Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling

Diego Marcheggiani and Ivan Titov

University of Amsterdam

University of Edinburgh

EMNLP 2017
Copenhagen

Contributions

- Syntactic Graph Convolutional Networks
- State-of-the-art semantic role labeling model
  - English and Chinese

Semantic Role Labeling

- Predicting the predicate-argument structure of a sentence

Sequa makes and repairs jet engines.

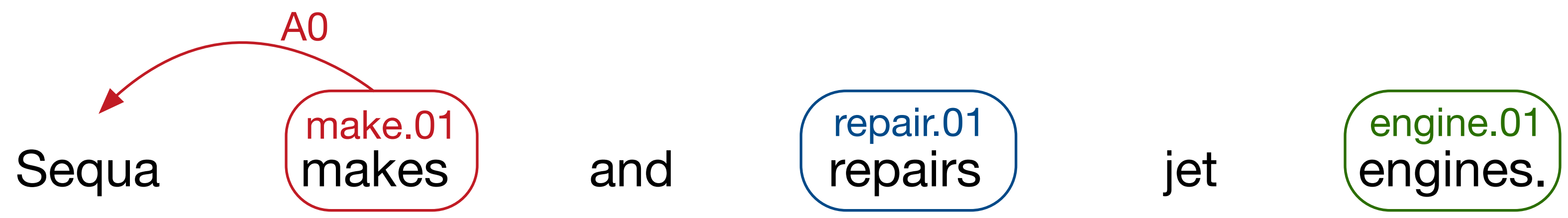
Semantic Role Labeling

- Predicting the predicate-argument structure of a sentence
  - Discover and disambiguate predicates

	make.01		repair.01		engine.01
Sequa	makes	and	repairs	jet	engines.

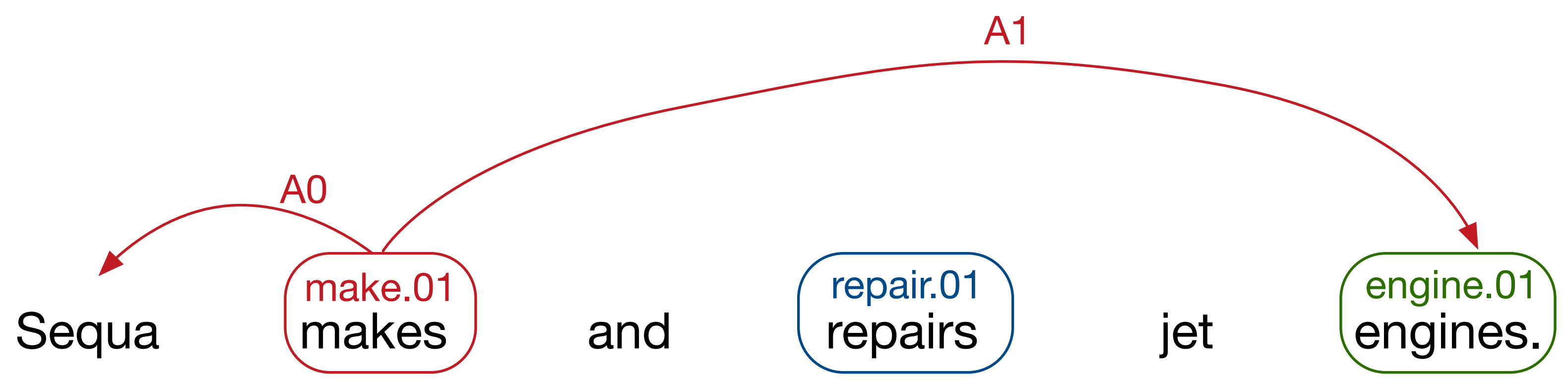
Semantic Role Labeling

- Predicting the predicate-argument structure of a sentence
  - Discover and disambiguate predicates
  - Identify arguments and label them with their semantic roles



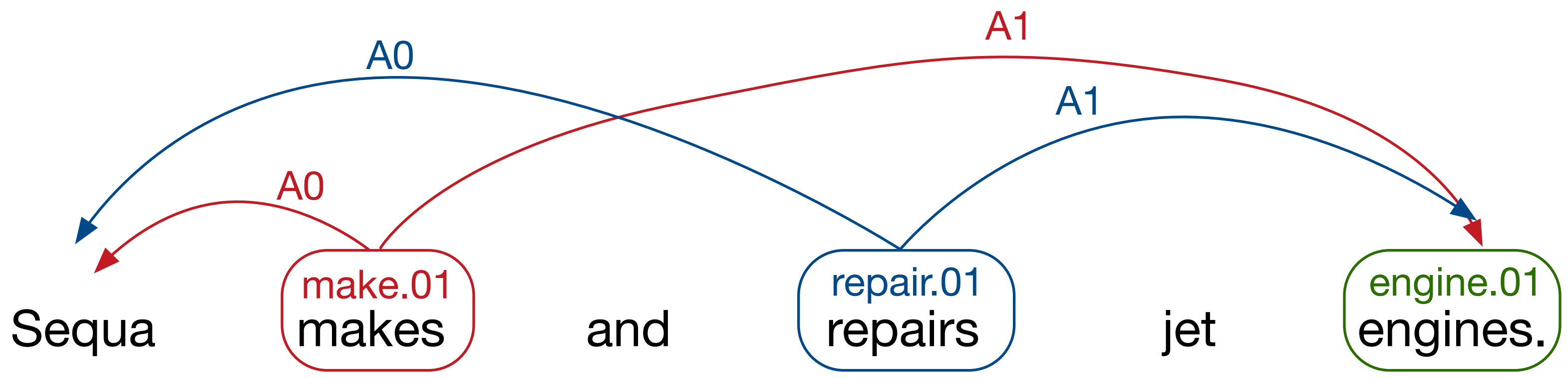
Semantic Role Labeling

- Predicting the predicate-argument structure of a sentence
  - Discover and disambiguate predicates
  - Identify arguments and label them with their semantic roles



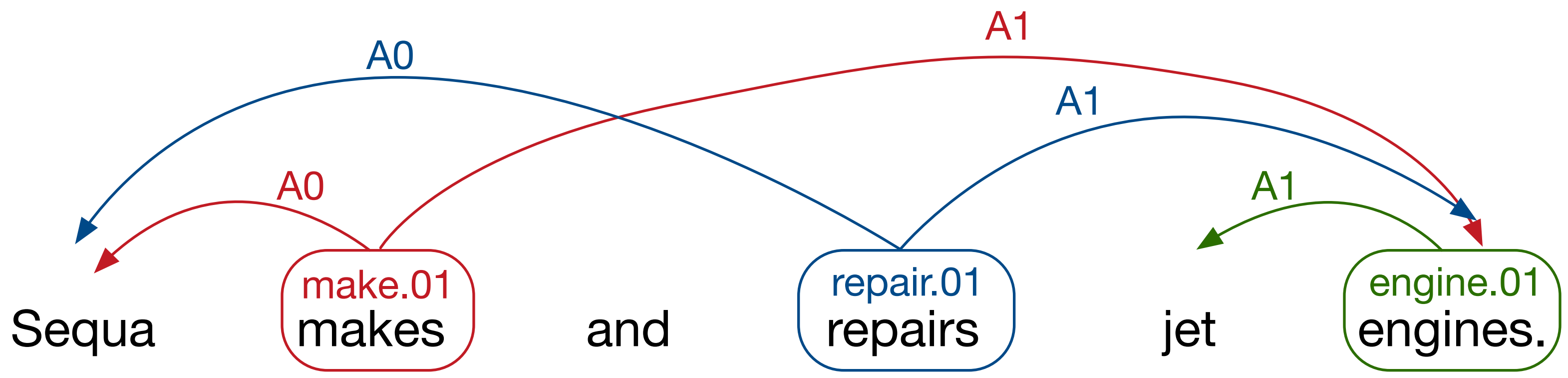
Semantic Role Labeling

- Predicting the predicate-argument structure of a sentence
  - Discover and disambiguate predicates
  - Identify arguments and label them with their semantic roles



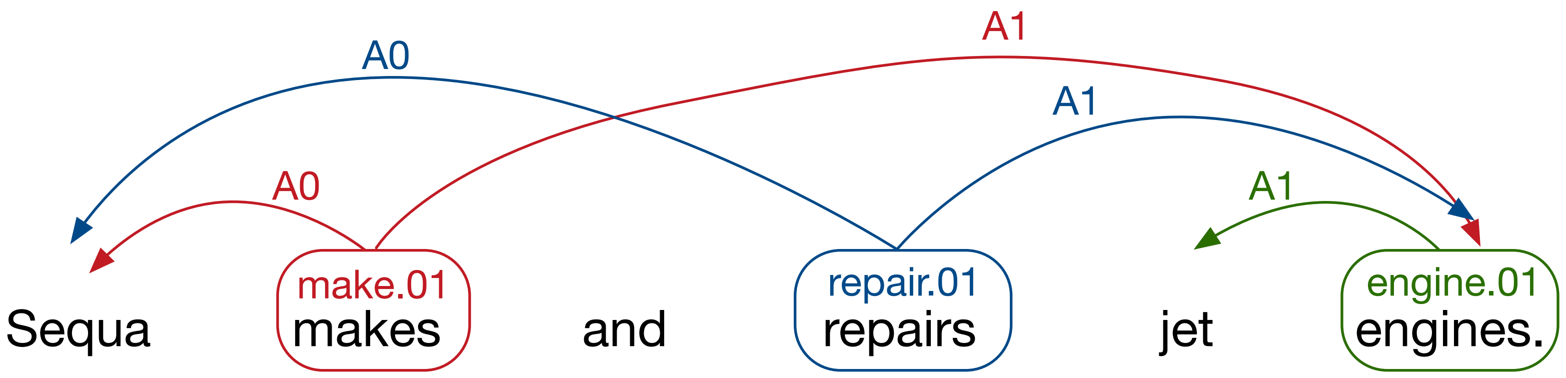
Semantic Role Labeling

- Predicting the predicate-argument structure of a sentence
  - Discover and disambiguate predicates
  - Identify arguments and label them with their semantic roles



Semantic Role Labeling

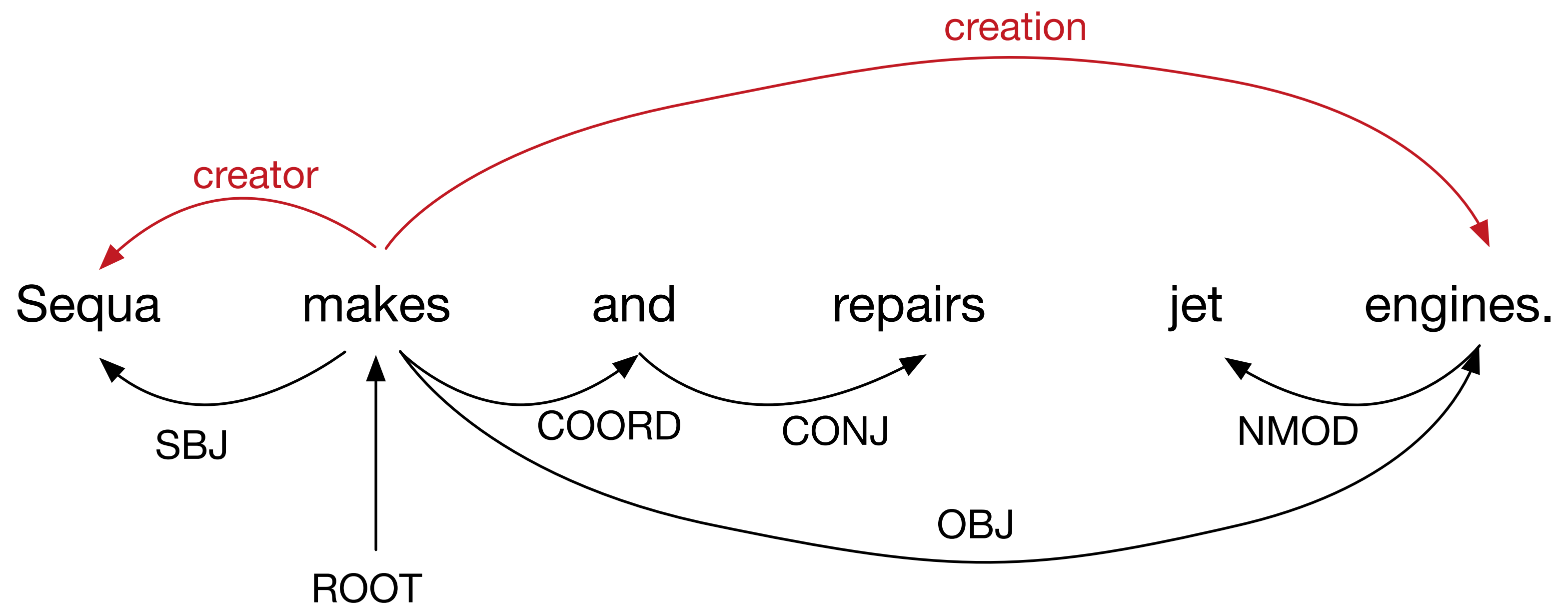
- Only the head of an argument is labeled
- Sequence labeling task for each predicate
- Focus on argument identification and labeling



Related work

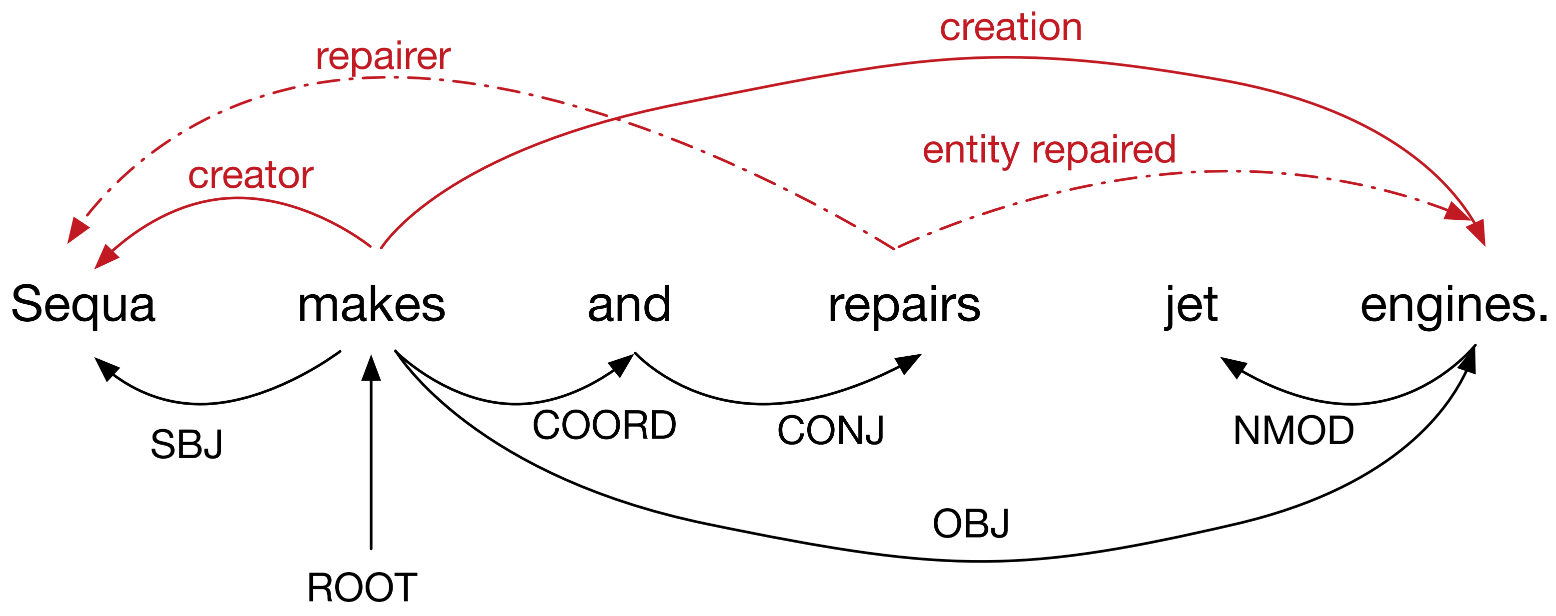
- SRL systems that use syntax with simple NN architectures
  - [FitzGerald et al., 2015]
  - [Roth and Lapata, 2016]
- Recent models ignore linguistic bias
  - [Zhou and Xu, 2014]
  - [He et al., 2017]
  - [Marcheggiani et al., 2017]

Motivations



- Some semantic dependencies are mirrored in the syntactic graph

Motivations



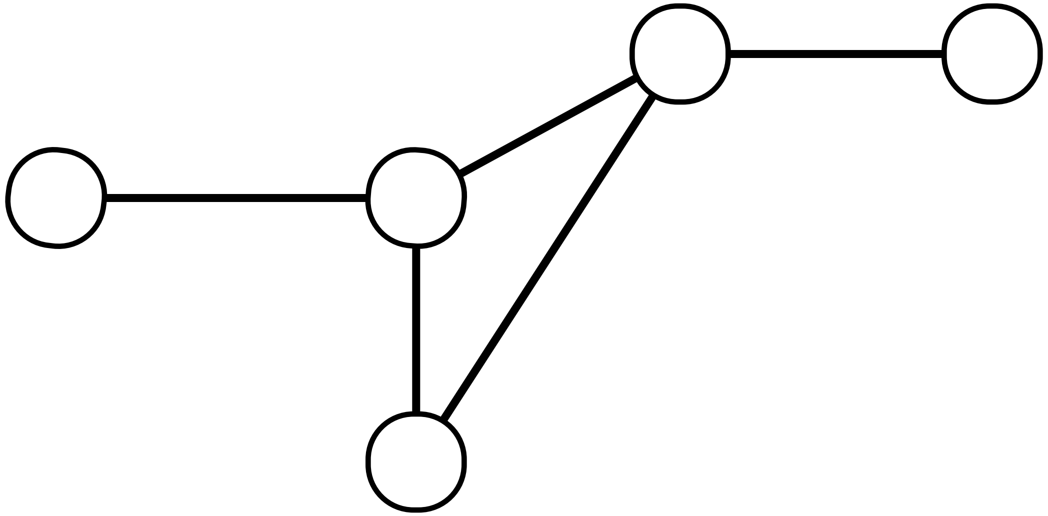
- Some semantic dependencies are mirrored in the syntactic graph
- Not all of them – syntax-semantic interface is not trivial

Encoding Sentences with Graph Convolutional Networks

- **Graph Convolutional Networks (GCNs)** [Kipf and Welling, 2017]
- Syntactic GCNs
- Semantic Role Labeling Model
- Experiments
- Conclusions

Graph Convolutional Networks (message passing)

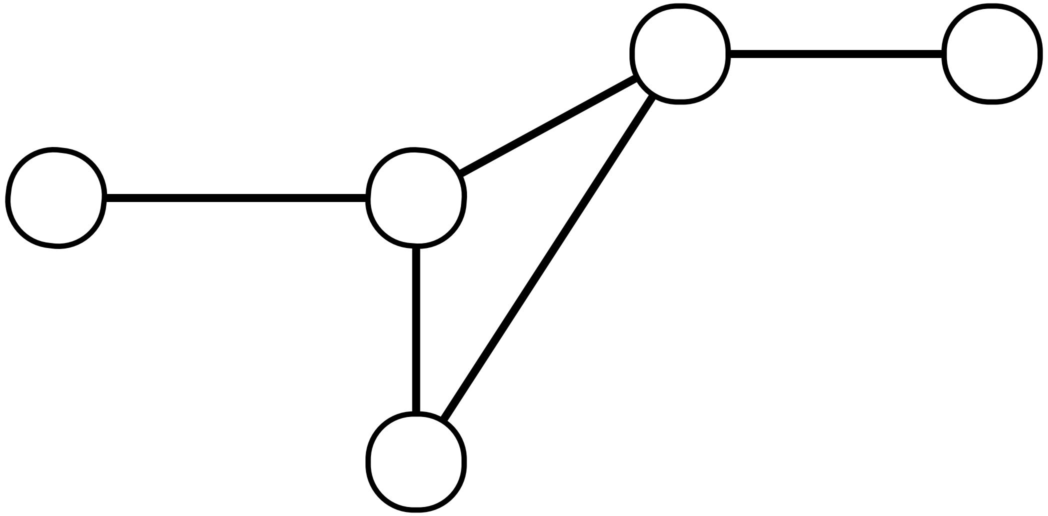
[Kipf and Welling, 2017]



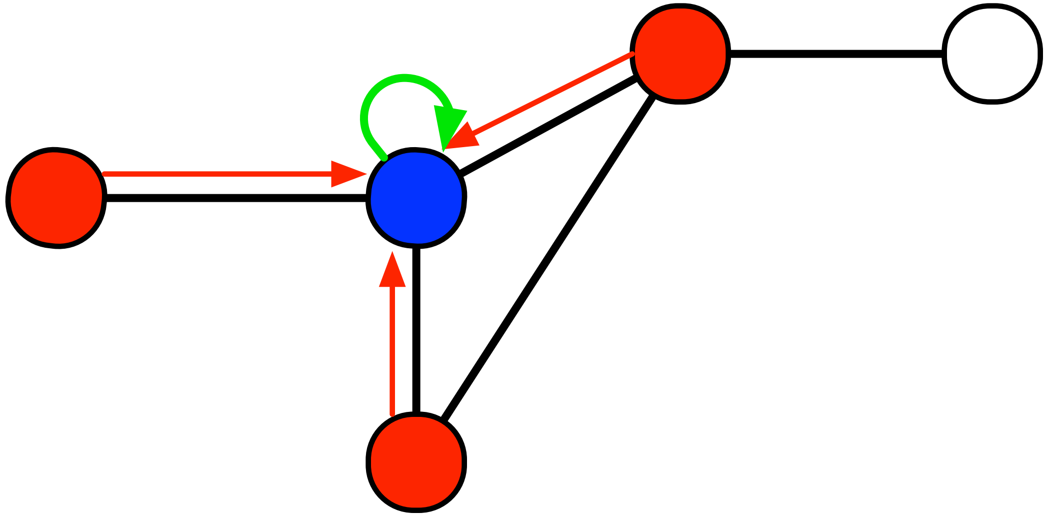
Undirected graph

Graph Convolutional Networks (message passing)

[Kipf and Welling, 2017]



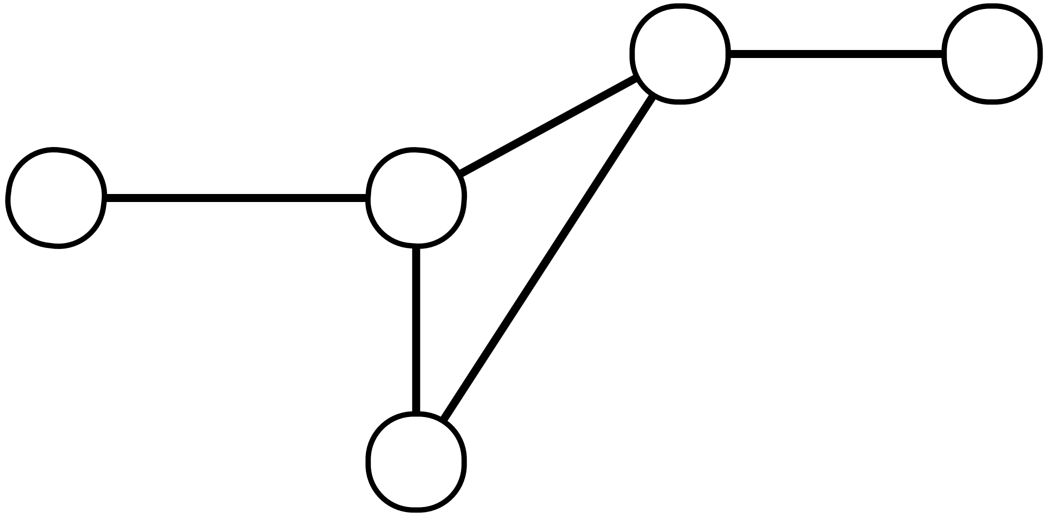
Undirected graph

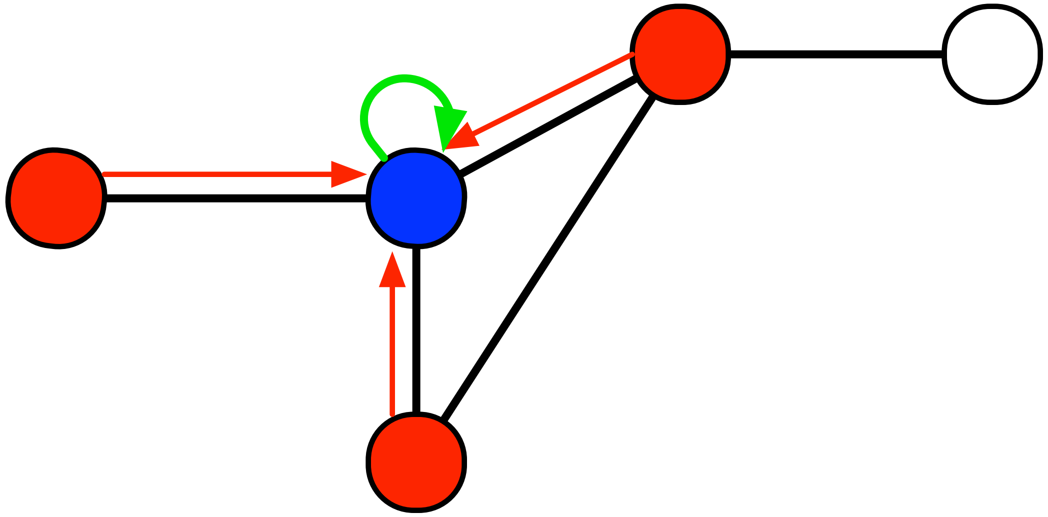


Update of the blue node

Graph Convolutional Networks (message passing)

[Kipf and Welling, 2017]





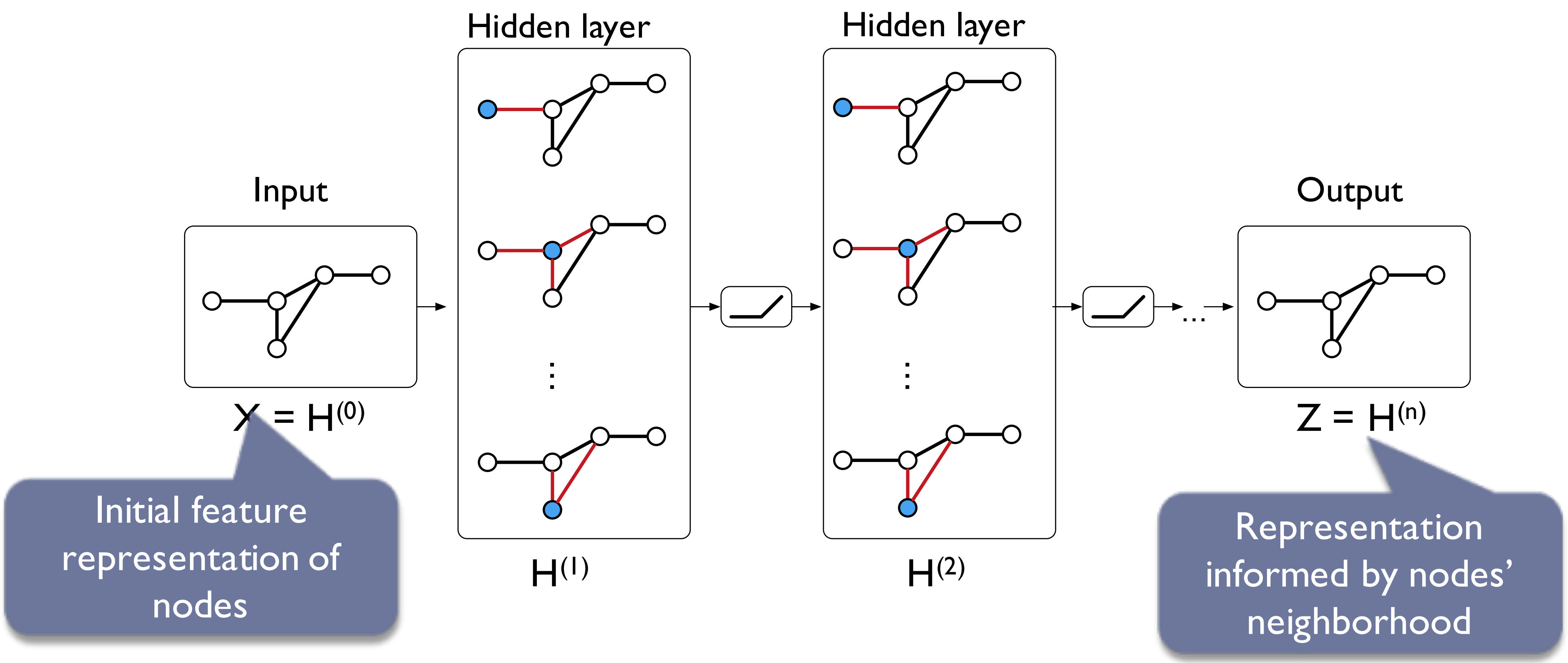
$$h_i = ReLU\left(W_0 h_i + \sum_{j \in \mathcal{N}(i)} W_1 h_j\right)$$

Self loop

Neighborhood

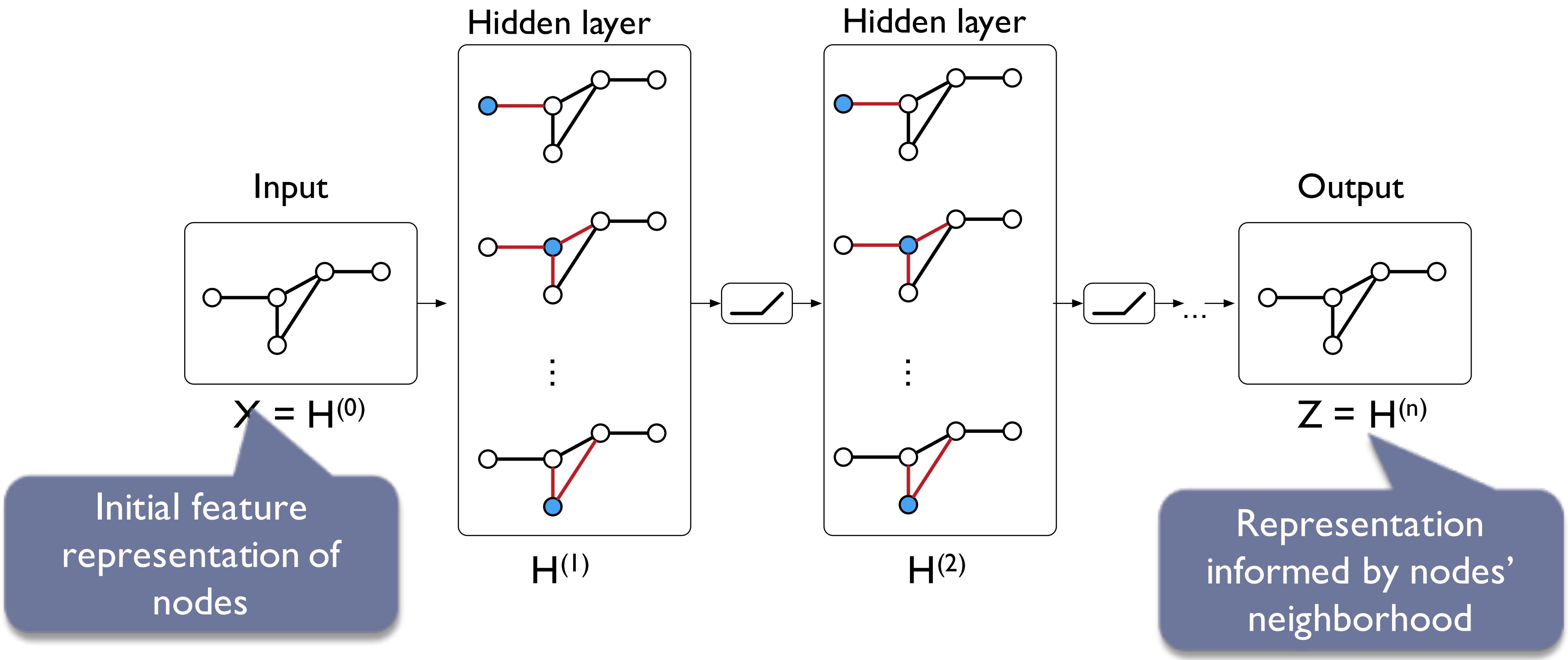
GCNs Pipeline

[Kipf and Welling, 2017]



GCNs Pipeline

[Kipf and Welling, 2017]

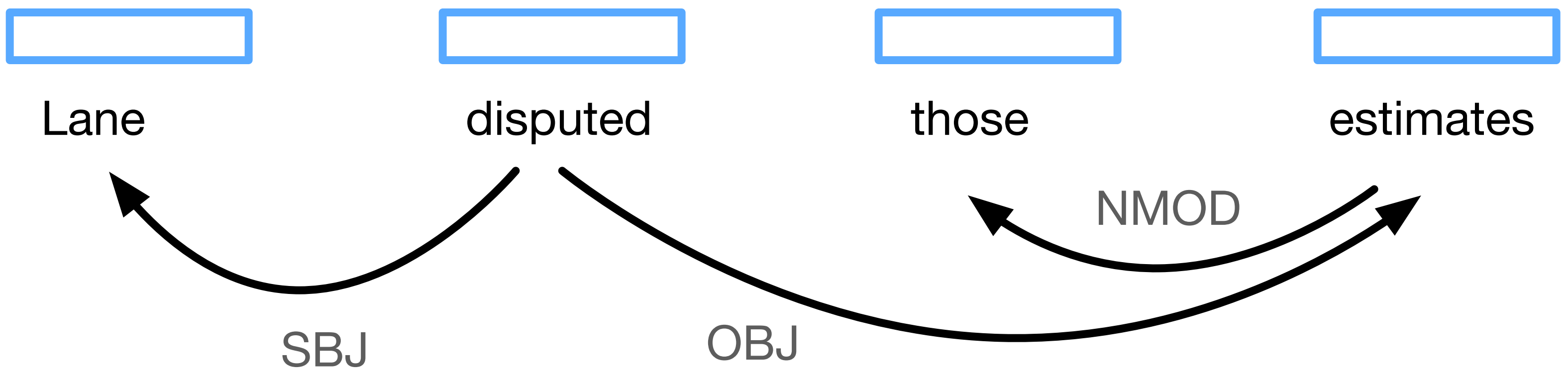


Extend GCNs for syntactic dependency trees

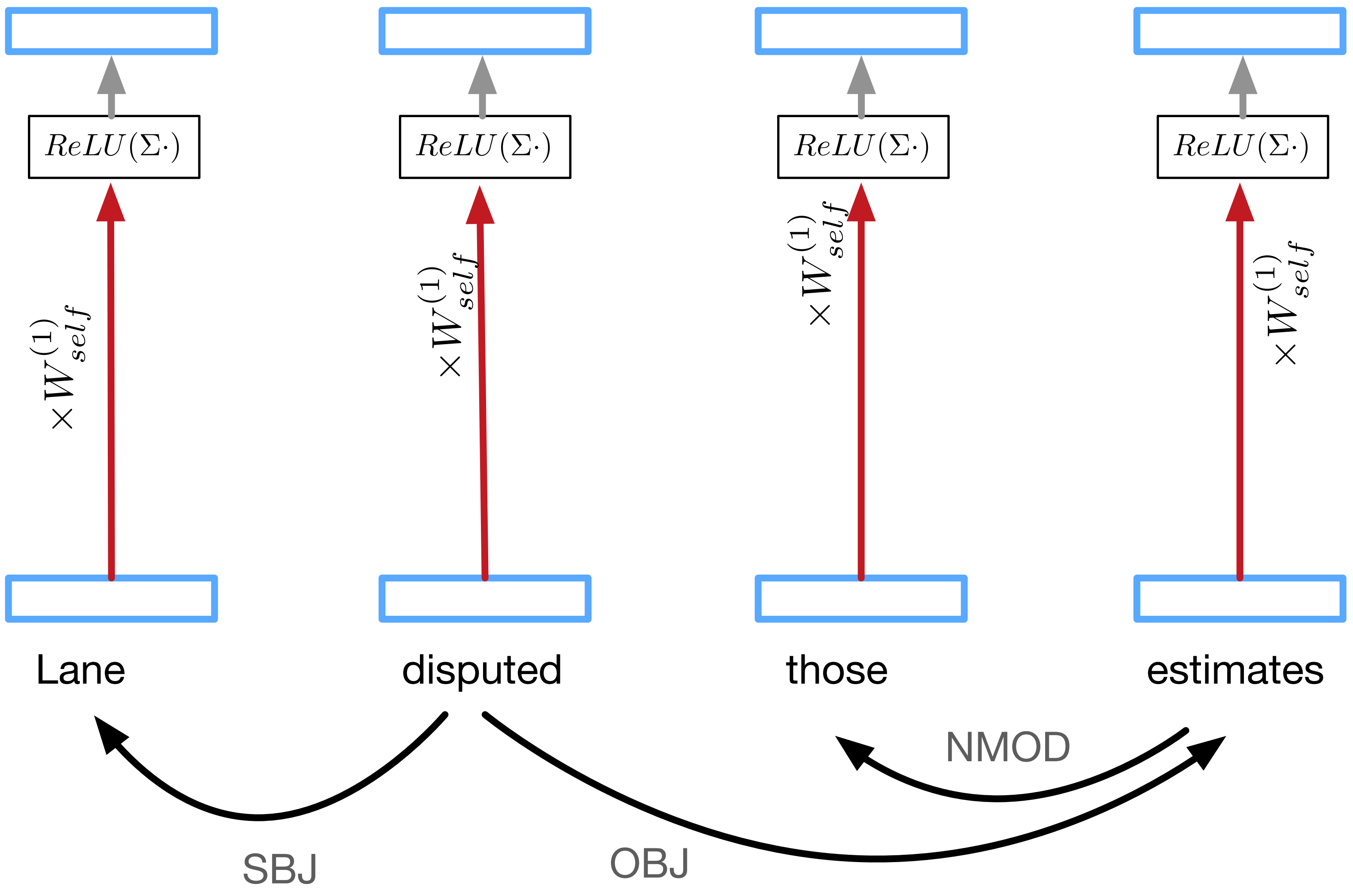
Encoding Sentences with Graph Convolutional Networks

- Graph Convolutional Networks (GCNs)
- **Syntactic GCNs**
- Semantic Role Labeling Model
- Experiments
- Conclusions

Example

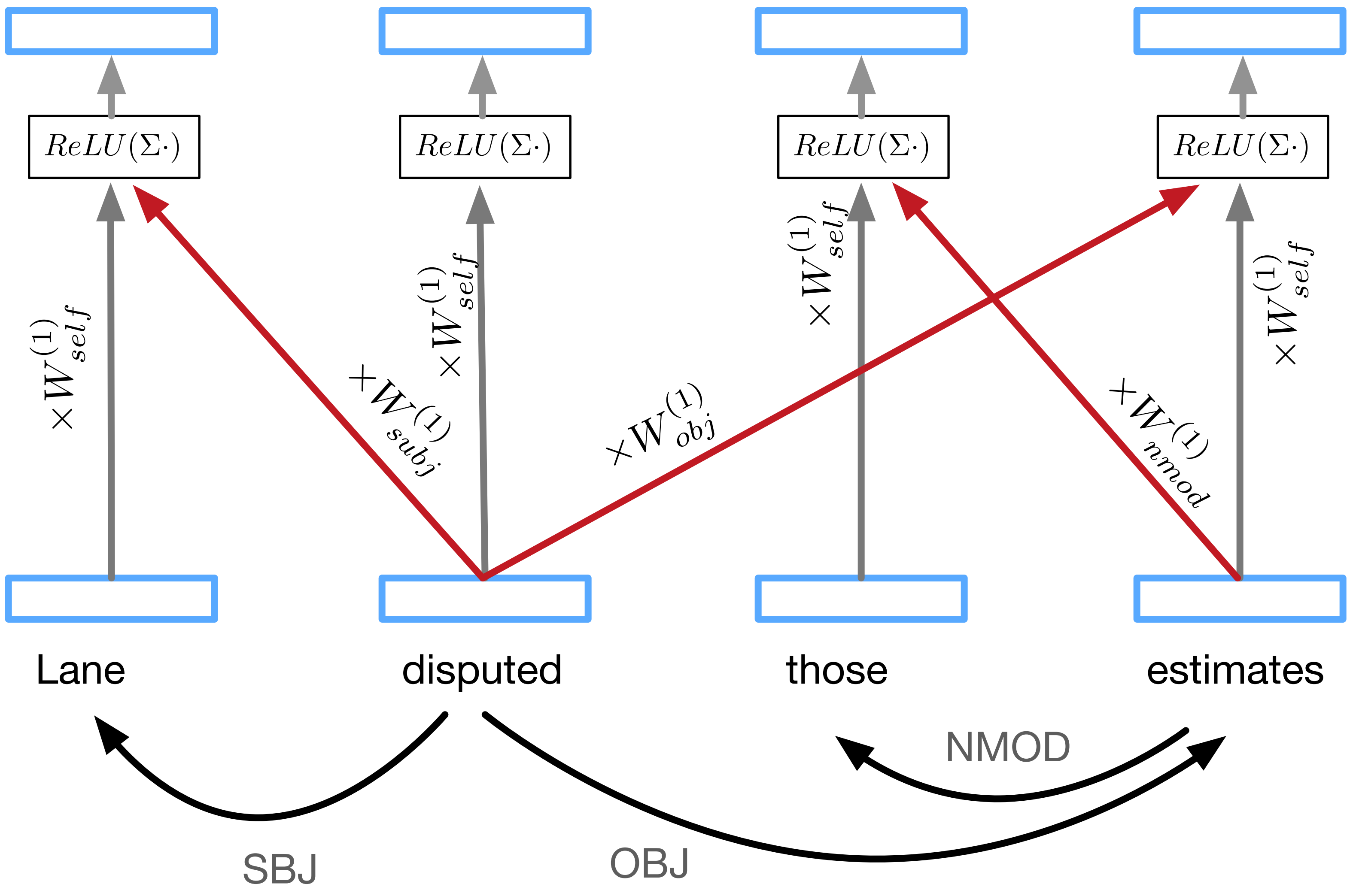


Example



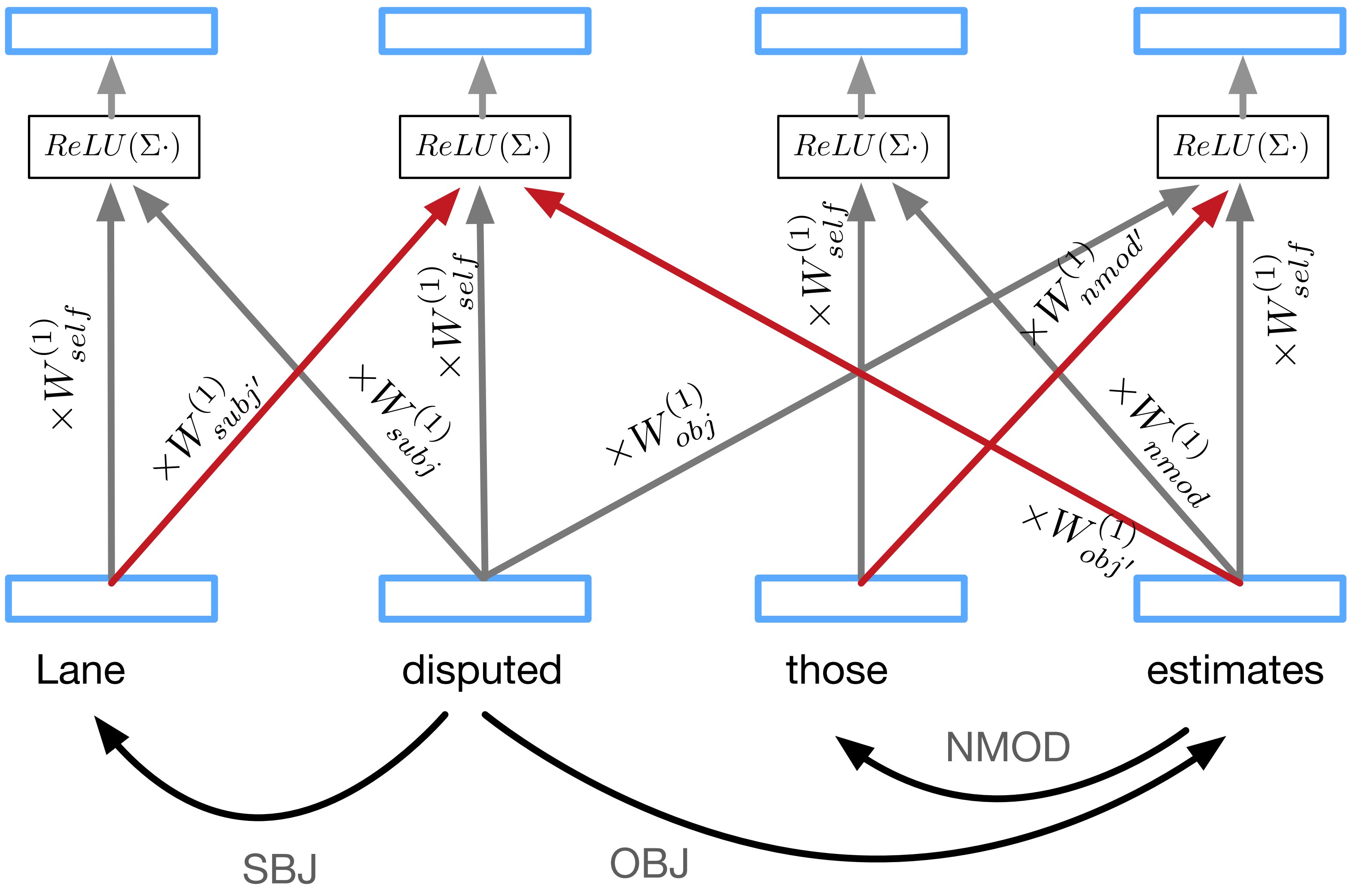
$ReLU(\Sigma\cdot)$
$ReLU(\Sigma\cdot)$
$ReLU(\Sigma\cdot)$
$ReLU(\Sigma\cdot)$
$\times W_{self}^{(1)}$
$\times W_{self}^{(1)}$
$\times W_{self}^{(1)}$
$\times W_{self}^{(1)}$
Lane
disputed
those
estimates
NMOD
SBJ
OBJ

Example



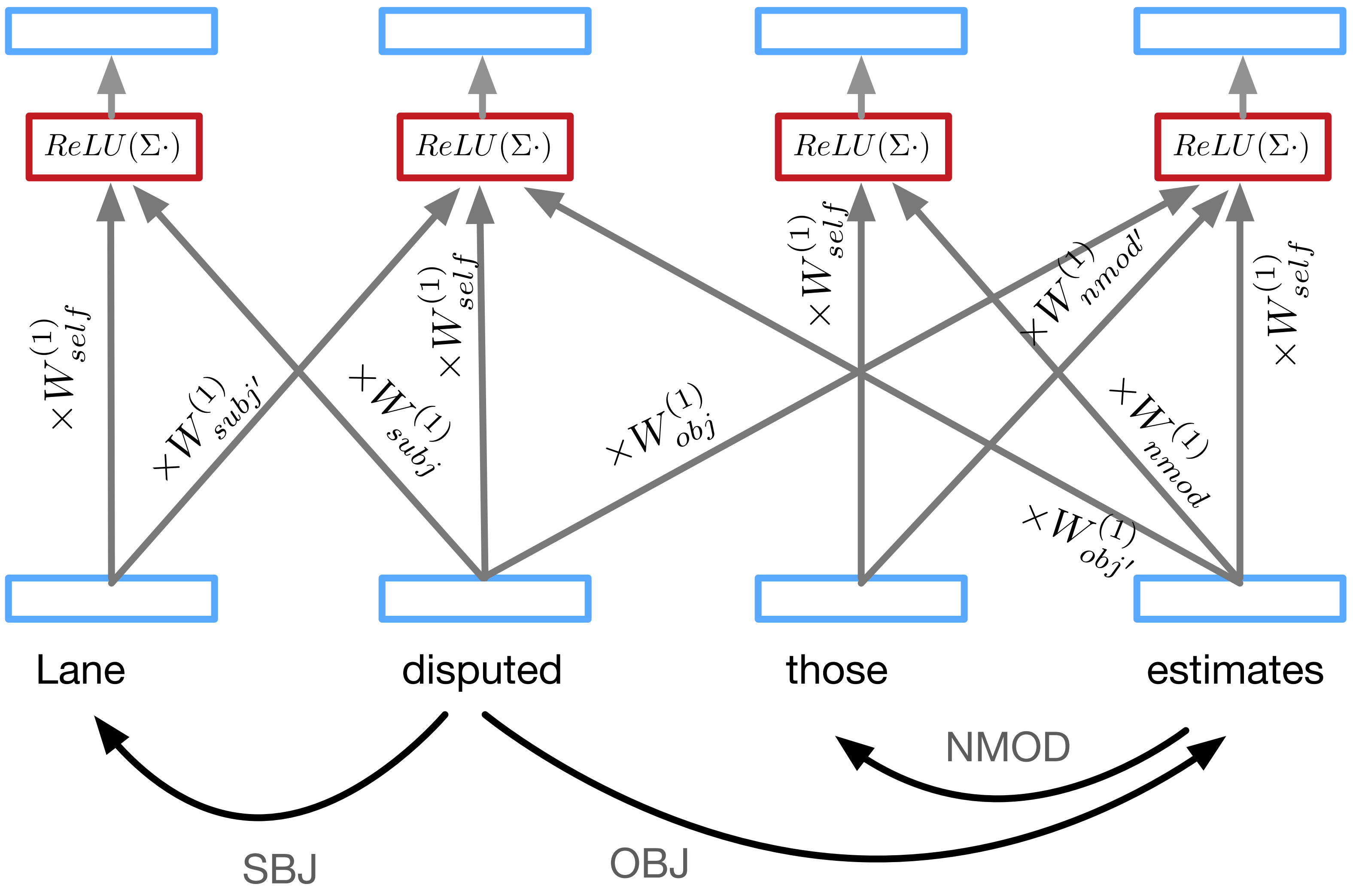
$ReLU(\Sigma\cdot)$
$ReLU(\Sigma\cdot)$
$ReLU(\Sigma\cdot)$
$ReLU(\Sigma\cdot)$
$\times W^{(1)}_{self}$
$\times W^{(1)}_{subj}$
$\times W^{(1)}_{self}$
$\times W^{(1)}_{obj}$
$\times W^{(1)}_{self}$
$\times W^{(1)}_{nmod}$
$\times W^{(1)}_{self}$
Lane
disputed
those
estimates
NMOD
SBJ
OBJ

Example



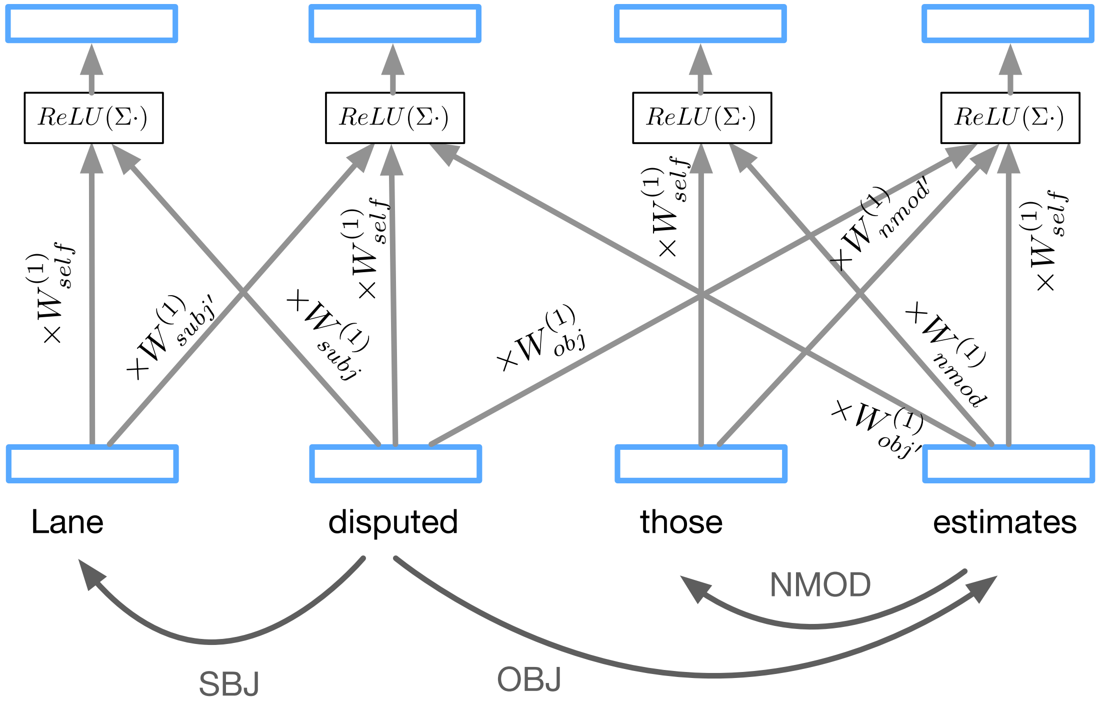
$ReLU(\Sigma\cdot)$
$\times W_{self}^{(1)}$
$\times W_{subj'}^{(1)}$
$\times W_{subj}^{(1)}$
$\times W_{obj}^{(1)}$
$\times W_{nmod'}^{(1)}$
$\times W_{nmod}^{(1)}$
$\times W_{obj'}^{(1)}$
Lane
disputed
those
estimates
SBJ
OBJ
NMOD

Example



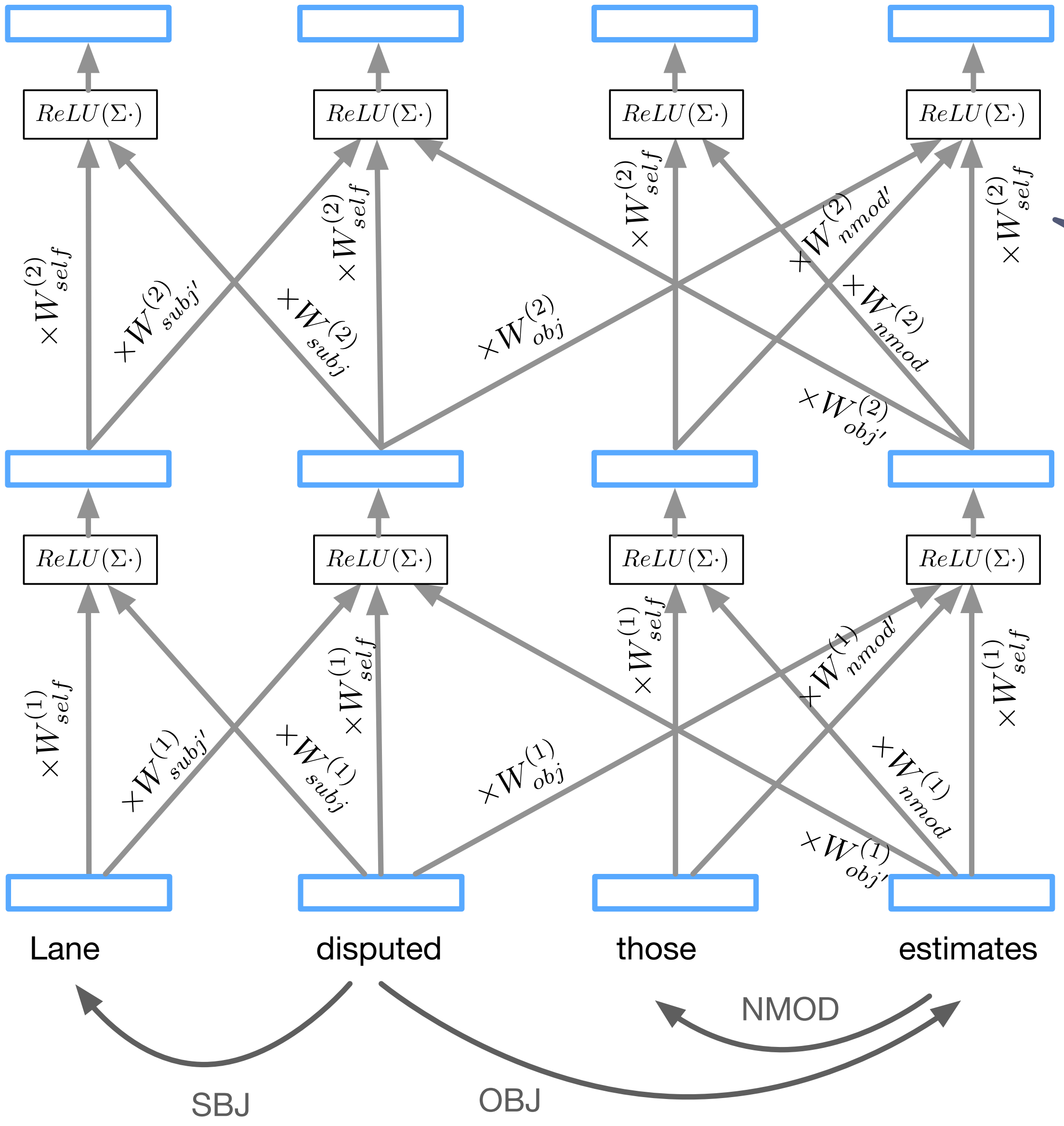
$ReLU(\Sigma\cdot)$
$ReLU(\Sigma\cdot)$
$ReLU(\Sigma\cdot)$
$ReLU(\Sigma\cdot)$
$\times W^{(1)}_{self}$
$\times W^{(1)}_{subj'}$
$\times W^{(1)}_{subj}$
$\times W^{(1)}_{self}$
$\times W^{(1)}_{obj}$
$\times W^{(1)}_{self}$
$\times W^{(1)}_{nmod'}$
$\times W^{(1)}_{nmod}$
$\times W^{(1)}_{obj'}$
$\times W^{(1)}_{self}$
Lane
disputed
those
estimates
NMOD
SBJ
OBJ

Example



$ReLU(\Sigma\cdot)$
$\times W^{(1)}_{self}$
$\times W^{(1)}_{subj'}$
$\times W^{(1)}_{subj}$
$\times W^{(1)}_{obj}$
$\times W^{(1)}_{nmod'}$
$\times W^{(1)}_{nmod}$
$\times W^{(1)}_{obj'}$
Lane
disputed
those
estimates
NMOD
SBJ
OBJ

Example



Stacking GCNs widens the syntactic neighborhood

Syntactic GCNs

$$h_v^{(k+1)} = ReLU\left(\sum_{u \in \mathcal{N}(v)} W_{L(u,v)}^{(k)} h_u^{(k)} + b_{L(u,v)}^{(k)}\right)$$

Syntactic GCNs

$$h_v^{(k+1)} = ReLU\left(\sum_{u\in\mathcal{N}(v)} W_{L(u,v)}^{(k)} h_u^{(k)} + b_{L(u,v)}^{(k)}\right)$$

Syntactic neighborhood

Syntactic GCNs

$$h_v^{(k+1)} = ReLU\left(\sum_{u \in \mathcal{N}(v)} \boxed{W_{L(u,v)}^{(k)} h_u^{(k)} + b_{L(u,v)}^{(k)}}\right)$$

Message

Syntactic neighborhood

Syntactic GCNs

Message

$$h_v^{(k+1)} = ReLU\left(\sum_{u \in \mathcal{N}(v)} \boxed{W_{L(u,v)}^{(k)} h_u^{(k)} + b_{L(u,v)}^{(k)}}\right)$$

Syntactic neighborhood

Messages are direction and label specific

Syntactic GCNs

Message

$$h_v^{(k+1)} = ReLU\left(\sum_{u \in \mathcal{N}(v)} \boxed{W_{L(u,v)}^{(k)} h_u^{(k)} + b_{L(u,v)}^{(k)}}\right)$$

Syntactic neighborhood

Messages are direction and label specific

- Overparametrized: one matrix for each label-direction pair
- $W_{L(u,v)}^{(k)} = V_{dir(u,v)}^{(k)}$

Edge-wise Gates

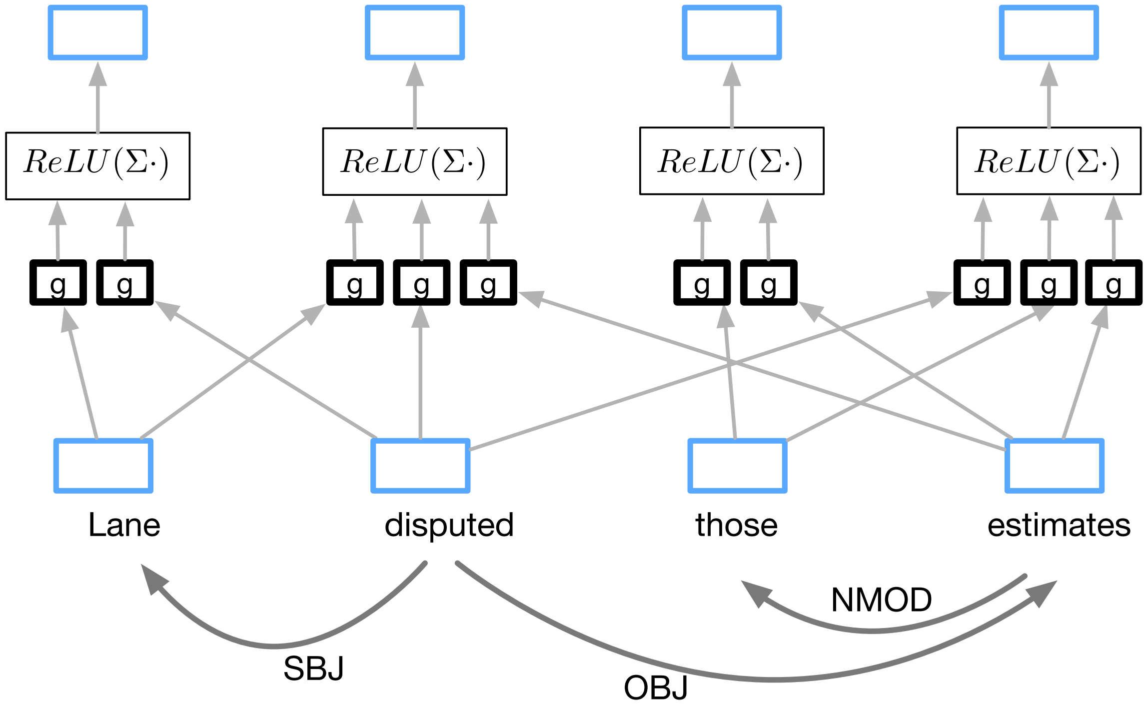
- Not all edges are equally important

Edge-wise Gates

- Not all edges are equally important
- We should not blindly rely on predicted syntax

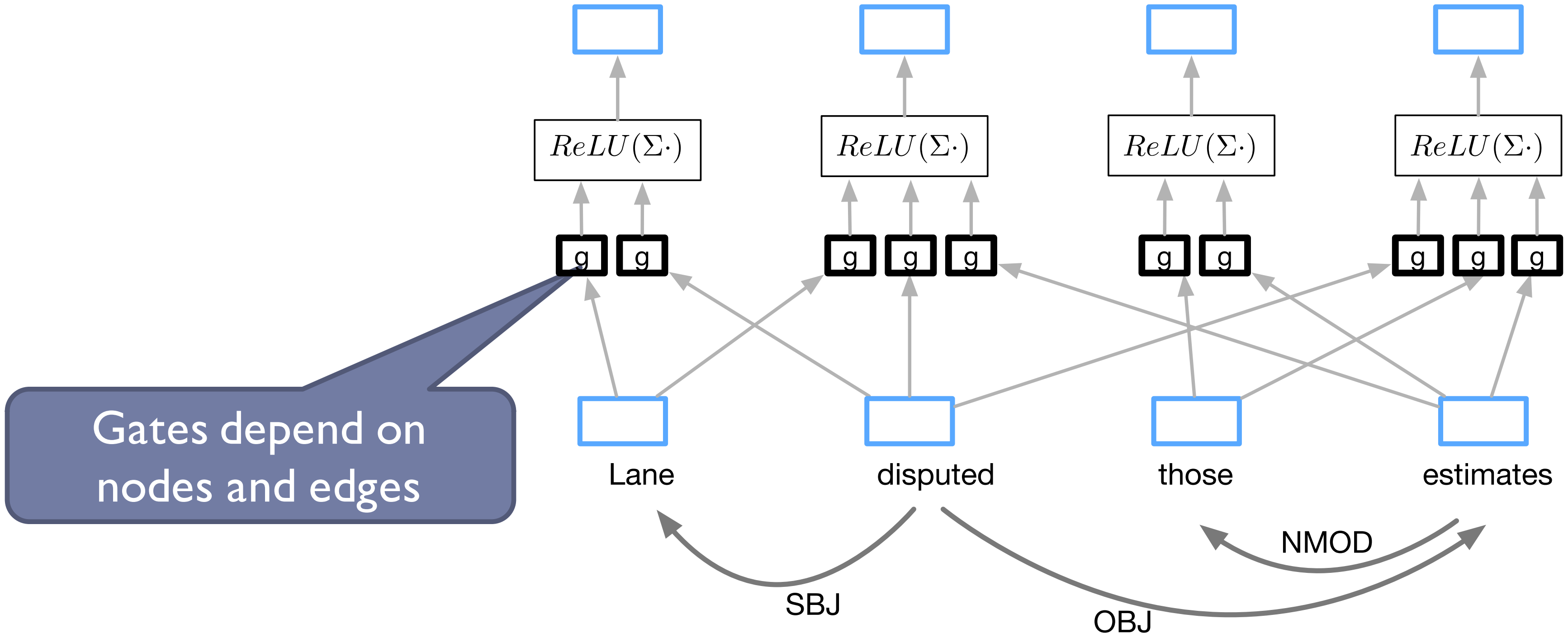
Edge-wise Gates

- Not all edges are equally important
- We should not blindly rely on predicted syntax
- Gates decide the "importance" of each message



Edge-wise Gates

- Not all edges are equally important
- We should not blindly rely on predicted syntax
- Gates decide the "importance" of each message



Encoding Sentences with Graph Convolutional Networks

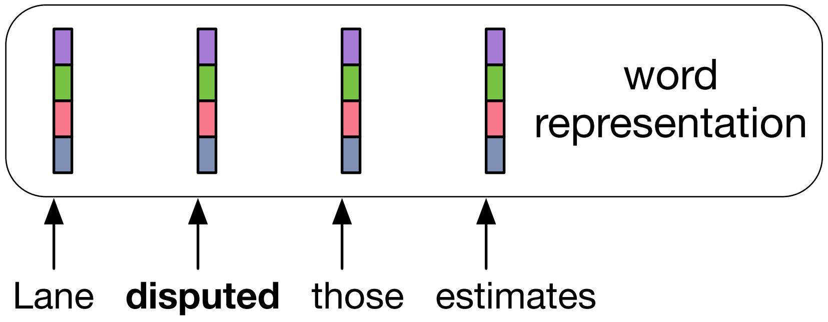
- Graph Convolutional Networks (GCNs)
- Syntactic GCNs
- **Semantic Role Labeling Model**
- Experiments
- Conclusions

Our Model

- Word representation
- Bidirectional LSTM encoder
- GCN Encoder
- Local role classifier

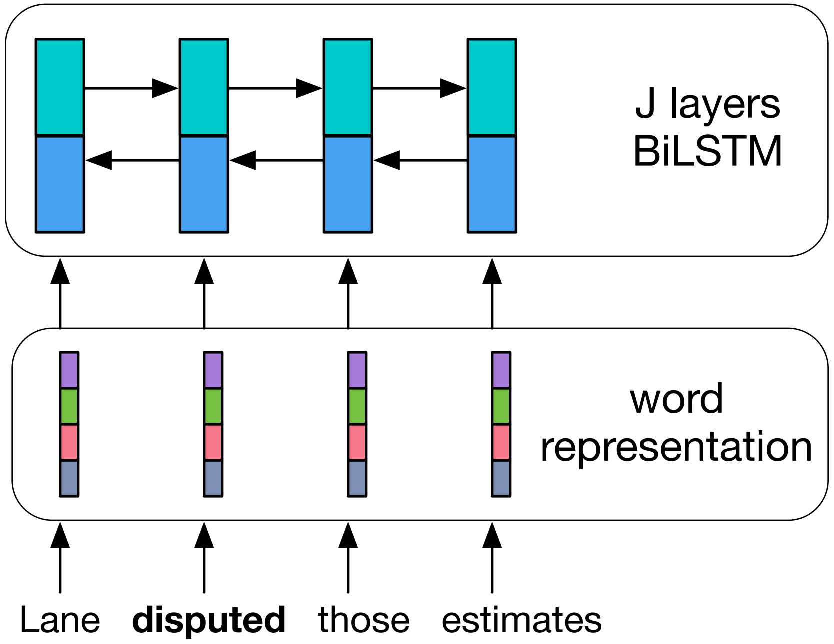
Word Representation

- Pretrained word embeddings
- Word embeddings
- POS tag embeddings
- Predicate lemma embeddings
- Predicate flag



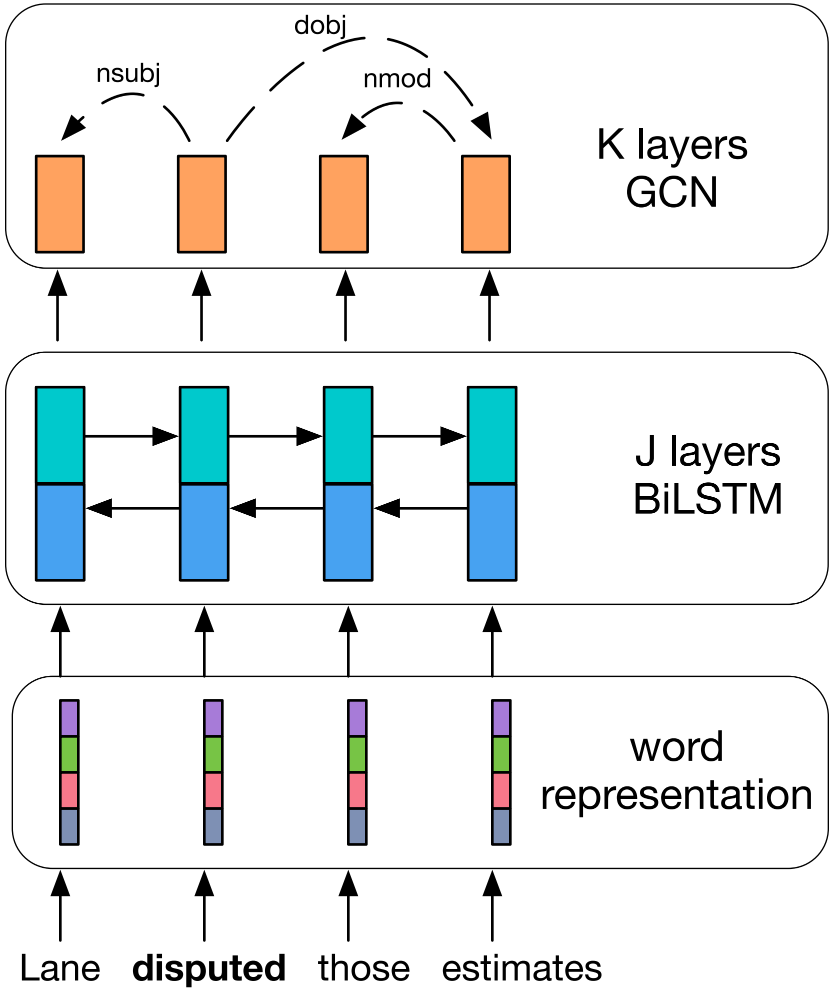
BiLSTM Encoder

- Encode each word with its left and right context
- Stacked BiLSTM



GCNs Encoder

- Syntactic GCNs after BiLSTM encoder
  - Add syntactic information
  - Skip connections
  - Longer dependencies are captured



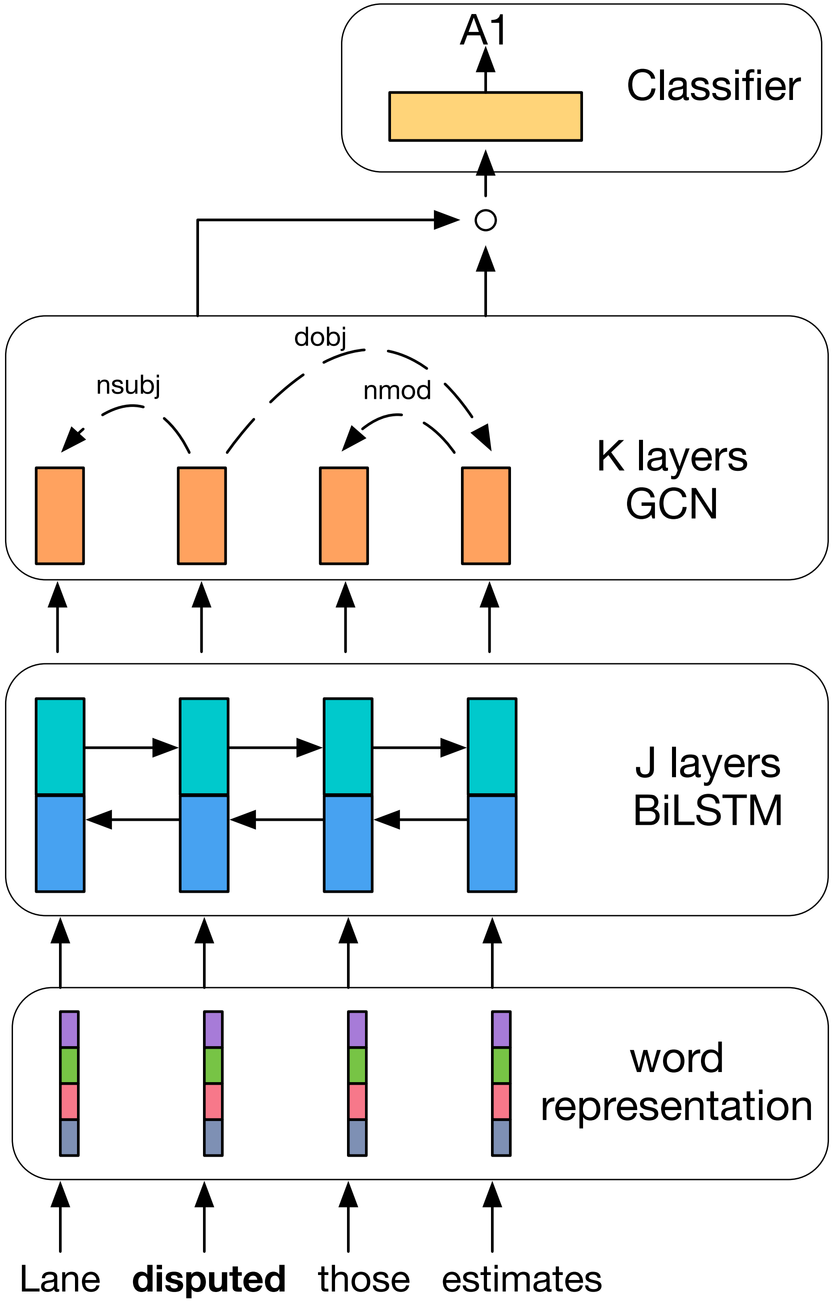
Semantic Role Classifier

- Local log-linear classifier

$$p(r|t_i, t_p, l) \propto \exp(W_{l,r}(t_i \circ t_p))$$

predicate representation

candidate argument representation



Encoding Sentences with Graph Convolutional Networks

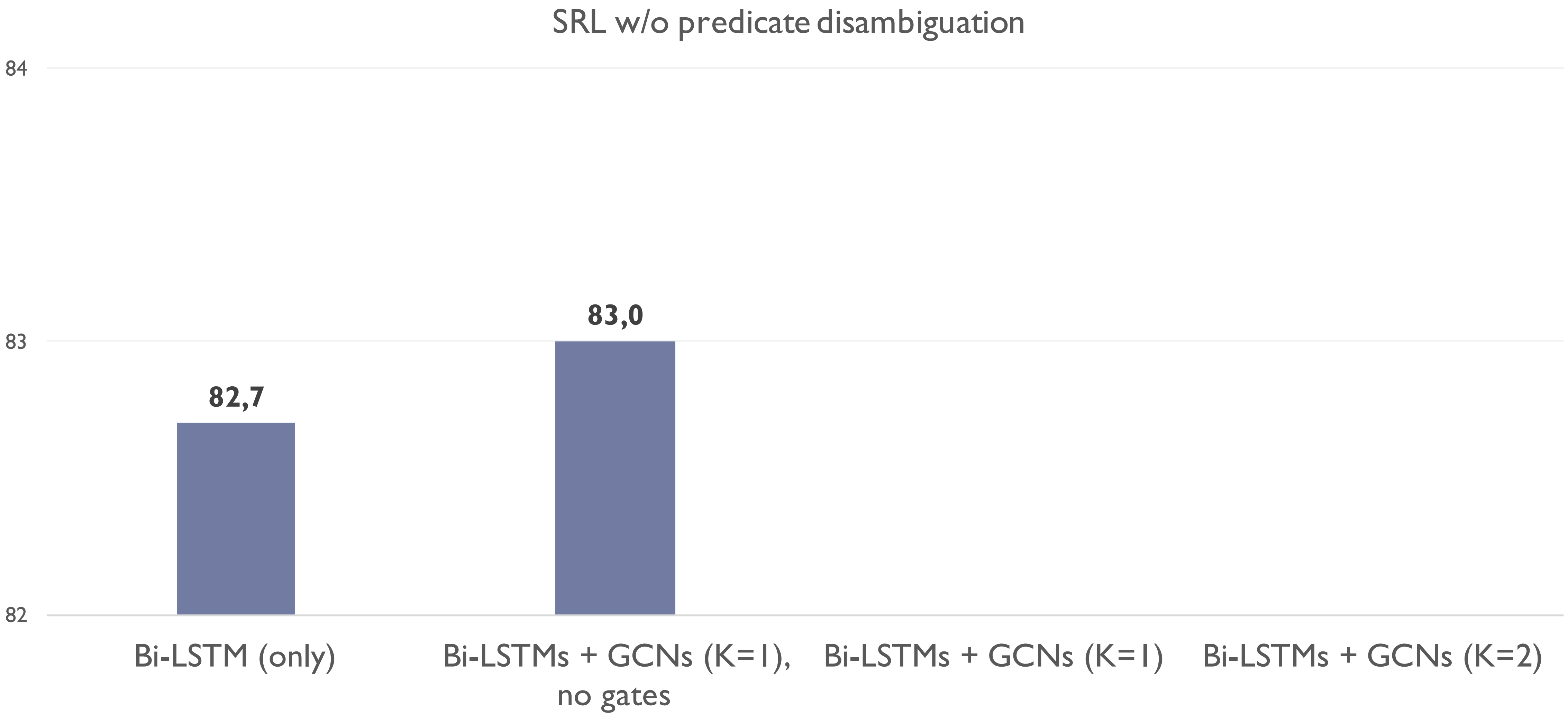
- Graph Convolutional Networks (GCNs)
- Syntactic GCNs
- Semantic Role Labeling Model
- **Experiments**
- Conclusions

Experiments

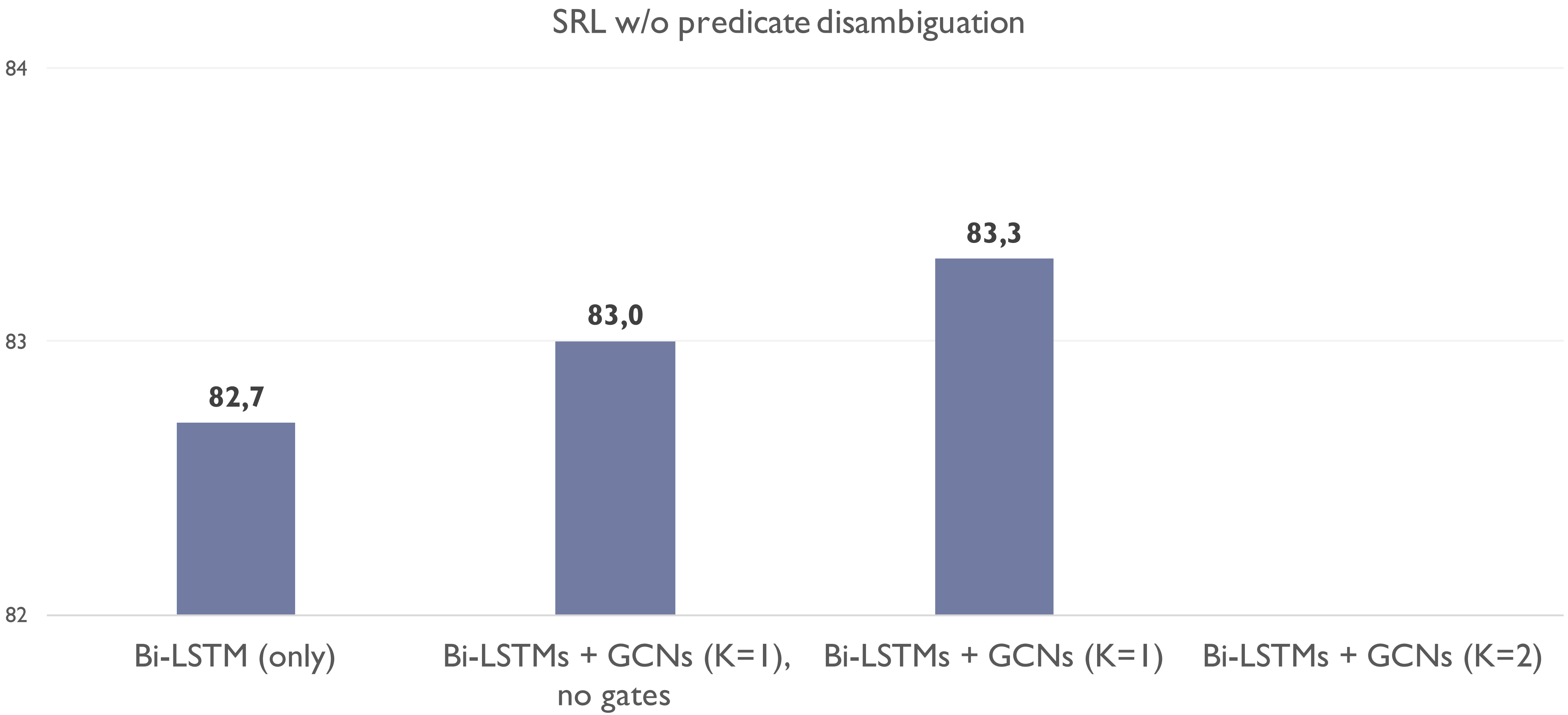
- Data
  - CoNLL-2009 dataset - English and Chinese
  - F1 evaluation measure
- Model
  - Hyperparameters tuned on English development set
  - State-of-the-art predicate disambiguation models

Ablation Experiments (English Dev set)

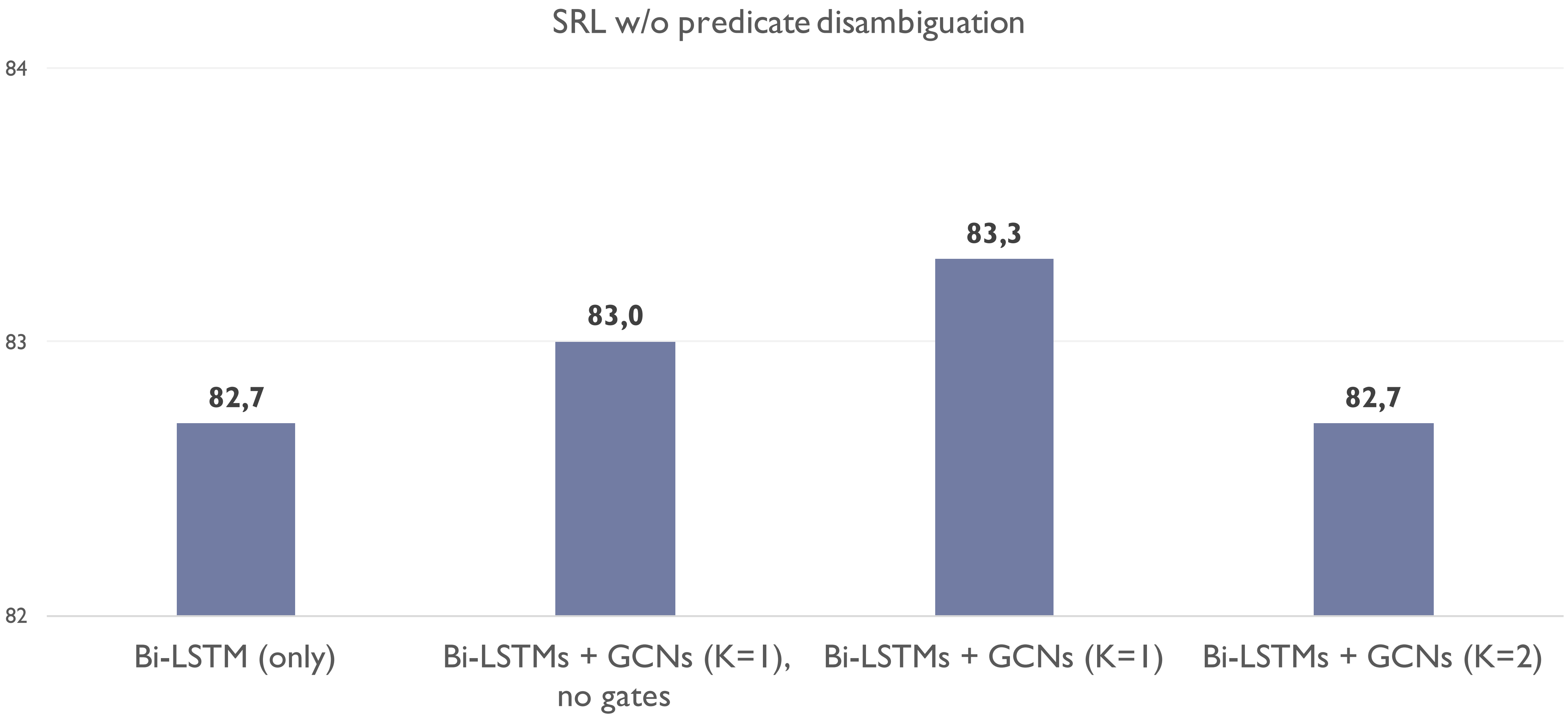
Ablation Experiments (English Dev set)



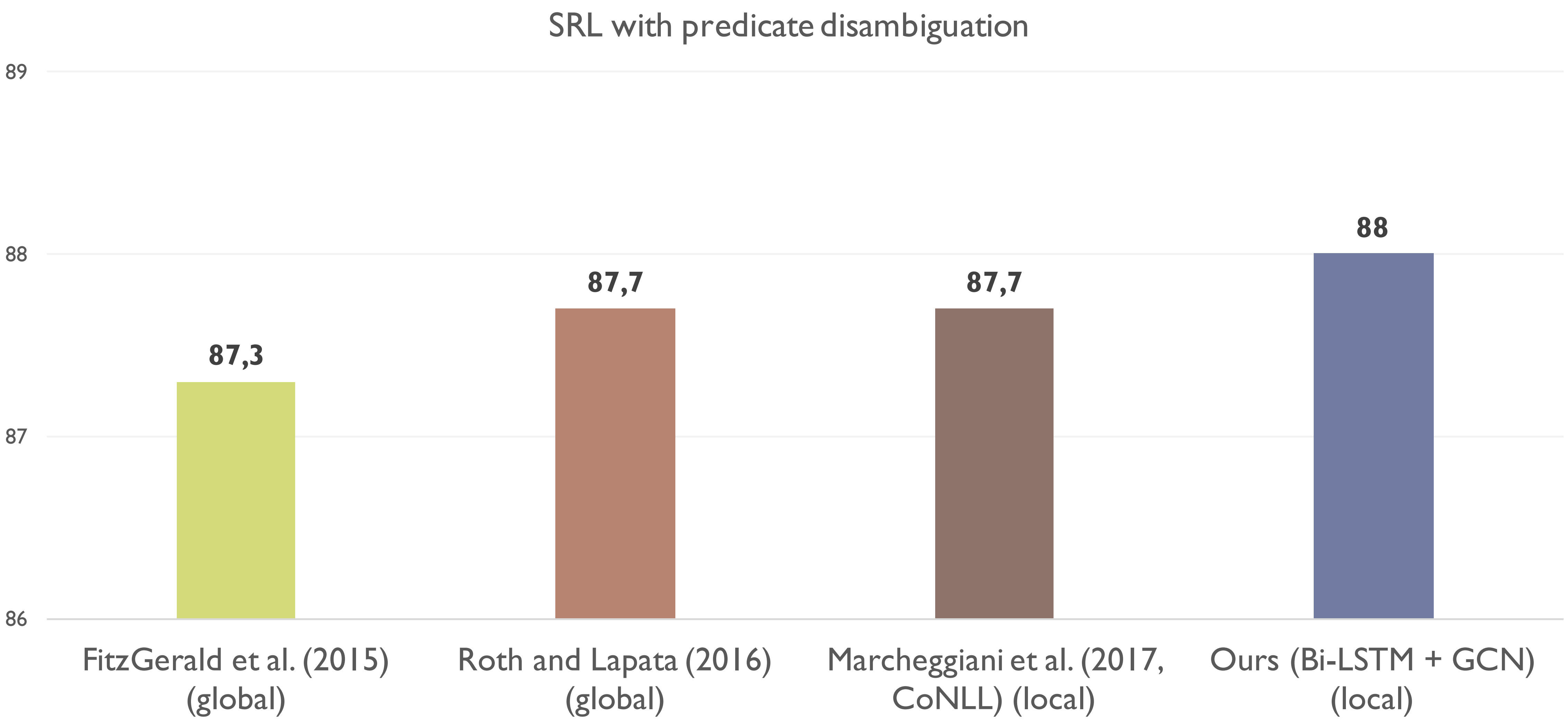
Ablation Experiments (English Dev set)



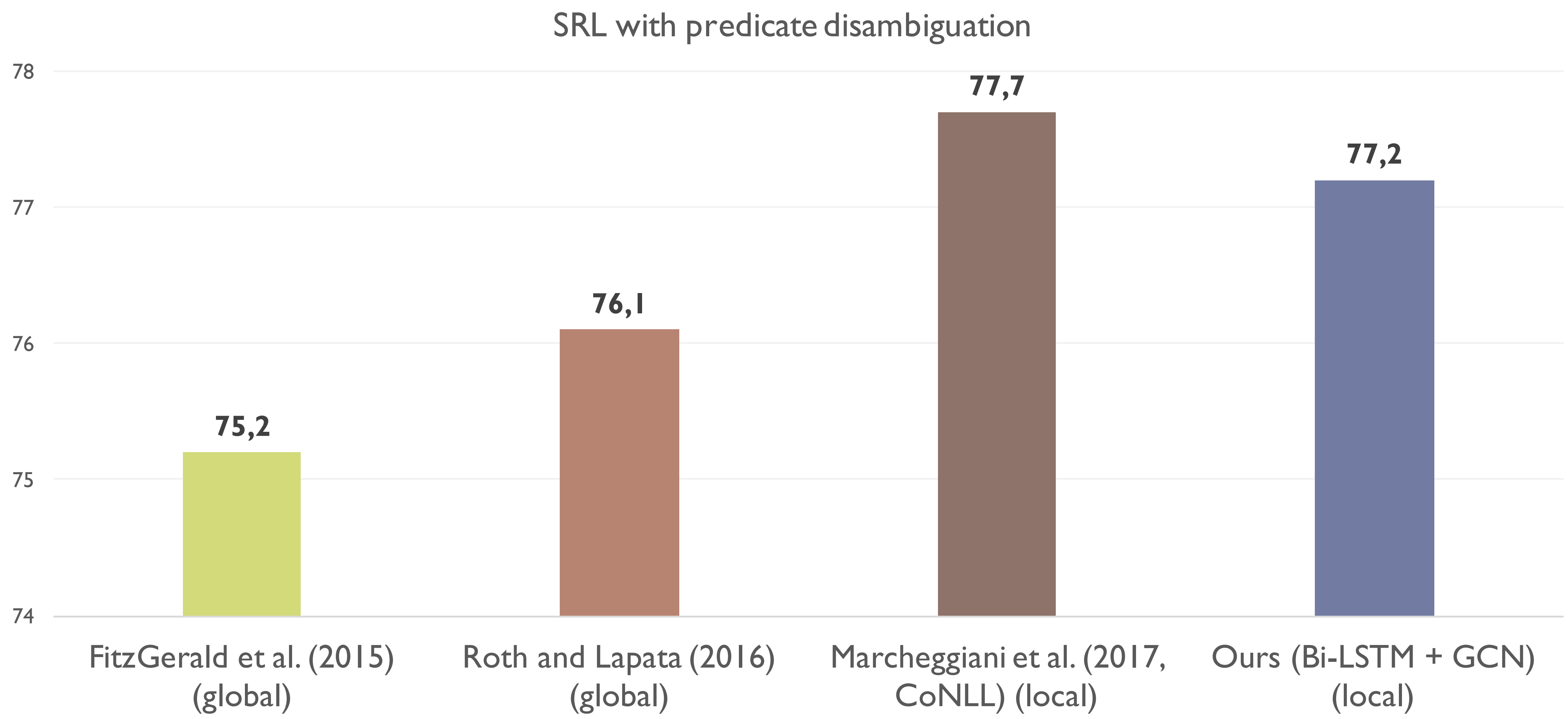
Ablation Experiments (English Dev set)



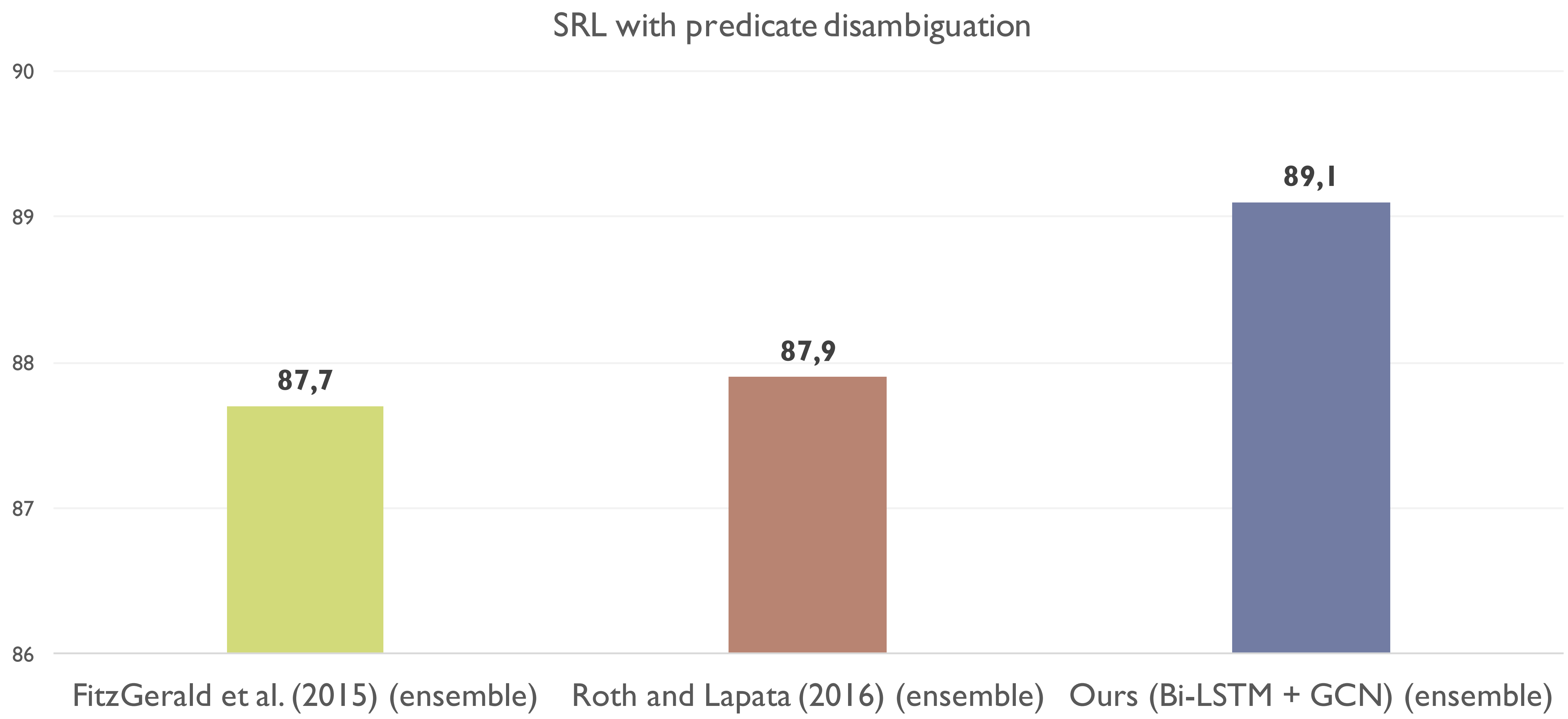
English Test Set



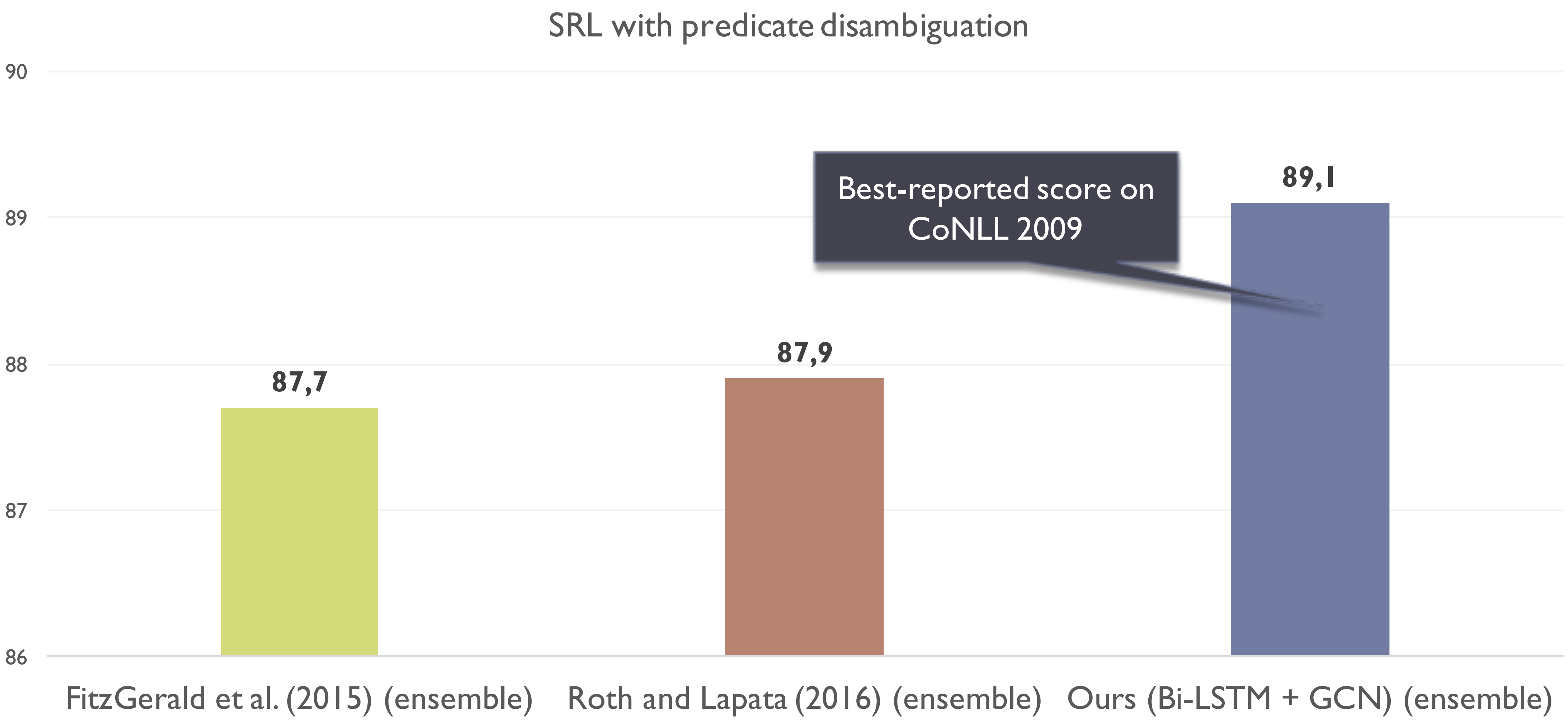
English Out of Domain



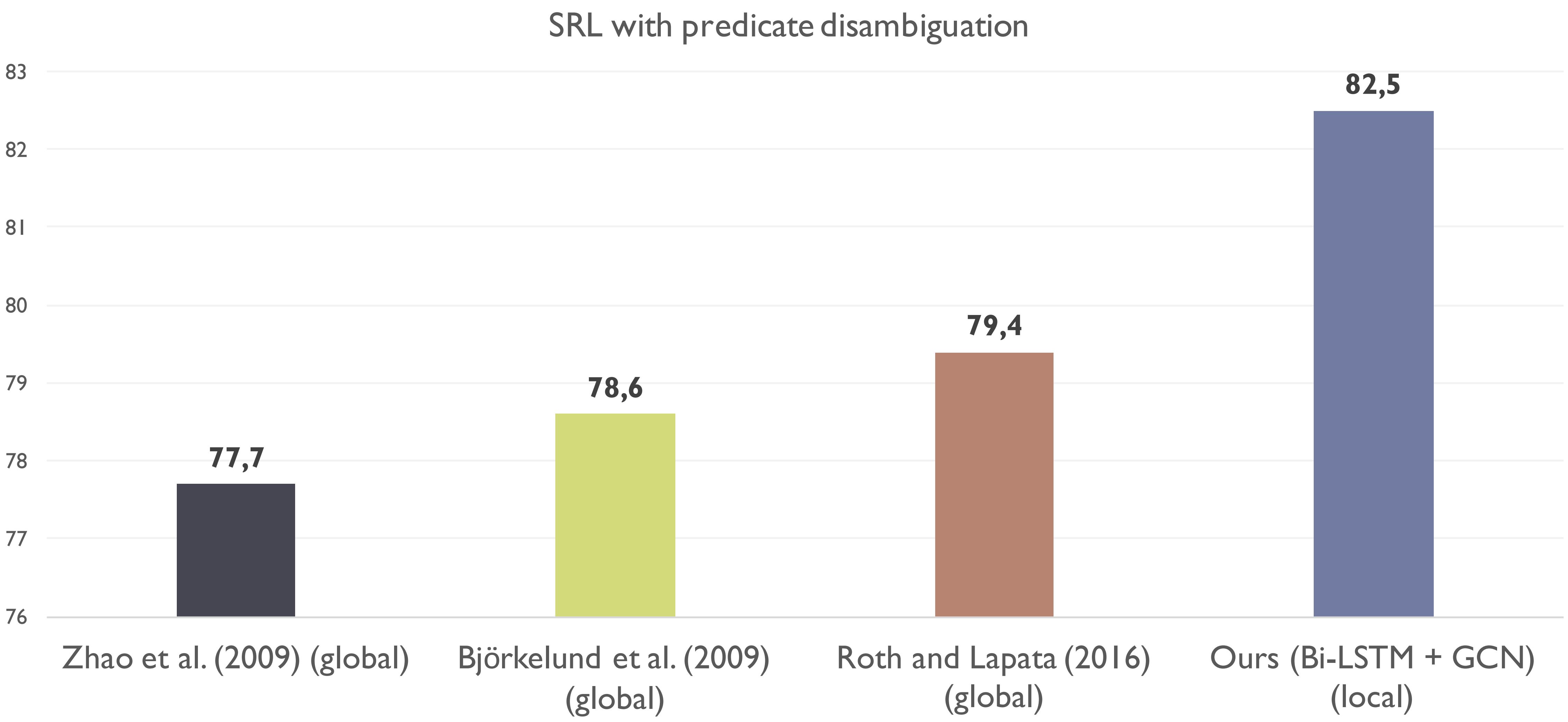
English Test Set (Ensemble)



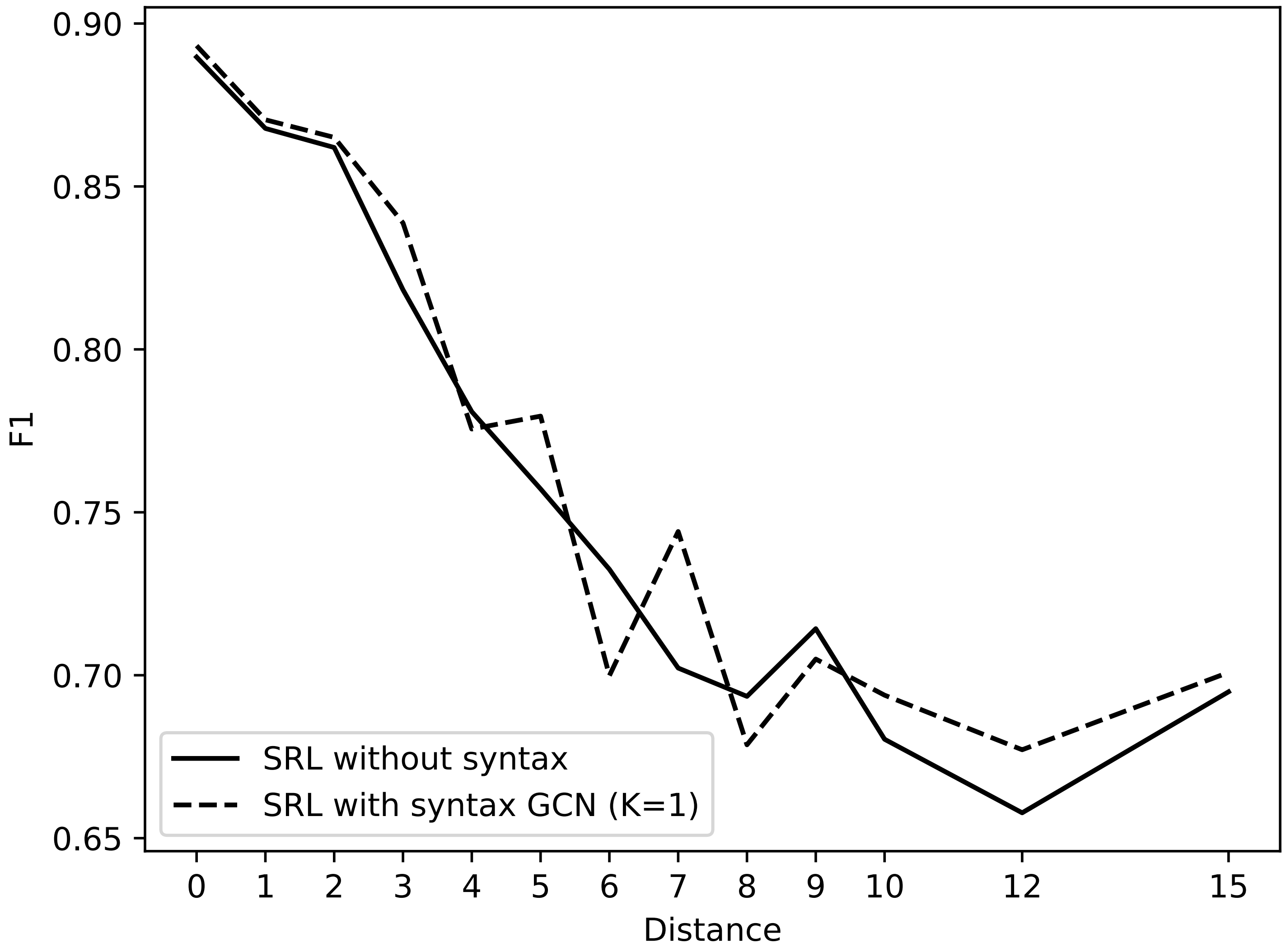
English Test Set (Ensemble)



Chinese Test Set



Long-range Dependencies (English Dev Set)



Conclusion

- Syntax-aware state-of-the-art model for dependency-based SRL
  - English and Chinese
- GCNs for encoding syntactic structures into NN
  - Semantics, coreference, discourse

Conclusion

- Syntax-aware state-of-the-art model for dependency-based SRL
  - English and Chinese
- GCNs for encoding syntactic structures into NN
  - Semantics, coreference, discourse

- Funding:
  - ERC StG BroadSem 678254
  - NWO VIDI 639.022.518
  - Amazon Web Services (AWS) grant

github.com/diegma/neural-dep-srl