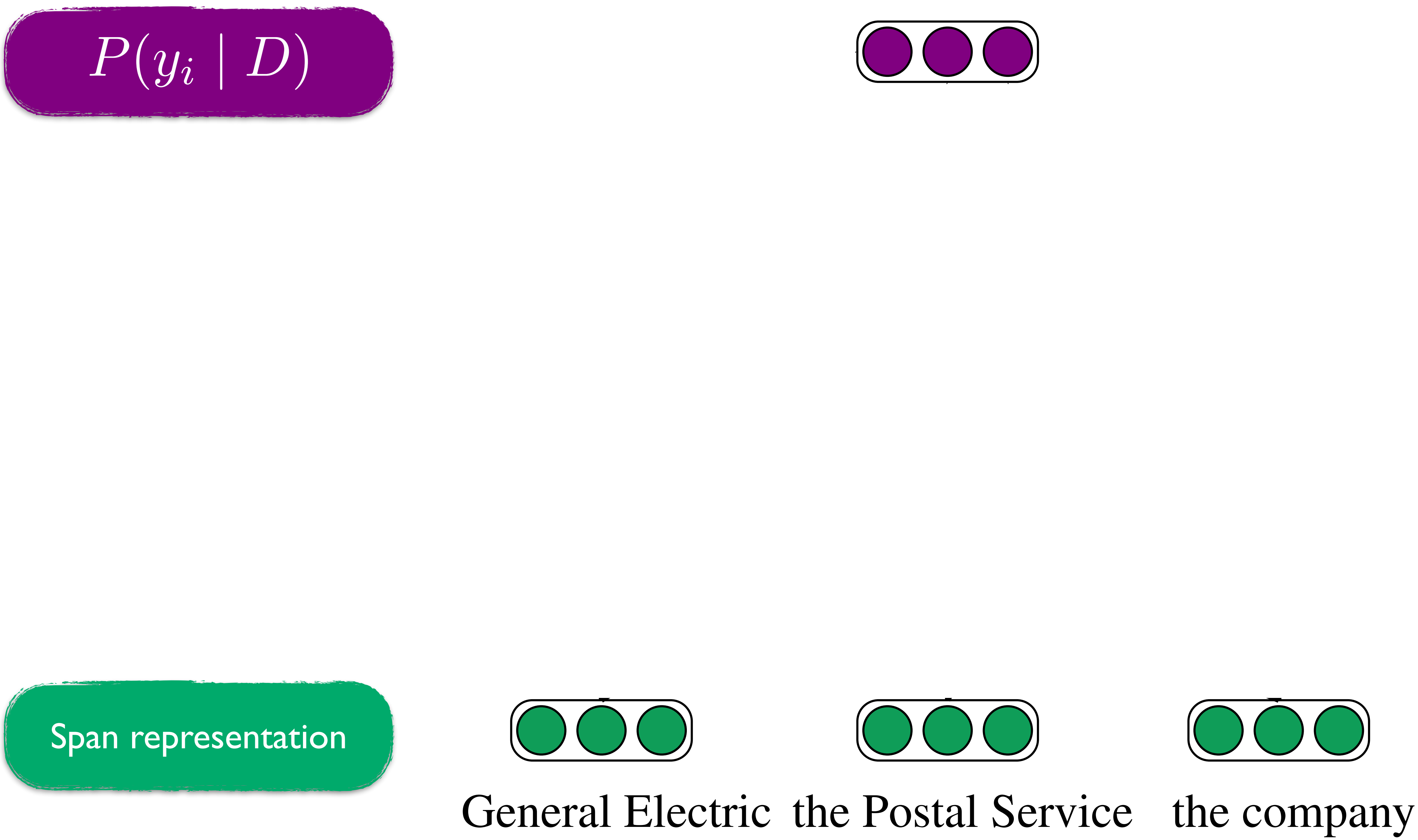
Neural Span Representations



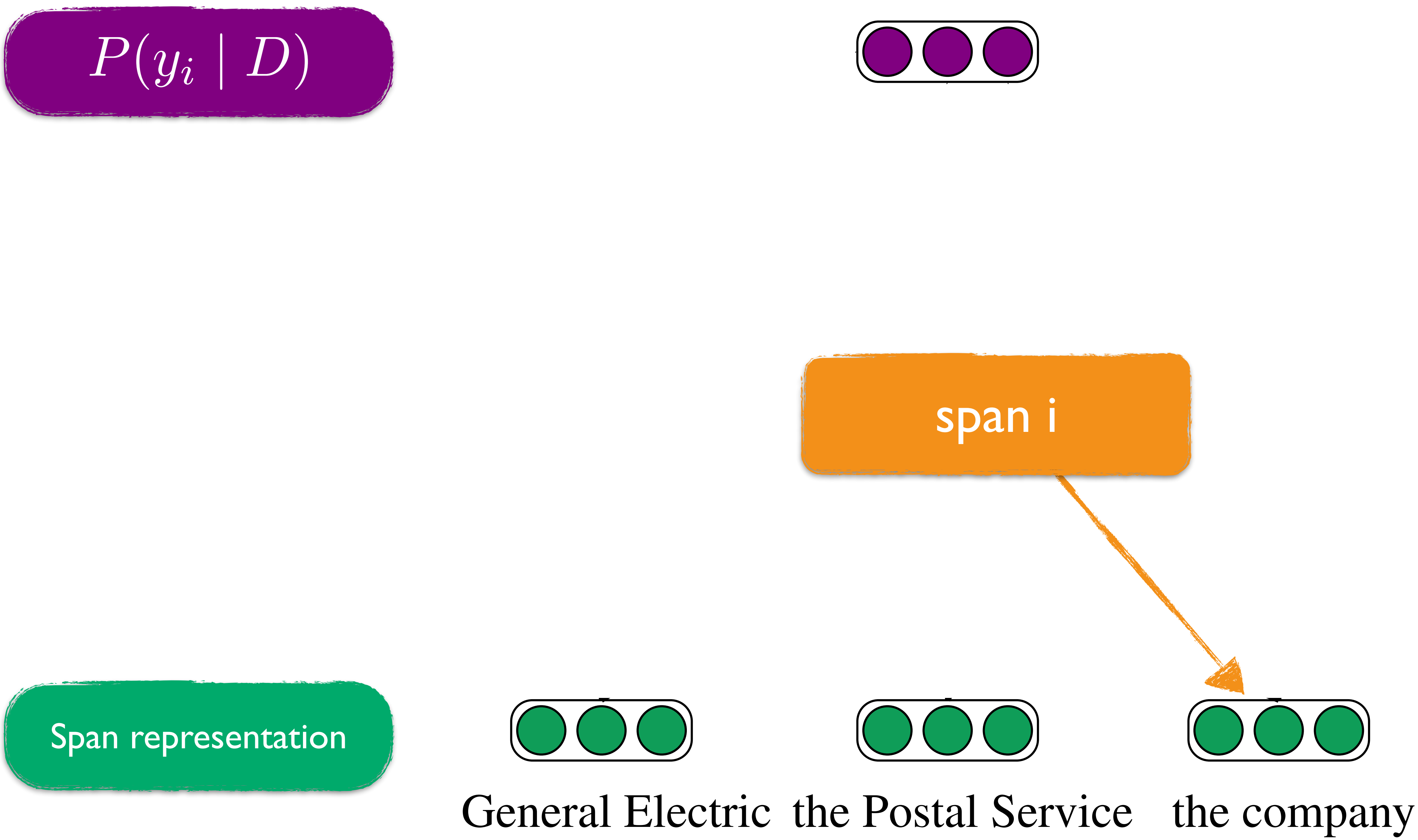
Neural Span Representations



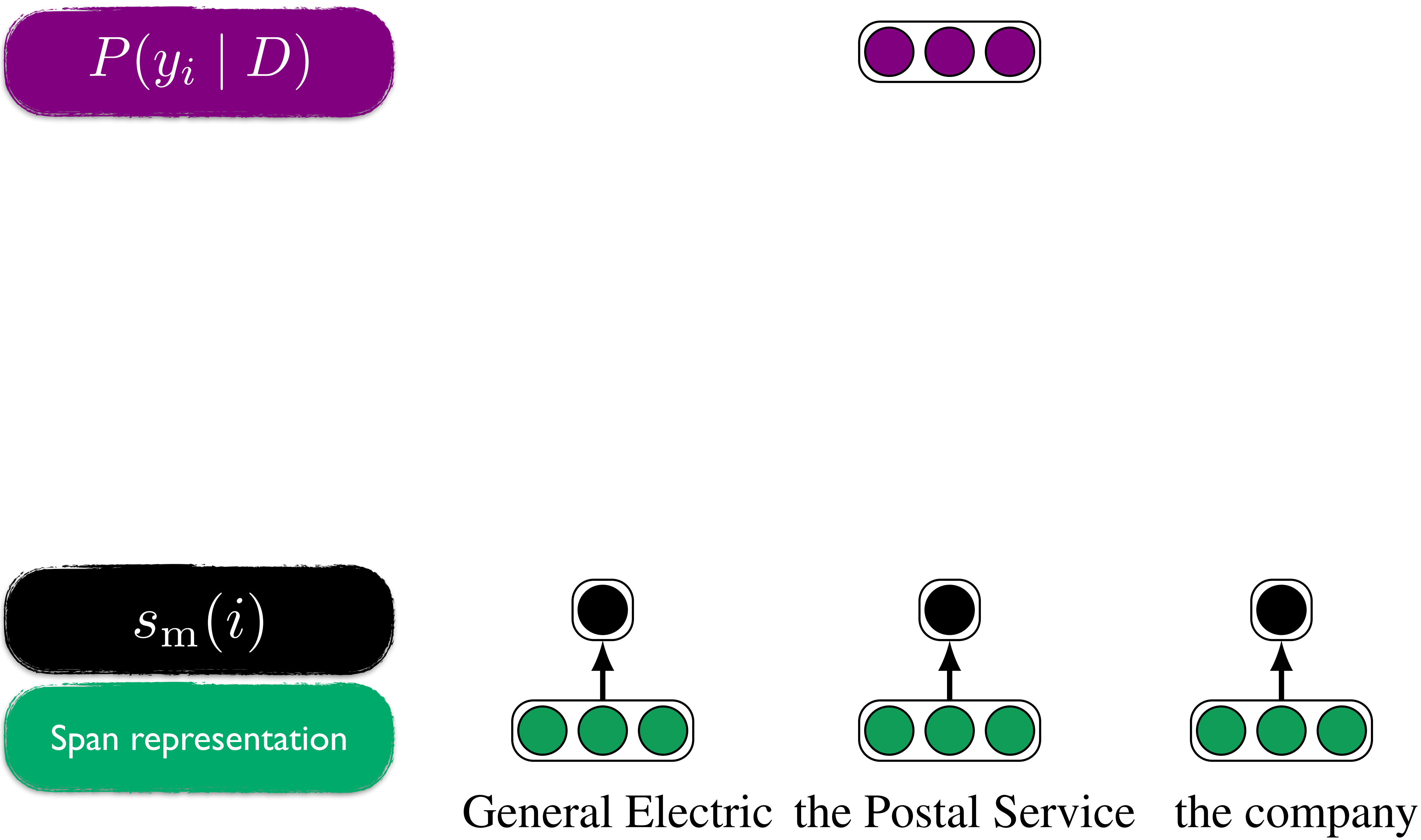
Coreference Architecture



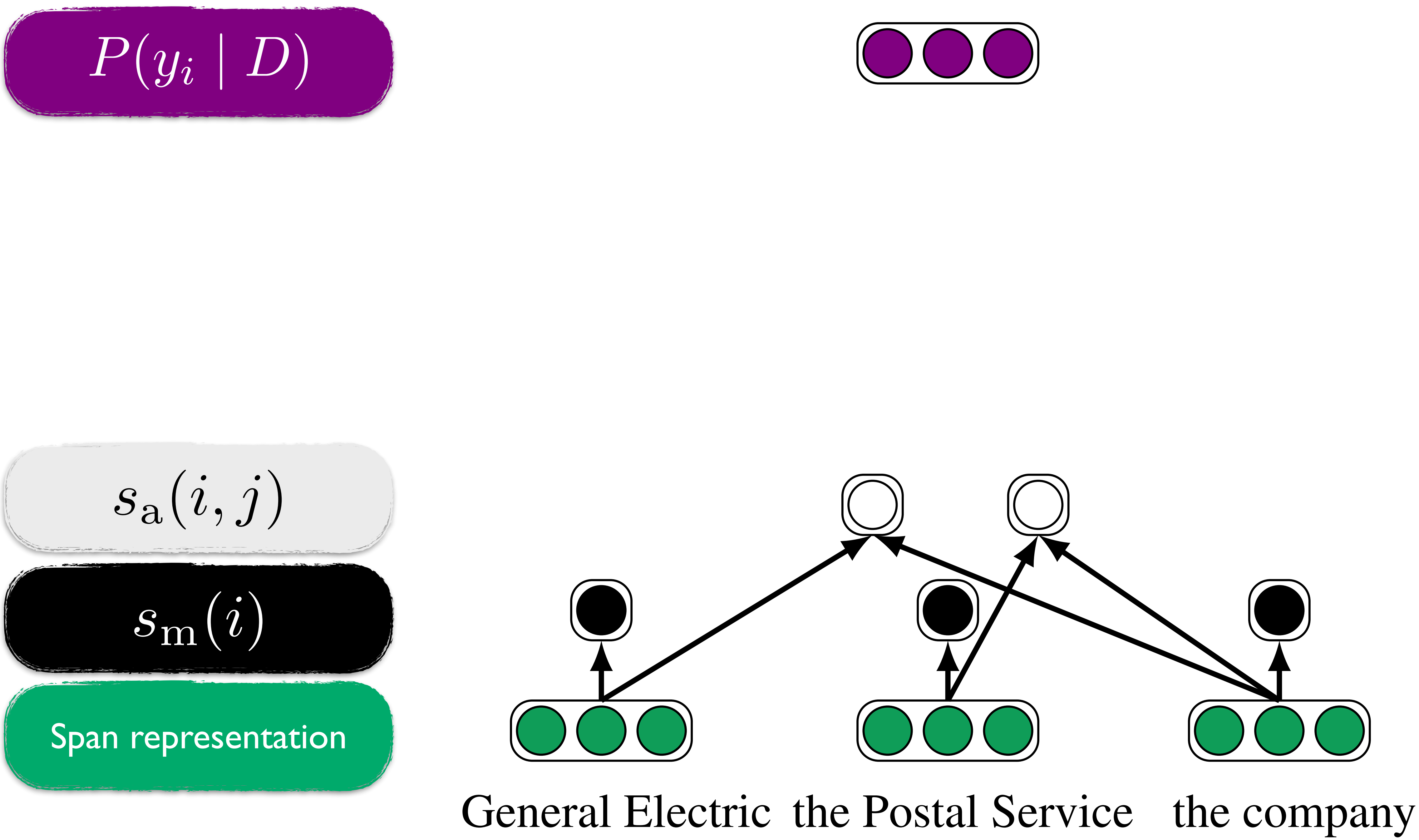
Coreference Architecture



Coreference Architecture



Coreference Architecture



Coreference Architecture

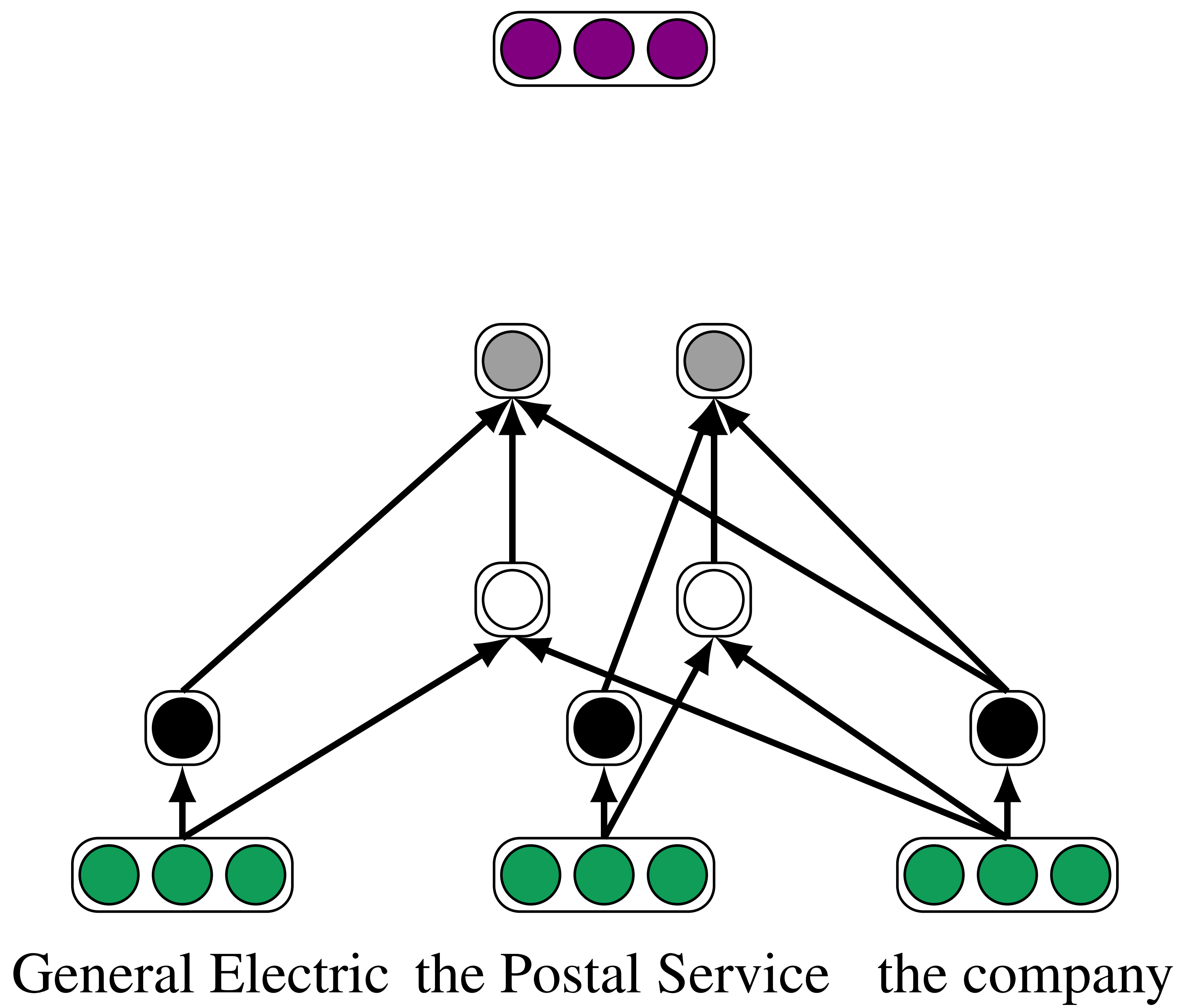
$P(y_i \mid D)$

$s(i, j)$

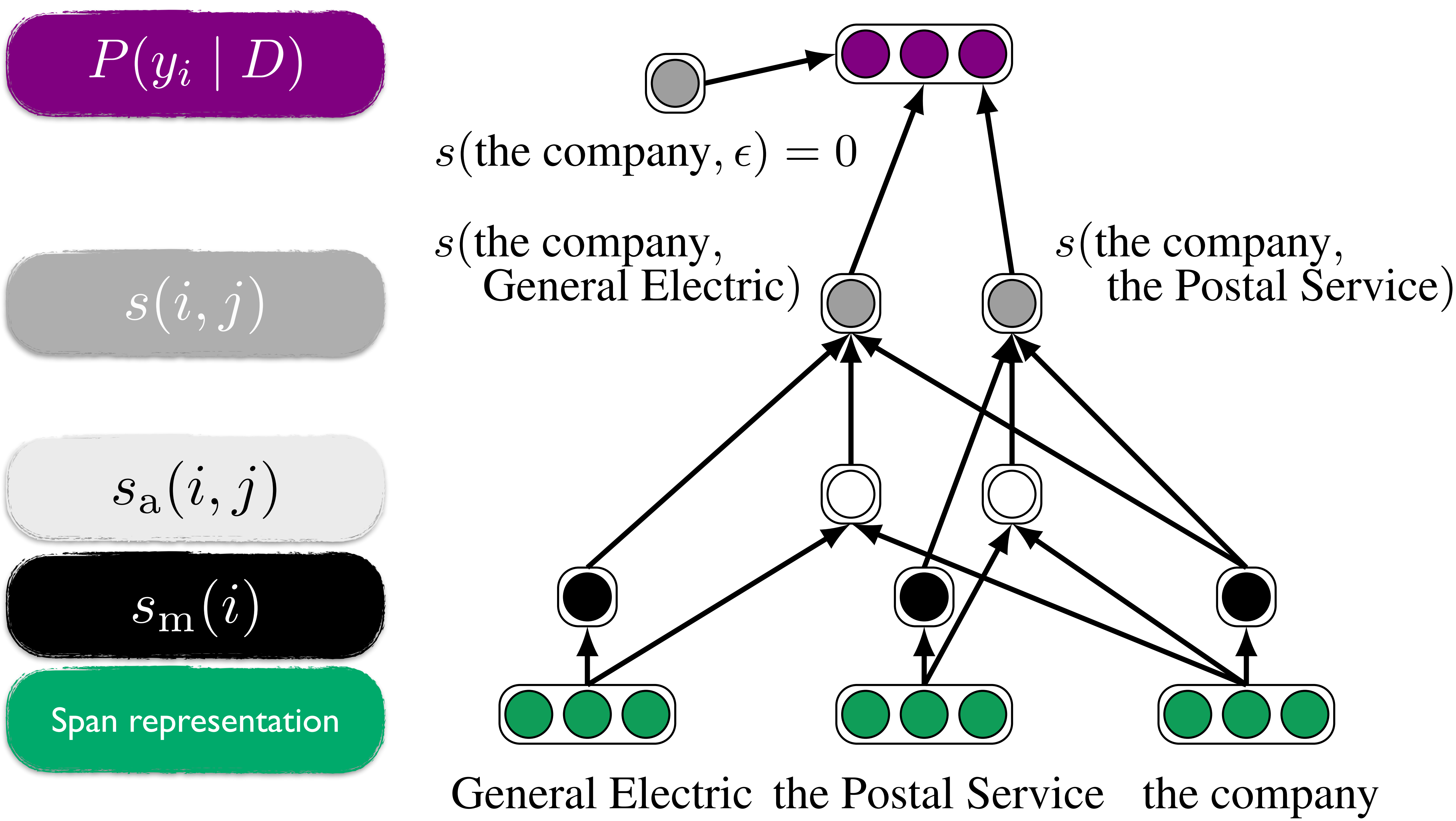
$s_{\mathrm{a}}(i, j)$

$s_{\mathrm{m}}(i)$

Span representation



Coreference Architecture



Experimental Setup

Dataset: English OntoNotes (CoNLL-2012)

Genres: Telephone conversations, newswire, newsgroups, broadcast conversation, broadcast news, weblogs

Documents: 2802 training, 343 development, 348 test

Experimental Setup

Dataset: English OntoNotes (CoNLL-2012)

Genres: Telephone conversations, newswire, newsgroups, broadcast conversation, broadcast news, weblogs

Documents: 2802 training, 343 development, 348 test

Longest document has 4009 words!

Additional pruning: Maximum span width, maximum sentence training, suppress spans with inconsistent bracketing, maximum number of antecedents

Experimental Setup

Dataset: English OntoNotes (CoNLL-2012)

Genres: Telephone conversations, newswire, newsgroups, broadcast conversation, broadcast news, weblogs

Documents: 2802 training, 343 development, 348 test

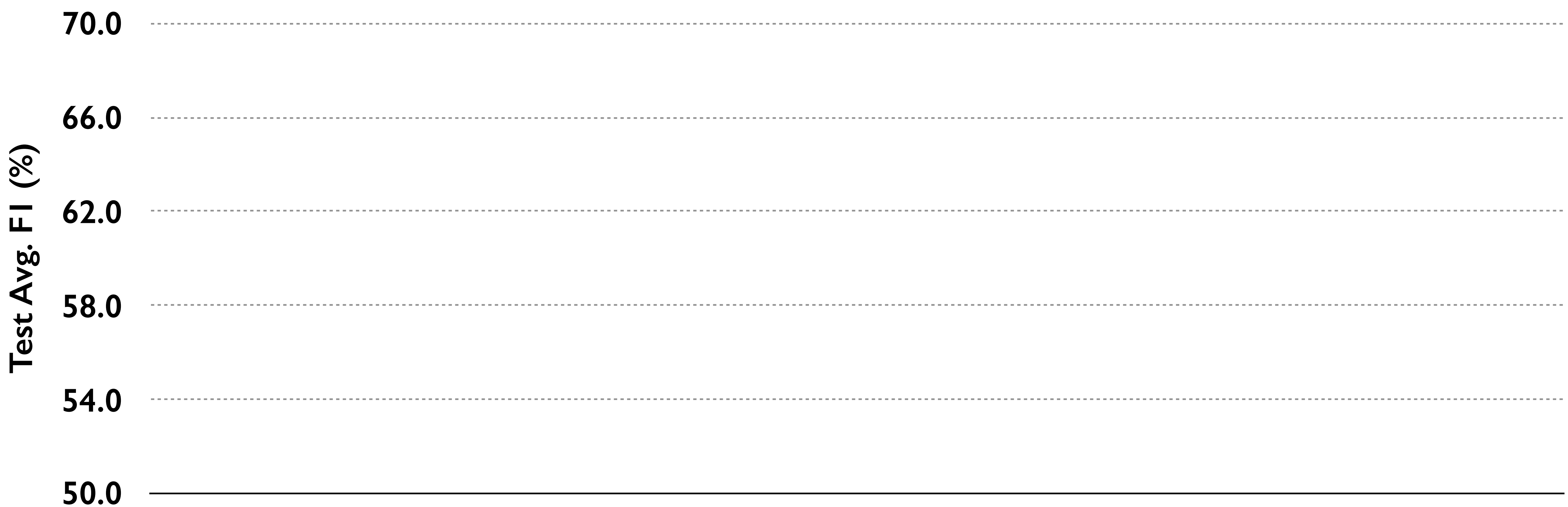
Longest document has 4009 words!

Additional pruning: Maximum span width, maximum sentence training, suppress spans with inconsistent bracketing, maximum number of antecedents

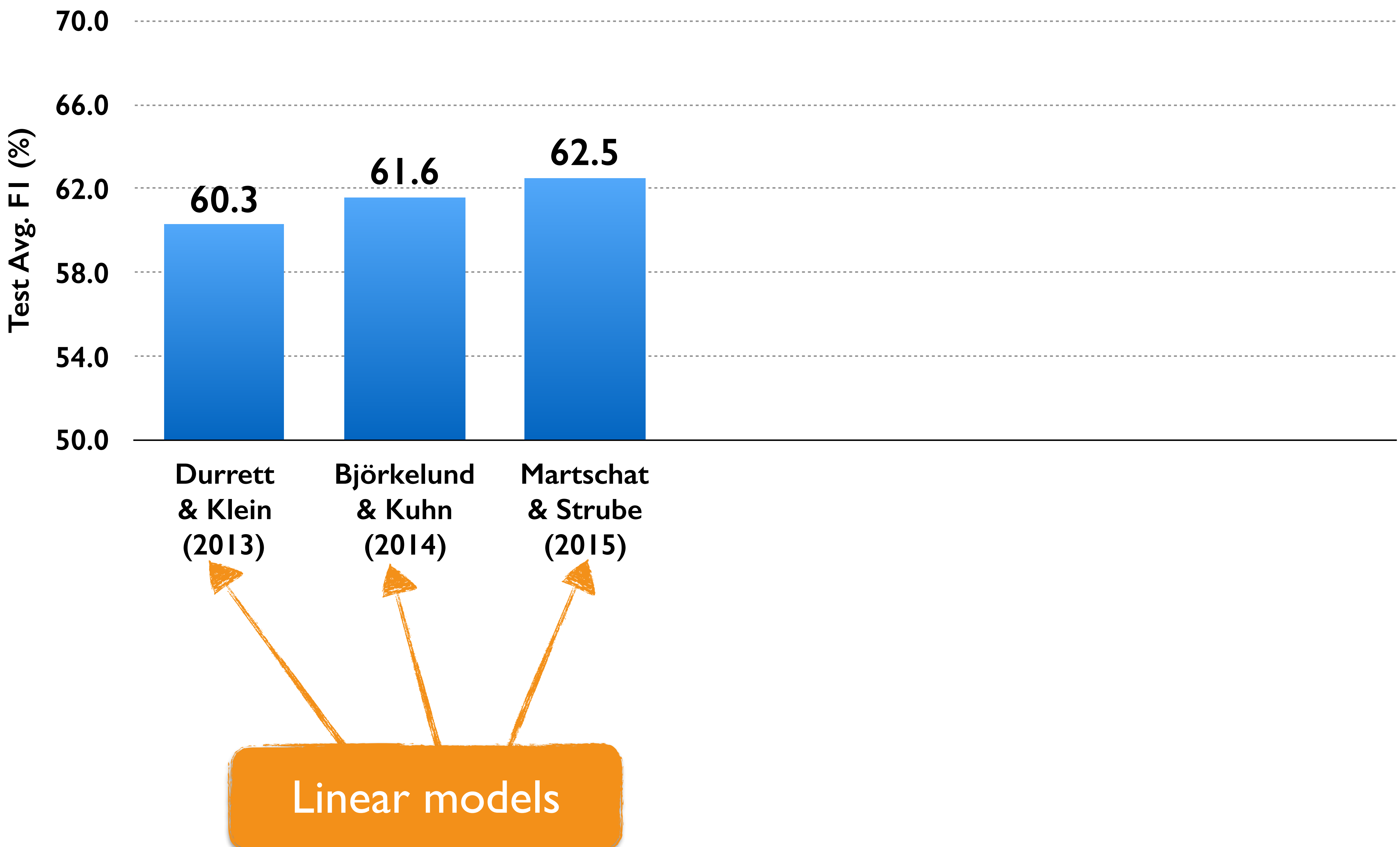
Features: distance between spans, span width

Metadata: speaker information, genre

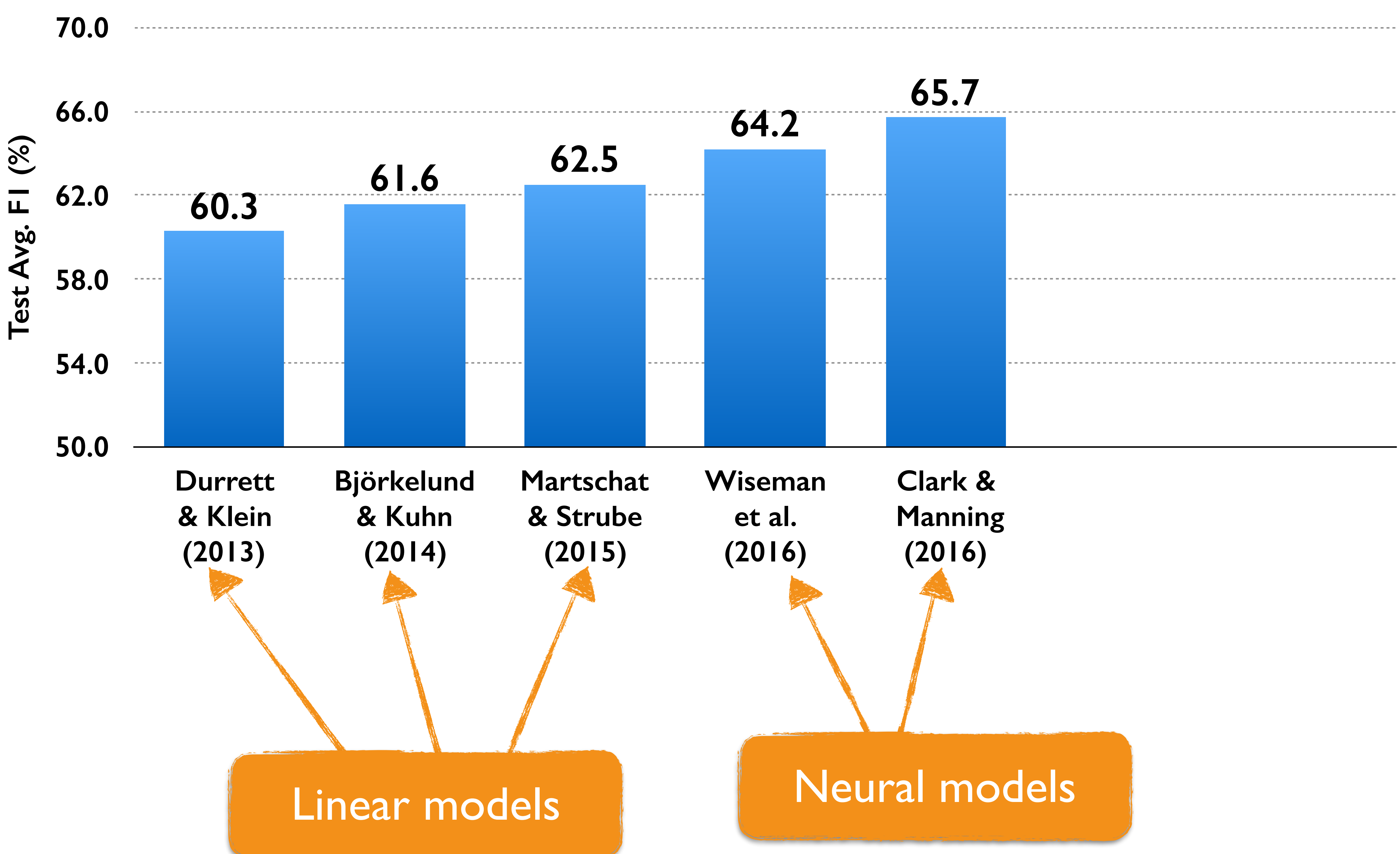
Coreference Results



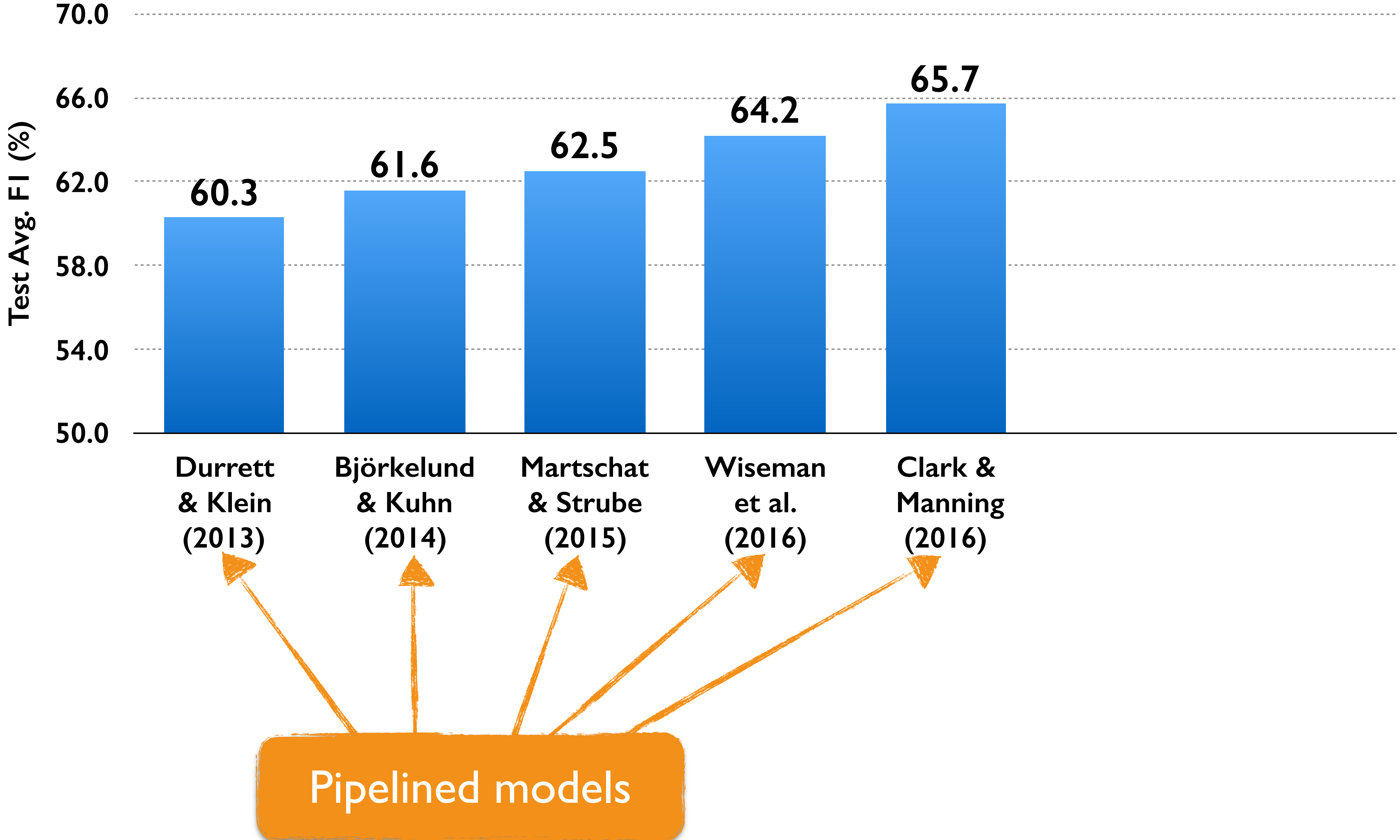
Coreference Results



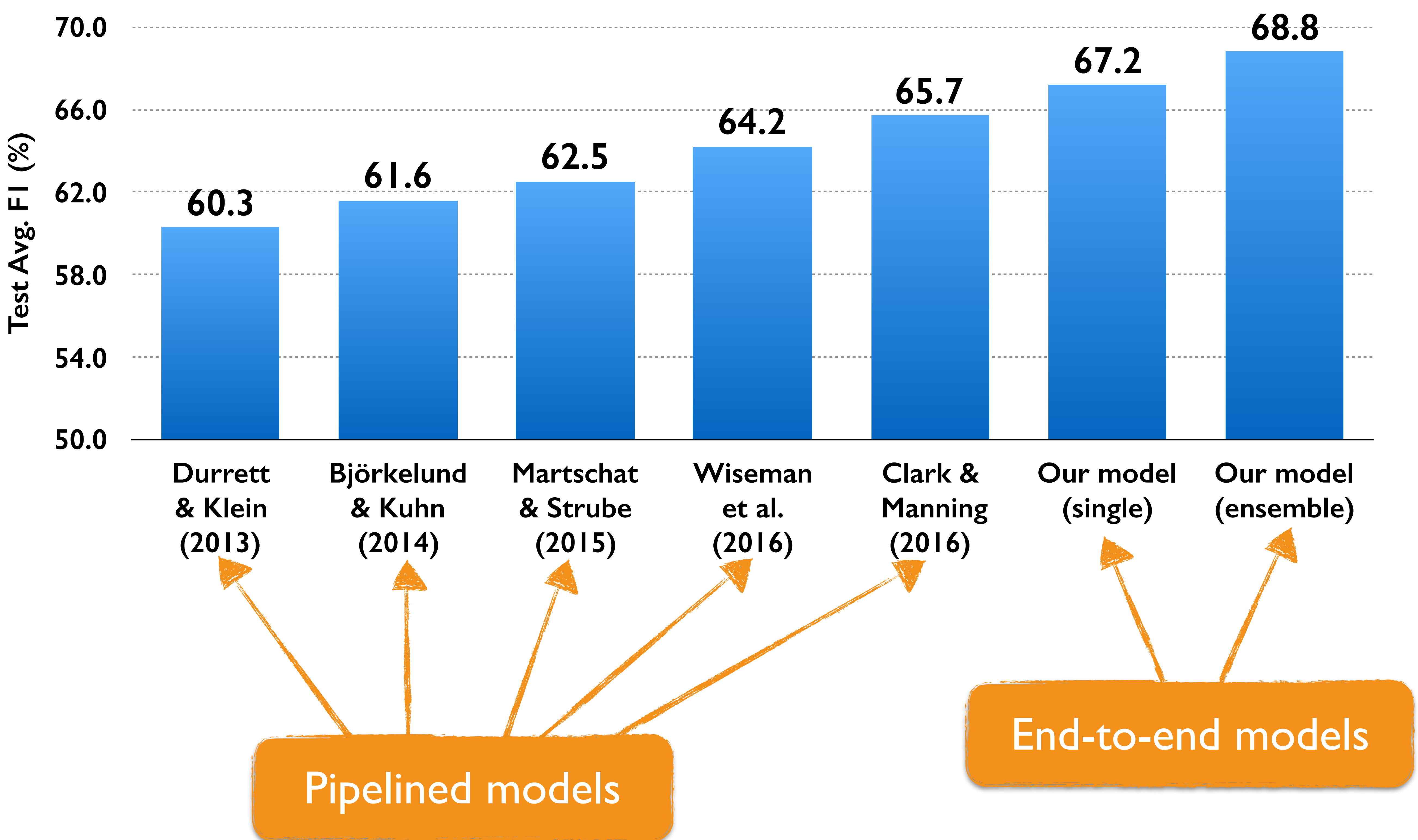
Coreference Results



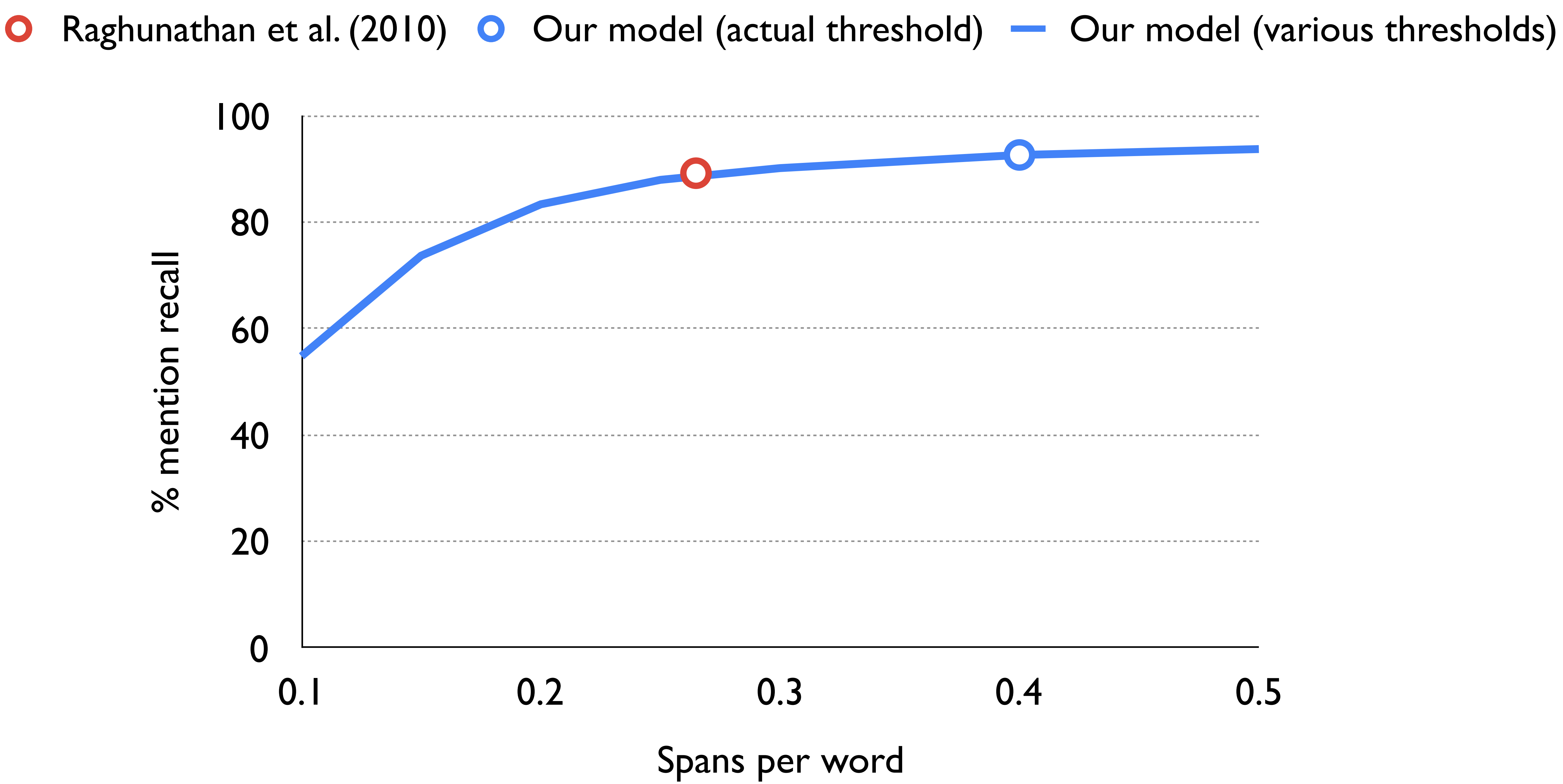
Coreference Results



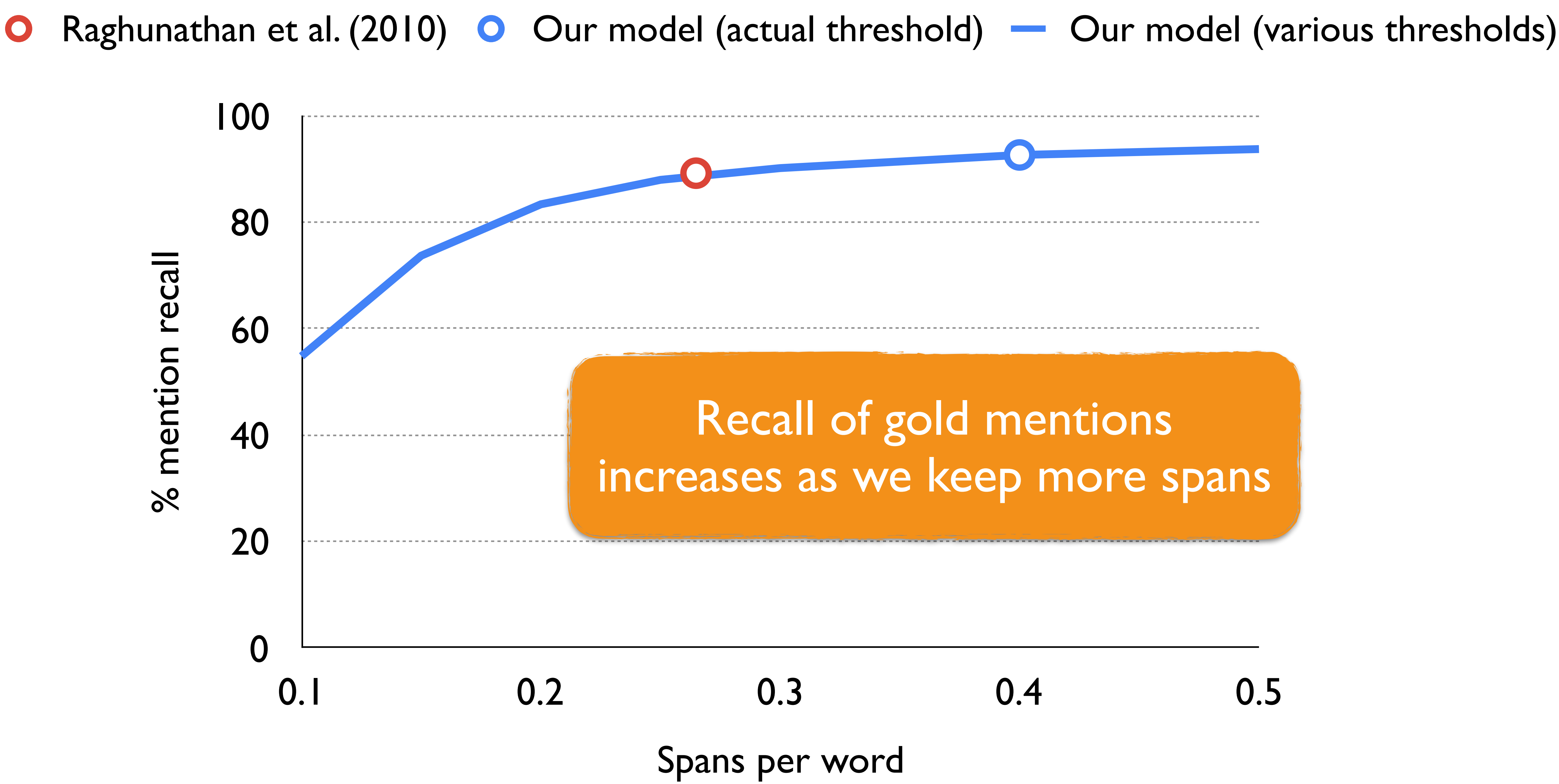
Coreference Results



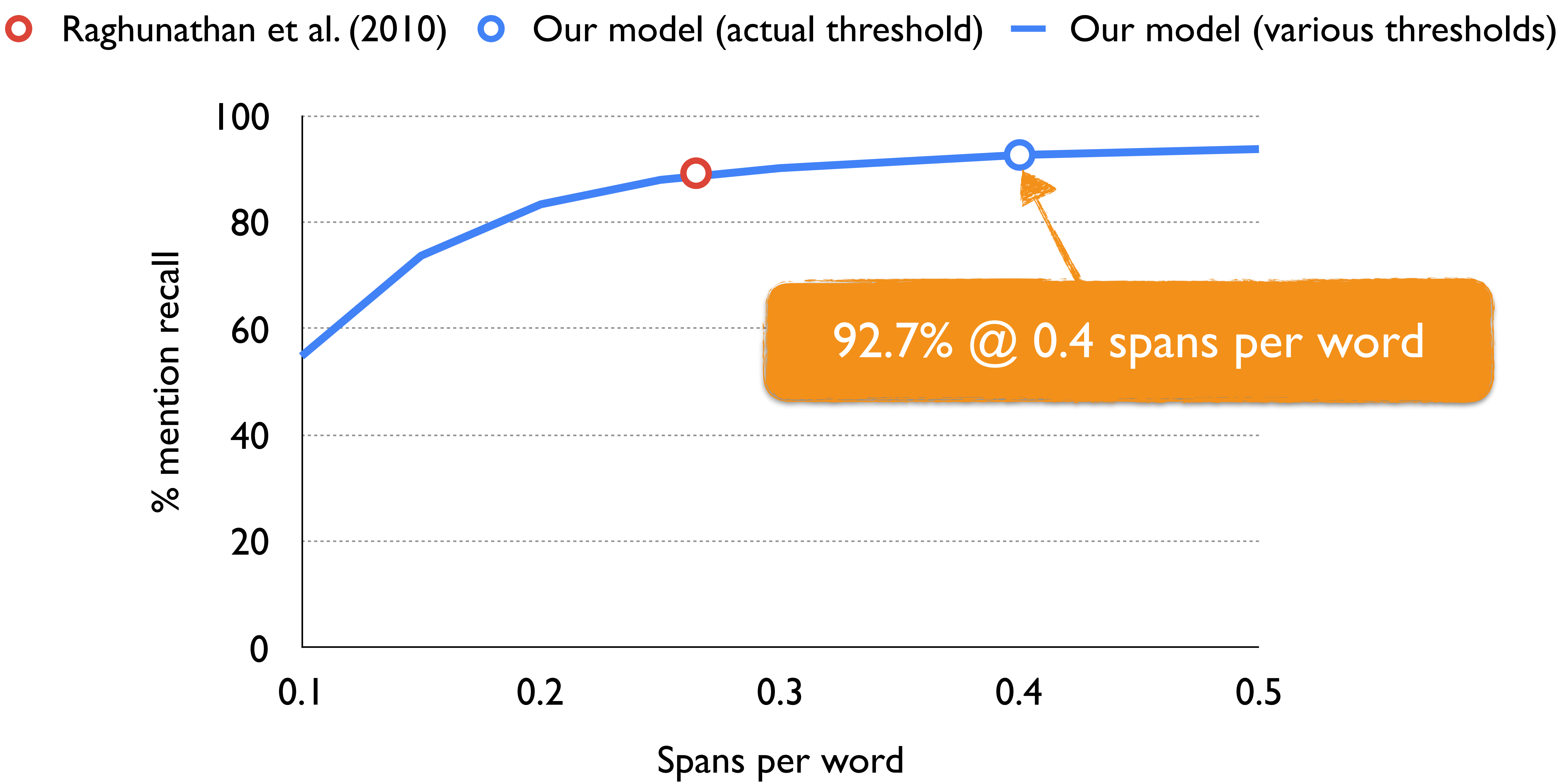
Mention Recall



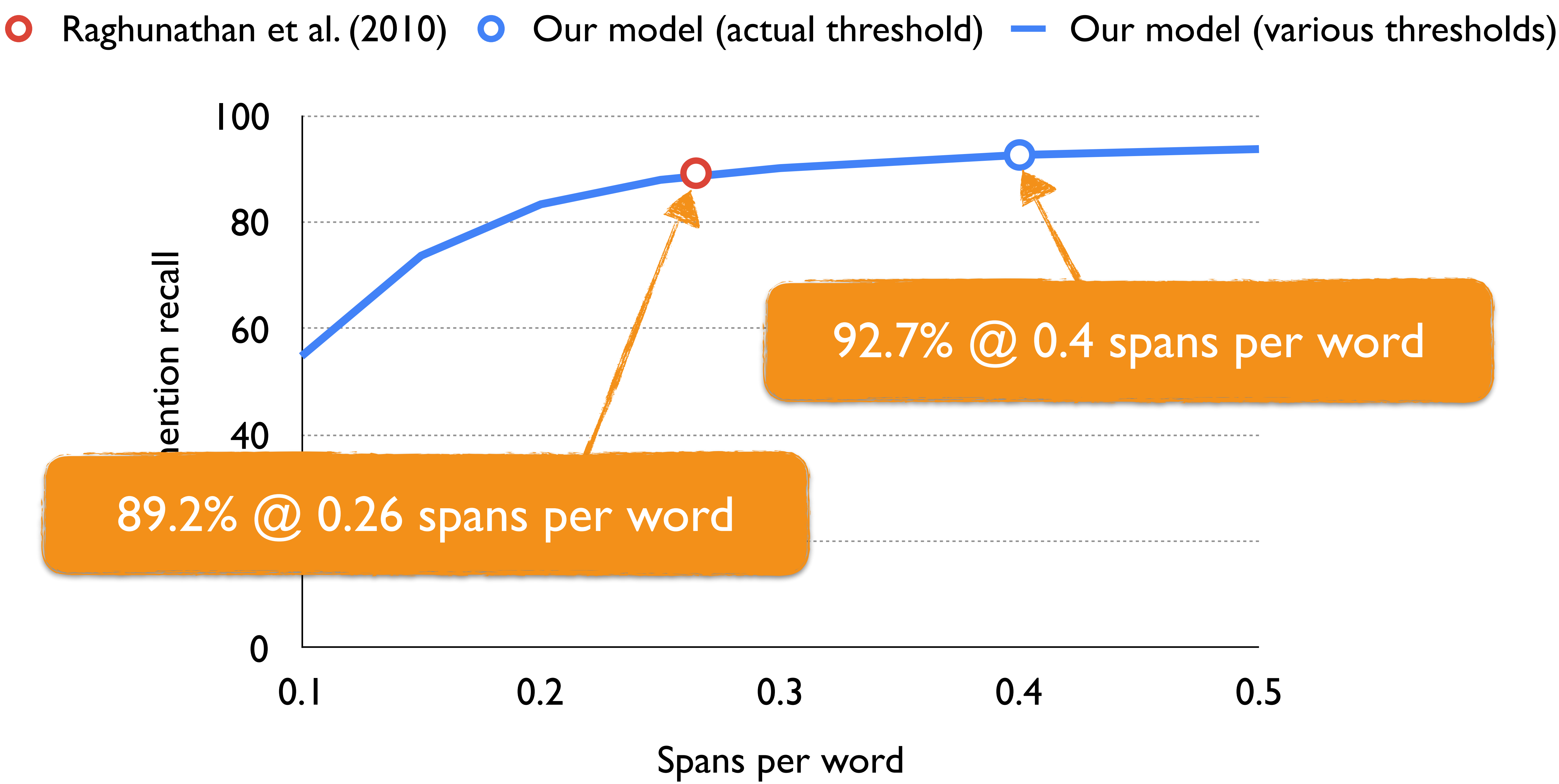
Mention Recall



Mention Recall



Mention Recall



Qualitative Analysis

A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building.

Qualitative Analysis

$\boxed{}$: Mention in a predicted cluster

$\boxed{\text{A fire in a Bangladeshi garment factory}}$ has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee $\boxed{\text{the blaze}}$ in the four-story building.

Qualitative Analysis

: Mention in a predicted cluster

: Head-finding attention weight

A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building.

Qualitative Analysis

: Mention in

: Head-findin

Attention-based head finder facilitates soft similarity cues

A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building.

Qualitative Analysis

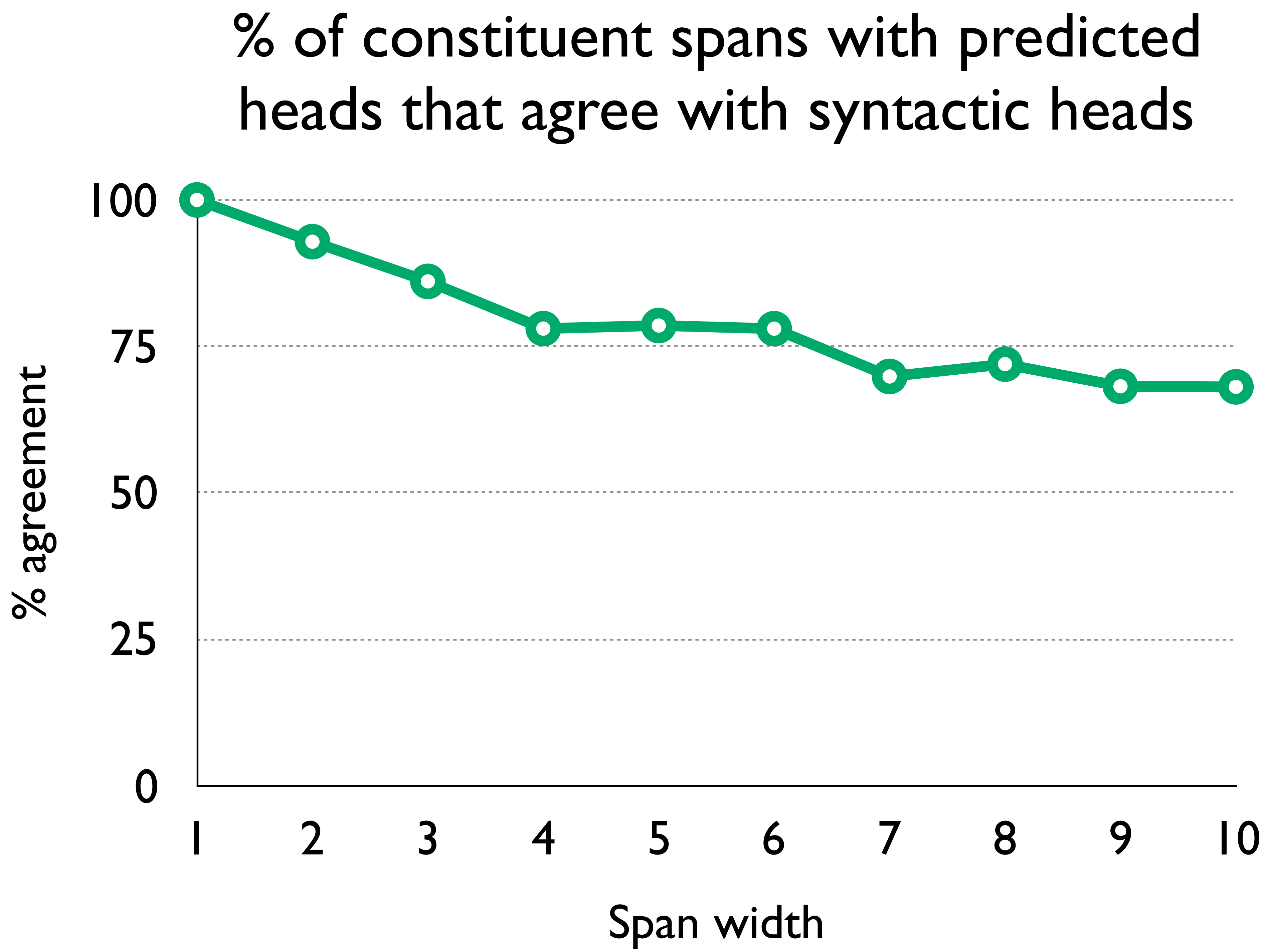
: Mention in a predicted cluster

: Hea

Good head-finding requires word-order information!

A fire in a Bangladeshi garment factory has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee the blaze in the four-story building.

Head-finding Agreement



Common Error Case

: Mention in a predicted cluster

: Head-finding attention weight

The flight attendants have until 6:00 today to ratify labor concessions. The pilots' union and ground crew did so yesterday.

Common Error Case

: Mention in a predicted cluster

: Head-finding attention weight

The flight attendants have until 6:00 today to ratify labor concessions. The pilots' union and ground crew did so yesterday.

Conflating **relatedness** with **paraphrasing**

Conclusion

- State-of-the-art end-to-end coreference resolver
  - Scalable inference
  - Learns latent mentions and heads
  - https://github.com/kentonl/e2e-coref

Conclusion

- State-of-the-art end-to-end coreference resolver
  - Scalable inference
  - Learns latent mentions and heads
  - https://github.com/kentonl/e2e-coref
- Relatively simplistic model:
  - Doesn't explicitly model clusters
  - Lacks discourse reasoning and world knowledge
  - Still a long way to go!