Encoding Linguistic Structures with Graph Convolutional Networks

Diego Marcheggiani

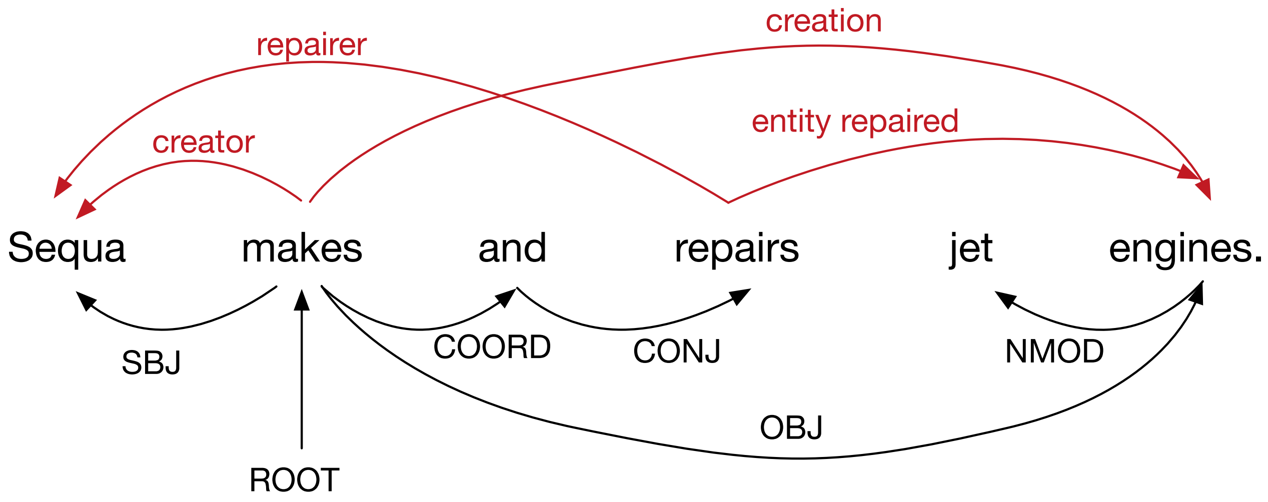
Joint work with Ivan Titov and Joost Bastings

University of Amsterdam

University of Edinburgh

@South England NLP Meetup

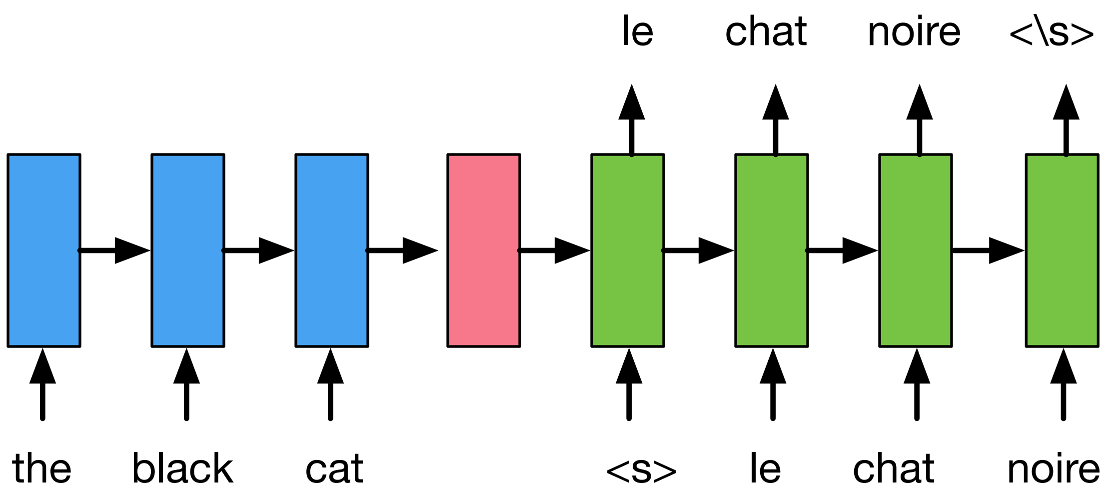
Structured (Linguistic) Priors



"I voted for Palpatine because he was most aligned with my values," she said.

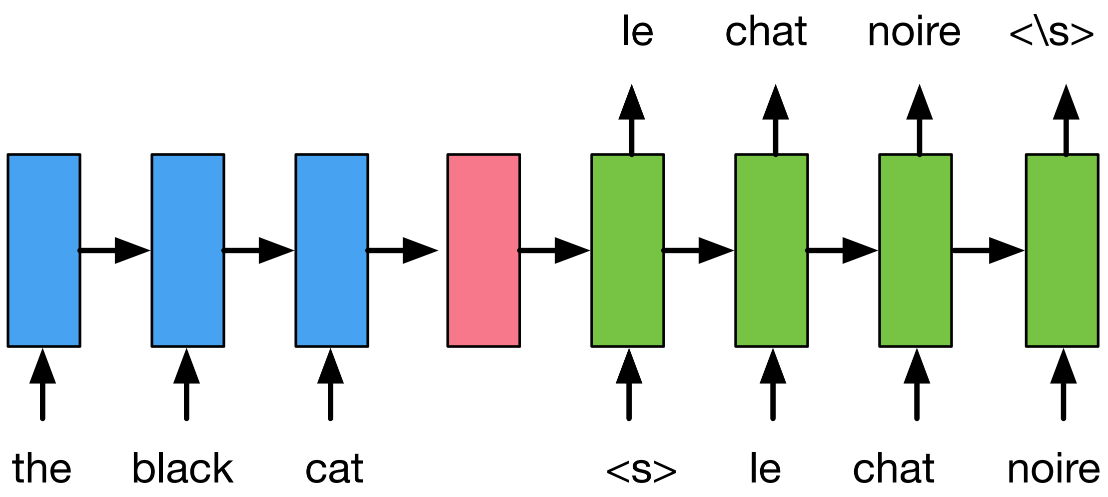
Sequence to Sequence

[Sutskever et al., 2014]



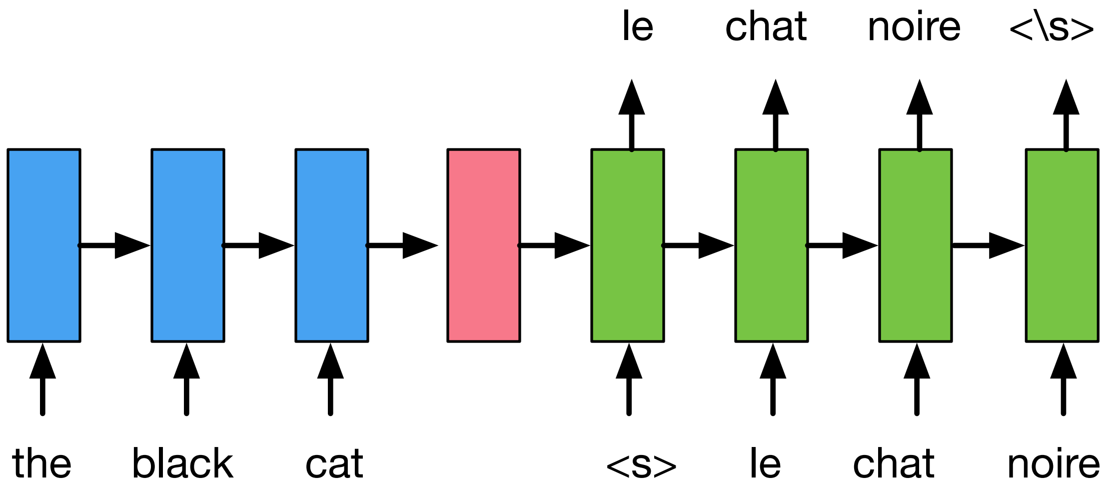
Sequence to Sequence

[Sutskever et al., 2014]



- Language is not (only) a sequence of words
- We have linguistic knowledge

Sequence to Sequence



- Language is not (only) a sequence of words
- We have linguistic knowledge

Encode structured linguistic knowledge into NN using Graph Convolutional Networks

Outline

- Semantic Role Labeling
- Graph Convolutional Networks (GCN)
- Syntactic GCN for Semantic Role Labeling (SRL)
- SRL Model
- Exploiting Semantics in Neural Machine Translation with GCNs

Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling
Diego Marcheggiani, Ivan Titov. In *Proceedings of EMNLP*, 2017.

Exploiting Semantics in Neural Machine Translation with Graph Convolutional Networks
Diego Marcheggiani, Joost Bastings, Ivan Titov. In *Proceedings of NAACL-HLT*, 2018.

Semantic Role Labeling

- Predicting the predicate-argument structure of a sentence

Sequa makes and repairs jet engines.

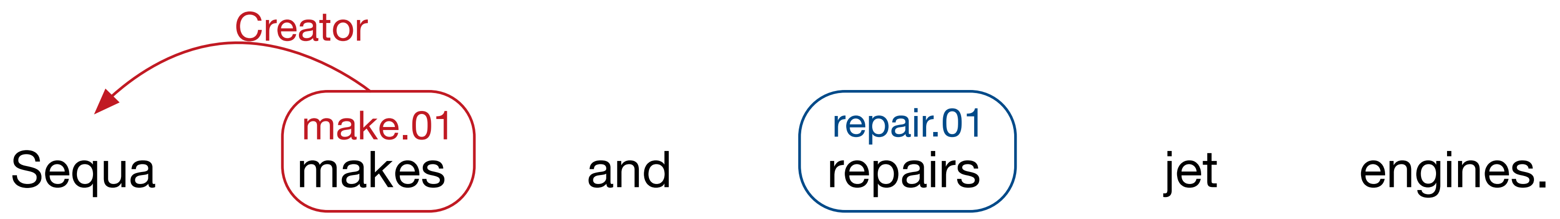
Semantic Role Labeling

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

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

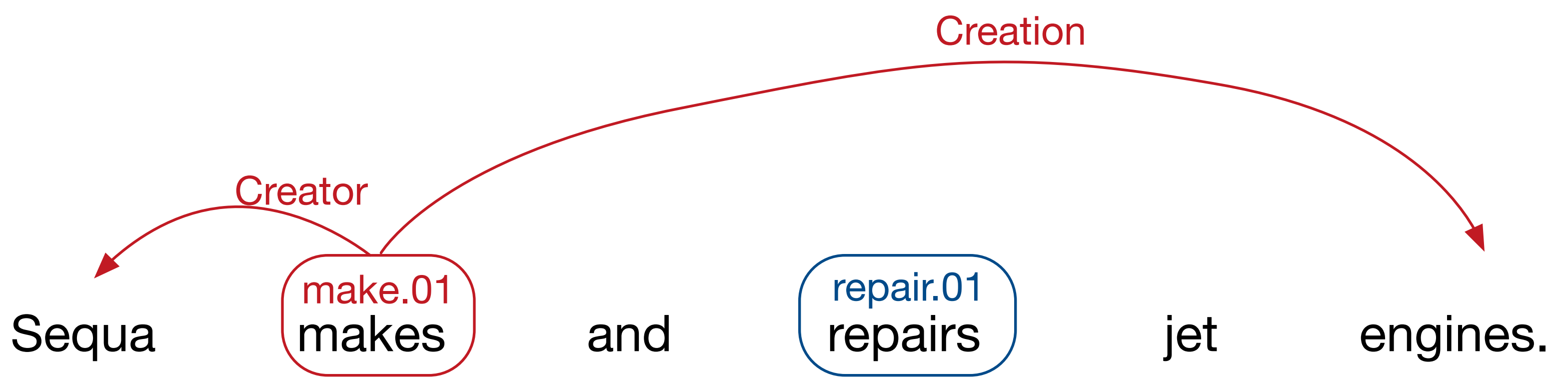
Semantic Role Labeling

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



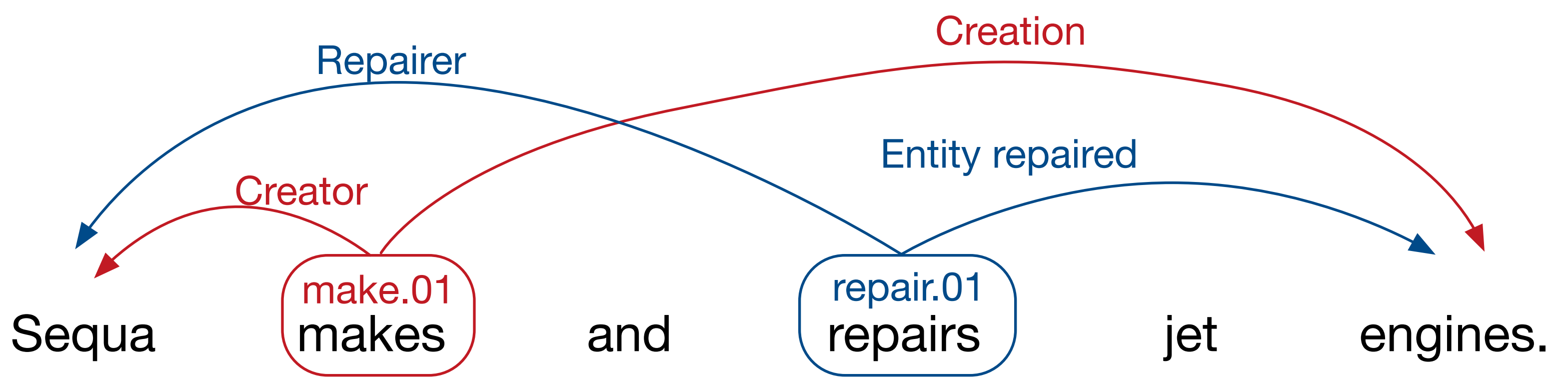
Semantic Role Labeling

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



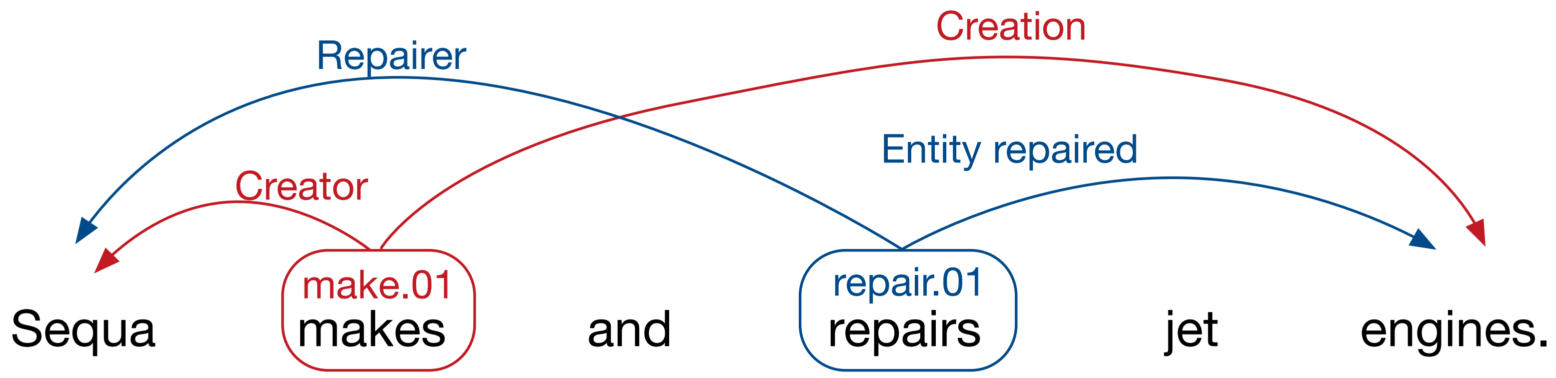
Semantic Role Labeling

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



Semantic Role Labeling

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



Semantic Role Labeling

Information extraction
Surdeanu et al. 2003
Christensen et al. 2010

Machine translation
Wu and Fung 2009
Aziz et al. 2011

Question answering
Narayanan and Harabagiu 2004
Shen and Lapata 2007
Khashabi et al. 2018

Related work

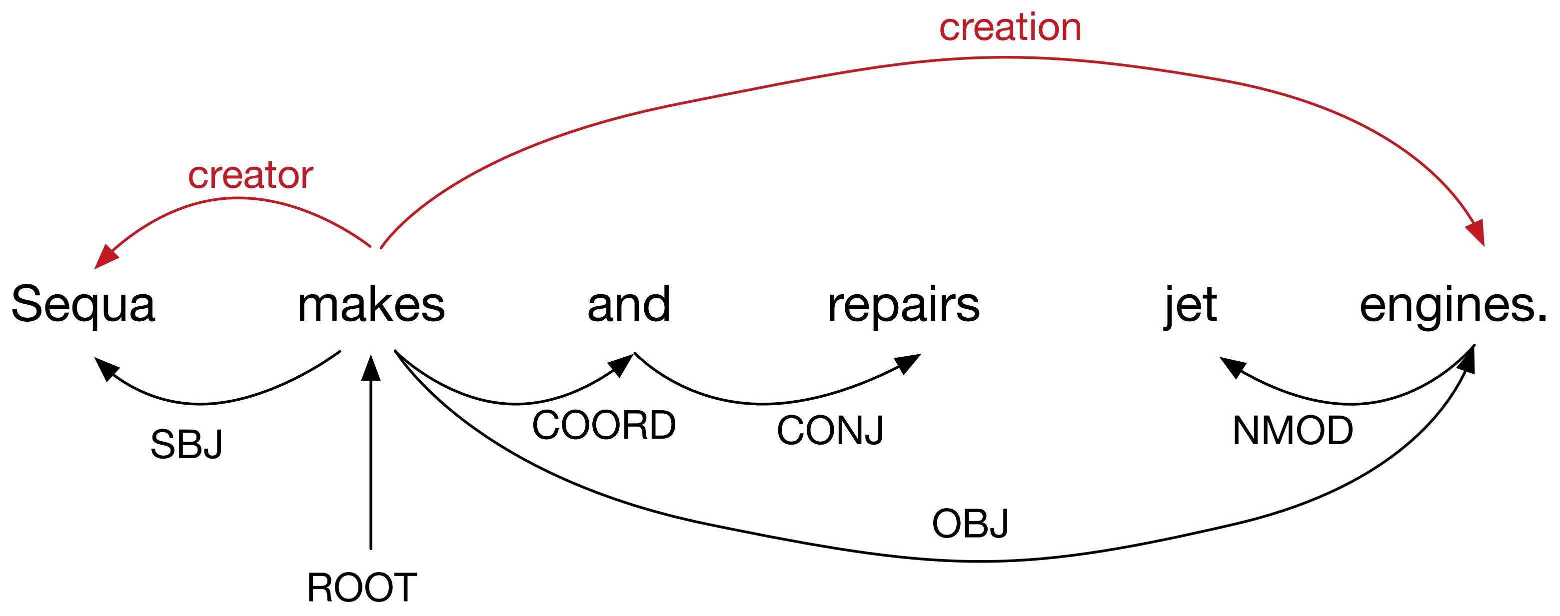
Tutorial on Semantic Role Labeling at EMNLP 2017

Related work

Tutorial on Semantic Role Labeling at EMNLP 2017

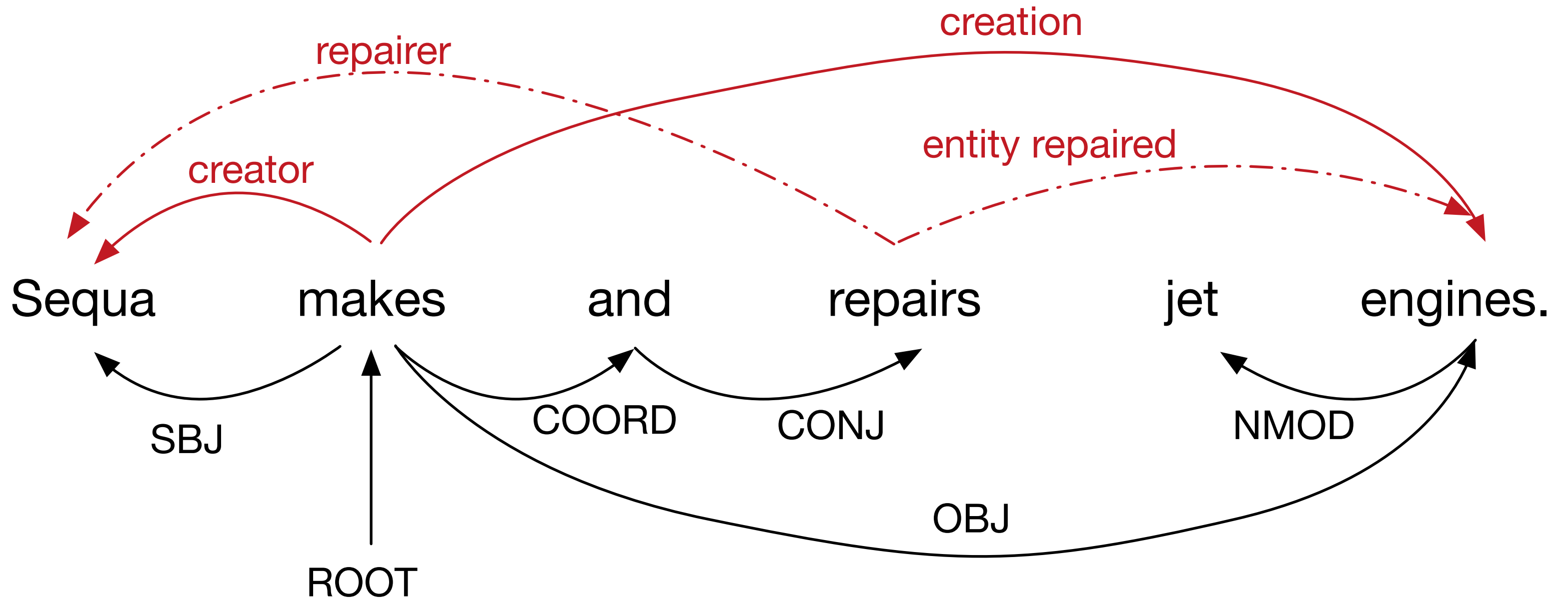
- SRL systems that use syntax with simple NN architectures
  - [FitzGerald et al., 2015]
  - [Roth and Lapata, 2016]
- Recent models ignore linguistic bias
  - [Zhou and Xu, 2014]
  - [He et al., 2017]
  - **[Marcheggiani et al., 2017]**

Motivations



- Some semantic dependencies are mirrored in the syntactic graph

Motivations



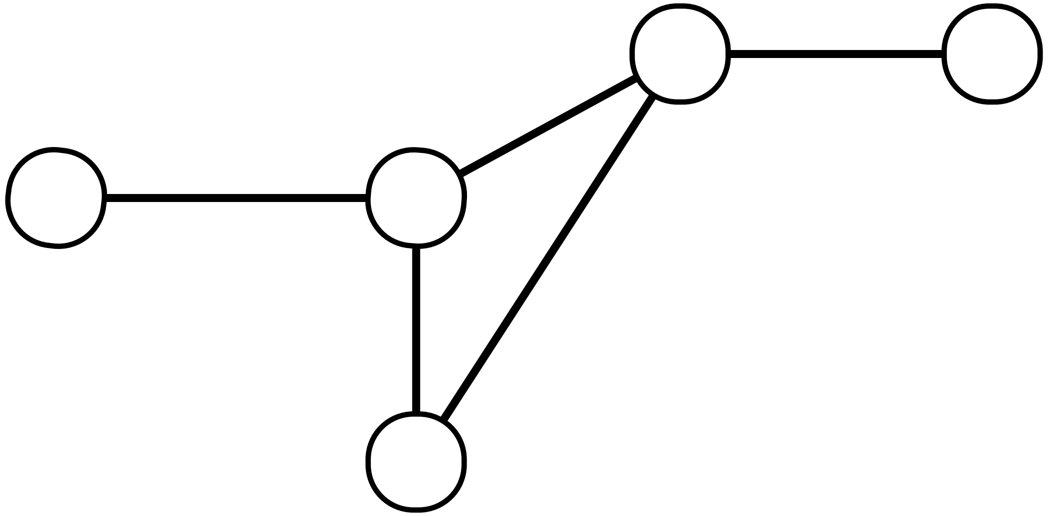
- Some semantic dependencies are mirrored in the syntactic graph
- Not all of them – syntax-semantics interface is not trivial

Outline

- Semantic Role Labeling
- **Graph Convolutional Networks (GCN)**
- Syntactic GCN for Semantic Role Labeling (SRL)
- SRL Model
- Exploiting Semantics in Neural Machine Translation with GCNs

Graph Convolutional Networks (message passing)

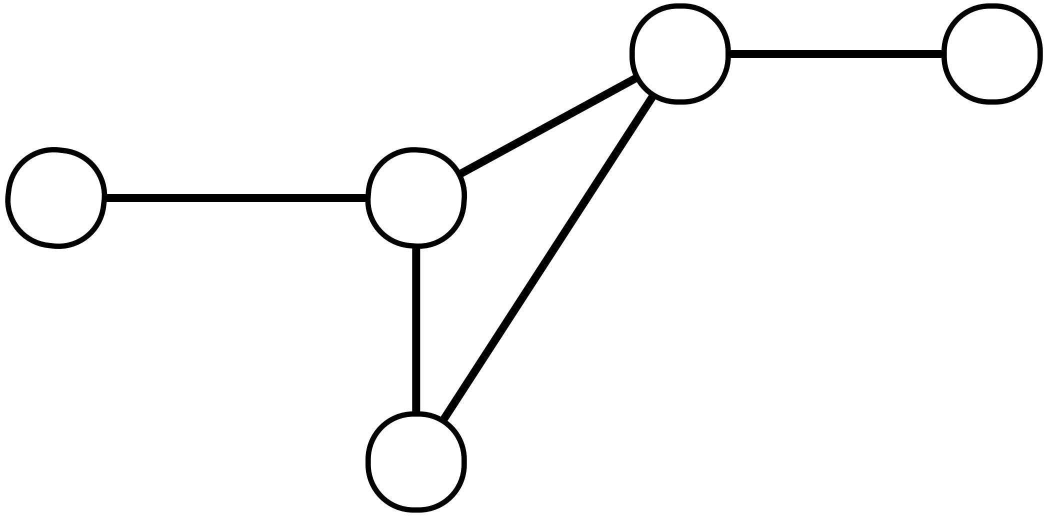
[Gori et al. 2005
Scarselli et al. 2009
Kipf and Welling, 2016]



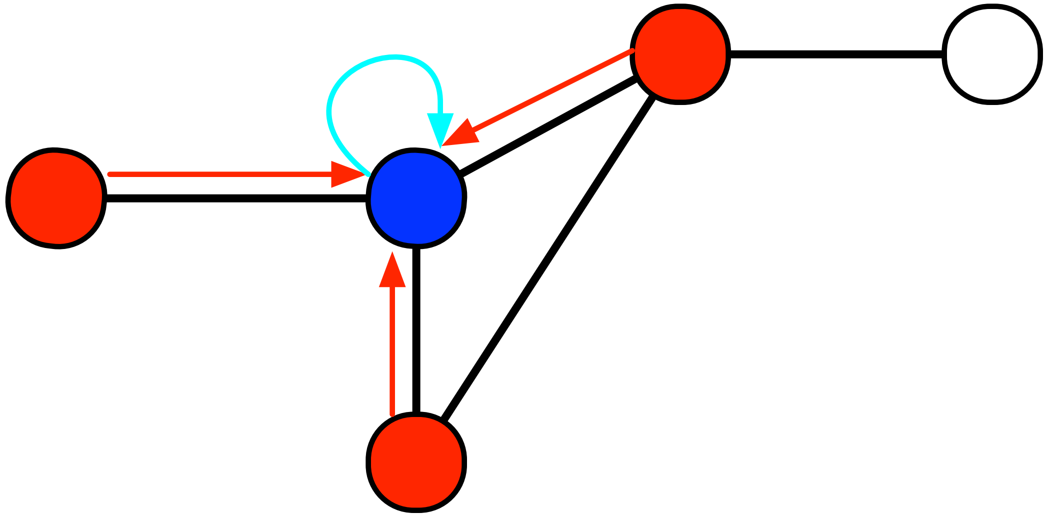
Undirected graph

Graph Convolutional Networks (message passing)

[Gori et al. 2005
Scarselli et al. 2009
Kipf and Welling, 2016]



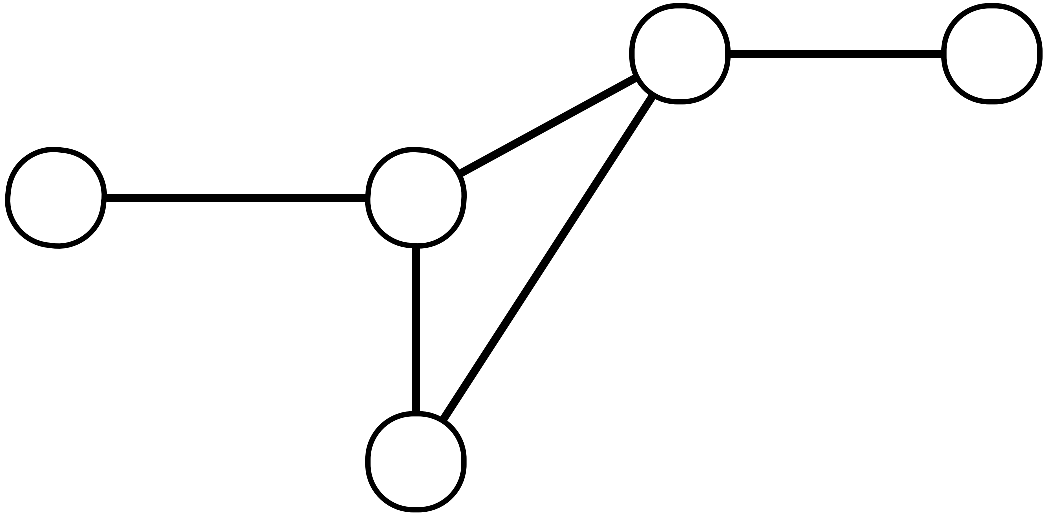
Undirected graph



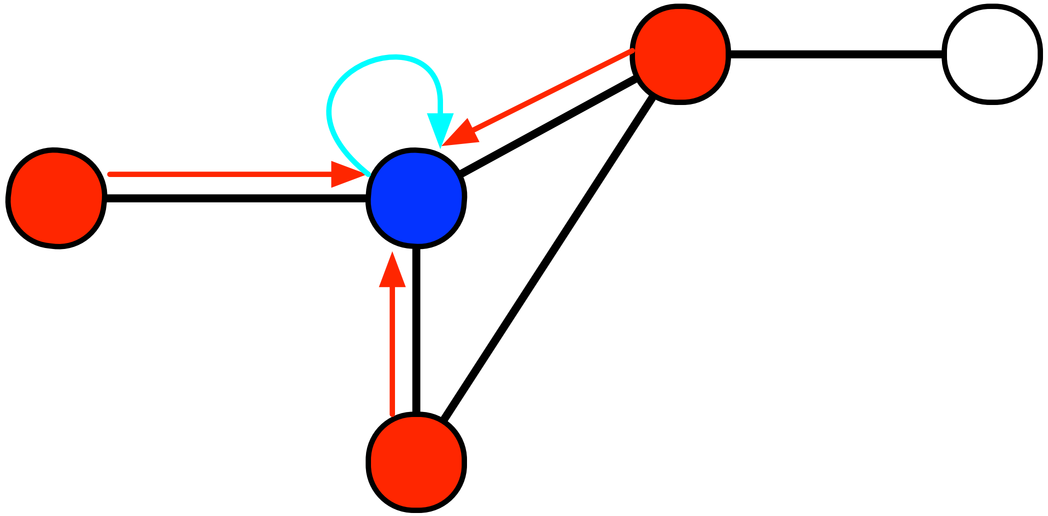
Update of the blue node

Graph Convolutional Networks (message passing)

[Kipf and Welling, 2016]



Undirected graph



Update of the blue node

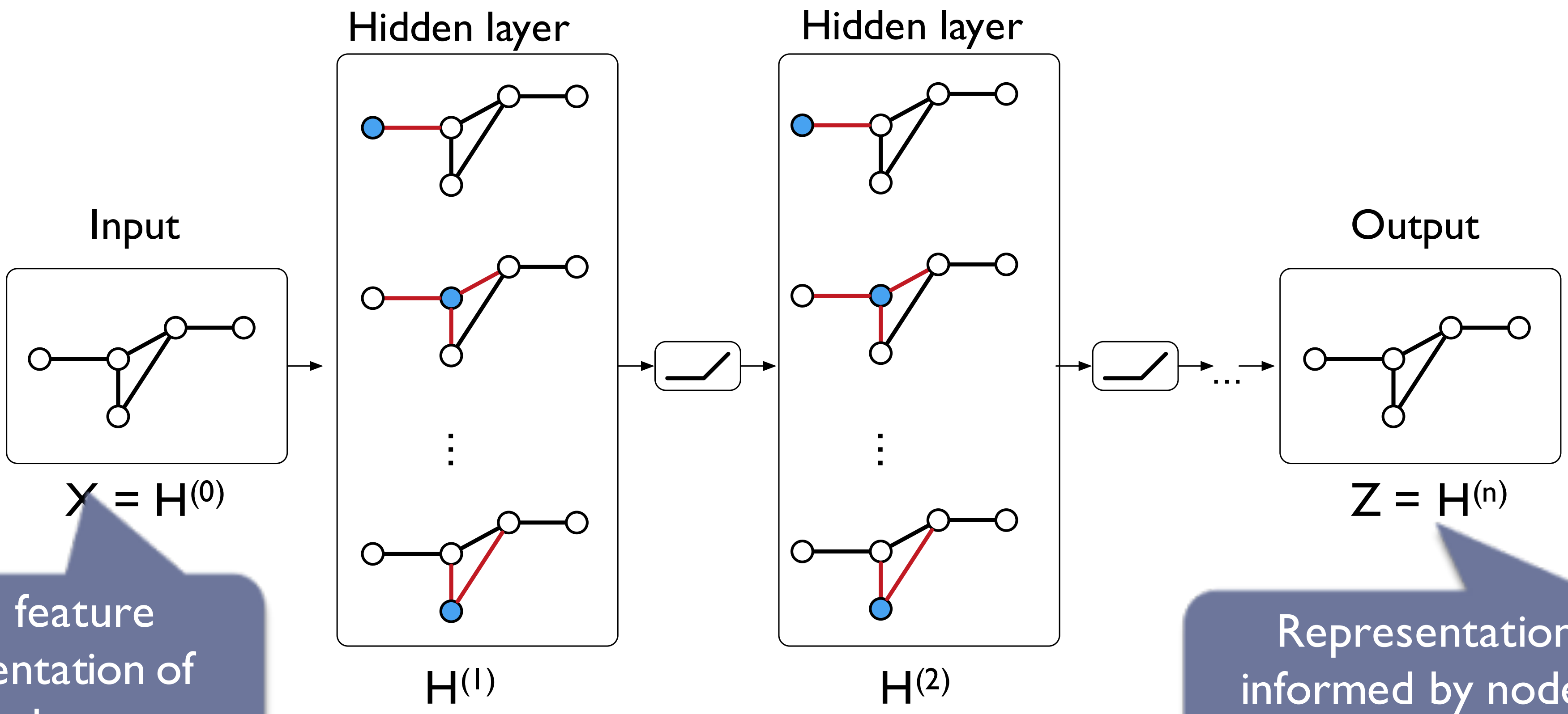
$$h_i = ReLU\left(W_0 h_i + \sum_{j \in \mathcal{N}(v)} W_1 h_j\right)$$

Self loop

Neighborhood

GCNs Pipeline

[Kipf and Welling, 2016]

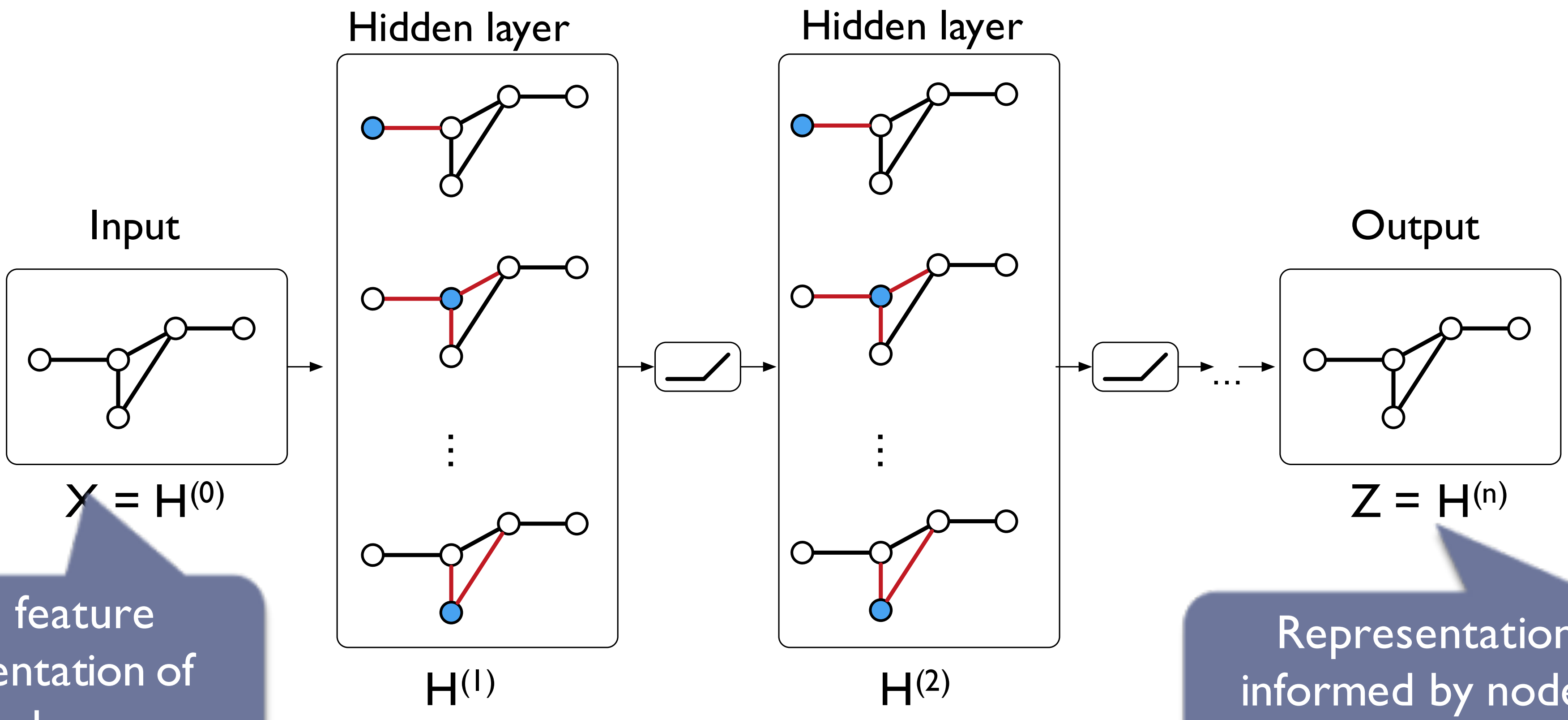


Initial feature representation of nodes

Representation informed by nodes' neighborhood

GCNs Pipeline

[Kipf and Welling, 2016]



Initial feature representation of nodes

Representation informed by nodes' neighborhood

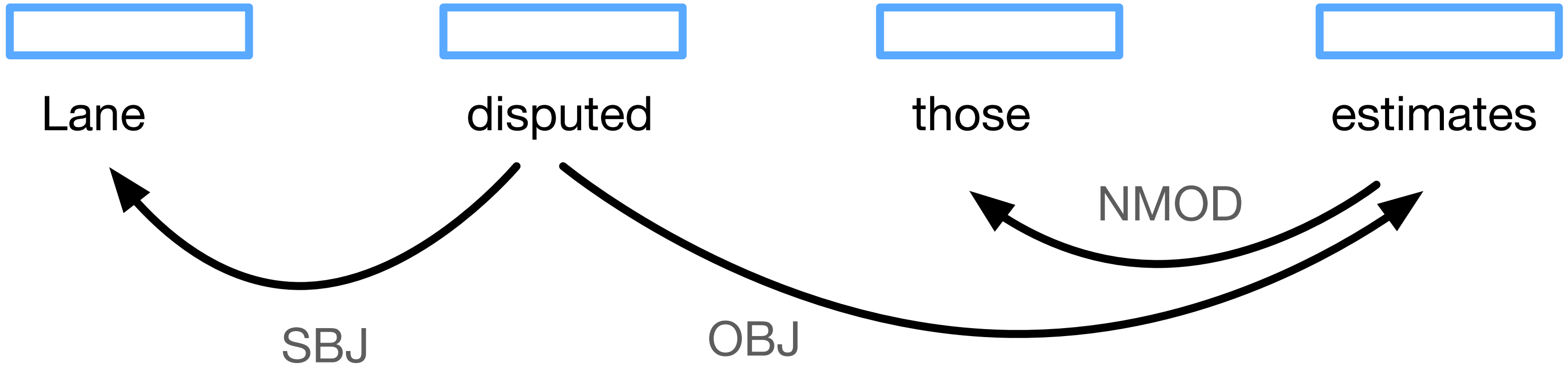
Extend GCNs for syntactic dependency trees

Outline

- Semantic Role Labeling
- Graph Convolutional Networks (GCN)
- **Syntactic GCN for Semantic Role Labeling (SRL)**
- SRL Model
- Exploiting Semantics in Neural Machine Translation with GCNs

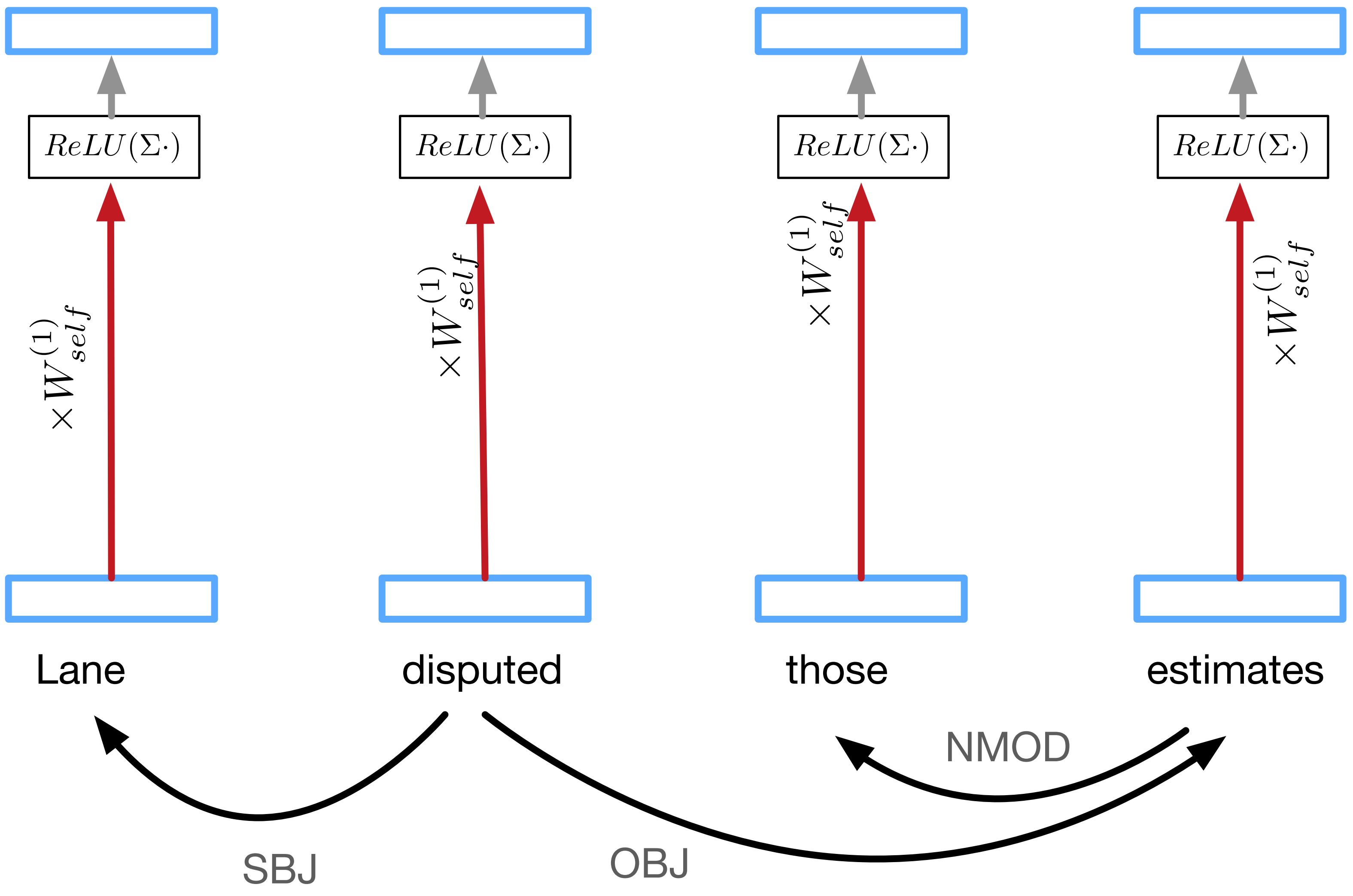
Example

[Marcheggiani and Titov, 2017]



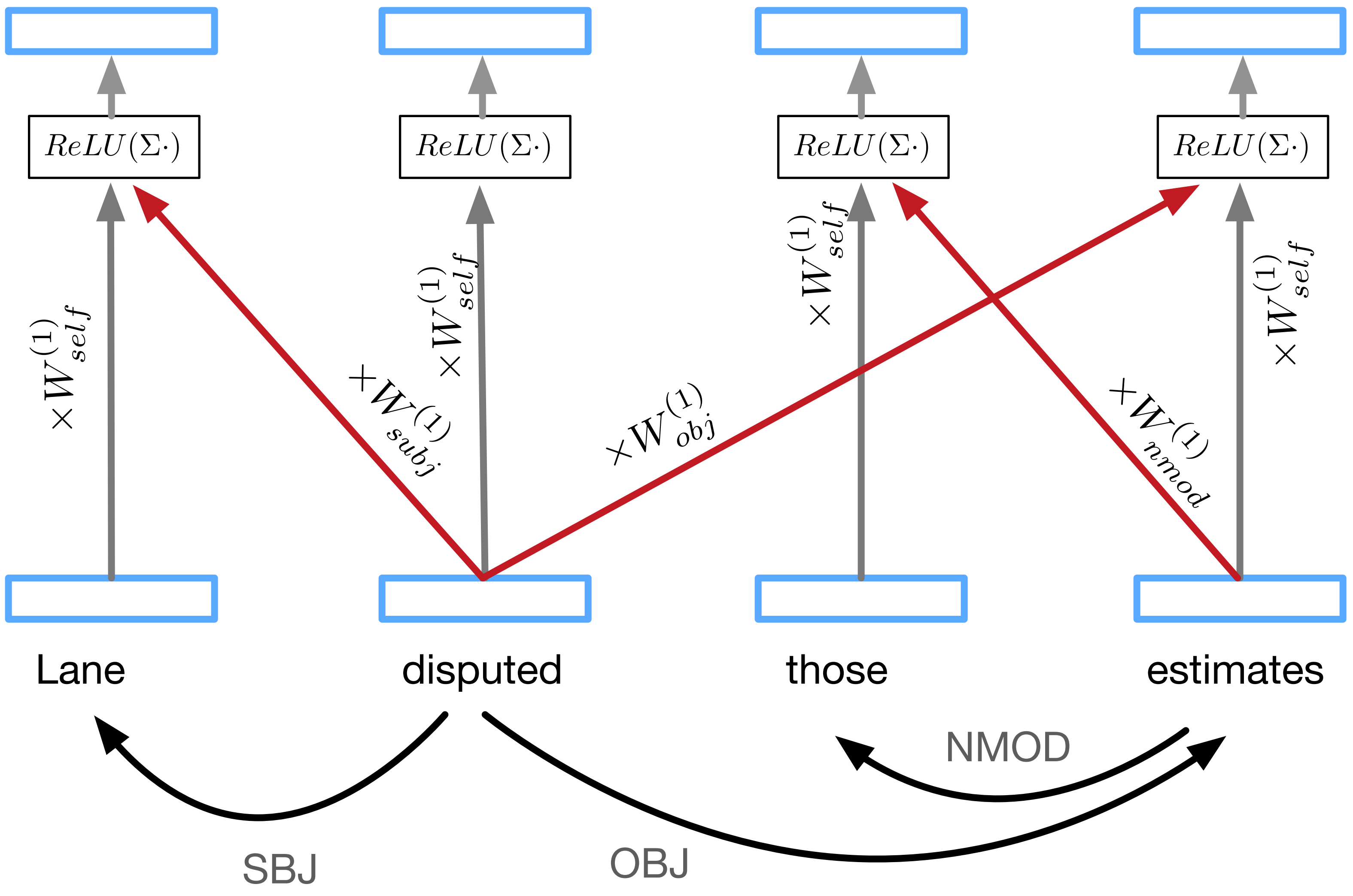
Example

[Marcheggiani and Titov, 2017]



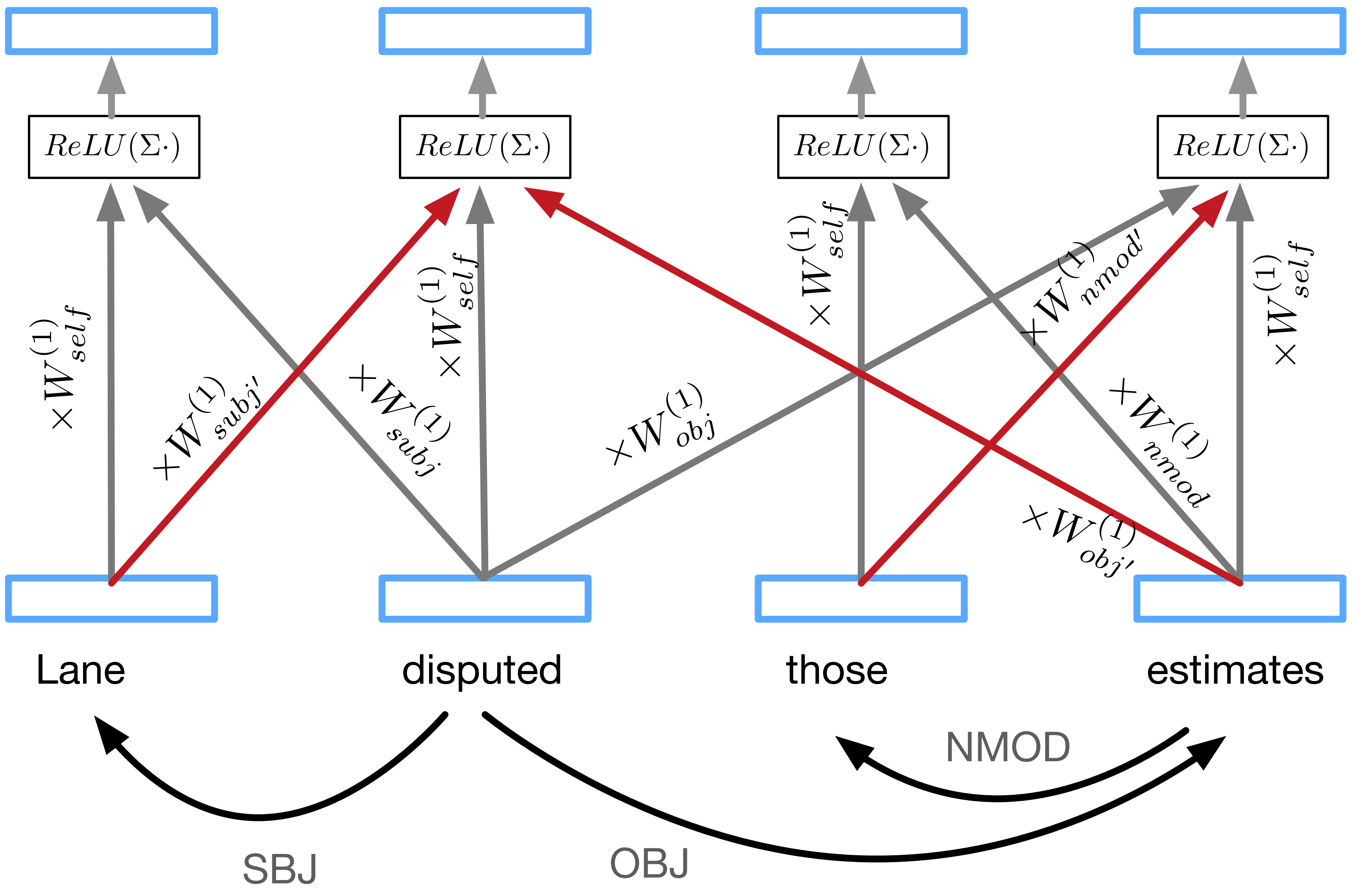
Example

[Marcheggiani and Titov, 2017]



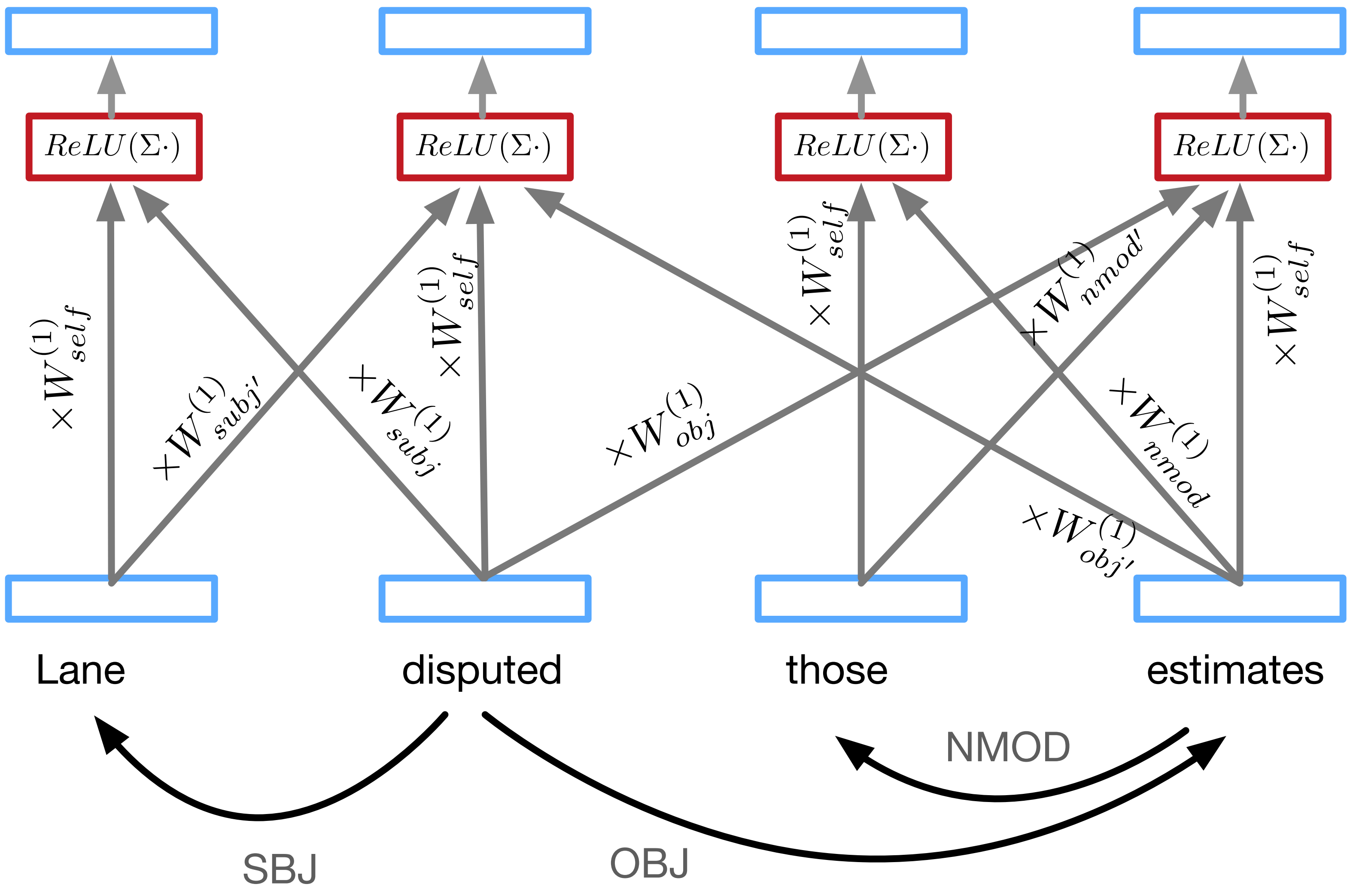
Example

[Marcheggiani and Titov, 2017]



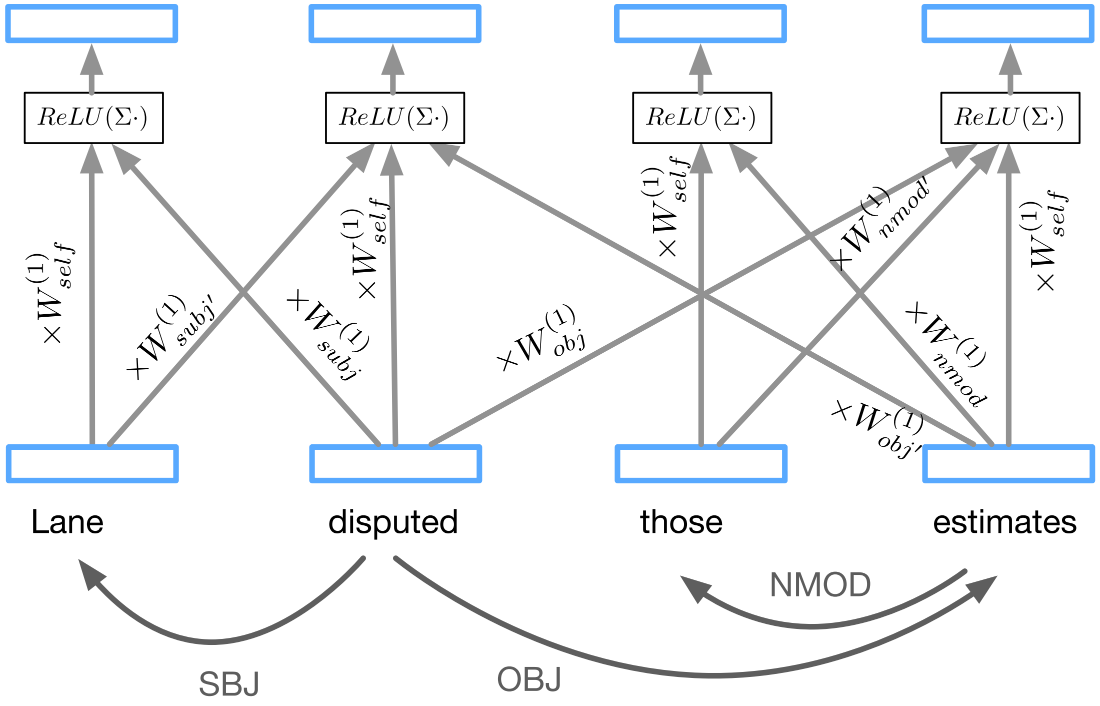
Example

[Marcheggiani and Titov, 2017]



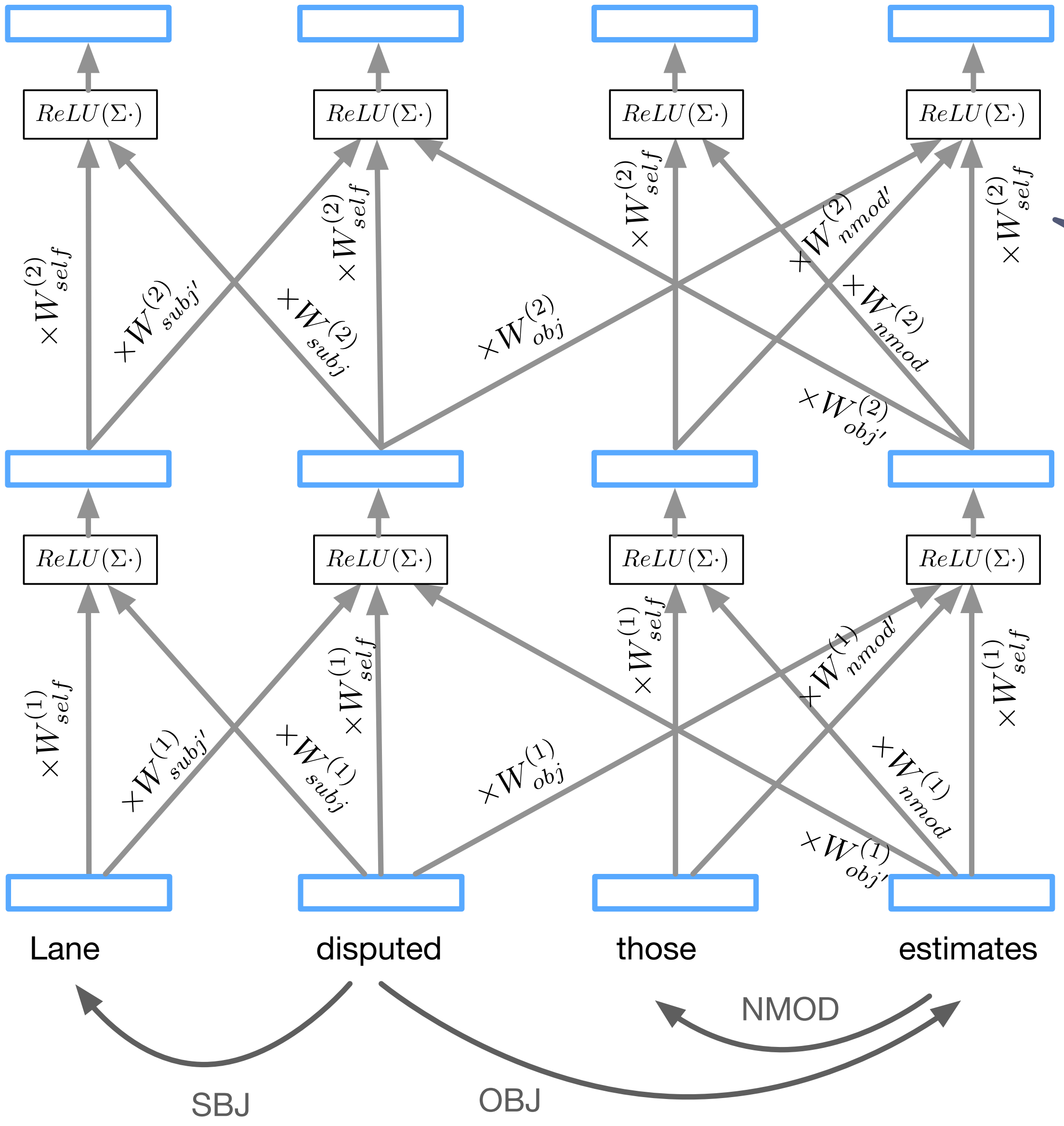
Example

[Marcheggiani and Titov, 2017]



Example

[Marcheggiani and Titov, 2017]



Stacking GCNs widens the syntactic neighborhood

Syntactic GCNs

[Marcheggiani and Titov, 2017]

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

Syntactic GCNs

[Marcheggiani and Titov, 2017]

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

Syntactic neighborhood

Syntactic GCNs

[Marcheggiani and Titov, 2017]

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

Syntactic GCNs

[Marcheggiani and Titov, 2017]

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

[Marcheggiani and Titov, 2017]

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

Message

Syntactic neighborhood

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

[Marcheggiani and Titov, 2017]

- Not all edges are equally important for the final task

Edge-wise Gates

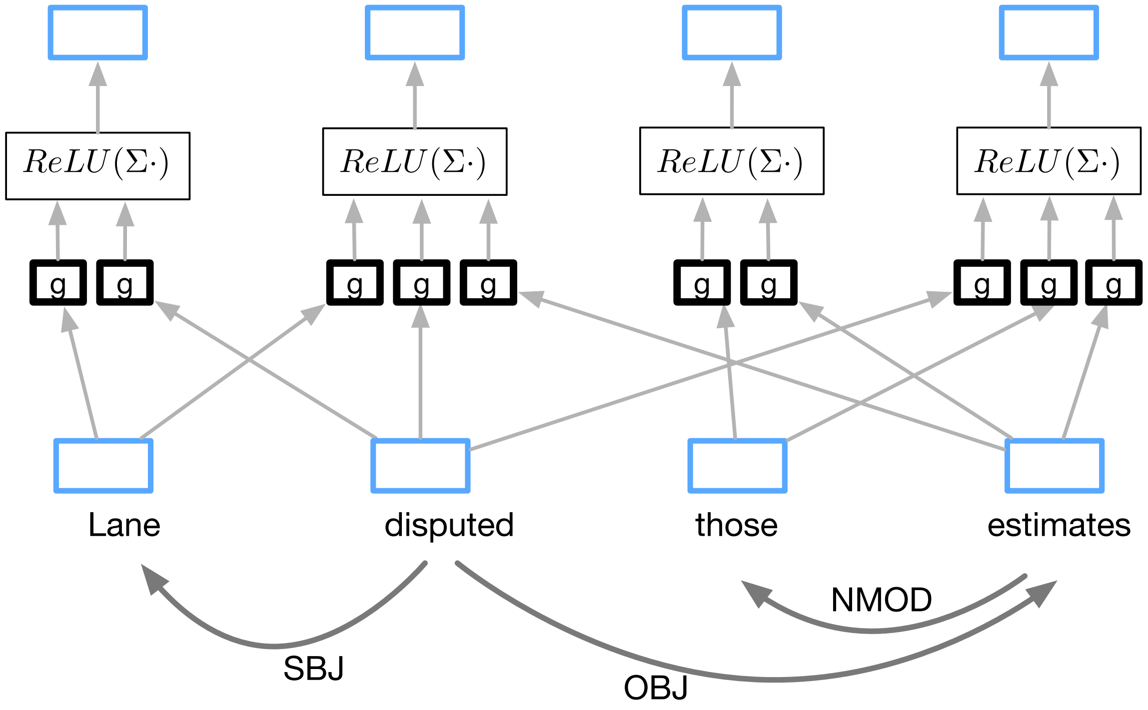
[Marcheggiani and Titov, 2017]

- Not all edges are equally important for the final task
- We should not blindly rely on predicted syntax

Edge-wise Gates

[Marcheggiani and Titov, 2017]

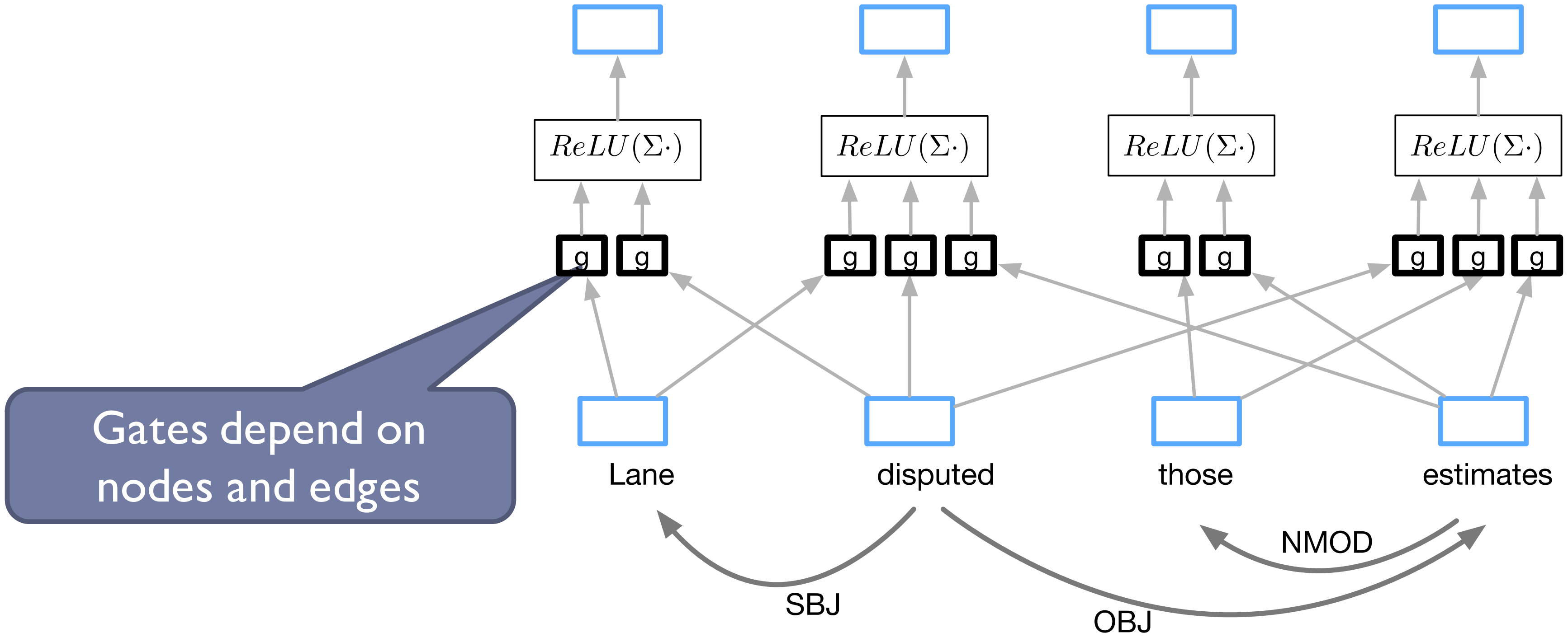
- Not all edges are equally important for the final task
- We should not blindly rely on predicted syntax
- Gates decide the "importance" of each message



Edge-wise Gates

[Marcheggiani and Titov, 2017]

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



Outline

- Semantic Role Labeling
- Graph Convolutional Networks (GCN)
- Syntactic GCN for Semantic Role Labeling (SRL)
- **SRL Model**
- Exploiting Semantics in Neural Machine Translation with GCNs

Our Model

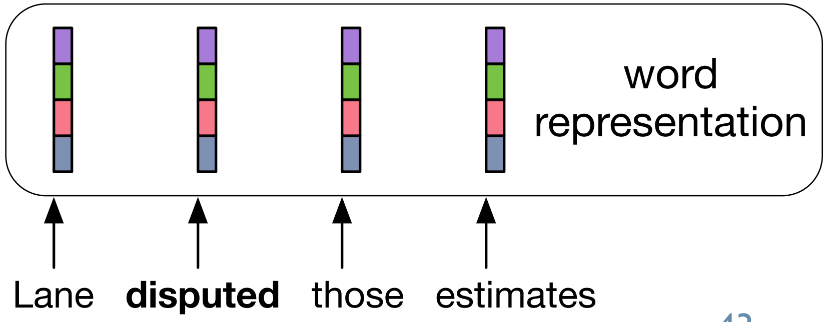
[Marcheggiani and Titov, 2017]

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

Word Representation

[Marcheggiani and Titov, 2017]

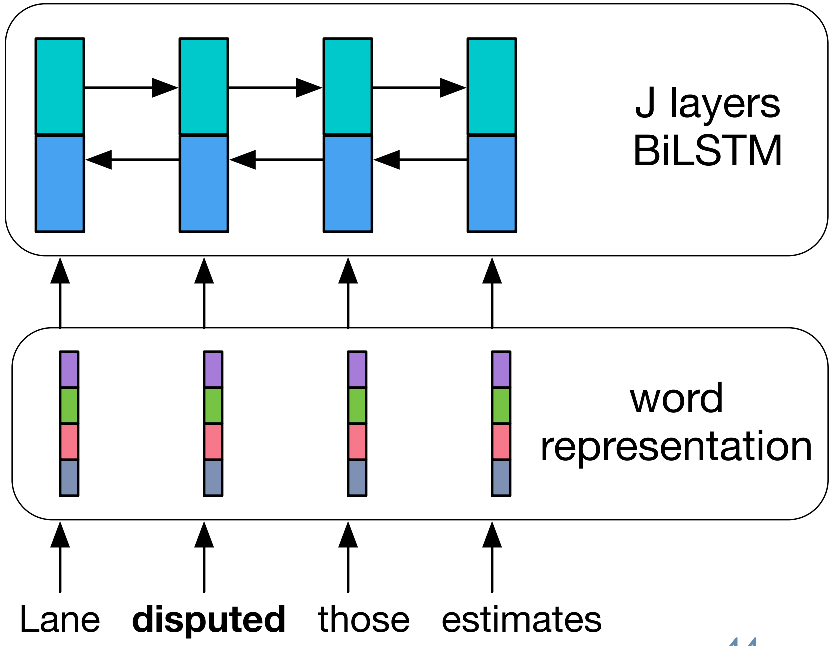
- Pretrained word embeddings
- Word embeddings
- POS tag embeddings
- Predicate lemma embeddings



BiLSTM Encoder

[Marcheggiani and Titov, 2017]

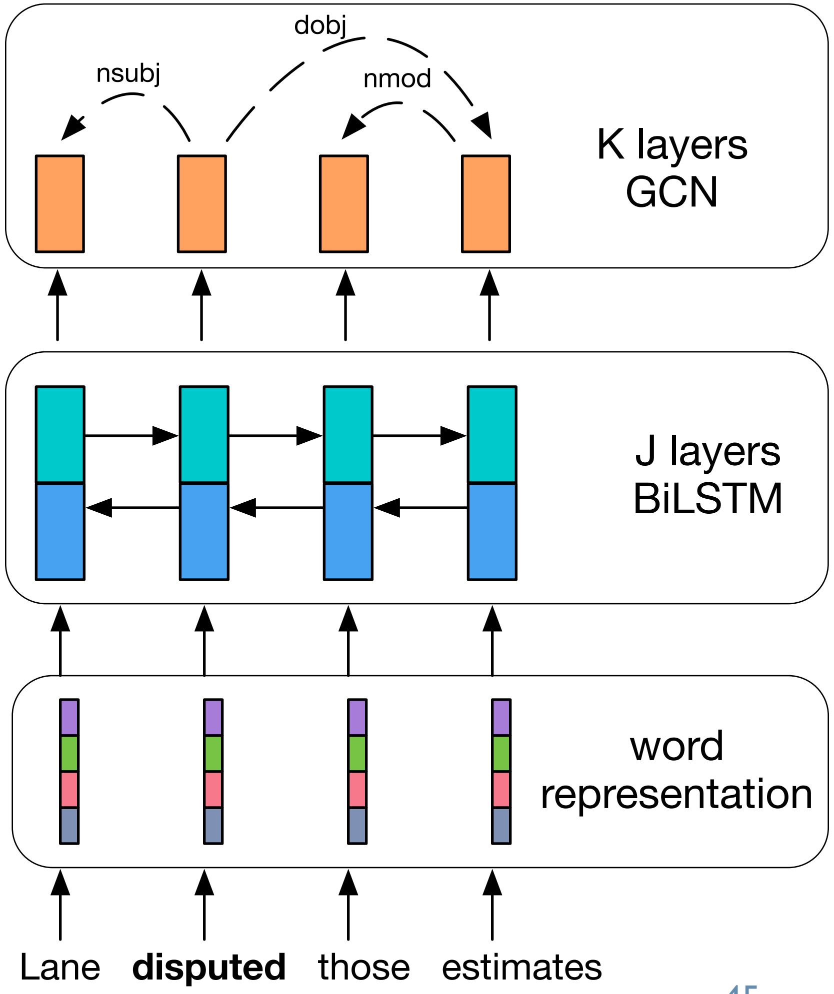
- Encode each word with its left and right context
- Stacked BiLSTM



GCNs Encoder

[Marcheggiani and Titov, 2017]

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



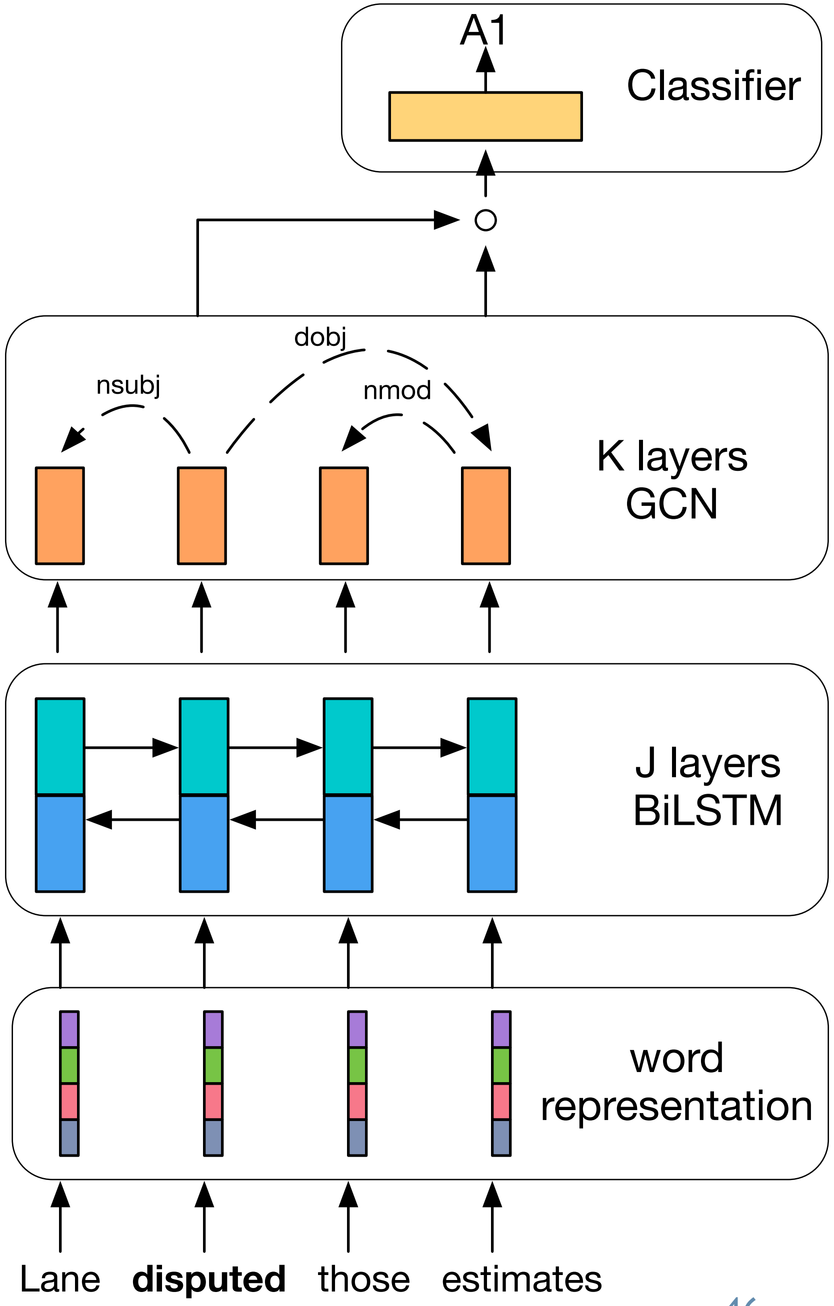
Semantic Role Classifier

- 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



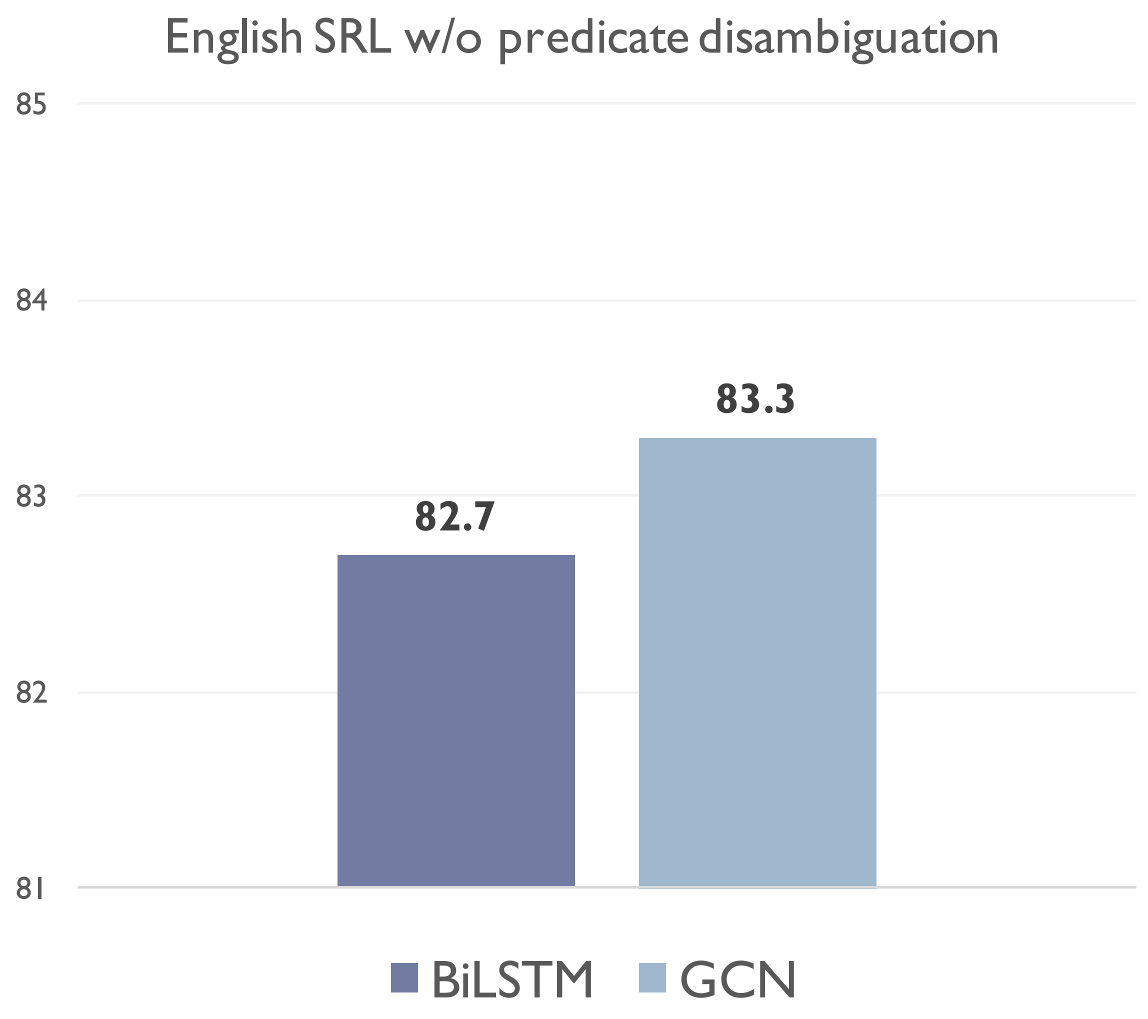
Experiments

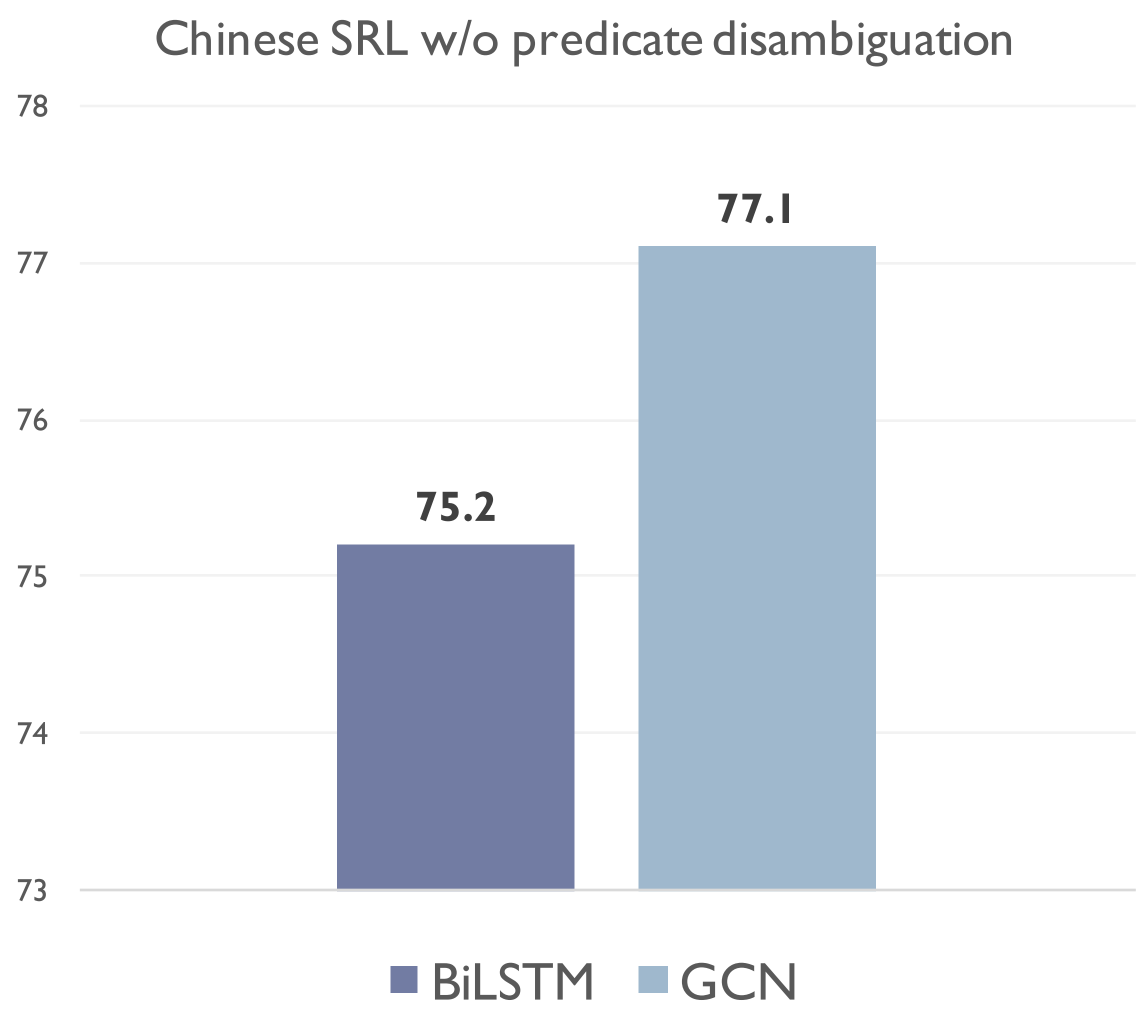
[Marcheggiani and Titov, 2017]

- 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 (Dev set)

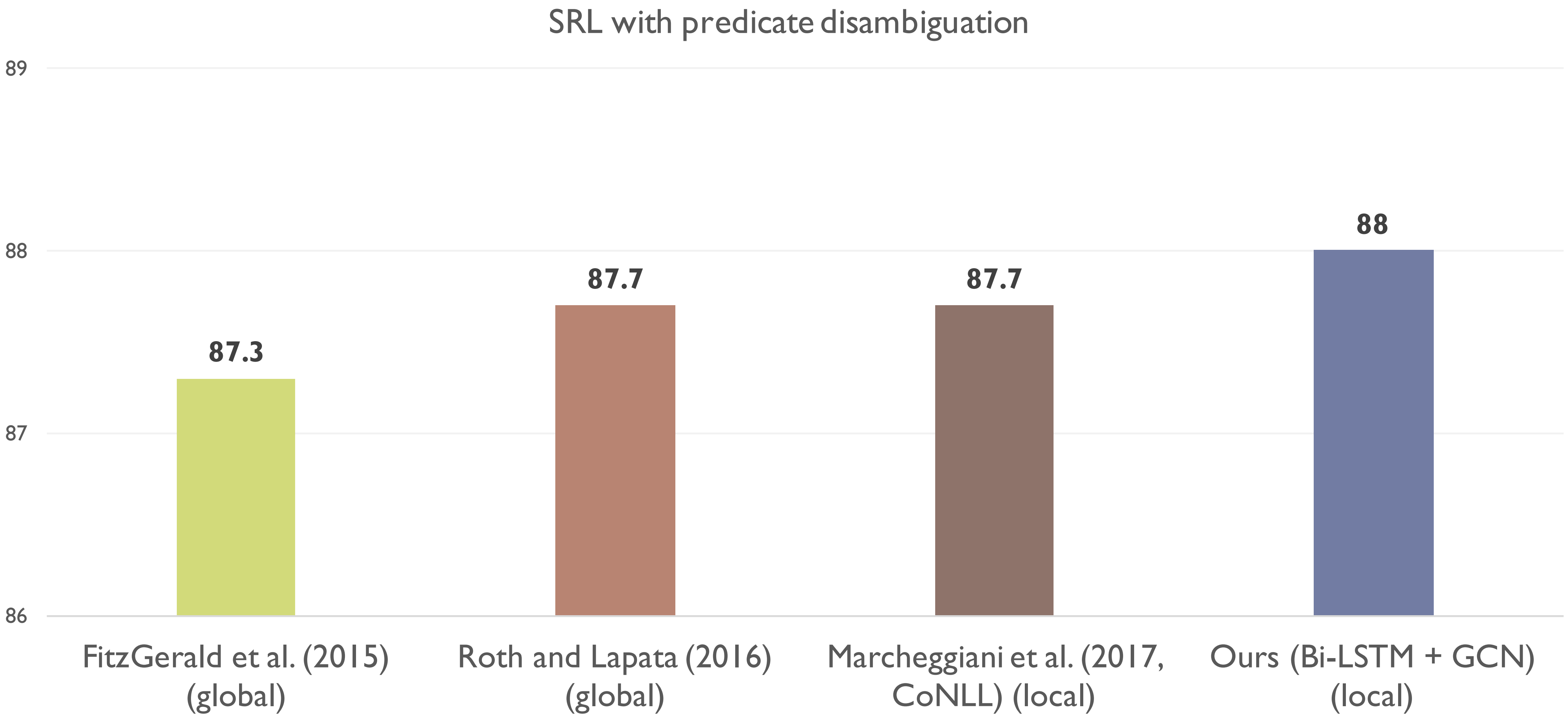
[Marcheggiani and Titov, 2017]





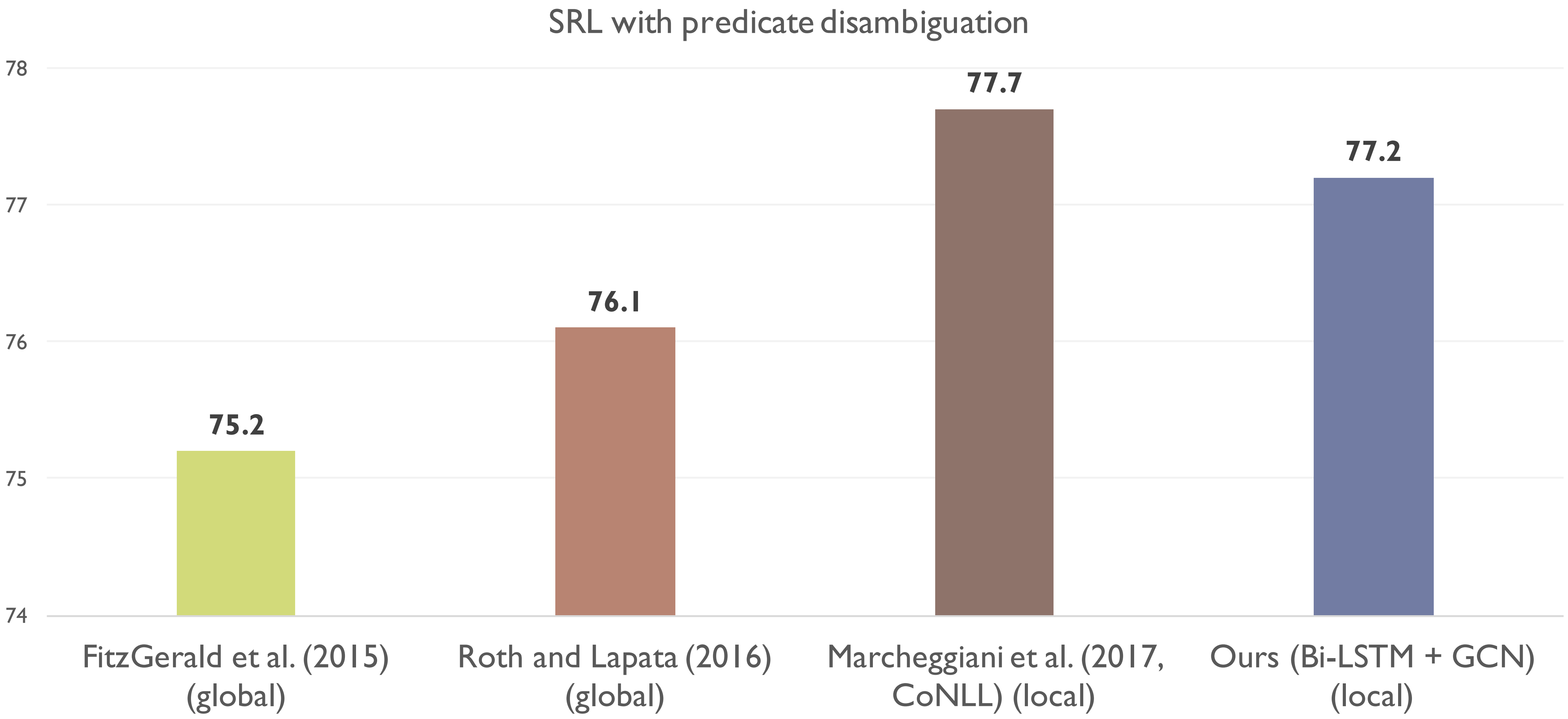
English Test Set

[Marcheggiani and Titov, 2017]



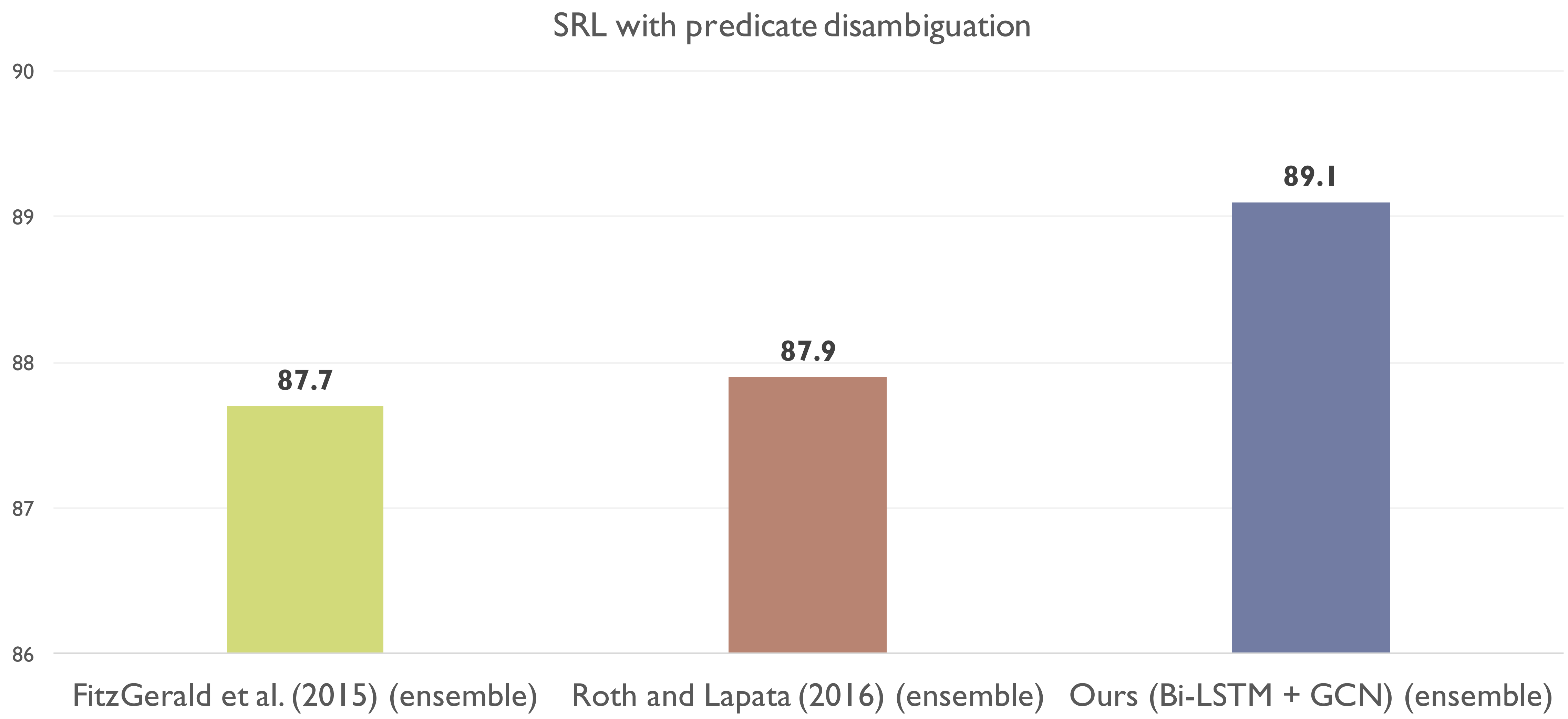
English Out of Domain

[Marcheggiani and Titov, 2017]



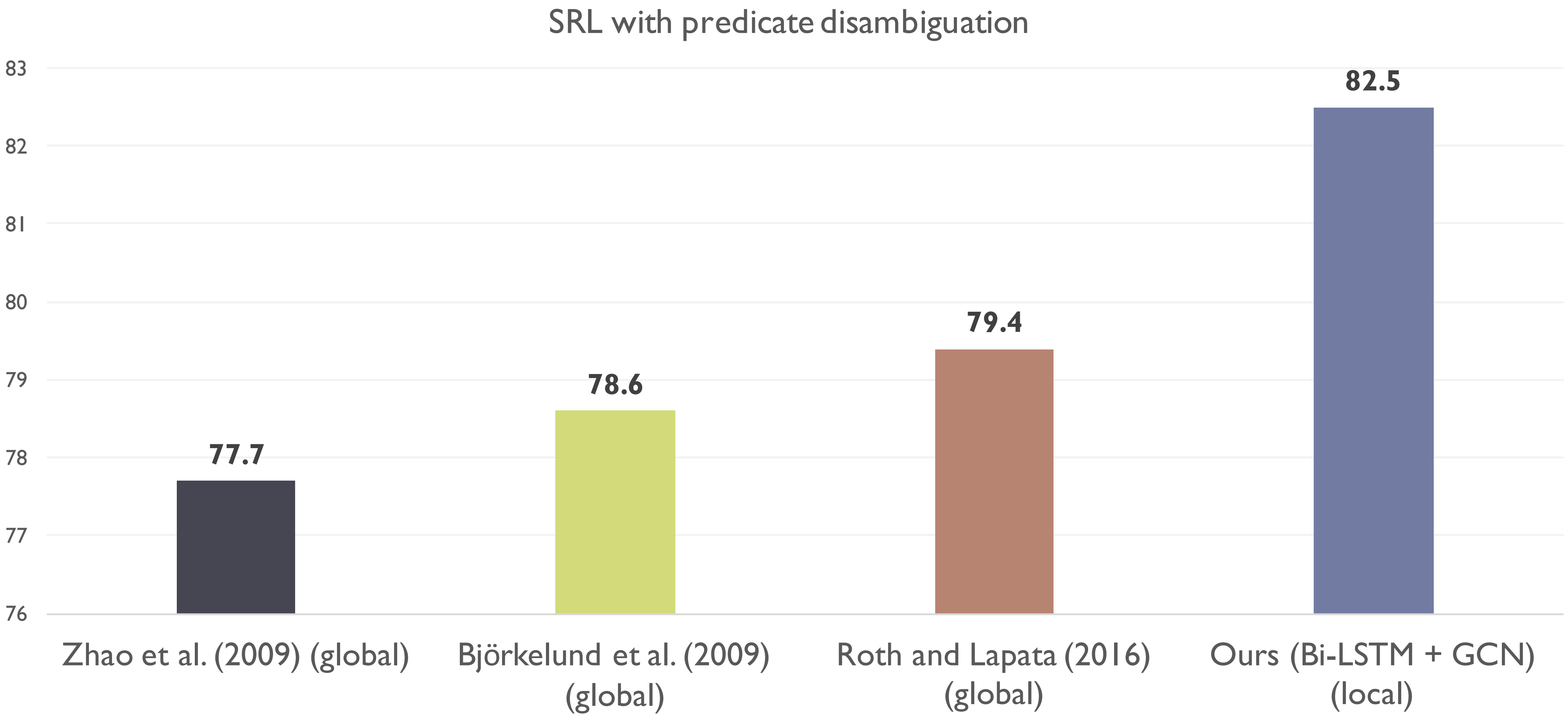
English Test Set (Ensemble)

[Marcheggiani and Titov, 2017]



Chinese Test Set

[Marcheggiani and Titov, 2017]



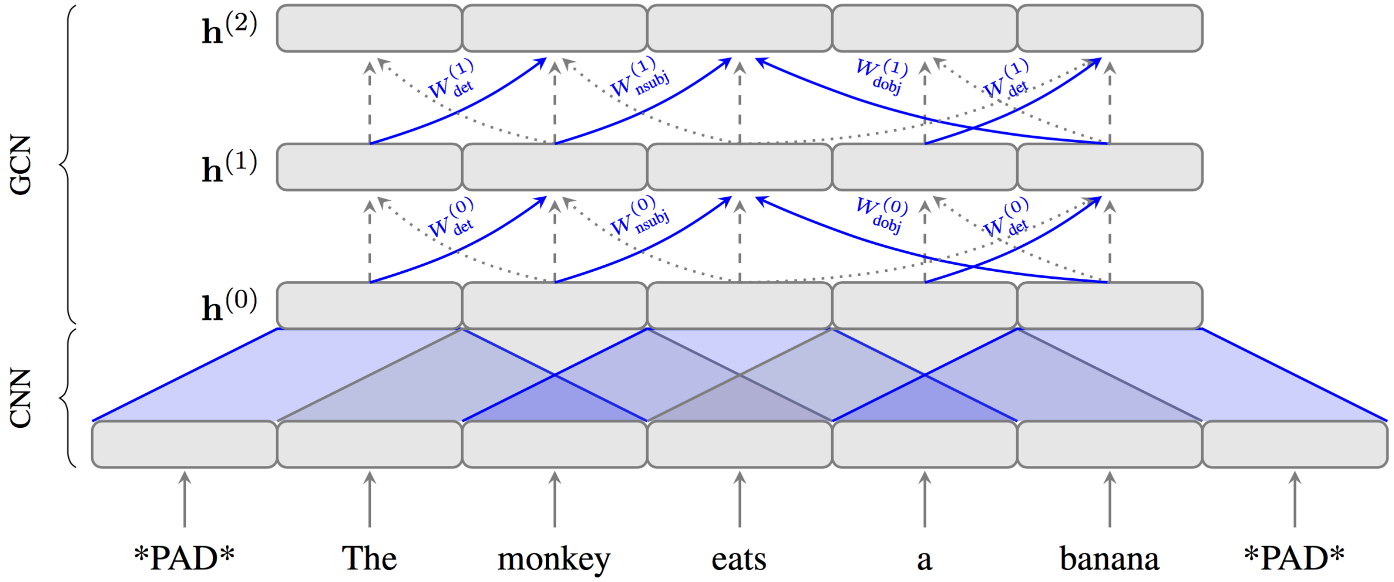
Syntactic Graph Convolutional Networks

- Fast and simple
- Can be seamlessly applied to other tasks

Syntactic Graph Convolutional Networks

- Fast and simple
- Can be seamlessly applied to other tasks

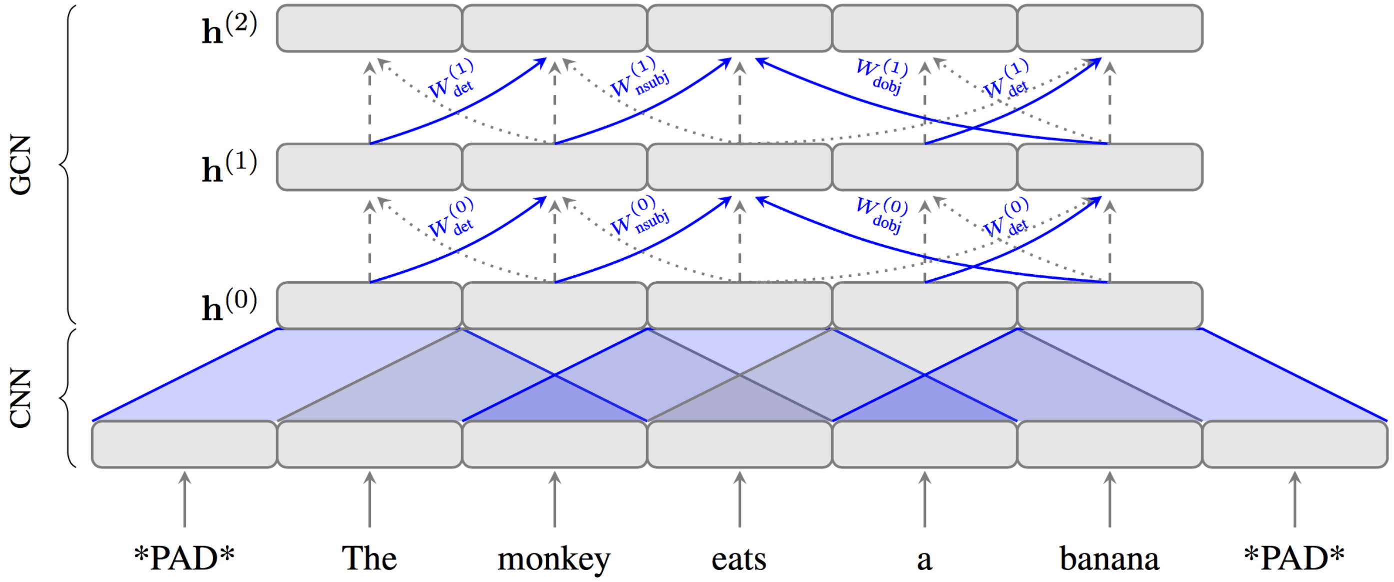
Graph Convolutional Encoders for Syntax-aware Machine Translation
Joost Bastings, Ivan Titov, Wilker Aziz, Diego Marcheggiani, Khalil Sima'an.
In *Proceedings of EMNLP*, 2017.



Syntactic Graph Convolutional Networks

- Fast and simple
- Can be seamlessly applied to other tasks

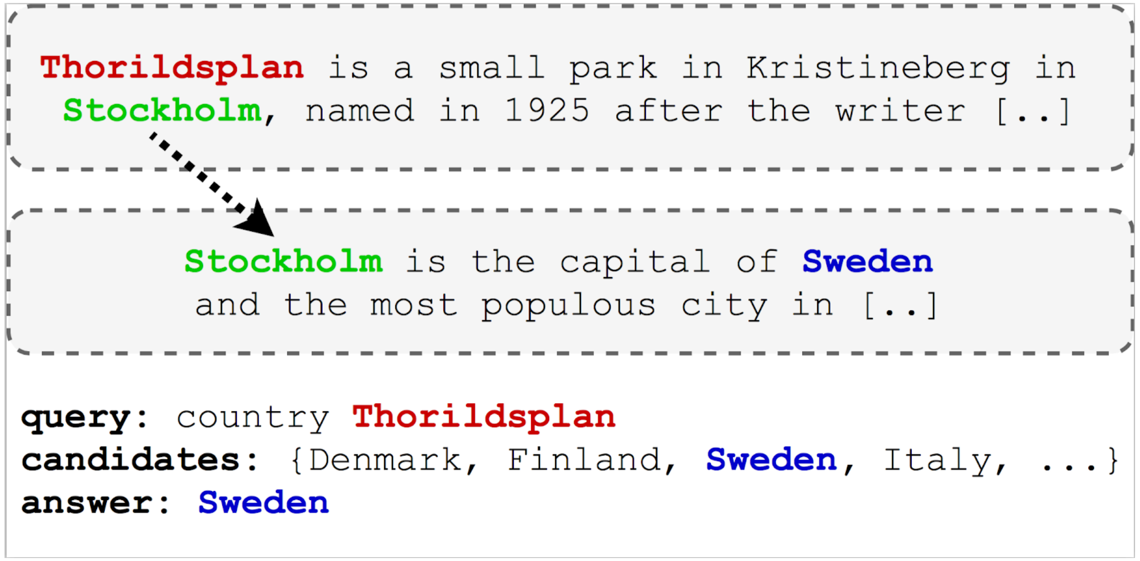
Graph Convolutional Encoders for Syntax-aware Machine Translation
Joost Bastings, Ivan Titov, Wilker Aziz, Diego Marcheggiani, Khalil Sima'an.
In *Proceedings of EMNLP*, 2017.

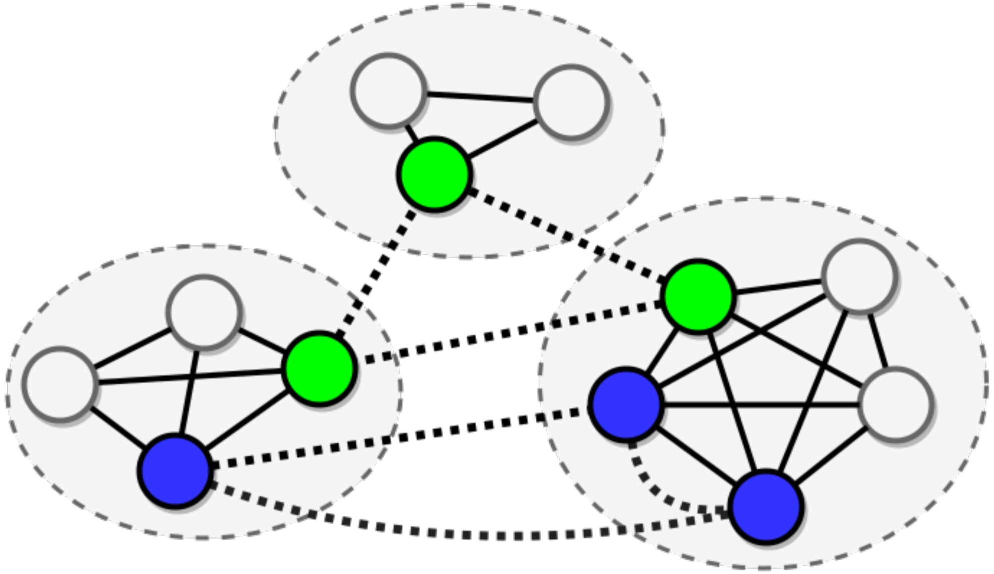


Improvements on
English to German and
English to Czech translations

Multi-document Question Answering

[De Cao et al., 2018]





- Nodes are entities and edges are co-reference links
- Inference on a graph representing the documents collection

Multi-document Question Answering

[De Cao et al., 2018]

WikiHop

#	Model / Reference	Affiliation	Date	Accuracy[%]
1	Entity-GCN	University of Amsterdam && University of Edinburgh	May 2018	67.6
2	MHQA-GRN	IBM && University of Rochester	August 2018	65.4
3	Jenga	Facebook AI Research	February 2018	65.3
4	[anonymized]	[anonymized]	May 2018	64.9
5	Vanilla CoAttention Model	Nanyang Technological University	December 2017	59.9
6	Coref-GRU	Carnegie Mellon University.	April 2018	59.3

Syntactic Graph Convolutional Networks

Graph Convolutional Networks for Named Entity Recognition

Cetoli, Alberto **Bragaglia, Stefano** **O'Harney, Andrew Daniel** **Sloan, Marc**
Context Scout

Syntactic Graph Convolutional Networks

Graph Convolutional Networks for Named Entity Recognition

Cetoli, Alberto **Bragaglia, Stefano** **O'Harney, Andrew Daniel** **Sloan, Marc**
Context Scout

Graph Convolutional Networks with Argument-Aware Pooling for Event Detection

Thien Huu Nguyen
Department of Computer and Information Science
University of Oregon
Eugene, Oregon 97403, USA

Ralph Grishman
Computer Science Department
New York University
New York, NY 10003 USA

Syntactic Graph Convolutional Networks

Graph Convolutional Networks for Named Entity Recognition

Cetoli, Alberto   Bragaglia, Stefano   O'Harney, Andrew Daniel   Sloan, Marc

Graph Convolution over Pruned Dependency Trees Improves Relation Extraction

Pooling

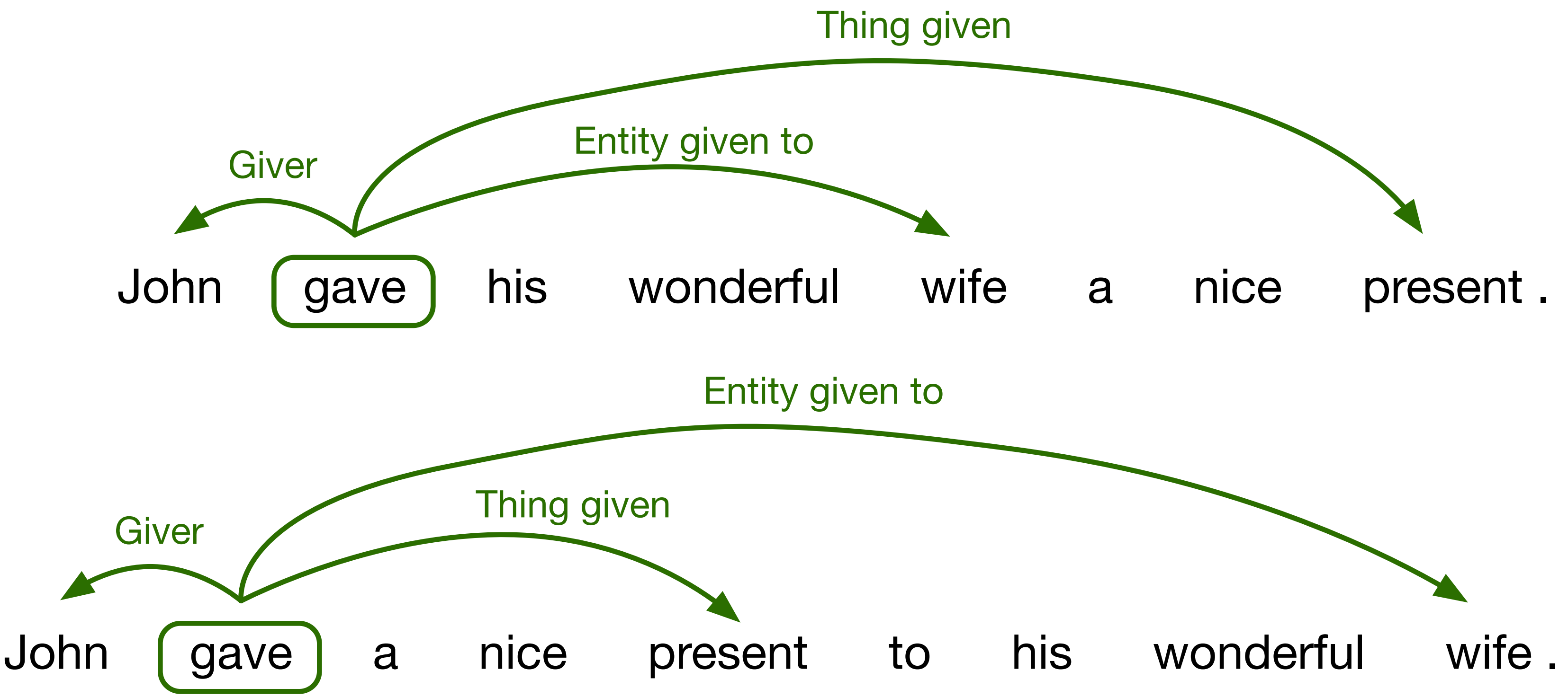
Yuhao Zhang,* Peng Qi,* Christopher D. Manning
Stanford University
Stanford, CA 94305
ın
ırtment
ty
Eugene, Oregon 97403, USA
New York, NY 10003 USA

Outline

- Semantic Role Labeling
- Graph Convolutional Networks (GCN)
- Syntactic GCN for Semantic Role Labeling (SRL)
- SRL Model
- **Exploiting Semantics in Neural Machine Translation with GCNs**

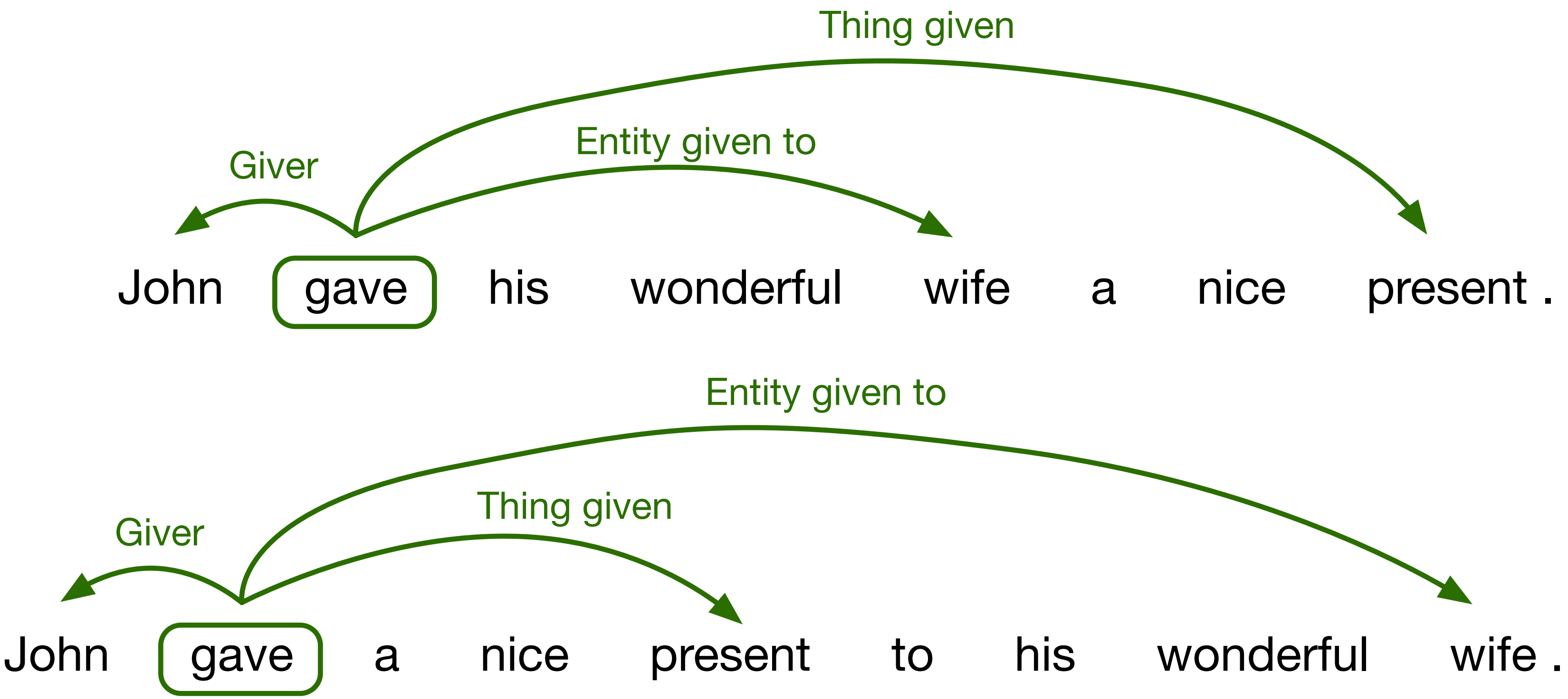
Motivations

[Marcheggiani at al., 2018]



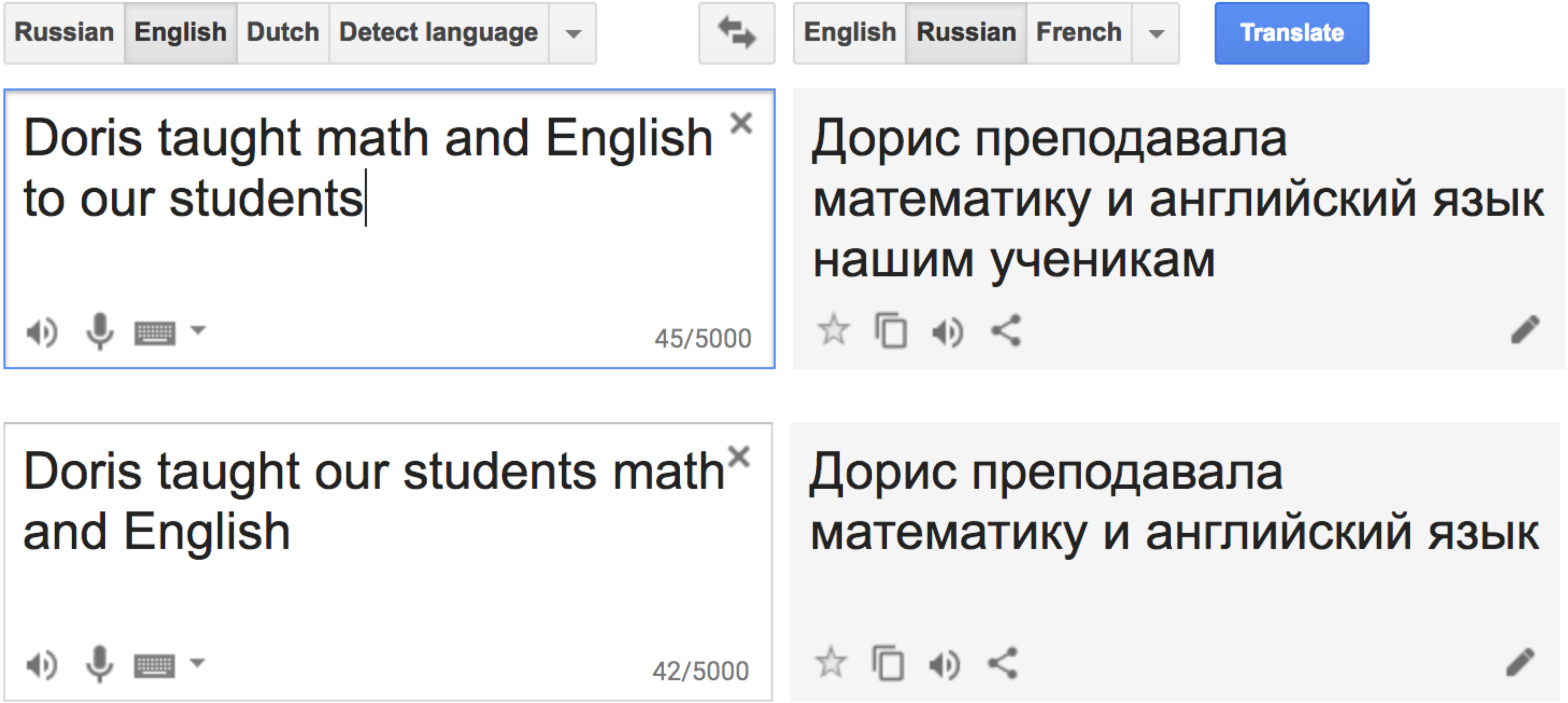
Motivations

[Marcheggiani at al., 2018]

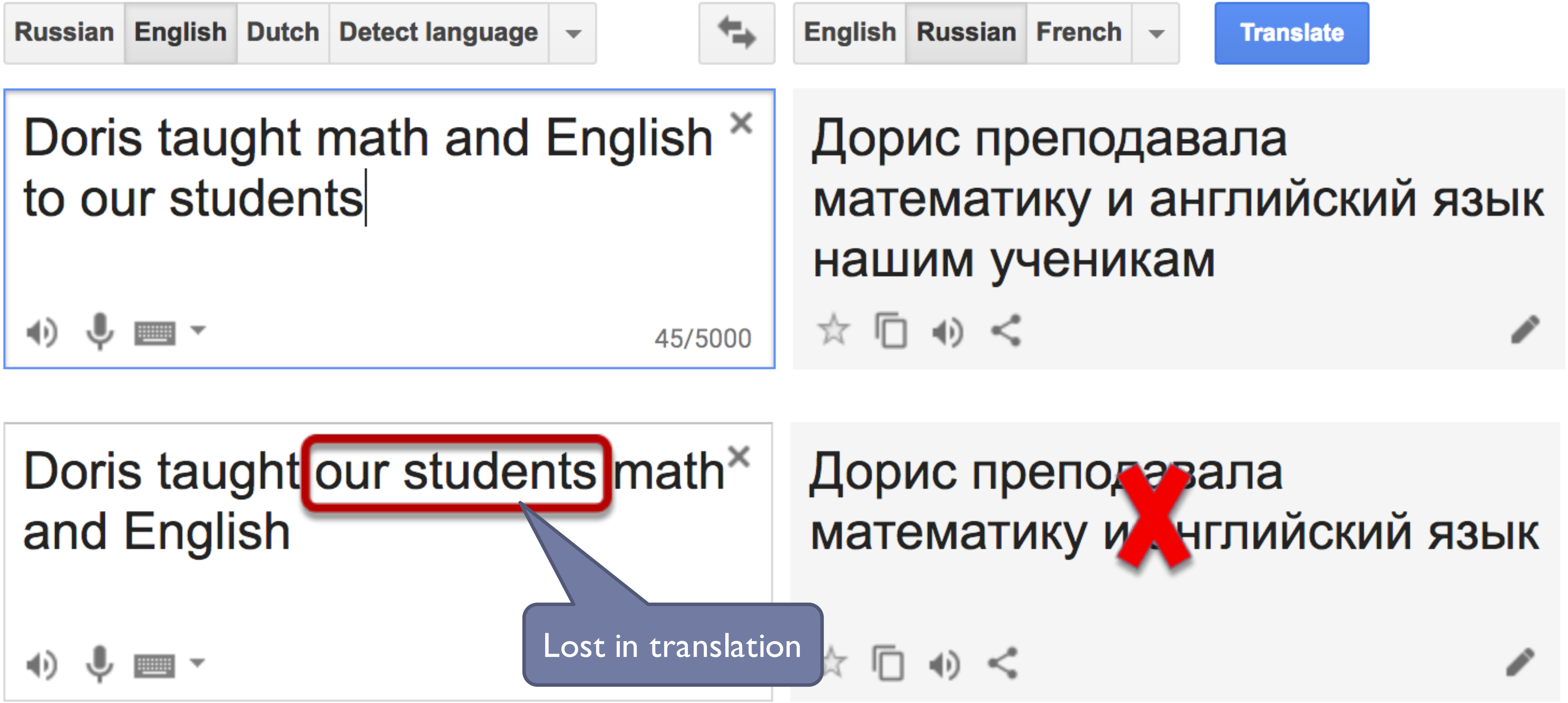


SRL helps to generalize over different surface realizations of the same underlying "meaning".

Motivations



Motivations

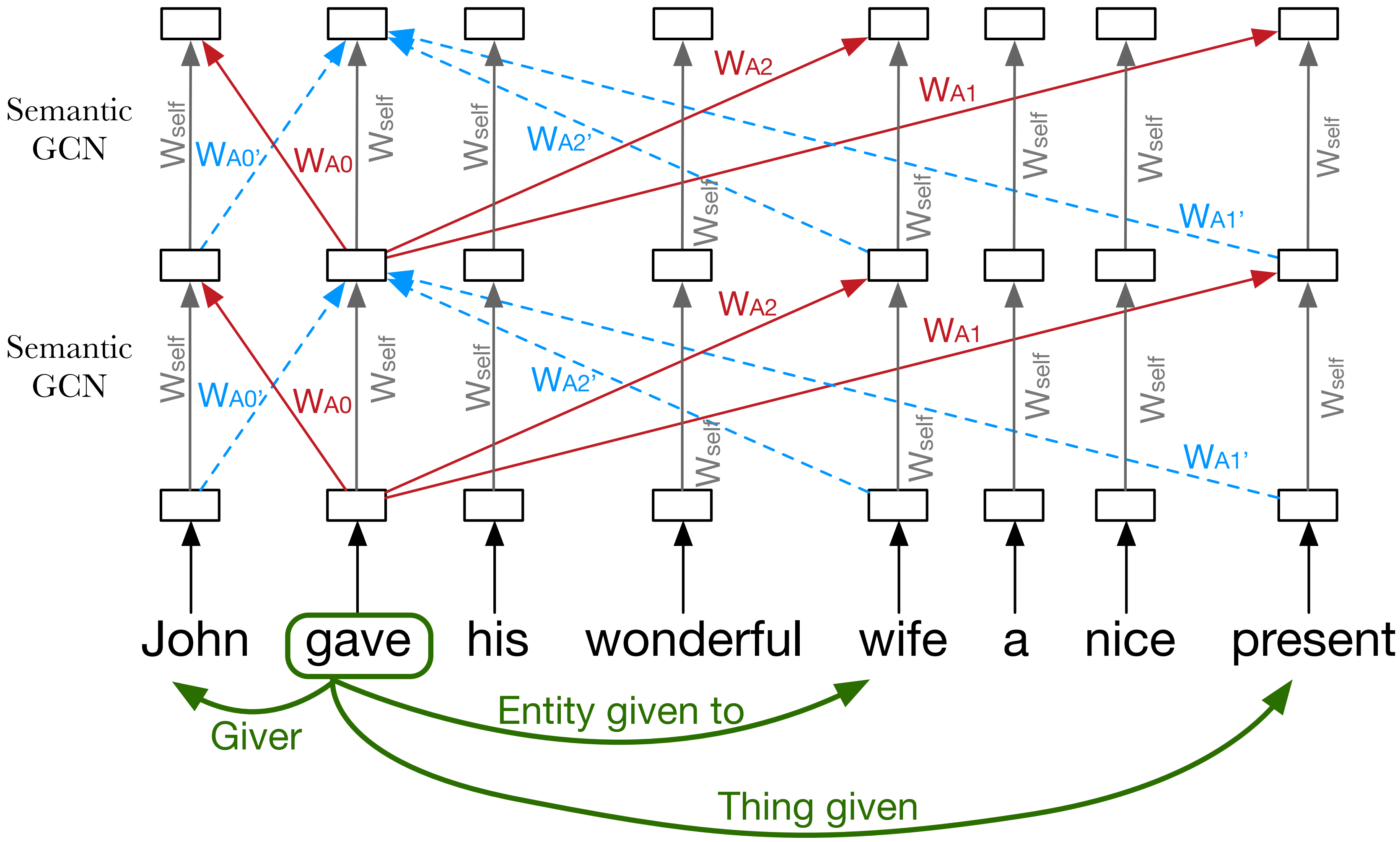


Related work

[Marcheggiani at al., 2018]

- Semantics in statistical MT
  - [Wu and Fung, 2009]
  - [Liu and Gildea, 2010]
  - [Aziz et al., 2011]
  - ...
- Syntax in neural MT
  - [Sennrich and Haddow, 2016]
  - [Aharoni and Goldberg, 2017]
  - **[Bastings et al., 2017]**
  - …
- Semantics in neural MT
  - ???

Predicate-argument encoding



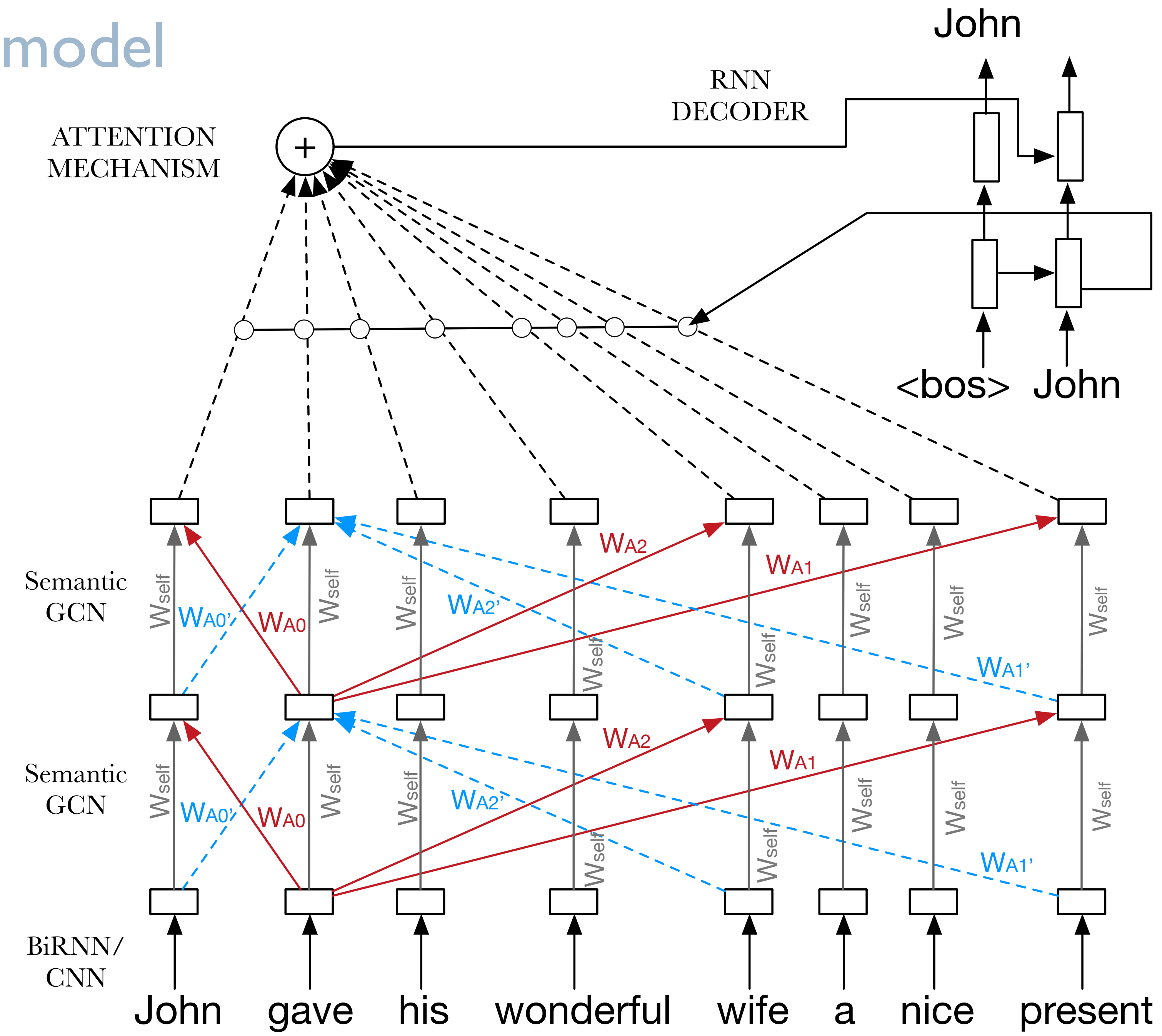
Our Model

[Marcheggiani at al., 2018]

- Standard sequence2sequence with attention
- Semantic GCN encoder on top of a bidirectional RNN
- RNN decoder

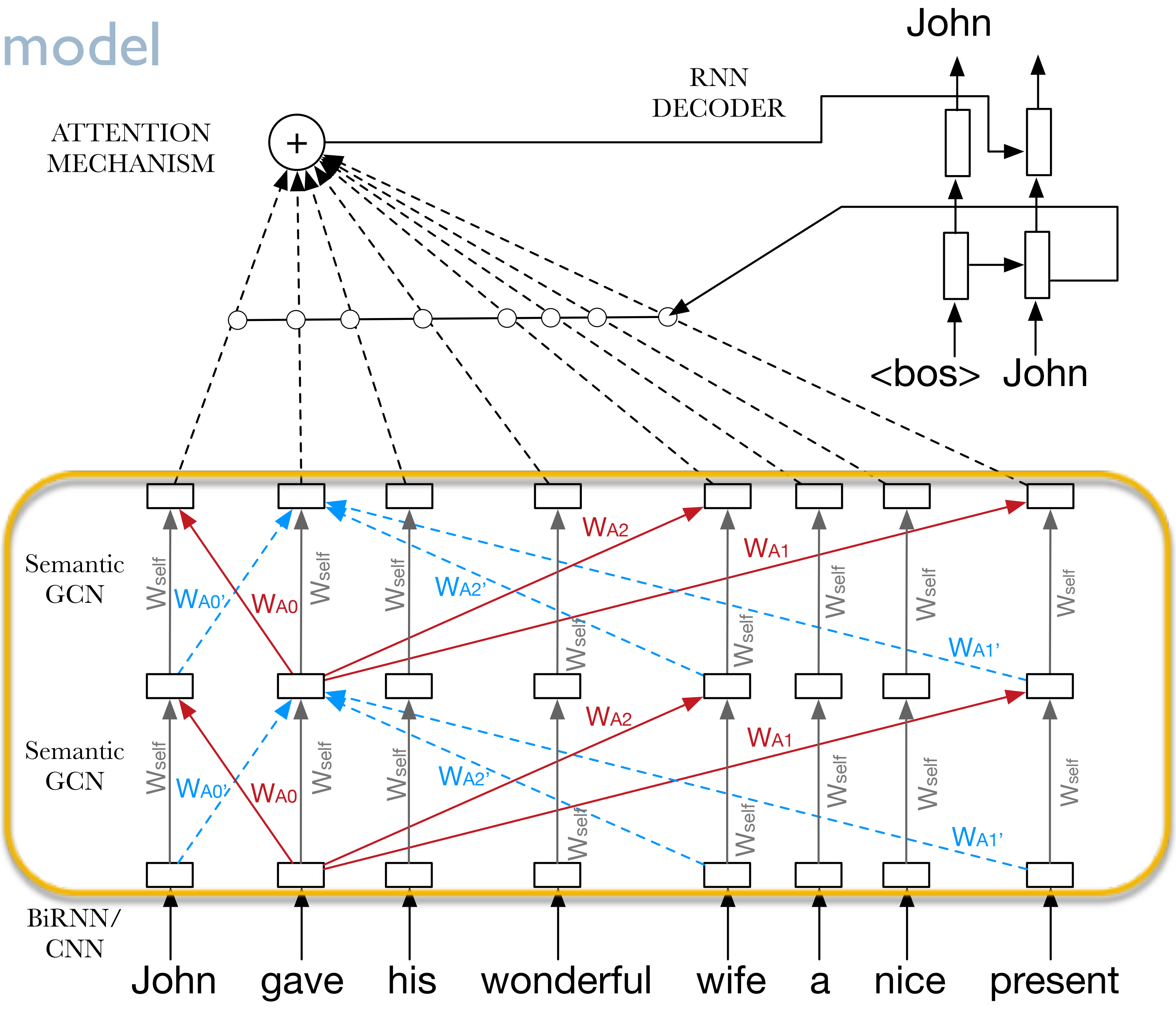
Our model

[Marcheggiani at al., 2018]



Our model

[Marcheggiani at al., 2018]



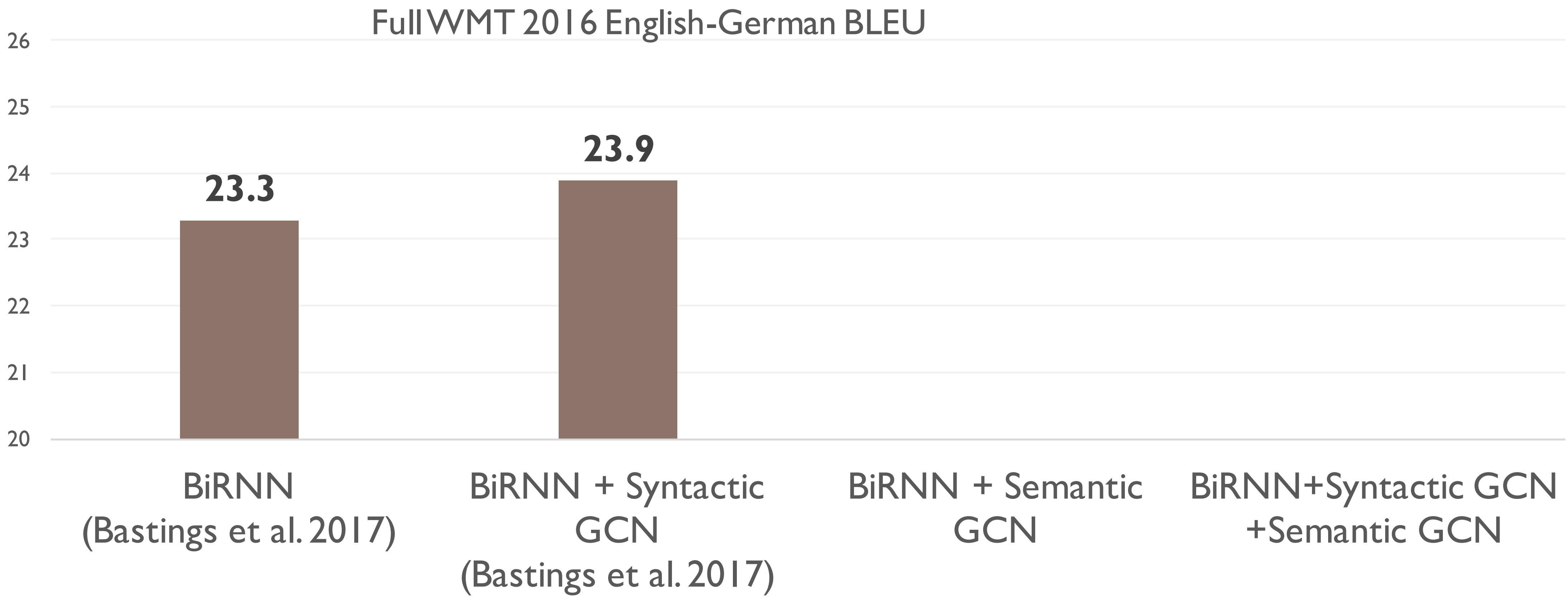
Experiments

[Marcheggiani at al., 2018]

- Data
  - WMT '16 English-German dataset (~4.5 million sentence pairs)
  - BLEU as evaluation measure
- Model
  - Hyperparameters tuned on News Commentary En-De (~226K sentence pairs)
  - GRU as RNN

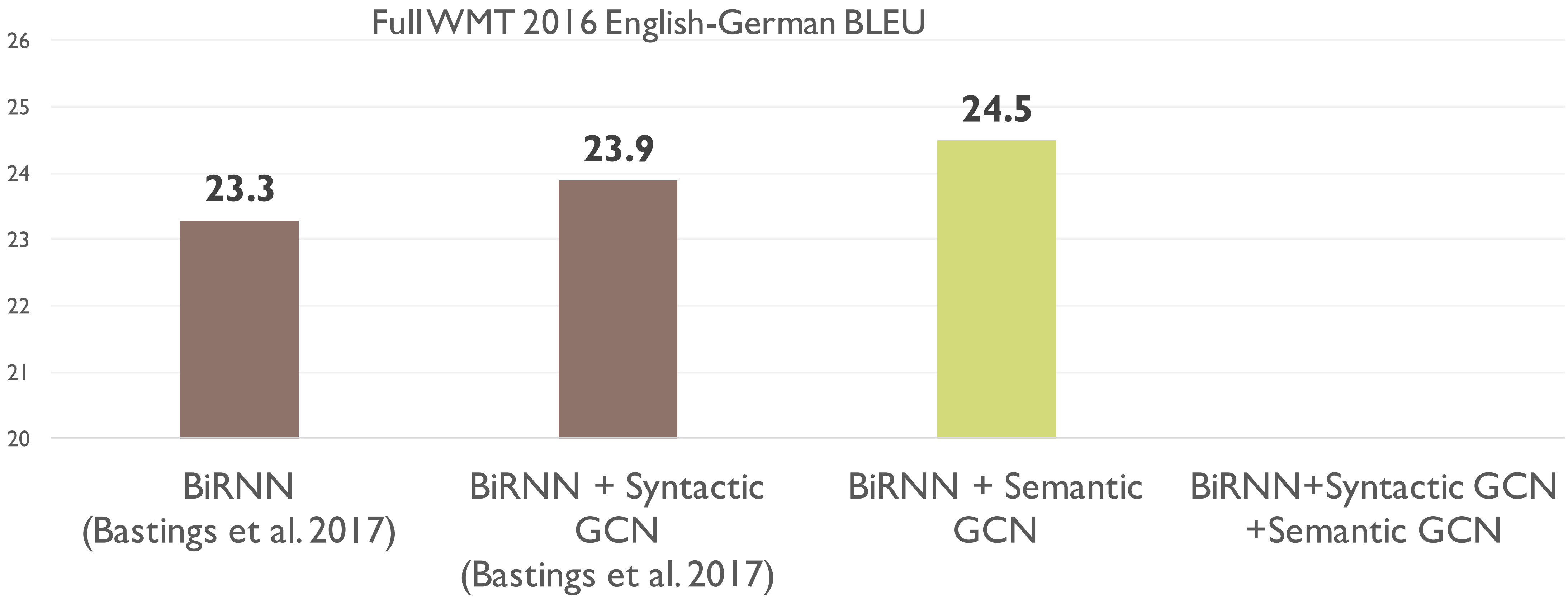
Results

[Marcheggiani at al., 2018]



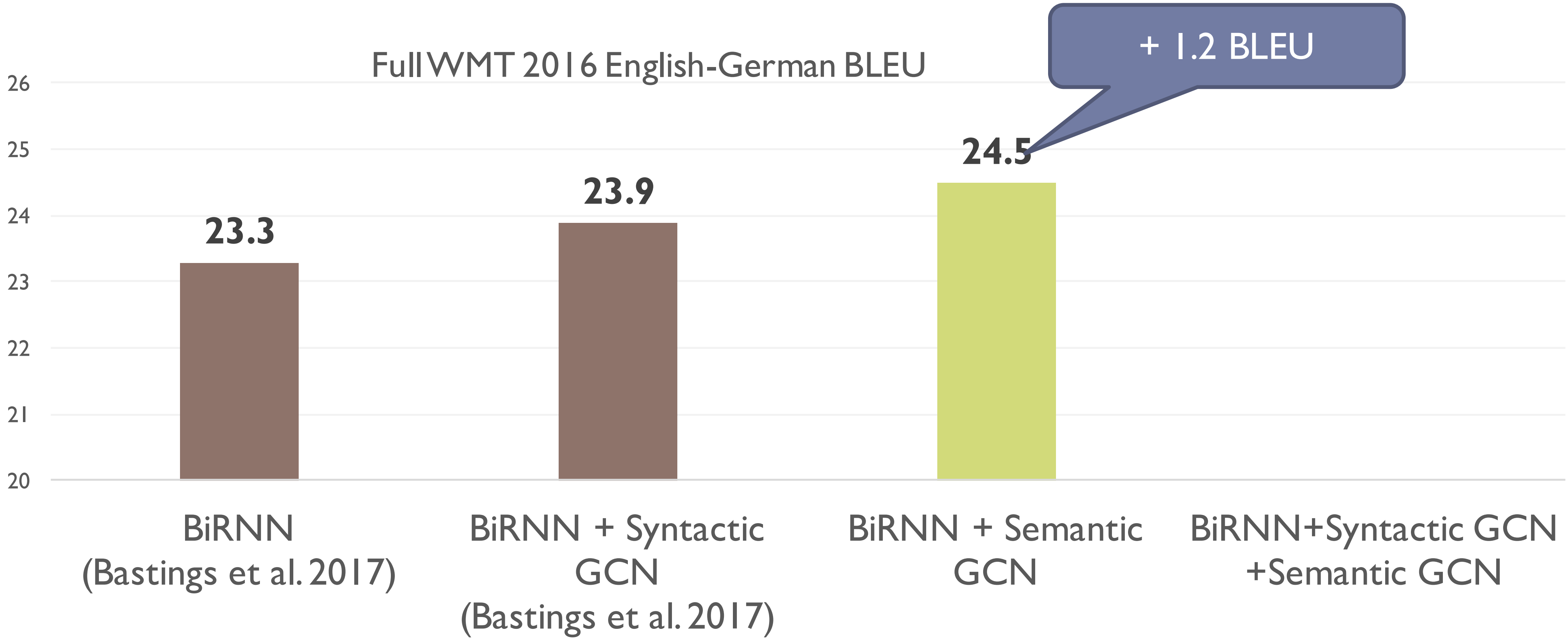
Results

[Marcheggiani at al., 2018]



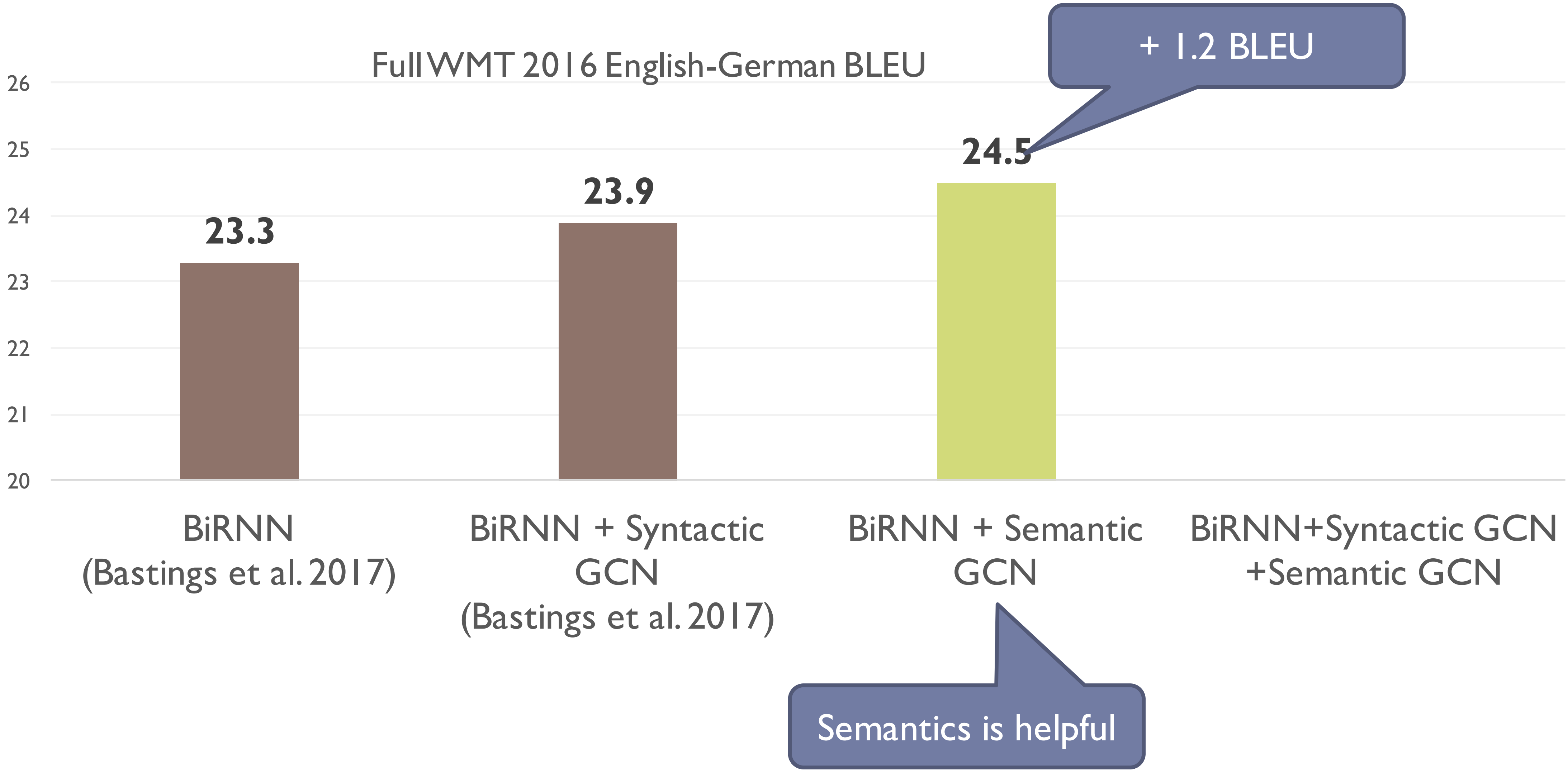
Results

[Marcheggiani at al., 2018]



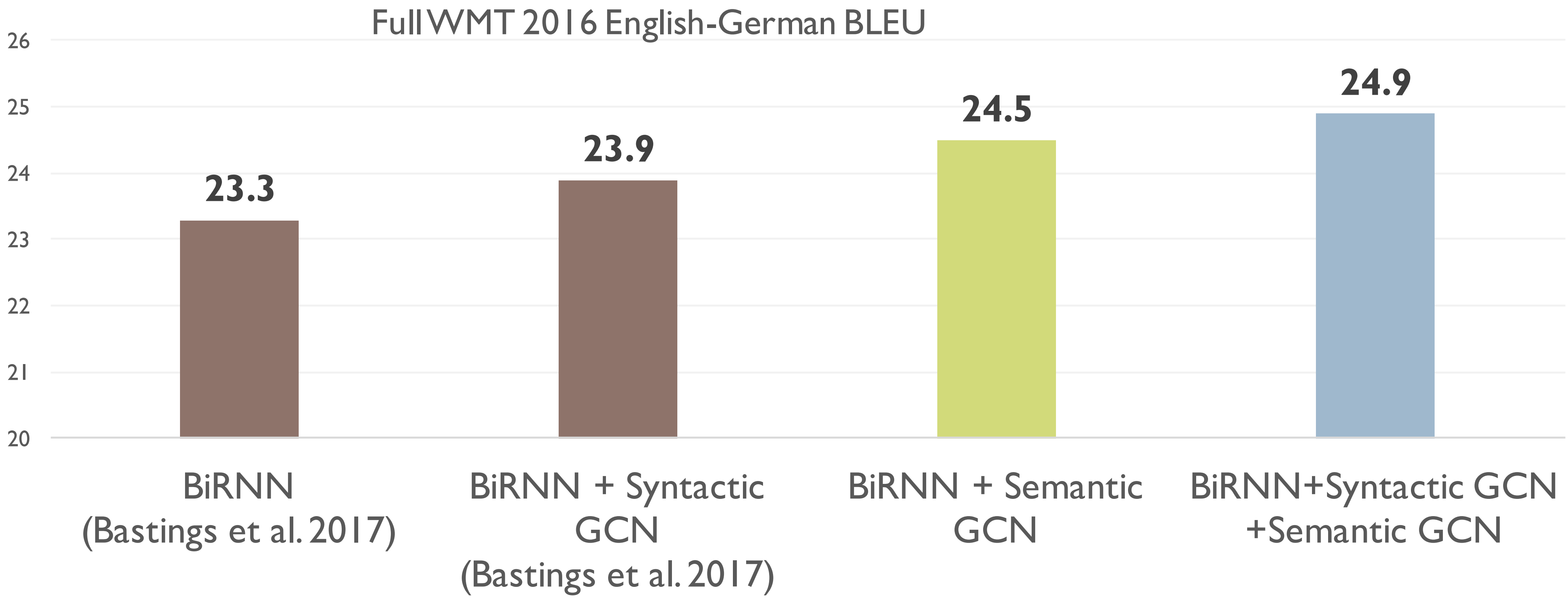
Results

[Marcheggiani at al., 2018]



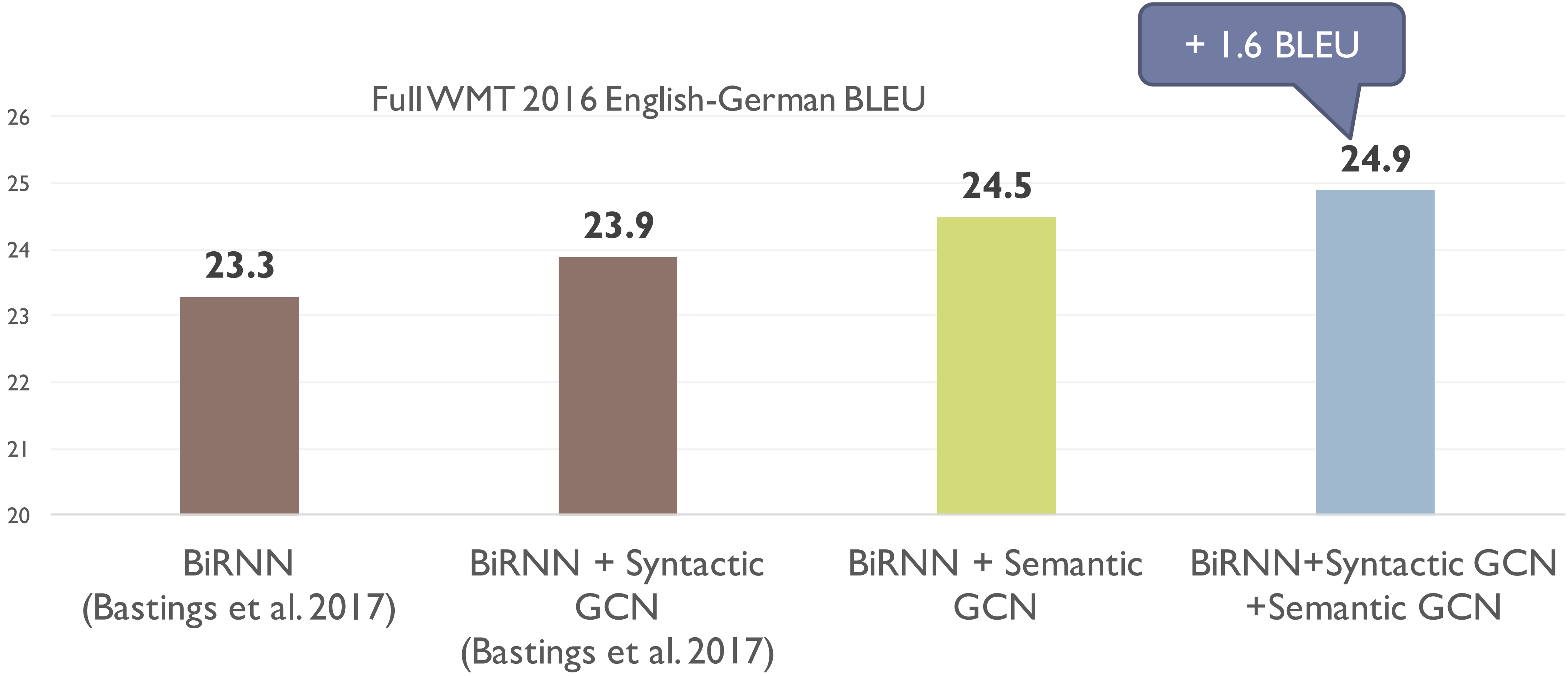
Results

[Marcheggiani at al., 2018]



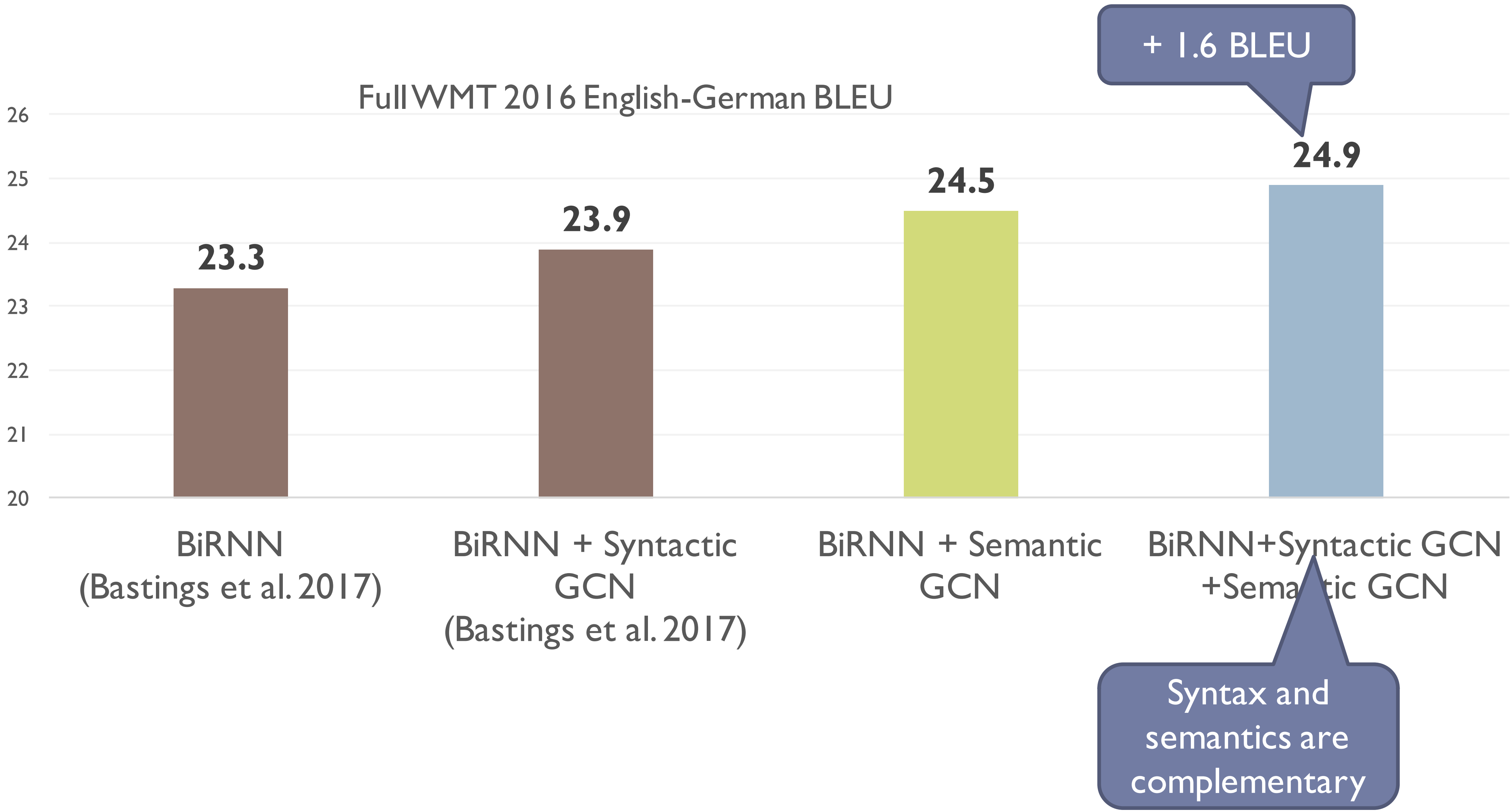
Results

[Marcheggiani at al., 2018]



Results

[Marcheggiani at al., 2018]



Analysis

[Marcheggiani at al., 2018]

Seller Thing sold Buyer

SOURCE	John sold the car to Mark .
BiRNN	John verkaufte das Auto nach Mark .
SEM GCN	John verkaufte das Auto an Mark .

BiRNN mistranslates “to” as “nach” (directionality)

Analysis

[Marcheggiani at al., 2018]

Seller Thing sold Buyer

SOURCE	John $\boxed{\text{sold}}$ the car to Mark .
BiRNN	John verkaufte das Auto nach Mark .
SEM GCN	John verkaufte das Auto an Mark .

BiRNN mistranslates “to” as “nach” (directionality)

Analysis

[Marcheggiani at al., 2018]

Seller Thing sold Buyer

SOURCE	John $\boxed{\text{sold}}$ the car to Mark .
BiRNN	John verkaufte das Auto nach Mark .
SEM GCN	John verkaufte das Auto <u>an</u> Mark .

BiRNN mistranslates “to” as “nach” (directionality)

Analysis

[Marcheggiani at al., 2018]

Thing sitting Location AM-LOC Player Game

SOURCE The boy sitting on a bench in the park plays chess .

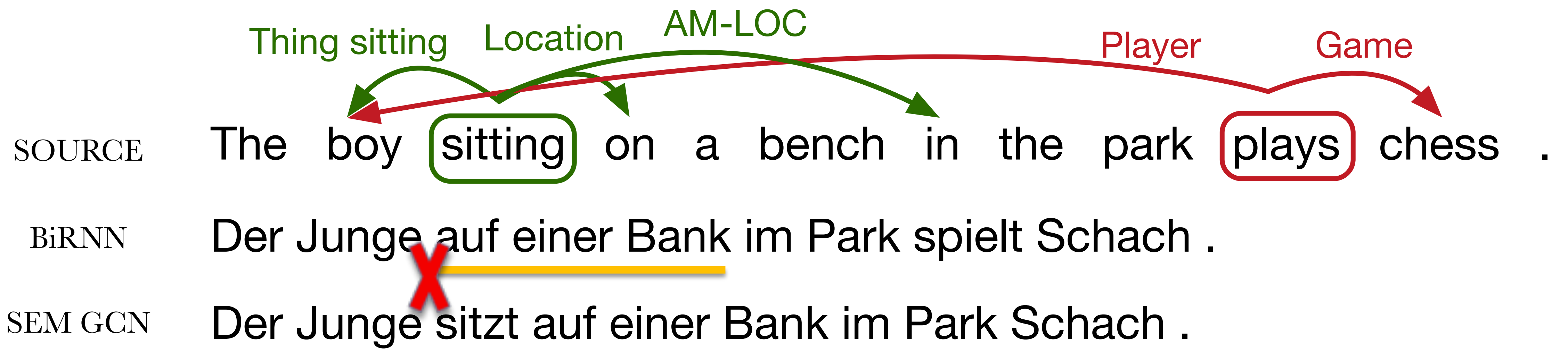
BiRNN Der Junge auf einer Bank im Park spielt Schach .

SEM GCN Der Junge sitzt auf einer Bank im Park Schach .

Both translations are wrong,
but the BiRNN's one is grammatically correct

Analysis

[Marcheggiani at al., 2018]



Both translations are wrong,
but the BiRNN's one is grammatically correct

Analysis

[Marcheggiani at al., 2018]

Thing sitting Location AM-LOC Player Game

SOURCE The boy sitting on a bench in the park plays chess .

BiRNN Der Junge auf einer Bank im Park spielt Schach .

SEM GCN Der Junge sitzt auf einer Bank im Park Schach .

Both translations are wrong,
but the BiRNN's one is grammatically correct

Conclusion

- GCNs for encoding linguistic structures into NN
  - Semantics, coreference, discourse
  - Fast
  - Cheap
- State-of-the-art model for dependency-based SRL
- First to exploit semantics in NMT

Roadmap

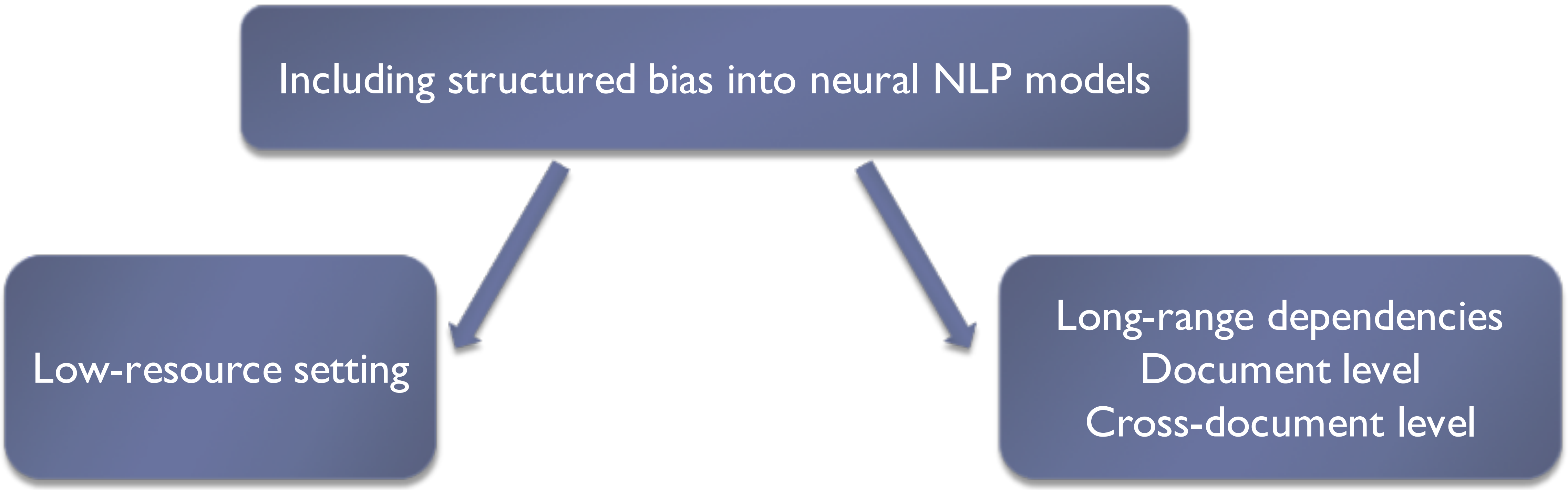
Including structured bias into neural NLP models

Roadmap

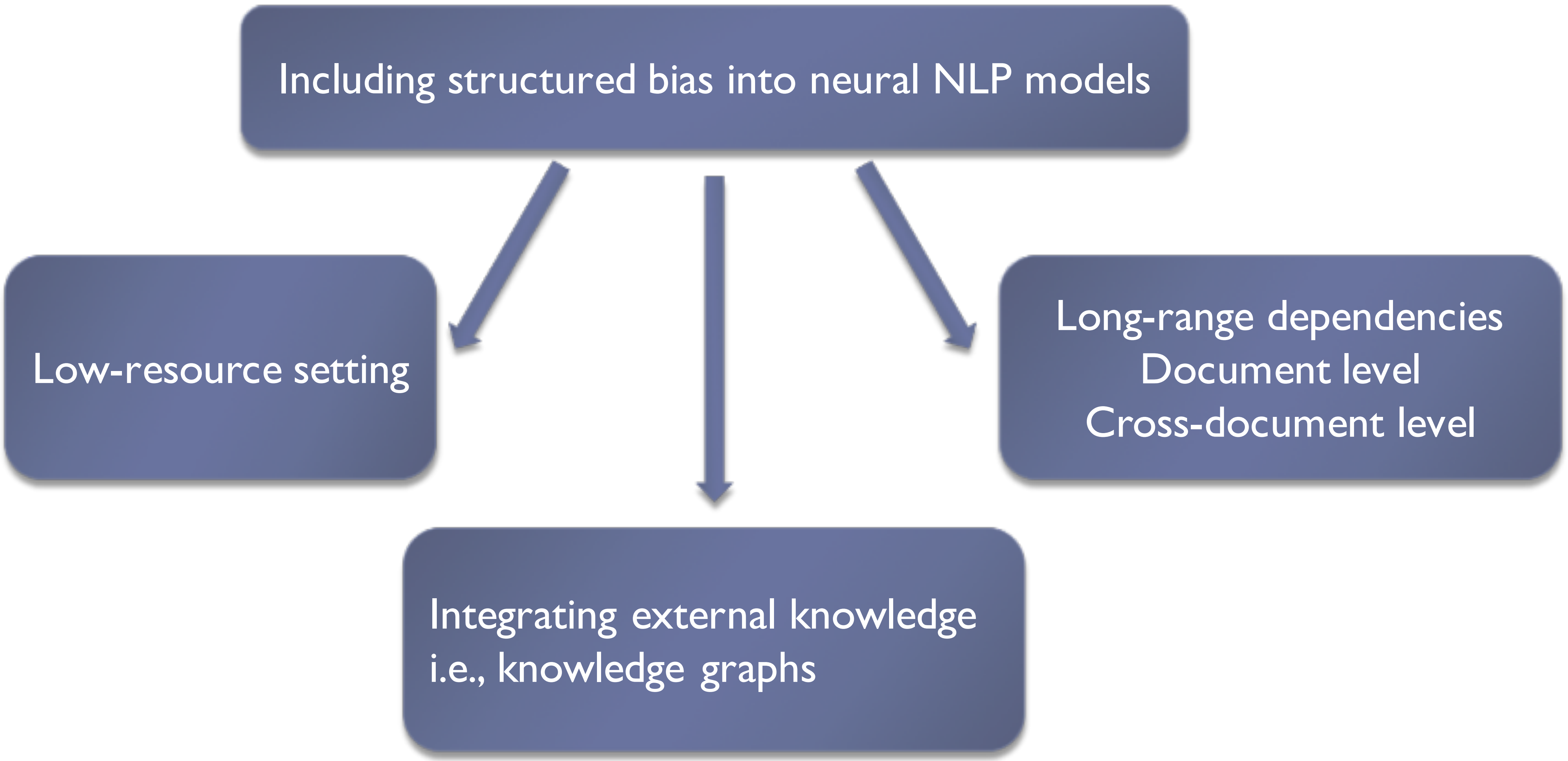
Including structured bias into neural NLP models

Low-resource setting

Roadmap


Including structured bias into neural NLP models
Low-resource setting
Long-range dependencies
Document level
Cross-document level

Roadmap


Including structured bias into neural NLP models
Low-resource setting
Integrating external knowledge
i.e., knowledge graphs
Long-range dependencies
Document level
Cross-document level

Roadmap

Thanks for your attention!

Including structured bias into neural NLP models

Low-resource setting

Long-range dependencies
Document level
Cross-document level

Integrating external knowledge
i.e., knowledge graphs