PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

UWNLP

Jointly Predicting **Predicates** and **Arguments** in Neural Semantic Role Labeling

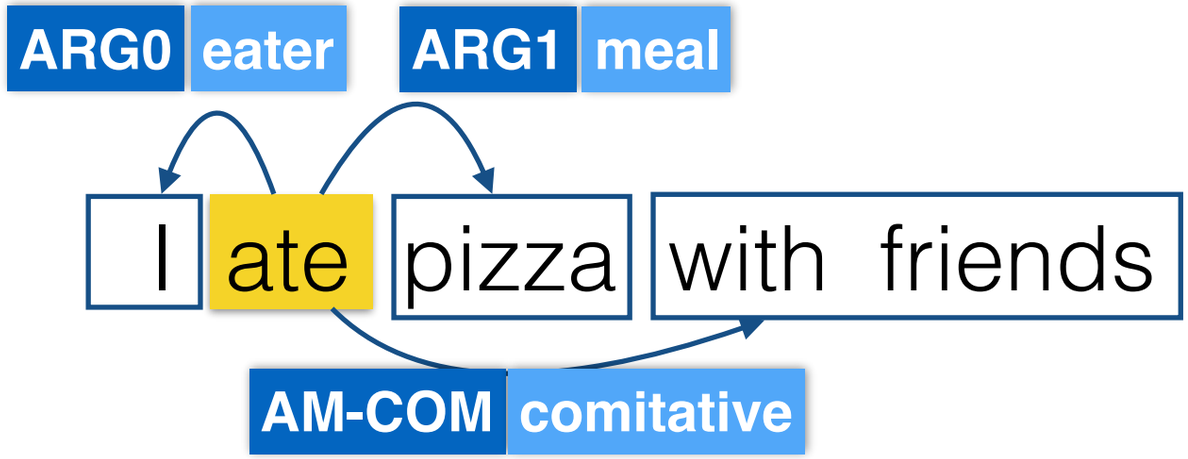
Luheng He*, Kenton Lee*, Omer Levy+, Luke Zettlemoyer
University of Washington

*Now at Google

+Now at Facebook AI Research

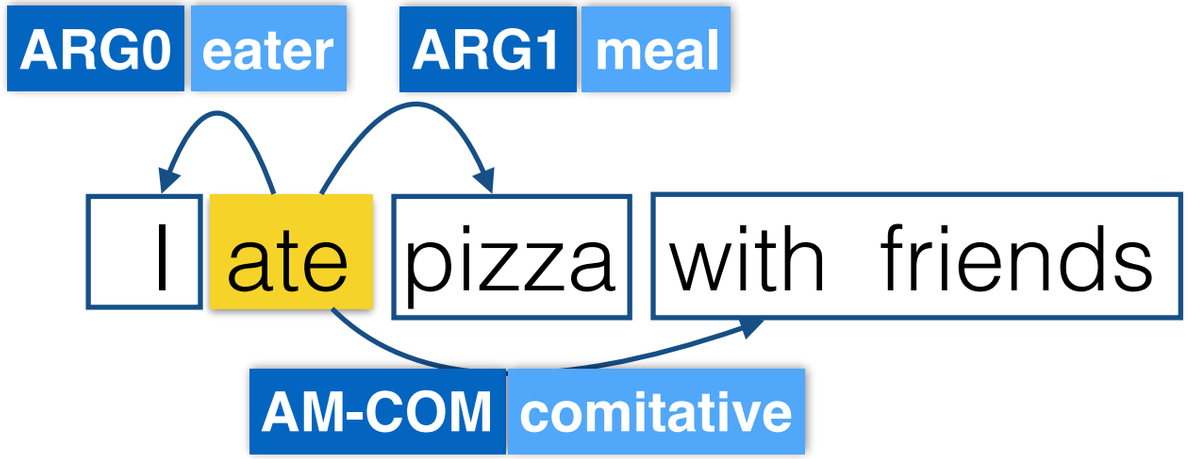
Semantic Role Labeling (SRL)

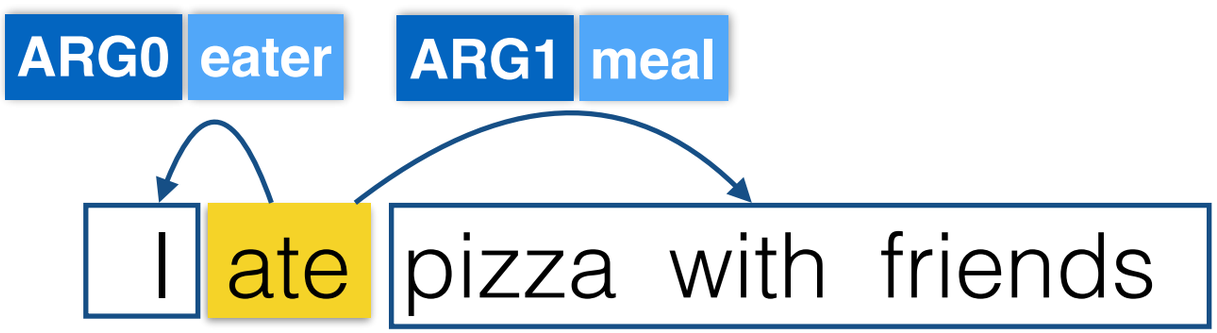
- Find out "**who did what to whom**" in text
- Capture **predicate-argument** structures



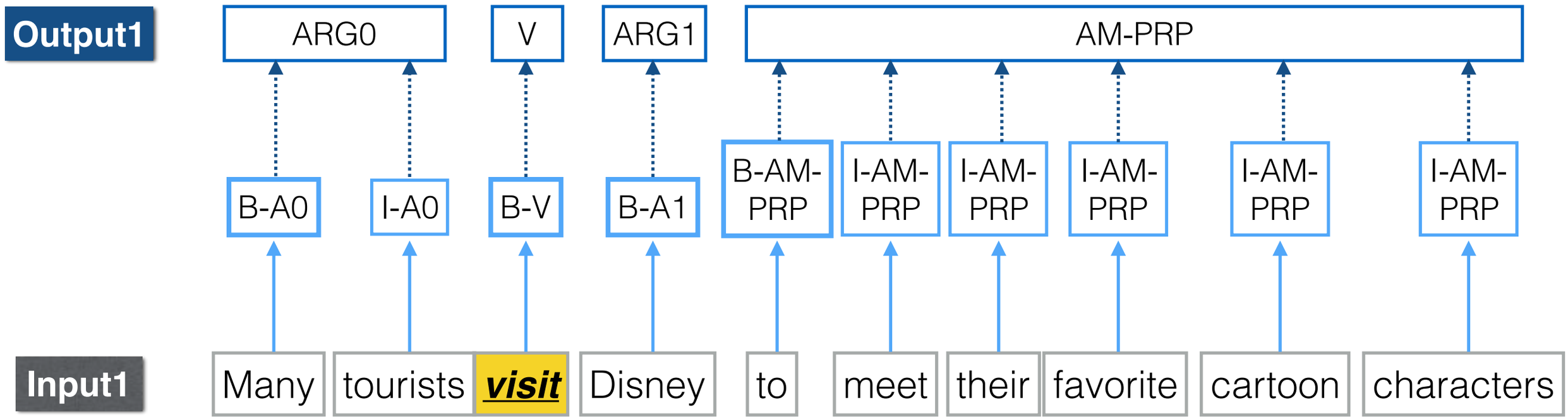
Semantic Role Labeling (SRL)

- Find out "**who did what to whom**" in text
- Capture **predicate-argument** structures





SRL as BIO Tagging

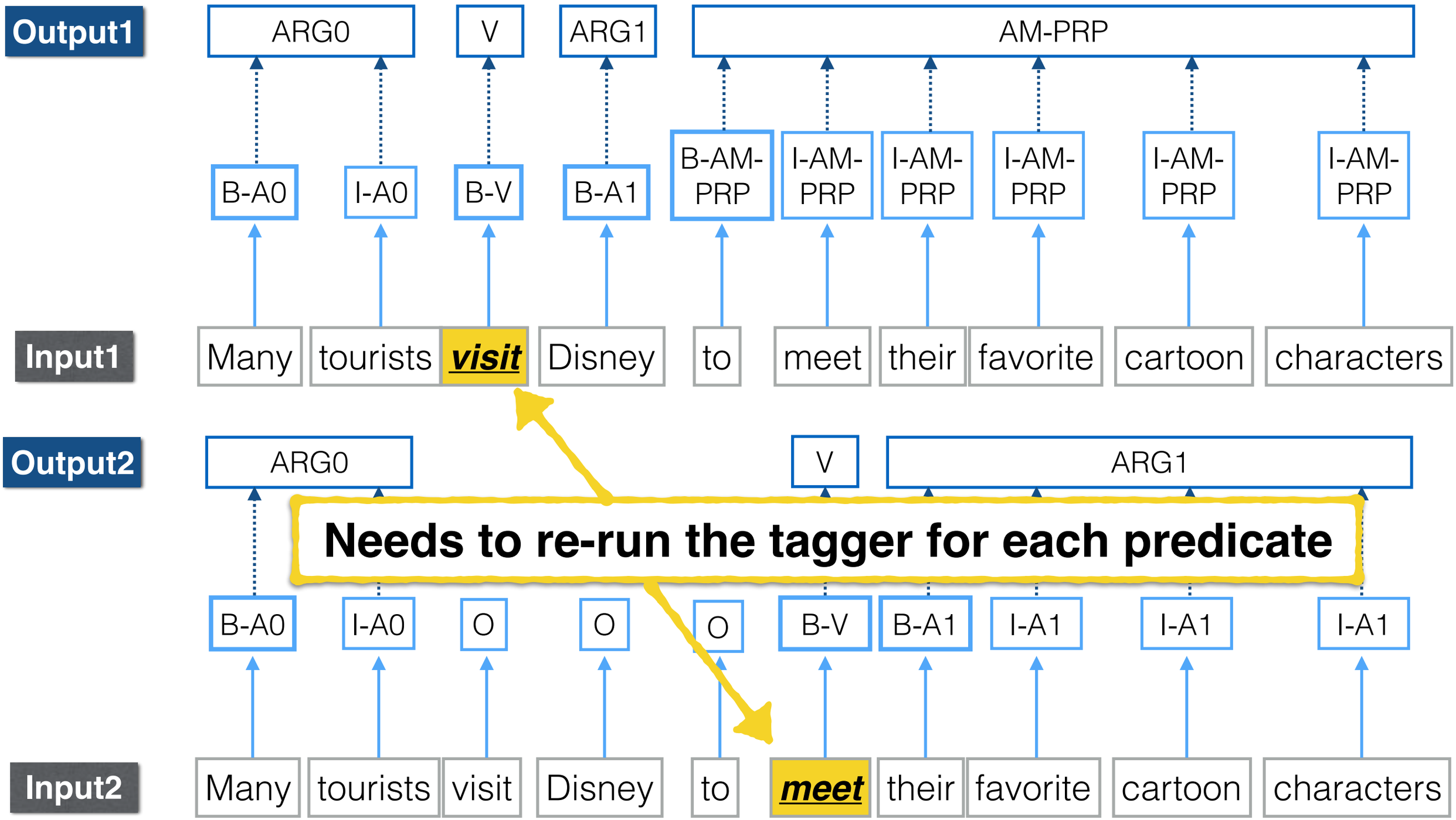


Needs target predicate as input!

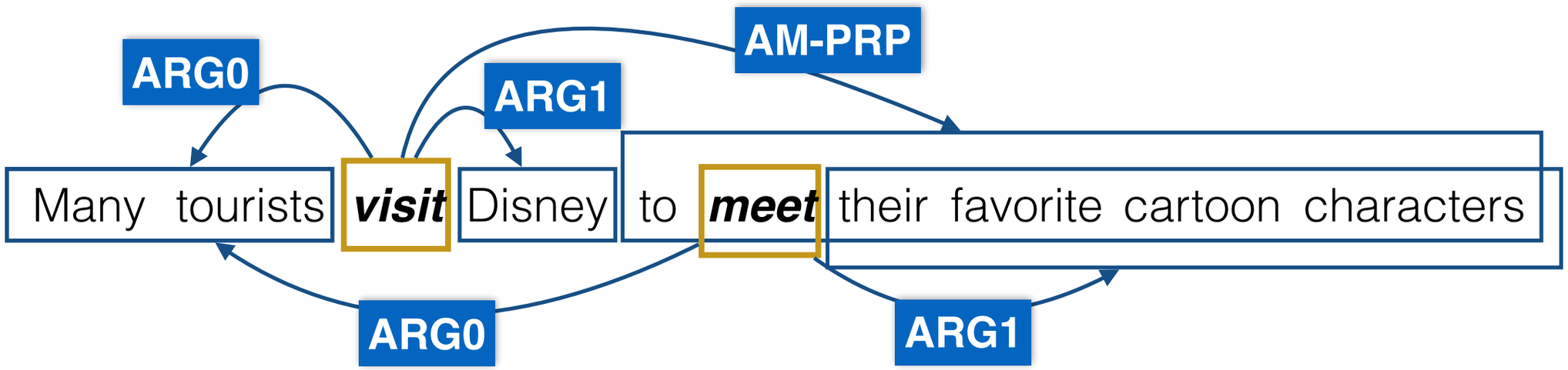
(Prior works typically used gold predicates)

Collobert et al., 2011,
Zhou and Xu, 2015,
He et. al, 2017,
inter alia

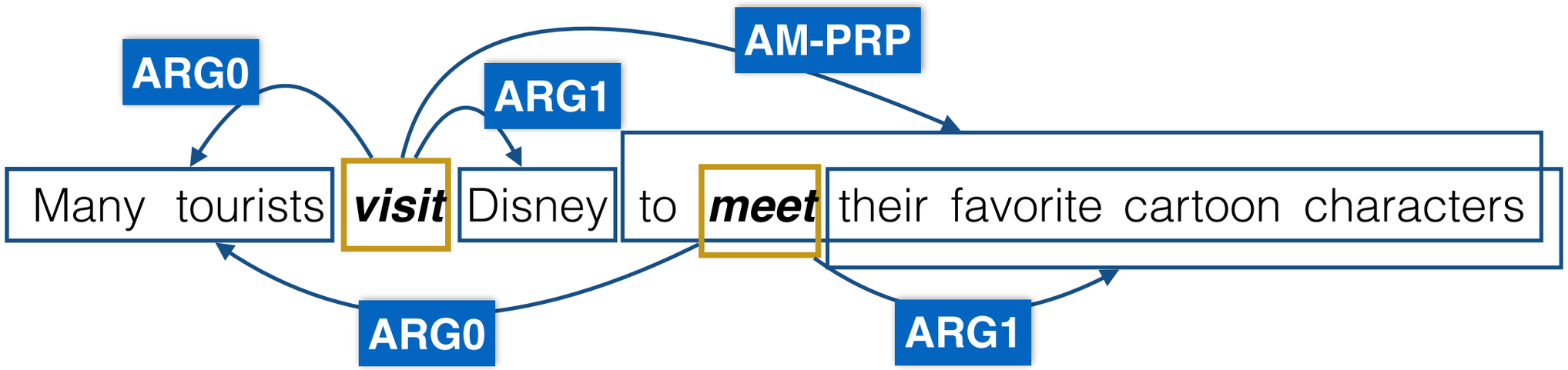
SRL as BIO Tagging



SRL as Predicting Word-Span Relations



SRL as Predicting Word-Span Relations



Advantages:

* Jointly predict predicates
* Span-level features
  (similar to Punyakanok08, FitzGerald15, inter alia)

Challenge:

* Too many possible edges (n^2 argument spans x n predicates)

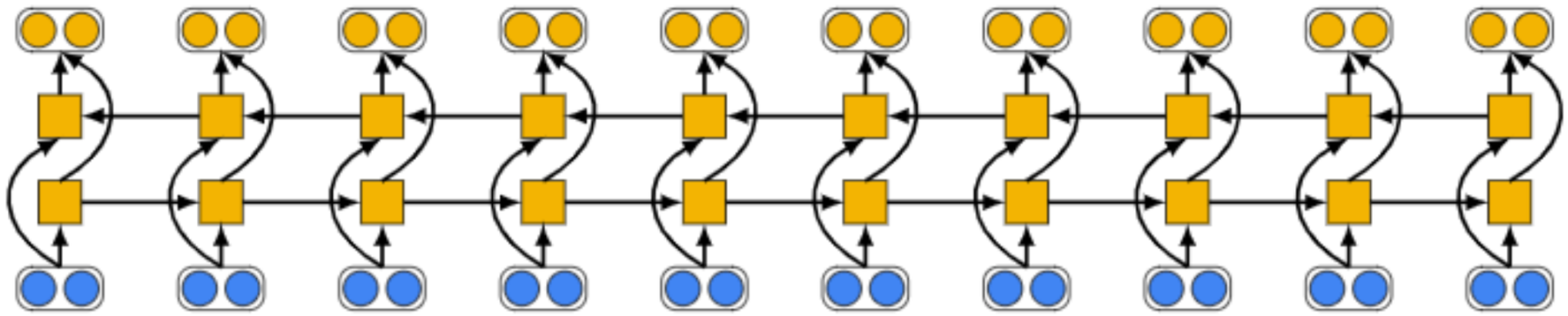
Our Model

Our Model: Overview

Highway BiLSTMs

Word & Char Embeddings

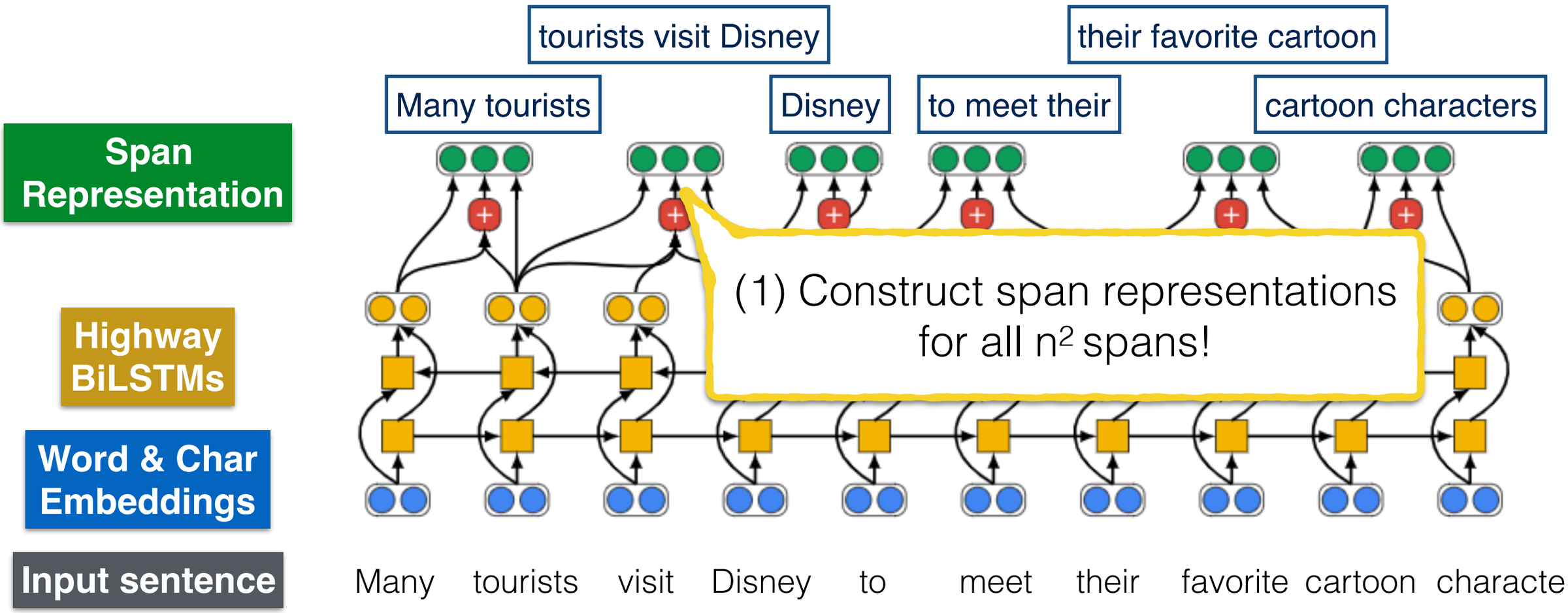
Input sentence



Many tourists visit Disney to meet their favorite cartoon characters

No predicate input!

Our Model: Overview



Our Model: Overview

Labeling Softmax

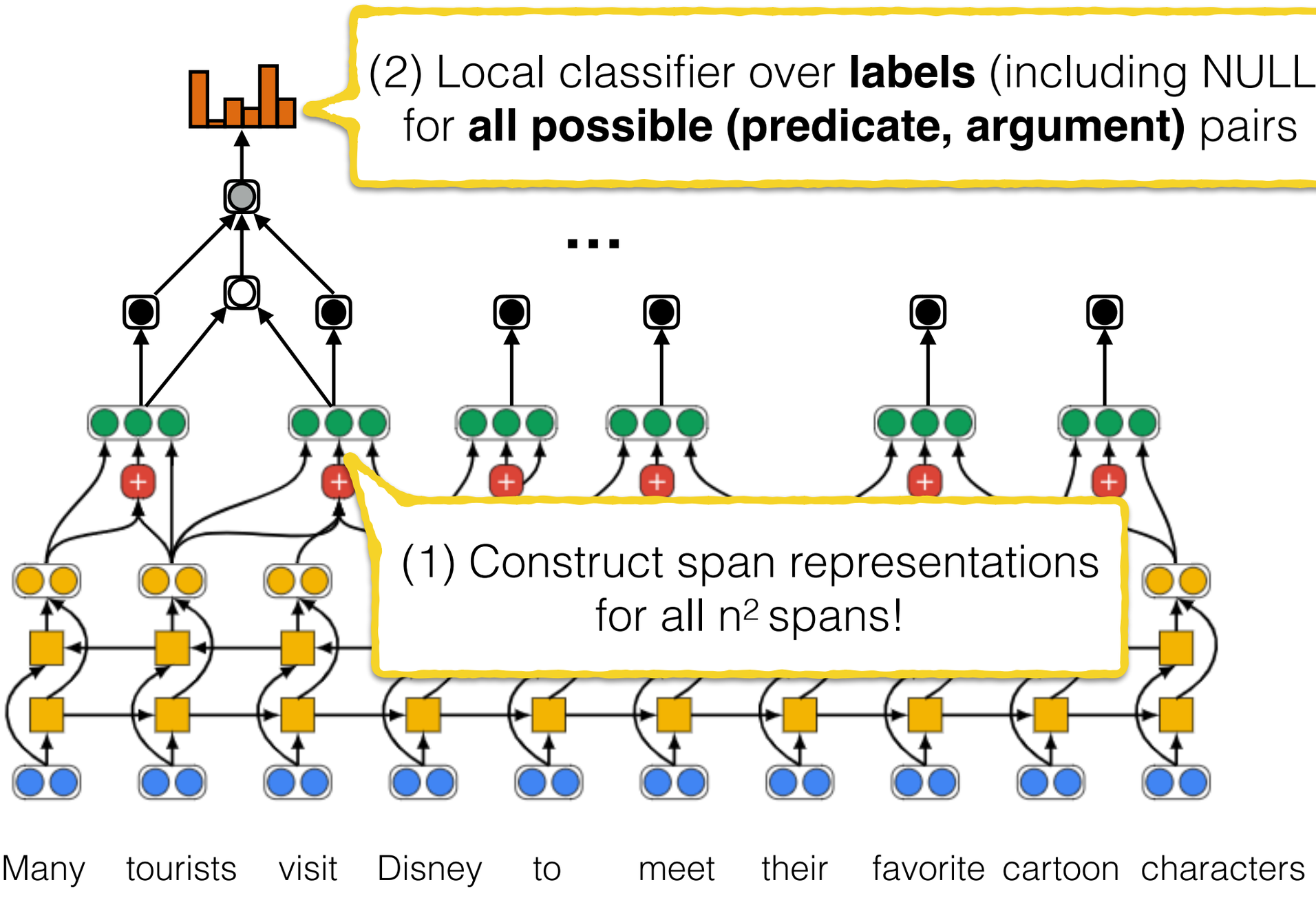
Node & Edge Scores

Span Representation

Highway BiLSTMs

Word & Char Embeddings

Input sentence



Our Model: Overview

Labeling Softmax

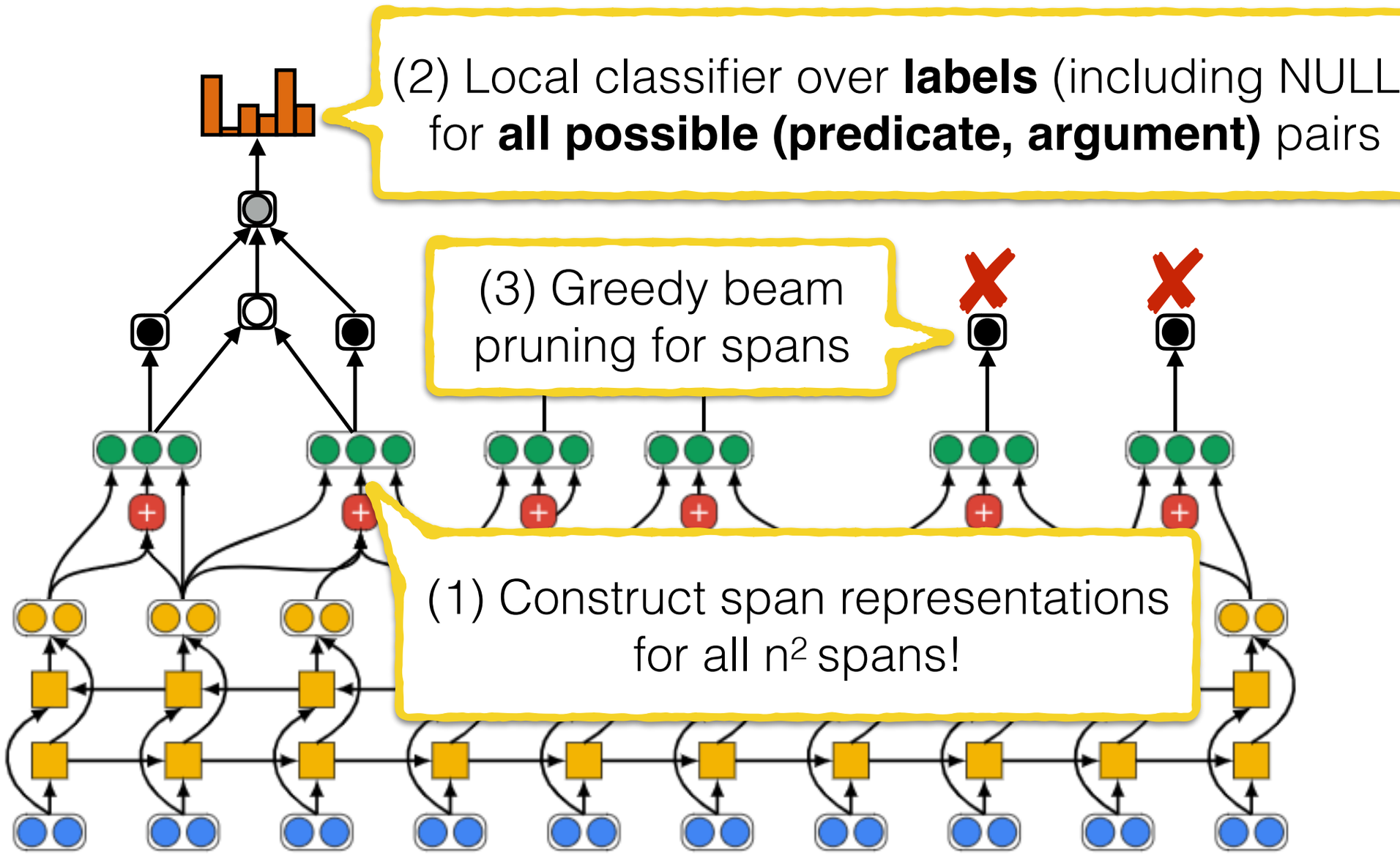
Node & Edge Scores

Span Representation

Highway BiLSTMs

Word & Char Embeddings

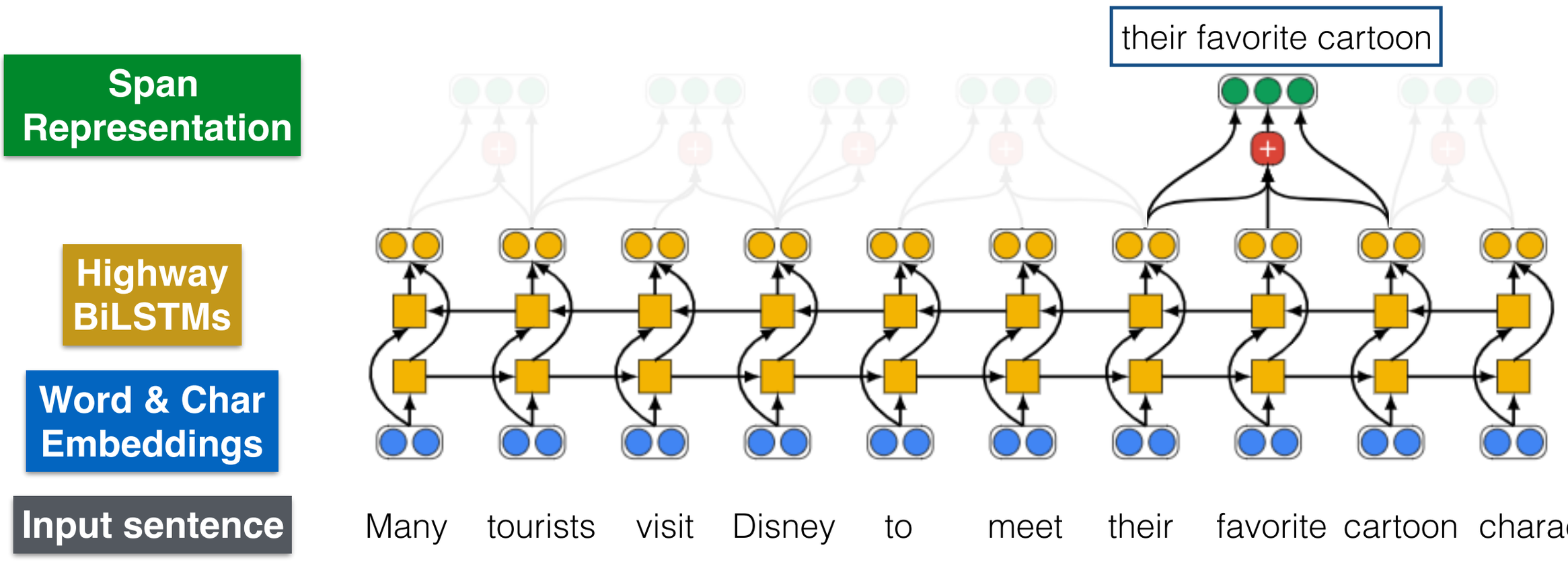
Input sentence



(1) Span Representations

(2) Local Label Classifiers

(3) Span Pruning



(Same as Lee et al., 2017)

(1) Span Representations

(2) Local Label Classifiers

(3) Span Pruning

LSTM boundary points

$$[\text{BiLSTM}(w_1 : w_n)_{\text{START}}, \text{BiLSTM}(w_1 : w_n)_{\text{END}}]$$

Span Representation

their favorite cartoon

left context

right context

Highway BiLSTMs

Word & Char Embeddings

Input sentence

Many tourists visit Disney to meet their favorite cartoon characters

(Same as Lee et al., 2017)

(1) Span Representations

(2) Local Label Classifiers

(3) Span Pruning

LSTM boundary points

Attention over words

$$[\textsc{BiLstm}(w_1 : w_n)_{\textsc{Start}}, \textsc{BiLstm}(w_1 : w_n)_{\textsc{End}}]$$

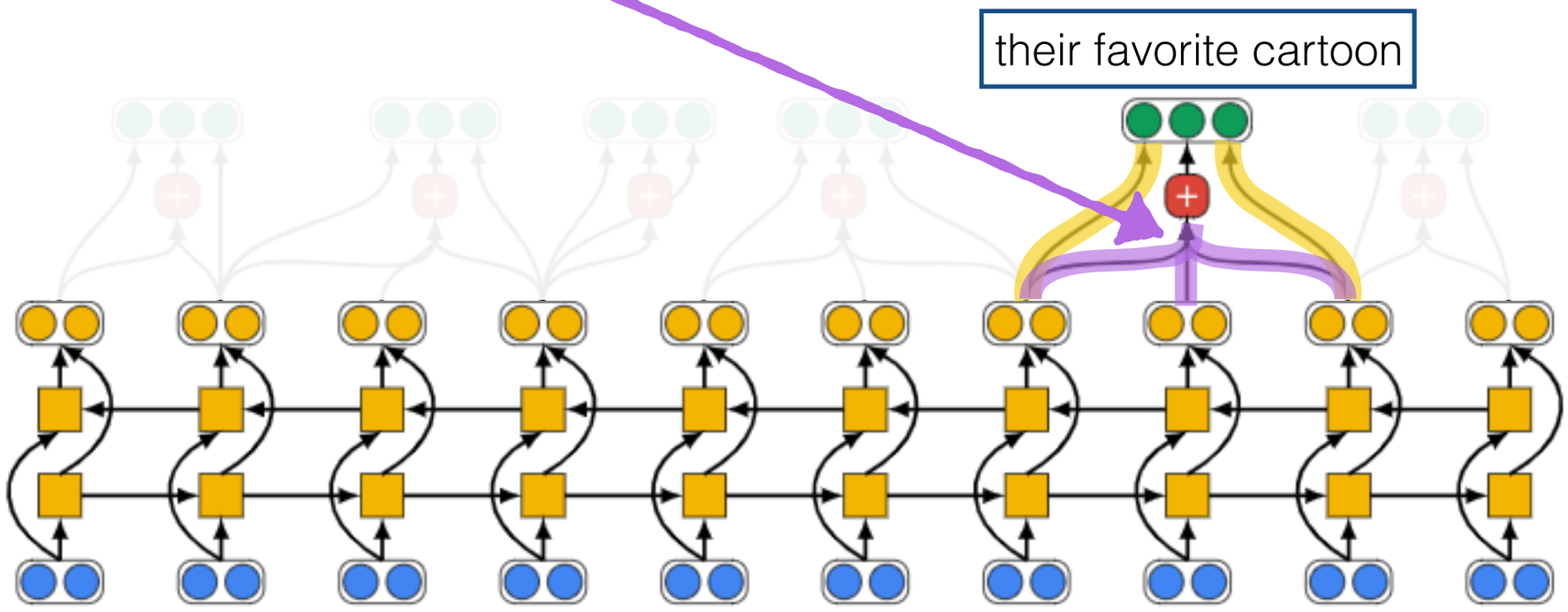
$$\sum_{i=\textsc{Start}}^{\textsc{End}} \textsc{SoftMax}(a_{\textsc{Start}} : a_{\textsc{End}})_i \; w_i$$

Span Representation

Highway BiLSTMs

Word & Char Embeddings

Input sentence



(Same as Lee et al., 2017)

(1) Span Representations

(2) Local Label Classifiers

(3) Span Pruning

Labeling Softmax

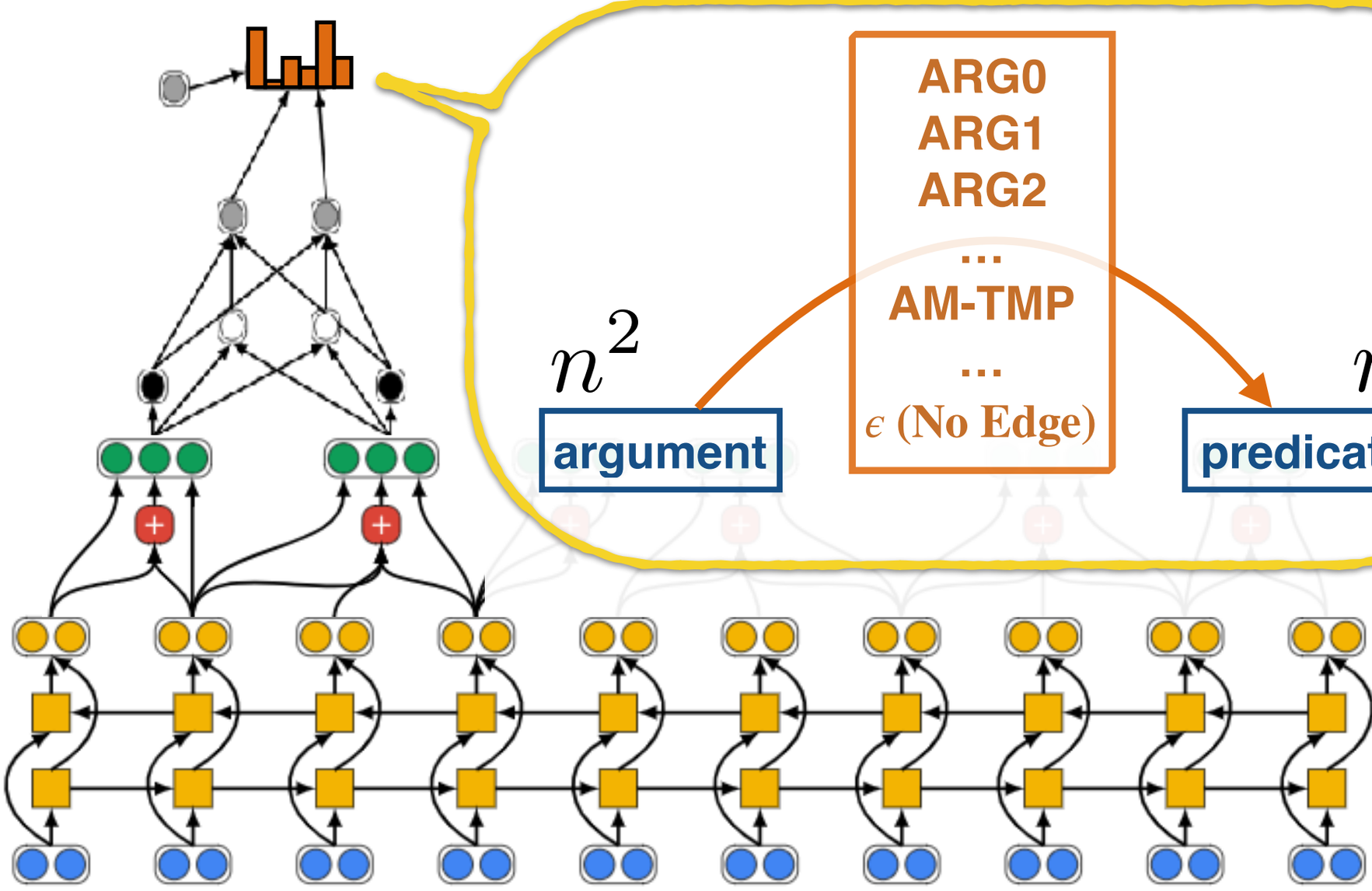
Node & Edge Scores

Span Representation

Highway BiLSTMs

Word & Char Embeddings

Input sentence



(1) Span Representations

(2) Local Label Classifiers

(3) Span Pruning

Labeling Softmax

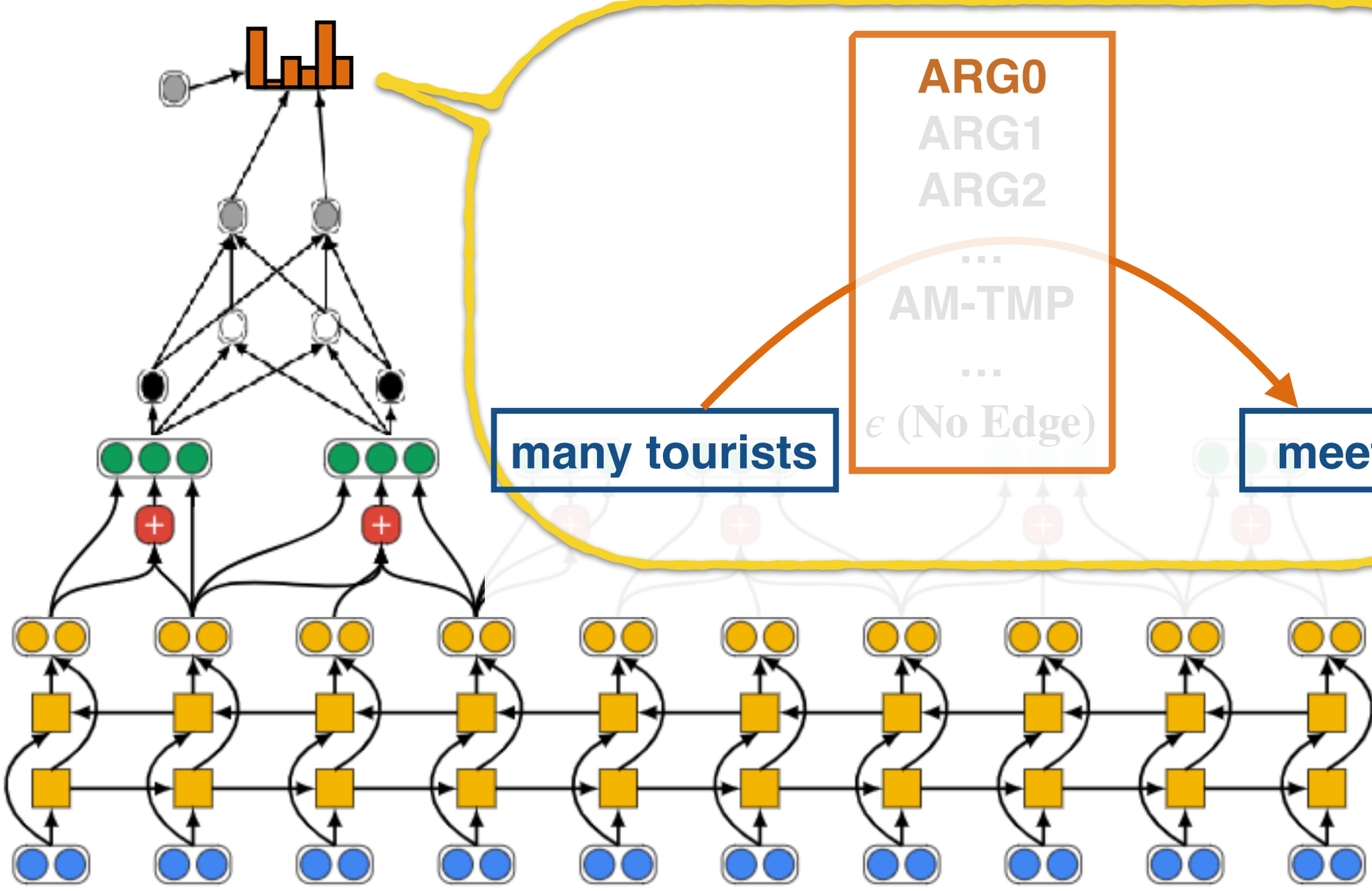
Node & Edge Scores

Span Representation

Highway BiLSTMs

Word & Char Embeddings

Input sentence



(1) Span Representations

(2) Local Label Classifiers

(3) Span Pruning

Labeling Softmax

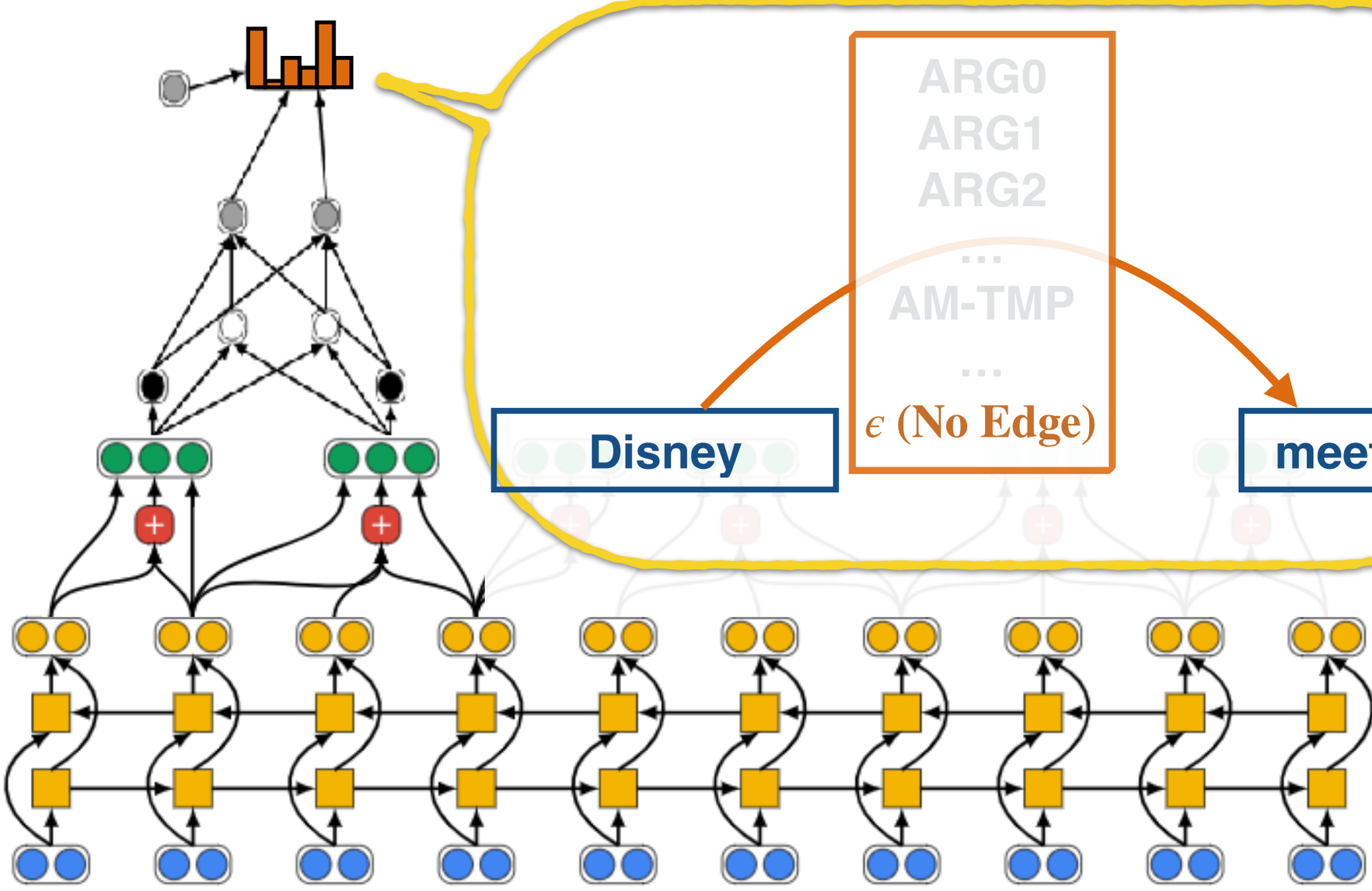
Node & Edge Scores

Span Representation

Highway BiLSTMs

Word & Char Embeddings

Input sentence



(1) Span Representations

(2) Local Label Classifiers

(3) Span Pruning

Labeling Softmax

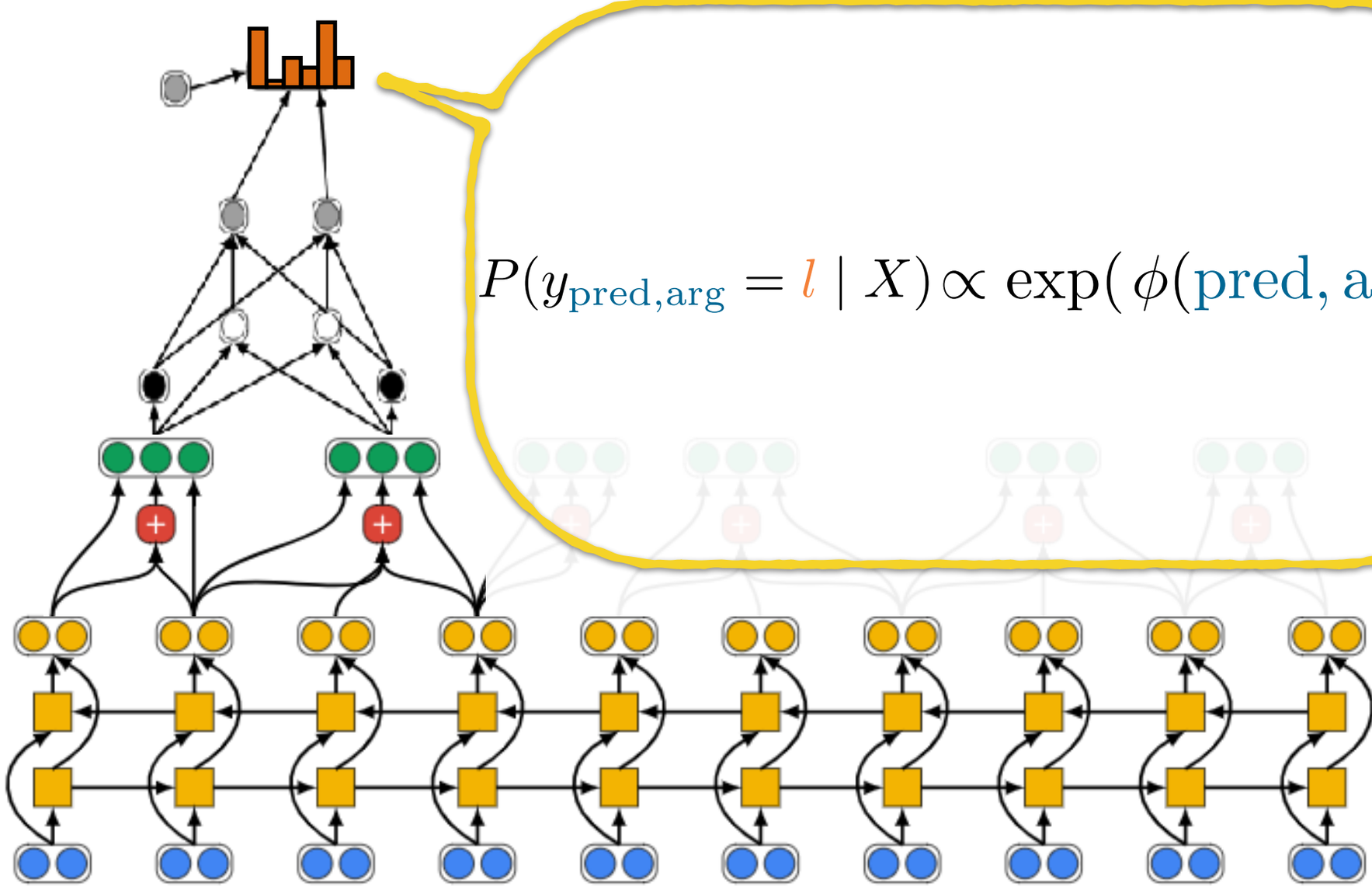
Node & Edge Scores

Span Representation

Highway BiLSTMs

Word & Char Embeddings

Input sentence



$$P(y_{\text{pred,arg}} = l \mid X) \propto \exp(\phi(\text{pred}, \text{arg}, l))$$

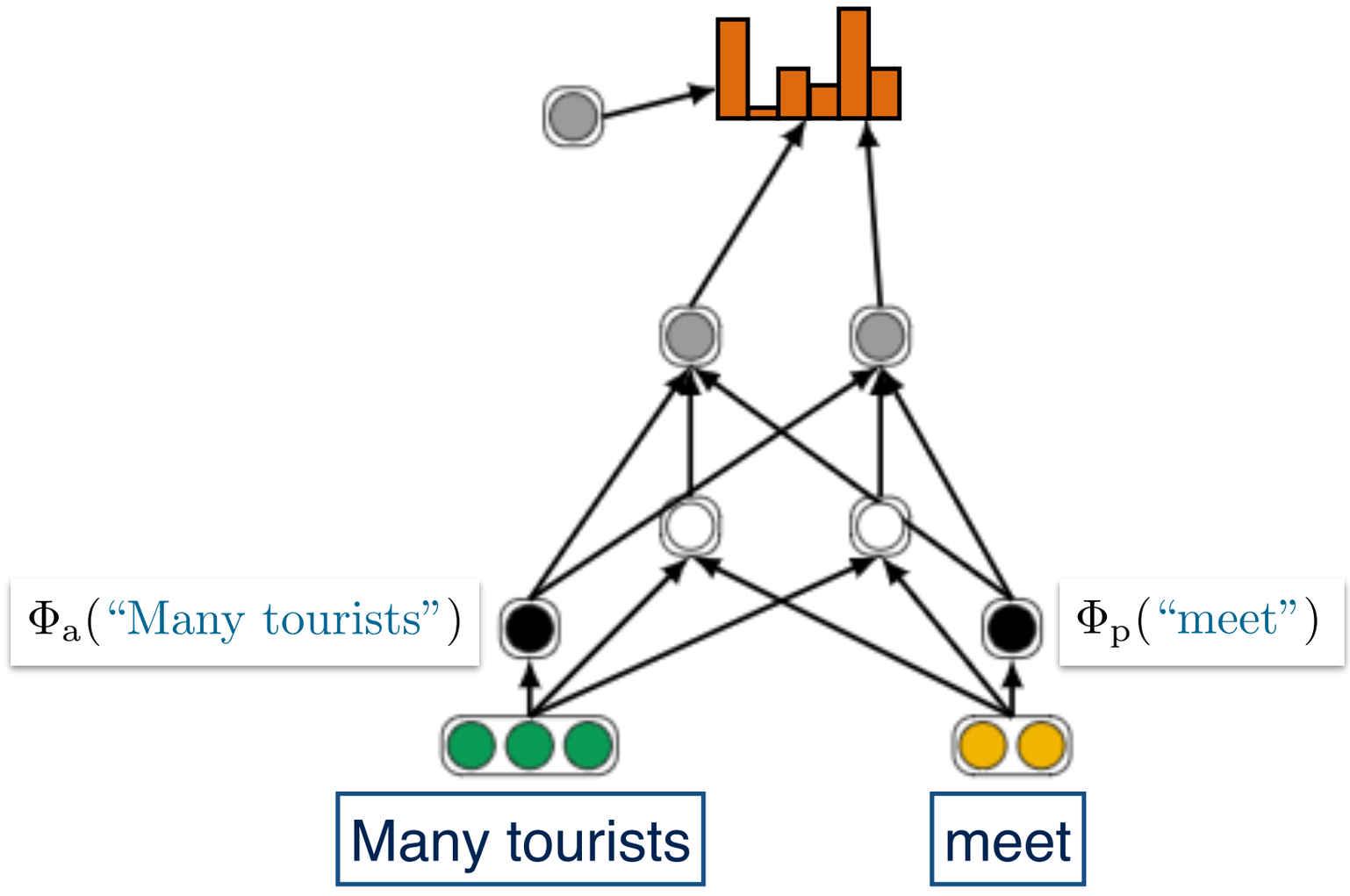
Many tourists visit Disney to meet their favorite cartoon characters

(1) Span Representations

(2) Local Label Classifiers

(3) Span Pruning

$$\phi(\text{pred}, \text{arg}, l) = \Phi_{\text{a}}(\text{arg}) + \Phi_{\text{p}}(\text{pred}) + \Phi^{(l)}_{\text{rel}}(\text{arg}, \text{pred})$$

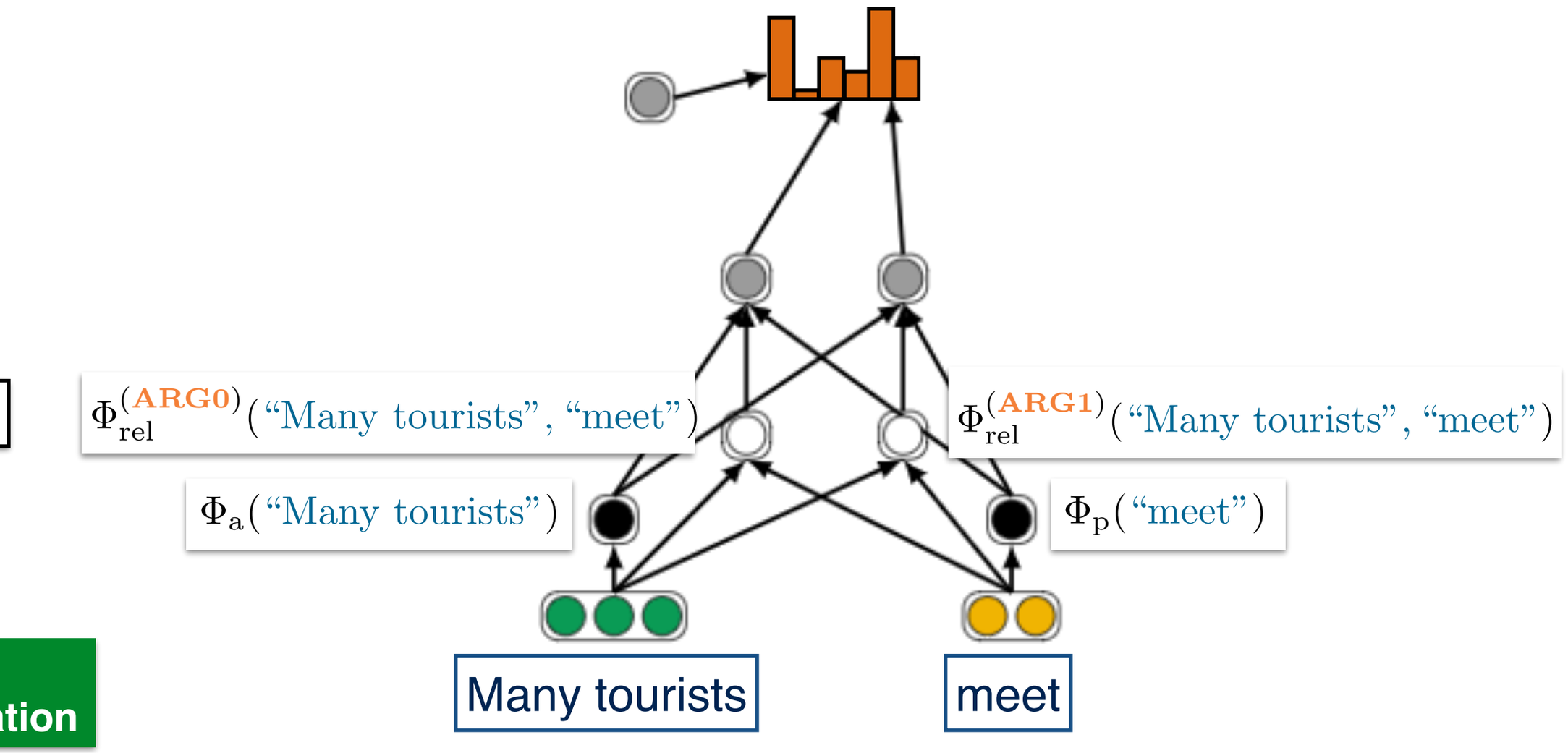


Pred./Arg. score

Span Representation

(2) Local Label Classifiers

$$\phi(\text{pred}, \text{arg}, l) = \Phi_{\text{a}}(\text{arg}) + \Phi_{\text{p}}(\text{pred}) + \Phi_{\text{rel}}^{(l)}(\text{arg}, \text{pred})$$



(2) Local Label Classifiers

$$\phi(\text{pred}, \text{arg}, l) = \Phi_{\text{a}}(\text{arg}) + \Phi_{\text{p}}(\text{pred}) + \Phi_{\text{rel}}^{(l)}(\text{arg}, \text{pred})$$

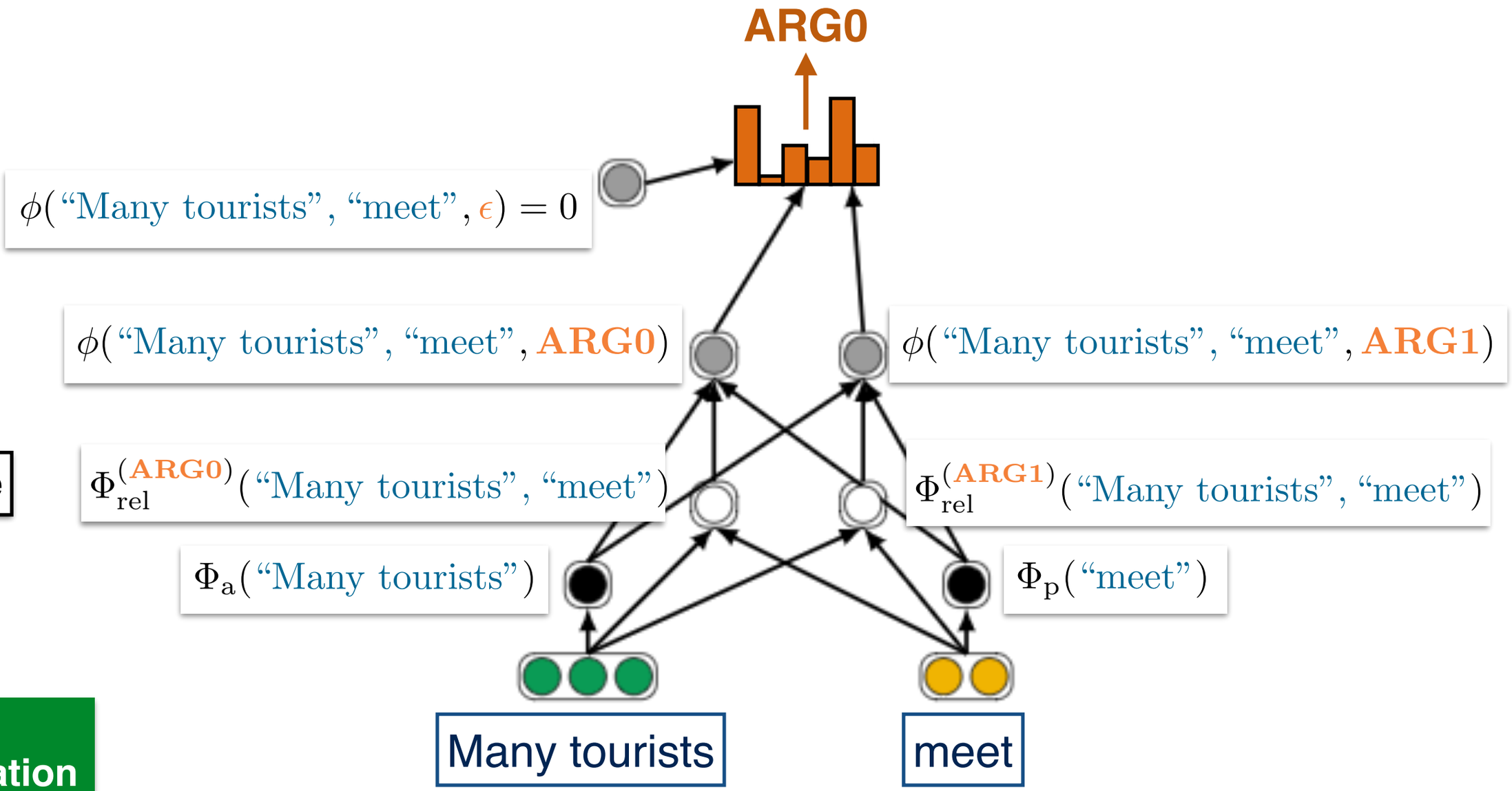
Softmax

Combined score

Edge score

Pred./Arg. score

Span Representation

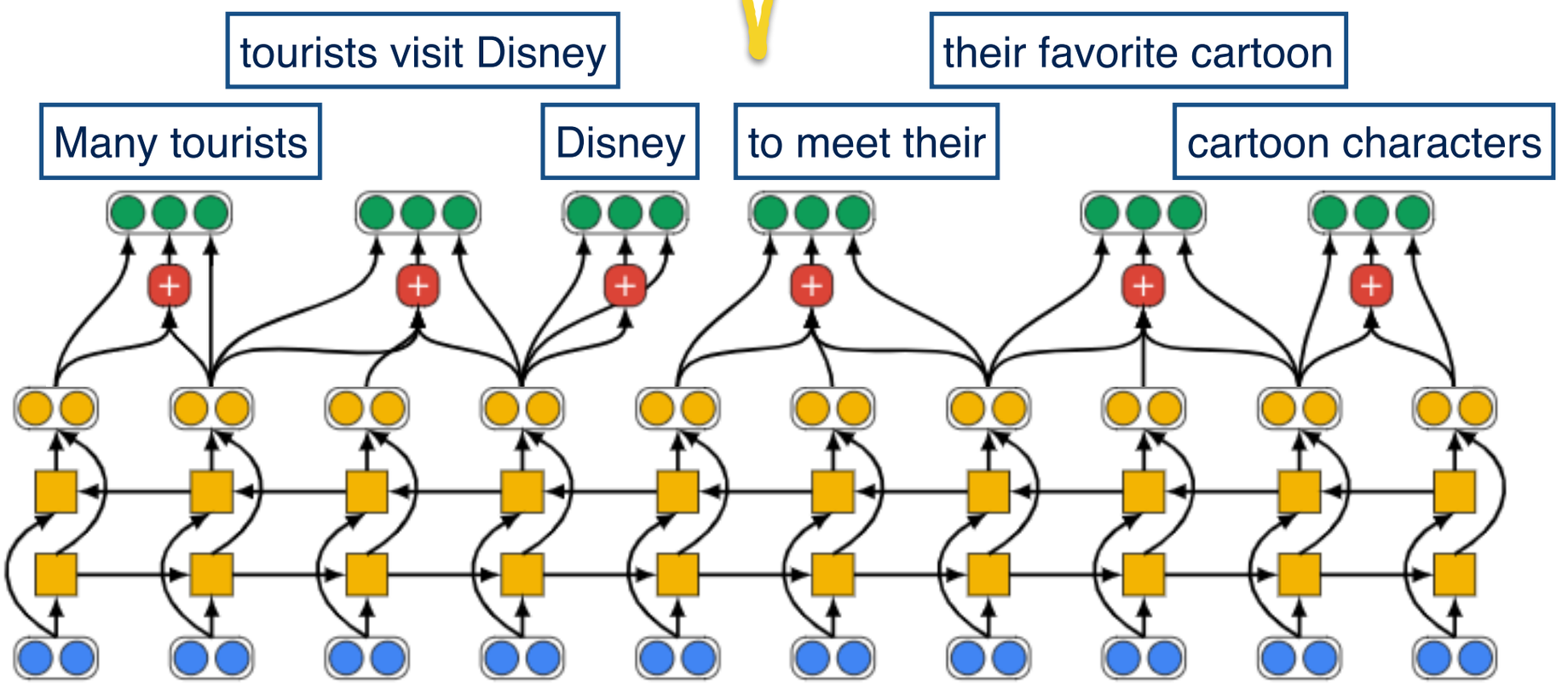


(1) Span Representations

(2) Local Label Classifiers

(3) Span Pruning

$O(n^2)$ arguments, $O(n)$ predicates,
—> $O(n^3)$ edges!



Span Representation

Highway BiLSTMs

Word & Char Embeddings

Input sentence

(1) Span Representations

(2) Local Label Classifiers

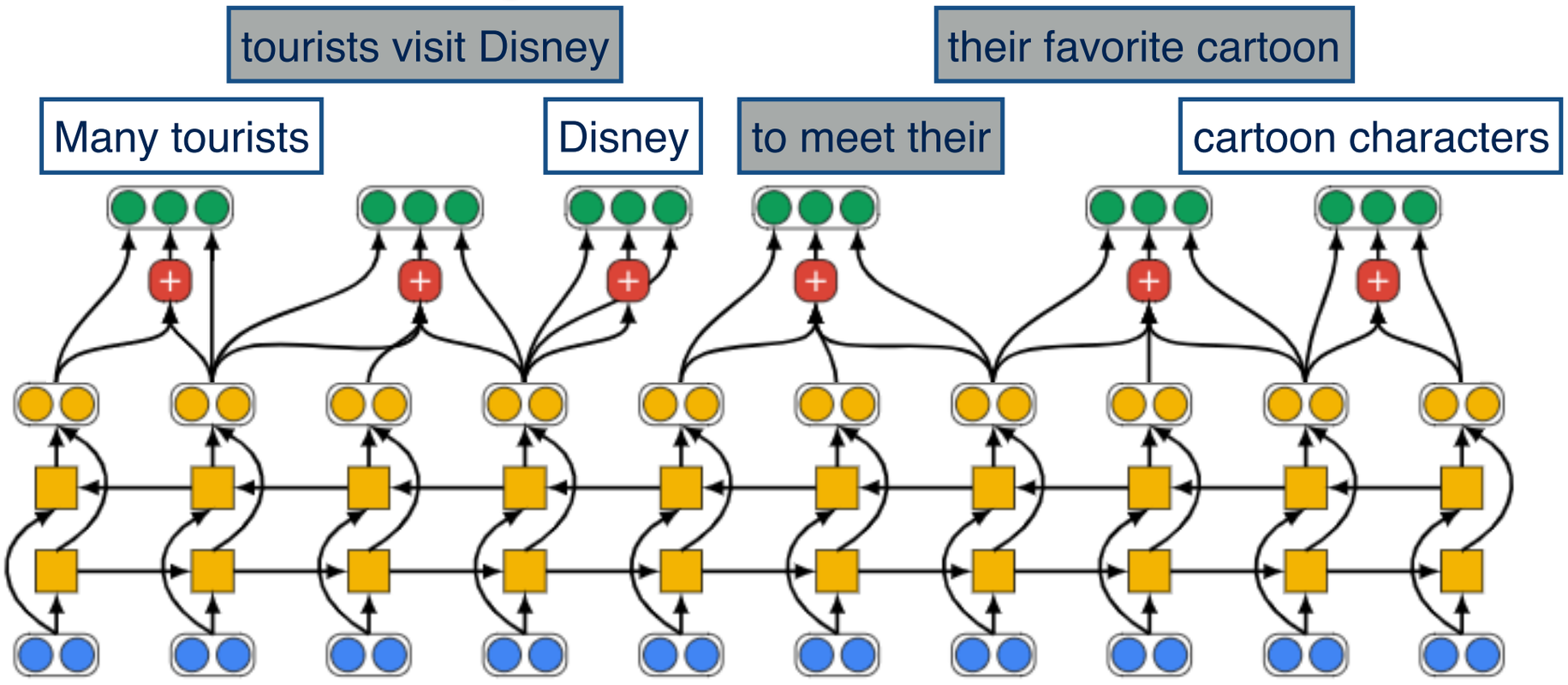
(3) Span Pruning

Only keep top **O(n)** spans using their unary scores

$$\Phi_{\mathrm{a}}(\text{“many tourists”}) = 2.5$$

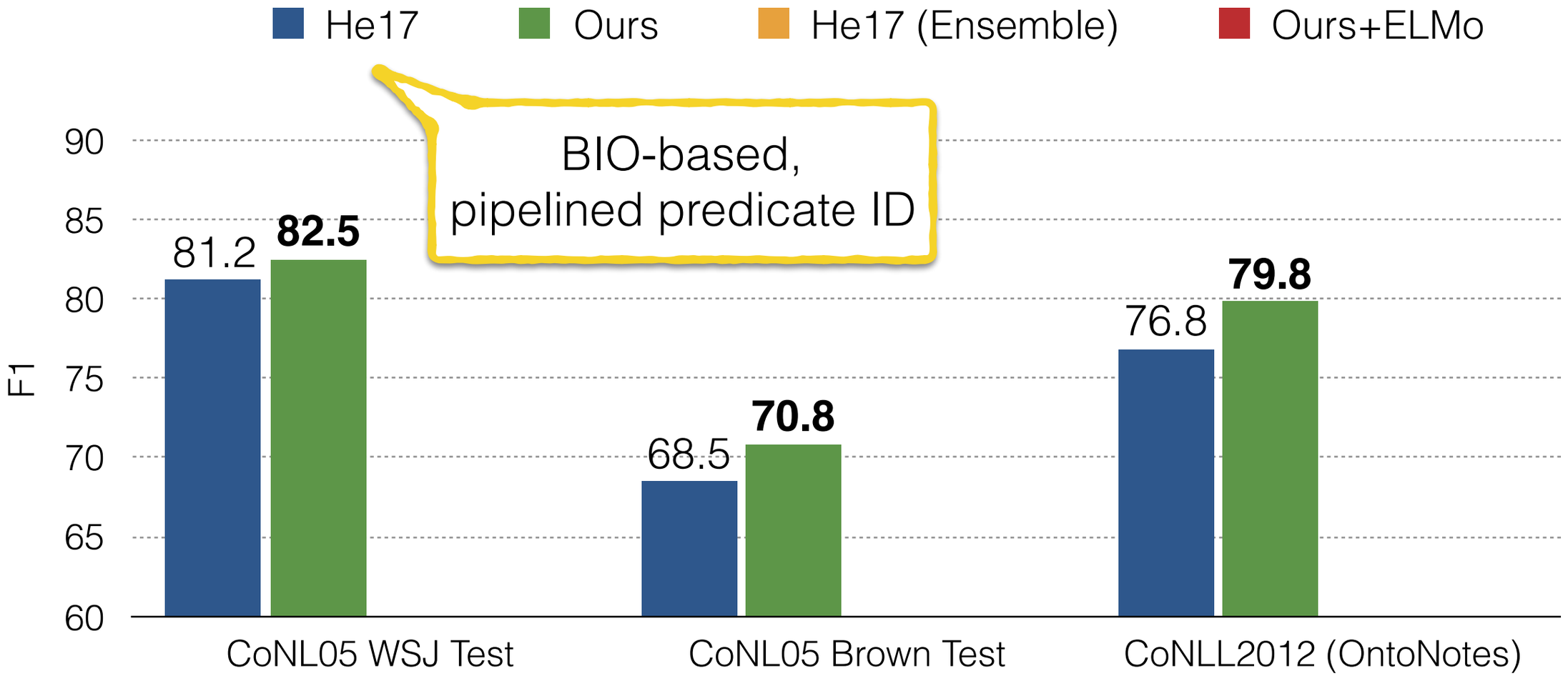
$$\Phi_{\mathrm{a}}(\text{“tourists visit Disney”}) = -0.8$$

$$\ldots$$

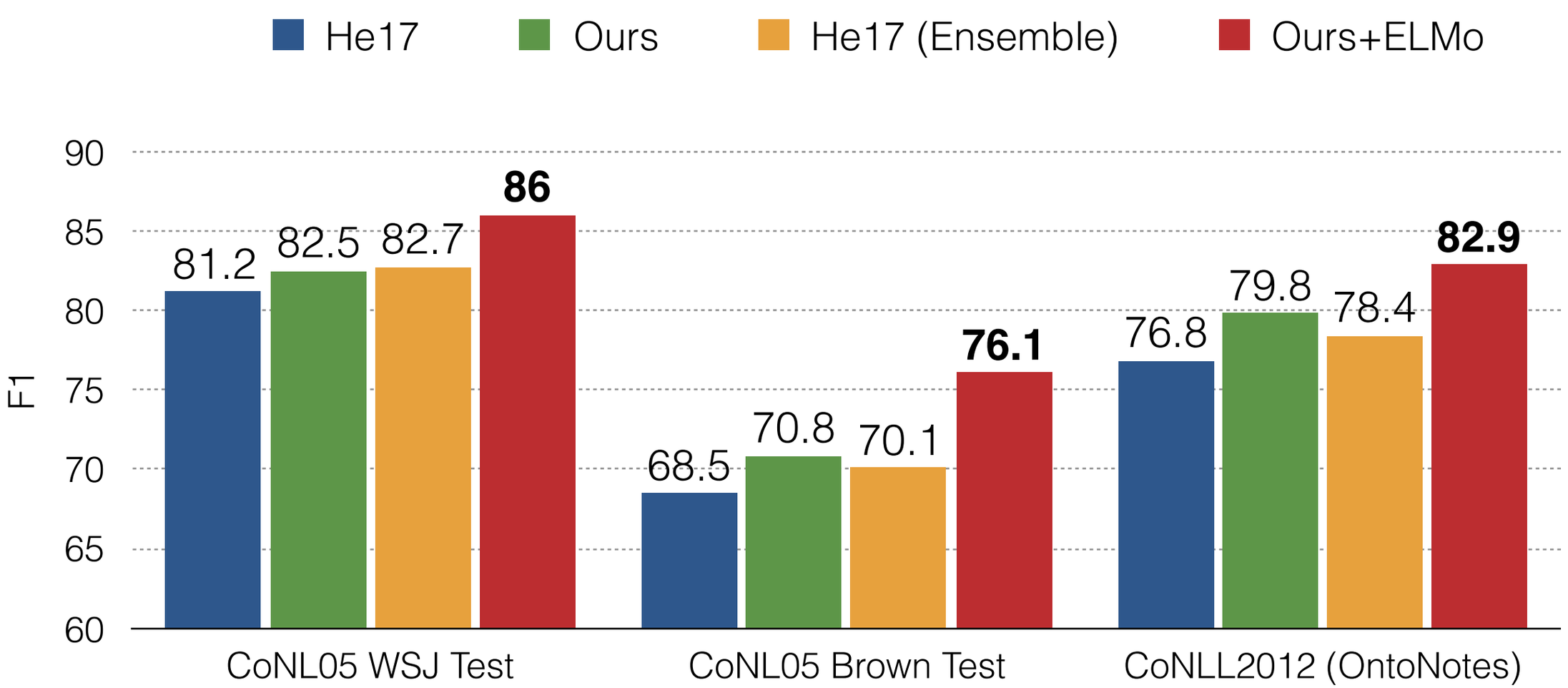


Results & Analysis

End-to-End SRL Results



End-to-End SRL Results



With **ELMo**, over 3 points improvement over SotA ensemble!

*ELMo: Deep Contextualized Word Representations, Peters et al., 2018

Span-based vs. BIO

	BIO (He17)	**Span-based (this work)**
Inputs	(Sentence, Predicate)	**Sentence**
Predicate Identification	Pipelined	**Joint**
Global Consistency		
Long-range Dependency		

Span-based vs. BIO

	BIO (He17)	**Span-based (this work)**
Inputs	(Sentence, Predicate)	**Sentence**
P	ipelined	**Joint**
Global Consistency		
Long-range Dependency		

Due to the strong independence assumption we make.

Span-based vs. BIO

	BIO (He17)	**Span-based (this work)**
Inputs	(Sentence, Predicate)	**Sentence**
P	ipelined	**Joint**
Global Cons		
Long-range Dependency		

Due to the strong independence assumption we make.

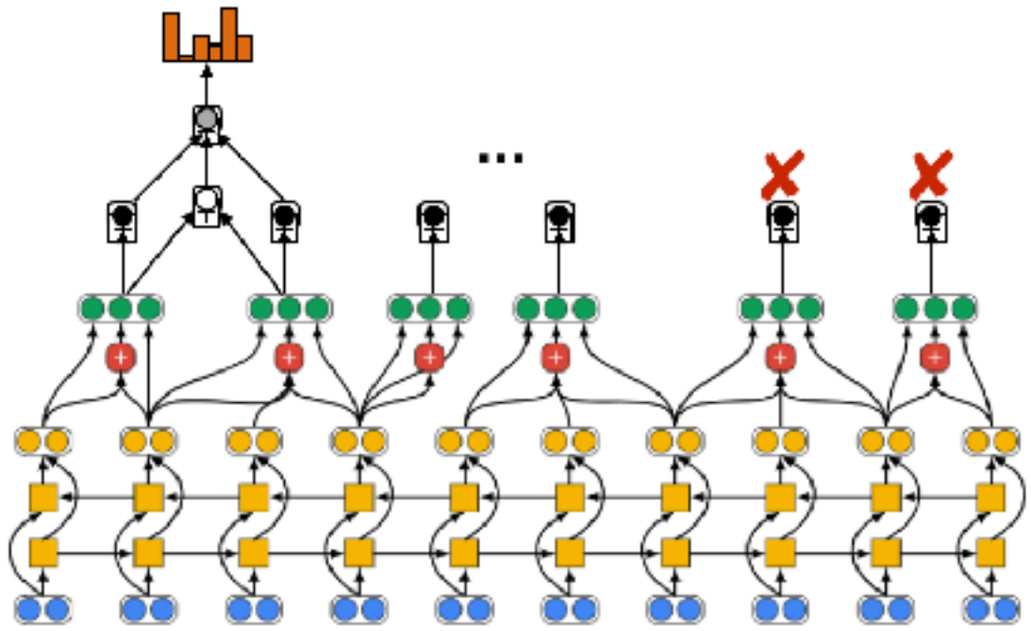
By allowing direct interaction between predicates and arguments

Conclusion

- Joint prediction of predicates and arguments

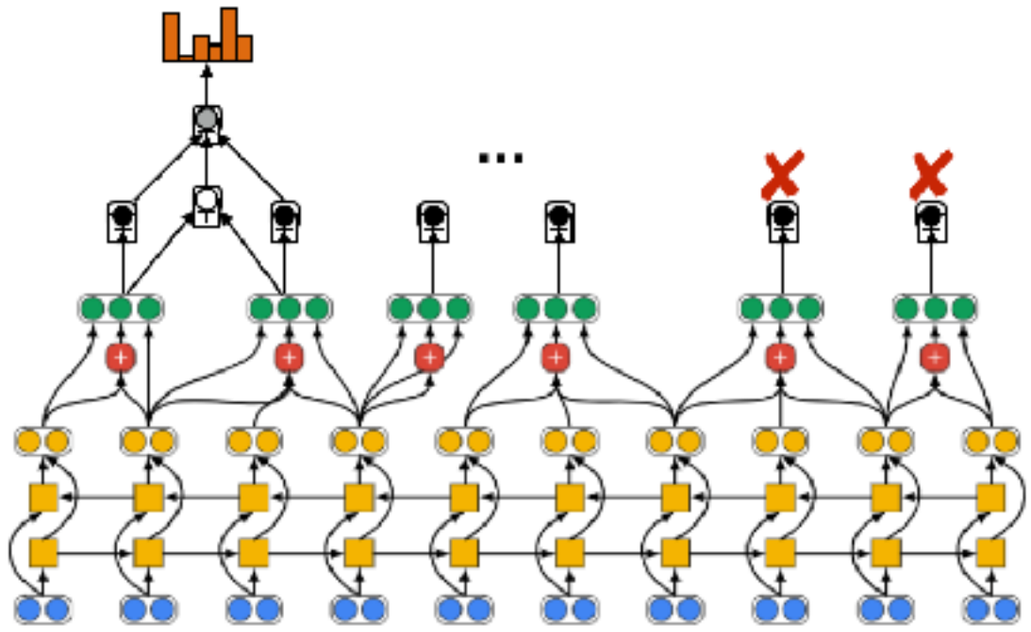
Conclusion

- Joint prediction of predicates and arguments
- Our recipe:
  1. Contextualized span representations
  2. Local label classifiers
  3. Greedy span pruning



Conclusion

- Joint prediction of predicates and arguments
- Our recipe:
  1. Contextualized span representations
  2. Local label classifiers
  3. Greedy span pruning



- Future work: Improve global consistency, use span representations for downstream tasks, etc.

THANK YOU!

Code and pertained models:
https://github.com/luheng/lsgn