Semantic Role Labeling Tutorial Part 2

Neural Methods for Semantic Role Labeling

Diego Marcheggiani, Michael Roth, Ivan Titov, Benjamin Van Durme

University of Amsterdam

University of Edinburgh

EMNLP 2017

Copenhagen

Outline: the fall and rise of syntax in SRL

- Early SRL methods
- Symbolic approaches + Neural networks (syntax-aware models)
- Syntax-agnostic neural methods
- Syntax-aware neural methods

Disclaimer

- Recent papers which involve neural networks and SRL
- English language
- Skip predicate identification and disambiguation methods
- Focus on labeling of semantic roles
- PropBank [Palmer et al. 2005]
  - CoNLL 2005 dataset (span-based SRL)
  - CoNLL 2009 dataset (dependency-based SRL)
- F1 measure for role labeling and predicate disambiguation

Outline: the fall and rise of syntax in SRL

- **Early SRL methods**
- Symbolic approaches + Neural networks (syntax-aware models)
- Syntax-agnostic neural methods
- Syntax-aware neural methods

General SRL Pipeline

- Given a predicate:

Sequa makes and repairs jet engines

repair.01

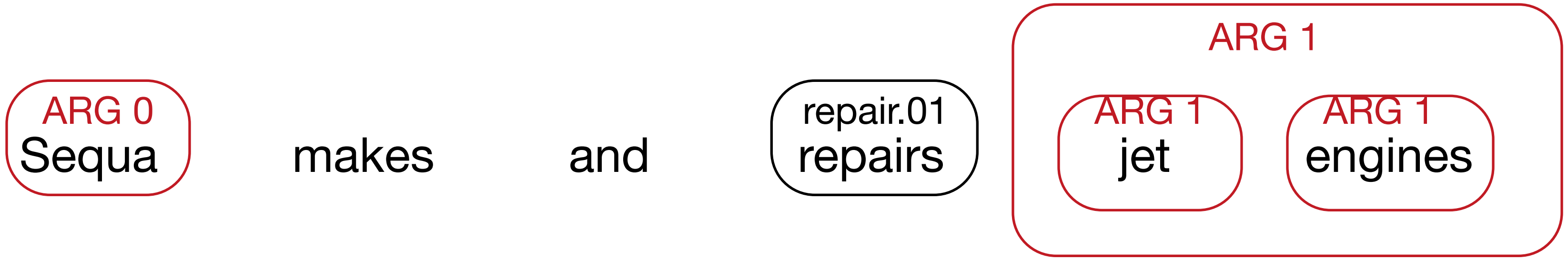
General SRL Pipeline

- Given a predicate:
  - Argument identification

Sequa makes and repair.01 repairs jet engines

General SRL Pipeline

- Given a predicate:
  - Argument identification
  - Role labeling



General SRL Pipeline

- Given a predicate:
  - Argument identification
  - Role labeling
  - Global and/or constrained inference

ARG 0			repair.01	ARG 1	
Sequa	makes	and	repairs	jet	engines

Argument identification

- Hand-crafted rules on the full syntactic tree [Xue and Palmer, 2004]
- Binary classifier [Pradhan et al., 2005; Toutanova et al., 2008]
- Both [Punyakanok et al., 2008]

Role labeling

- Labeling is performed using a classifier (SVM, logistic regression)
- For each argument we get a label distribution
- Argmax over roles will result in a local assignment
- No guarantee the labeling is well formed
  - overlapping arguments, duplicate core roles, etc.

Inference

- Enforce linguistic and structural constraint (e.g., no overlaps, discontinuous arguments, reference arguments, …)
- Viterbi decoding (k-best list with constraints) [Täckström et al., 2015]
- Dynamic programming [Täckström et al., 2015; Toutanova et al., 2008]
- Integer linear programming [Punyakanok et al., 2008]
- Re-ranking [Toutanova et al., 2008; Björkelund et al., 2009]

Early symbolic models

- 3 steps pipeline
- Massive feature engineering
  - argument identification
  - role labeling
  - re-ranking
- Most of the features are syntactic [Gildea and Jurafsky, 2002]

Outline: the fall and rise of syntax in SRL

- Early SRL framework
- **Symbolic approaches + Neural networks (syntax-aware models)**
- Syntax-agnostic neural methods
- Syntax-Aware neural methods

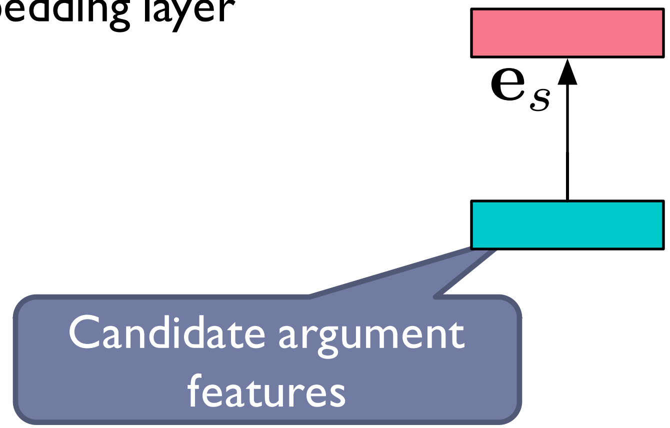
Fitzgerald et al., 2015

- Rule based argument identification
  - as in [Xue and Palmer, 2004] but for dependency parsing
- **Neural network** for local role labeling
- Global structural inference based on dynamic programming
  - [Täckström et al., 2015]

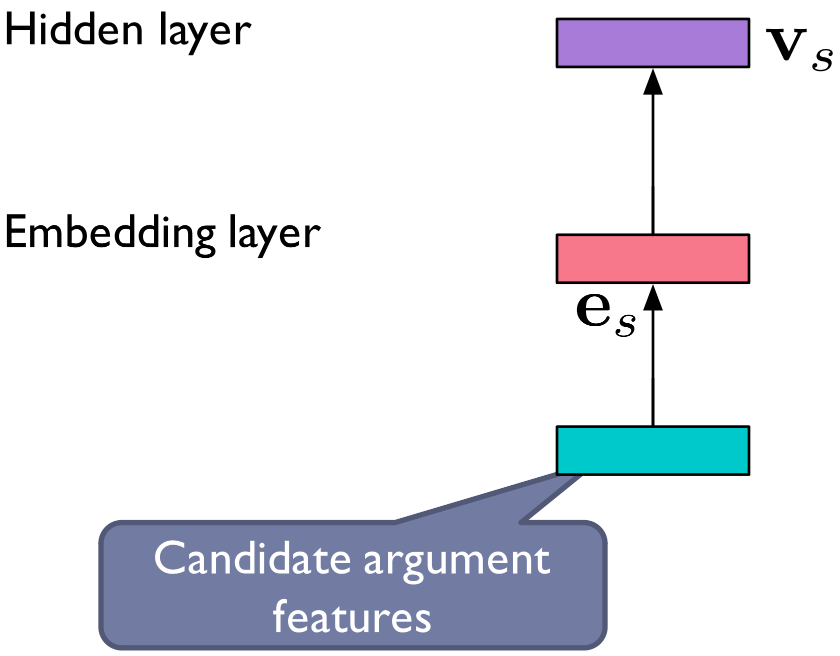
Fitzgerald et al., 2015: Architecture

Hidden layer

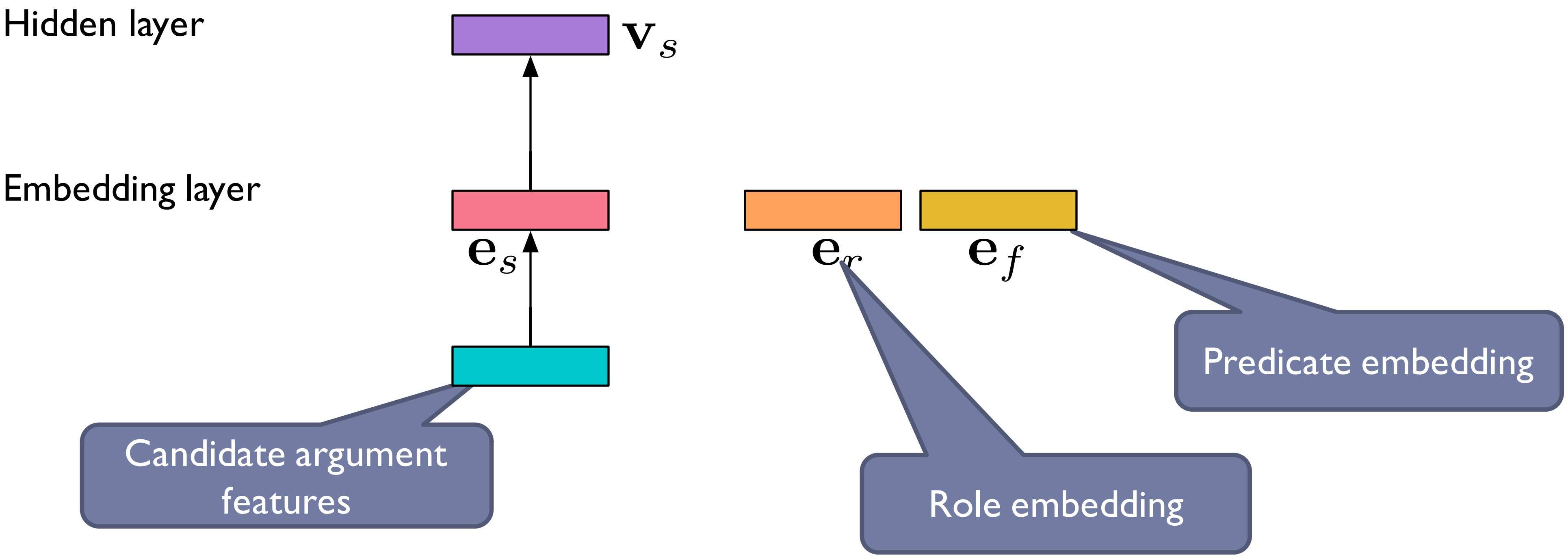
Embedding layer



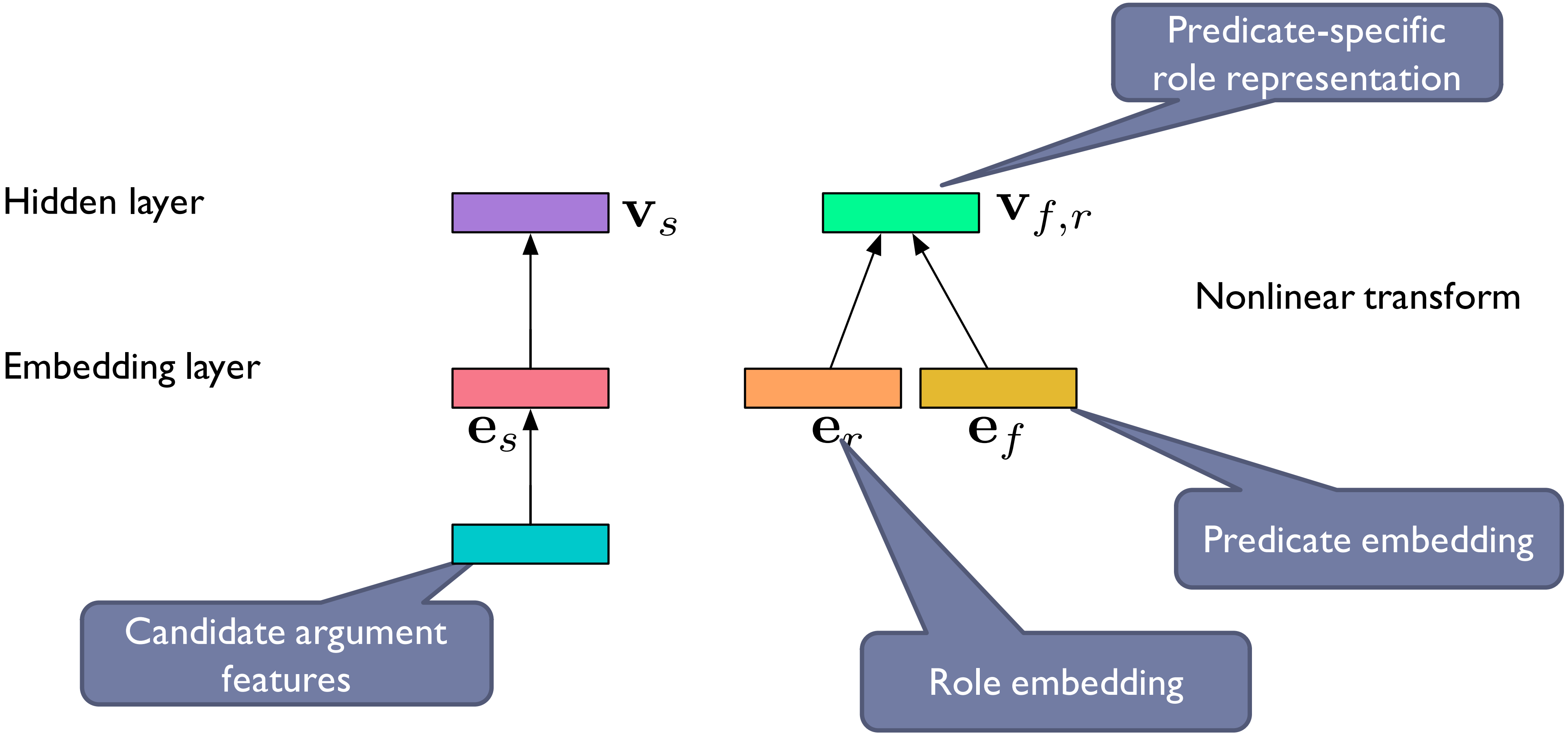
Fitzgerald et al., 2015: Architecture



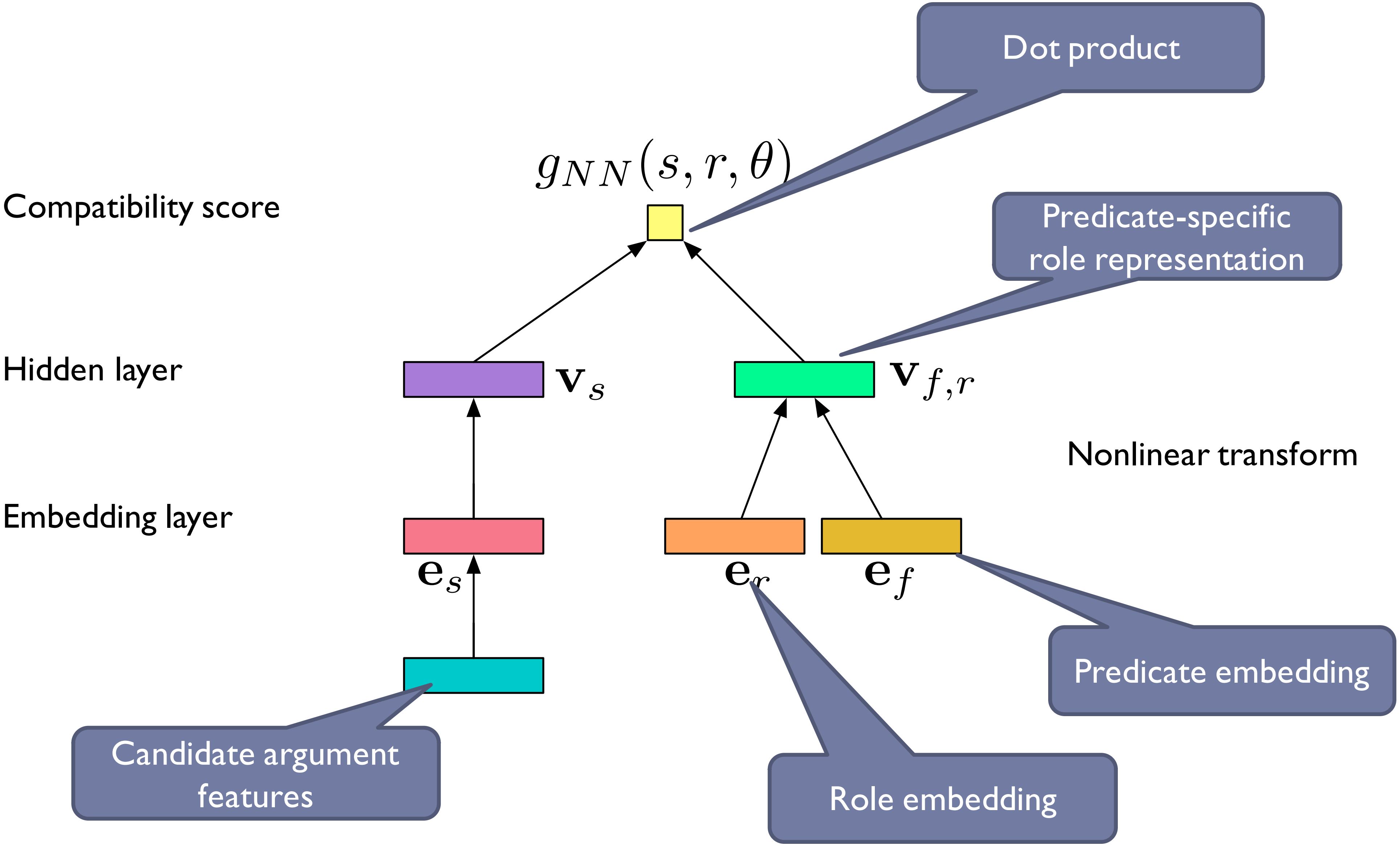
Fitzgerald et al., 2015: Architecture



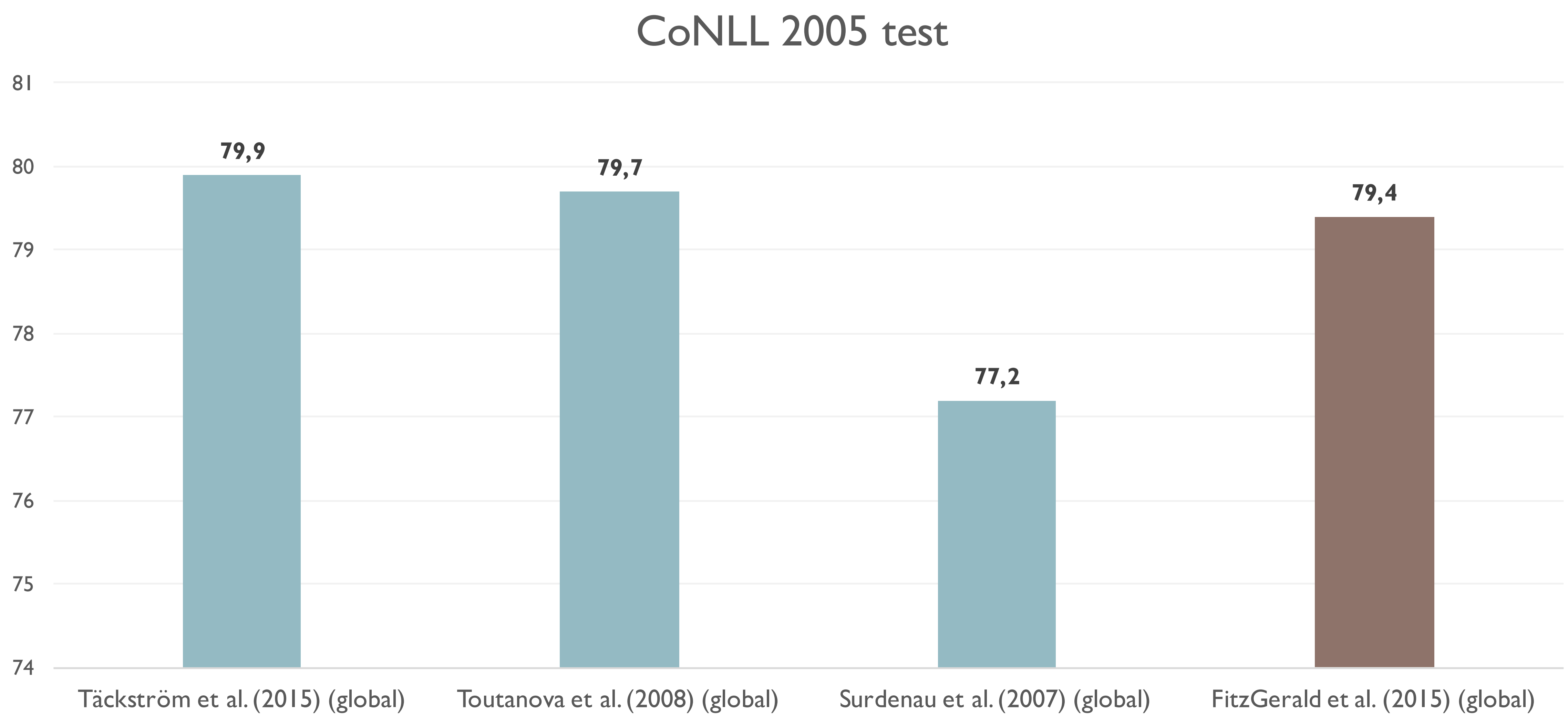
Fitzgerald et al., 2015: Architecture



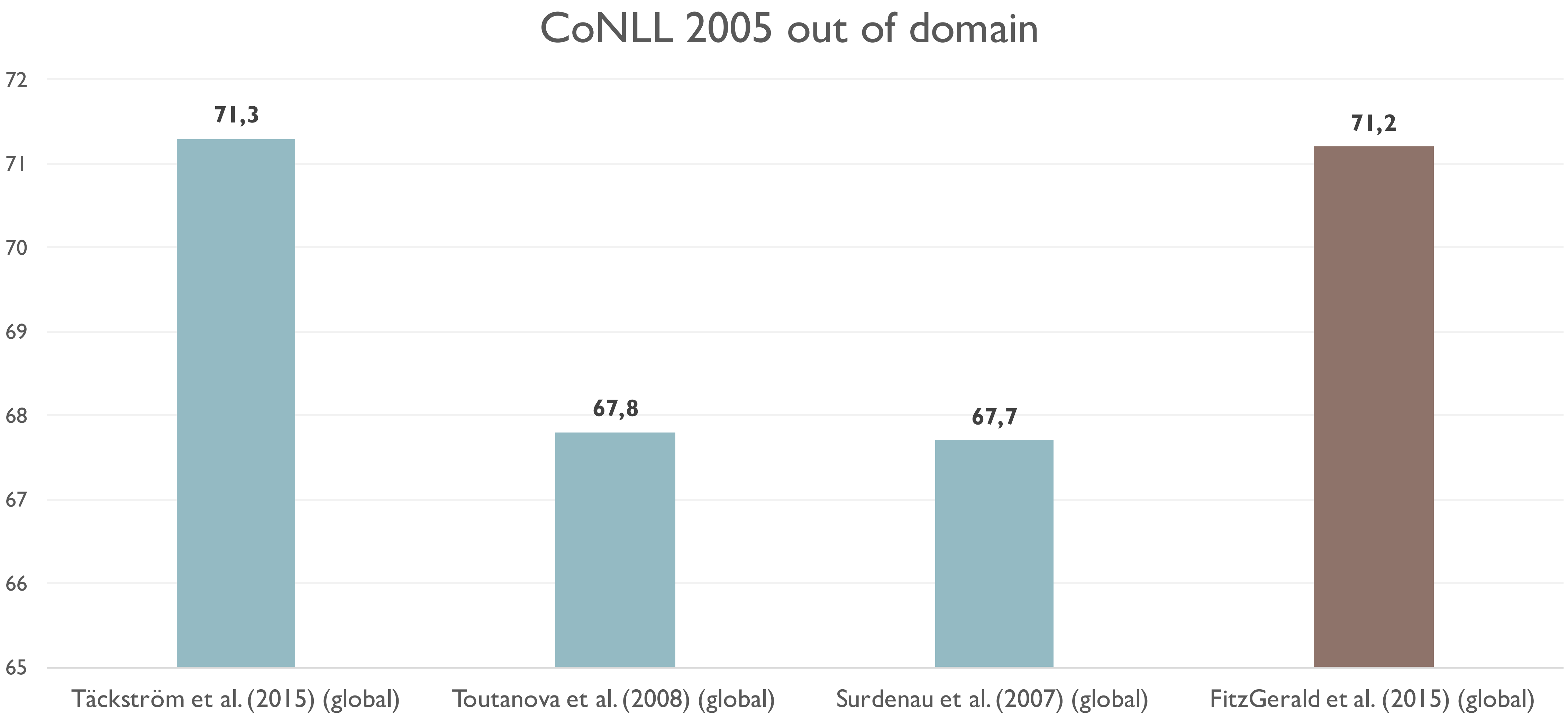
Fitzgerald et al., 2015: Architecture



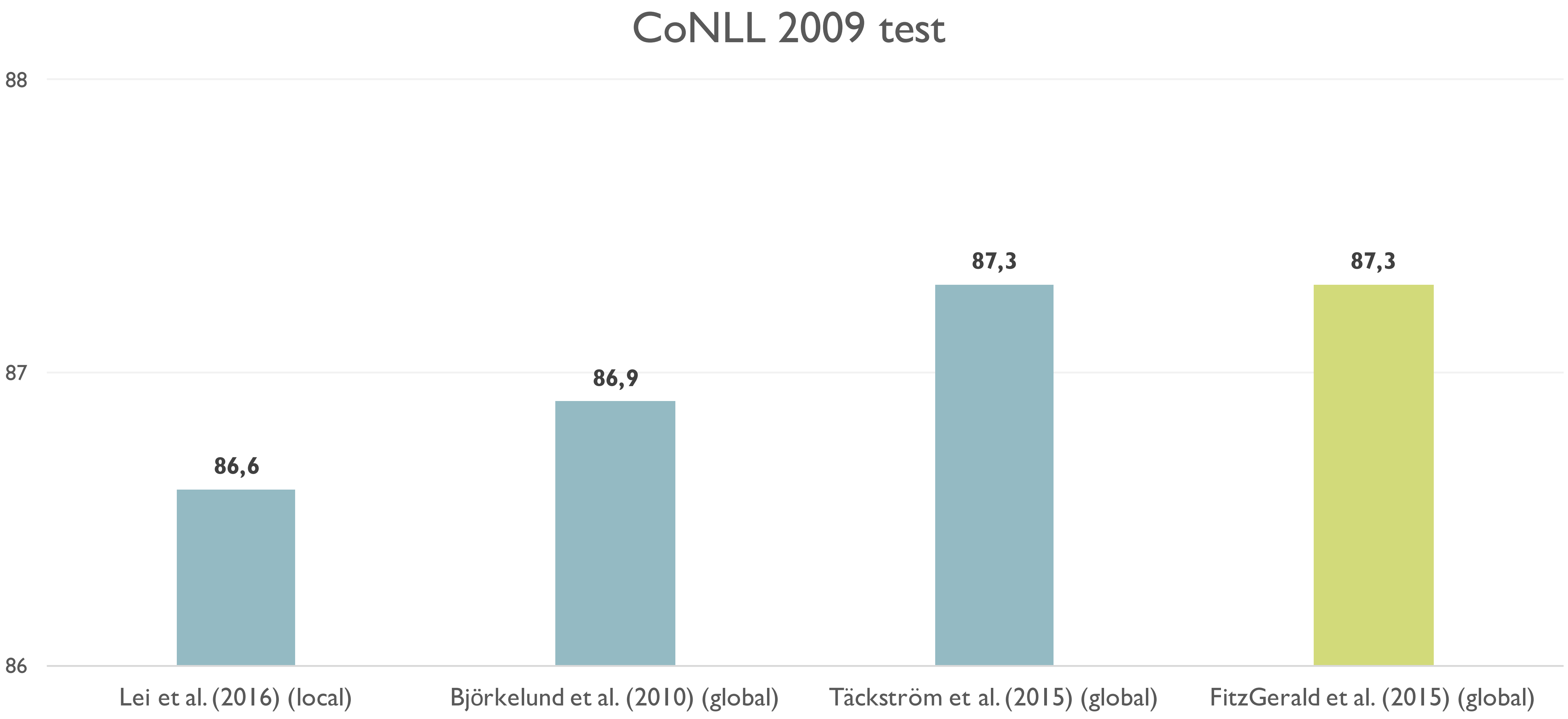
Fitzgerald et al., 2015: Span-based SRL results



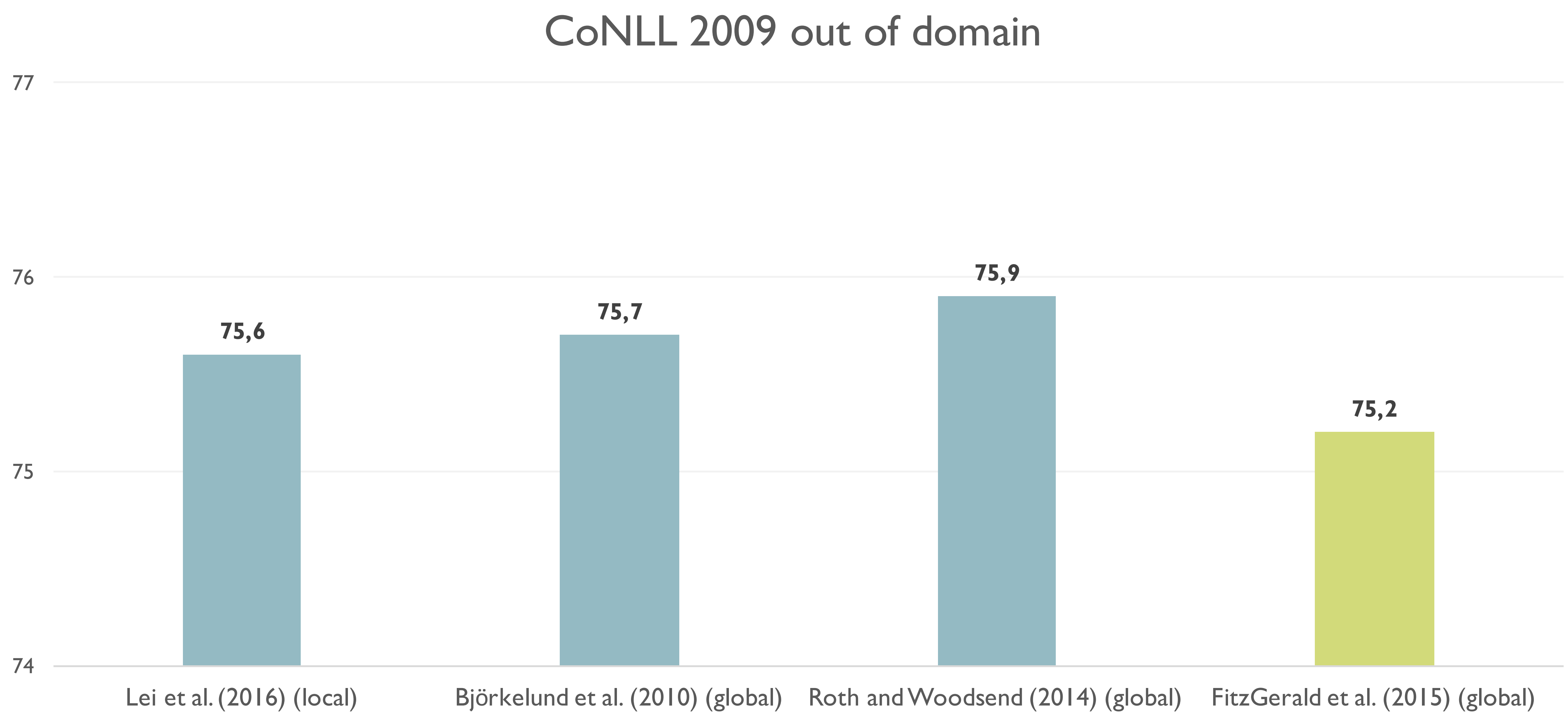
Fitzgerald et al., 2015: Span-based SRL results



Fitzgerald et al., 2015: Dependency-based SRL results



Fitzgerald et al., 2015: Dependency-based SRL results



Fitzgerald et al., 2015

- Predicate-role composition
  - Predicate-specific role representation
  - Learning distributed predicate representation across different formalisms
  - State of the art on FrameNet dataset
- Feature embeddings
  - Use “simple” span features
  - Let the network figure out how to compose them
  - Reduced feature engineering

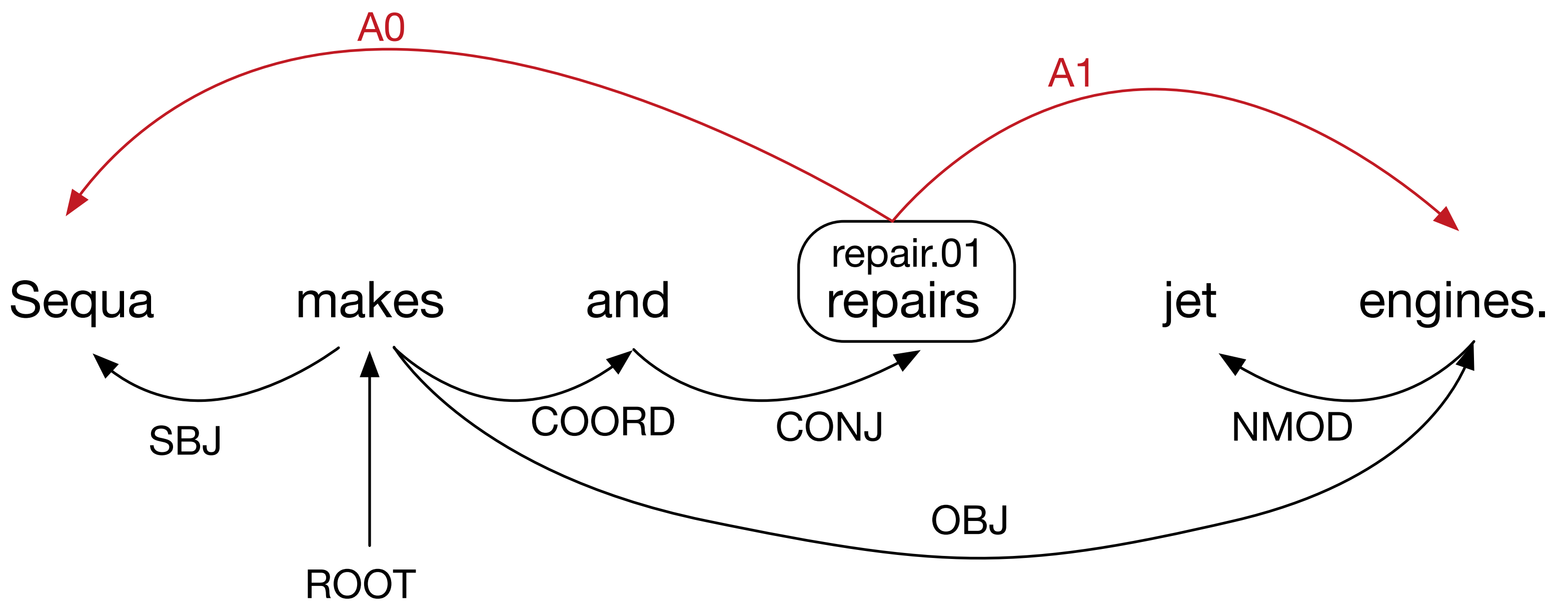
Roth and Lapata, 2016

- Dependency-based SRL
- Neural network with **dependency path embeddings** as local classifier
  - Argument identification
  - Role labeling
- Global re-ranking of k-best local assignments

Roth and Lapata, 2016: Dependency path embeddings

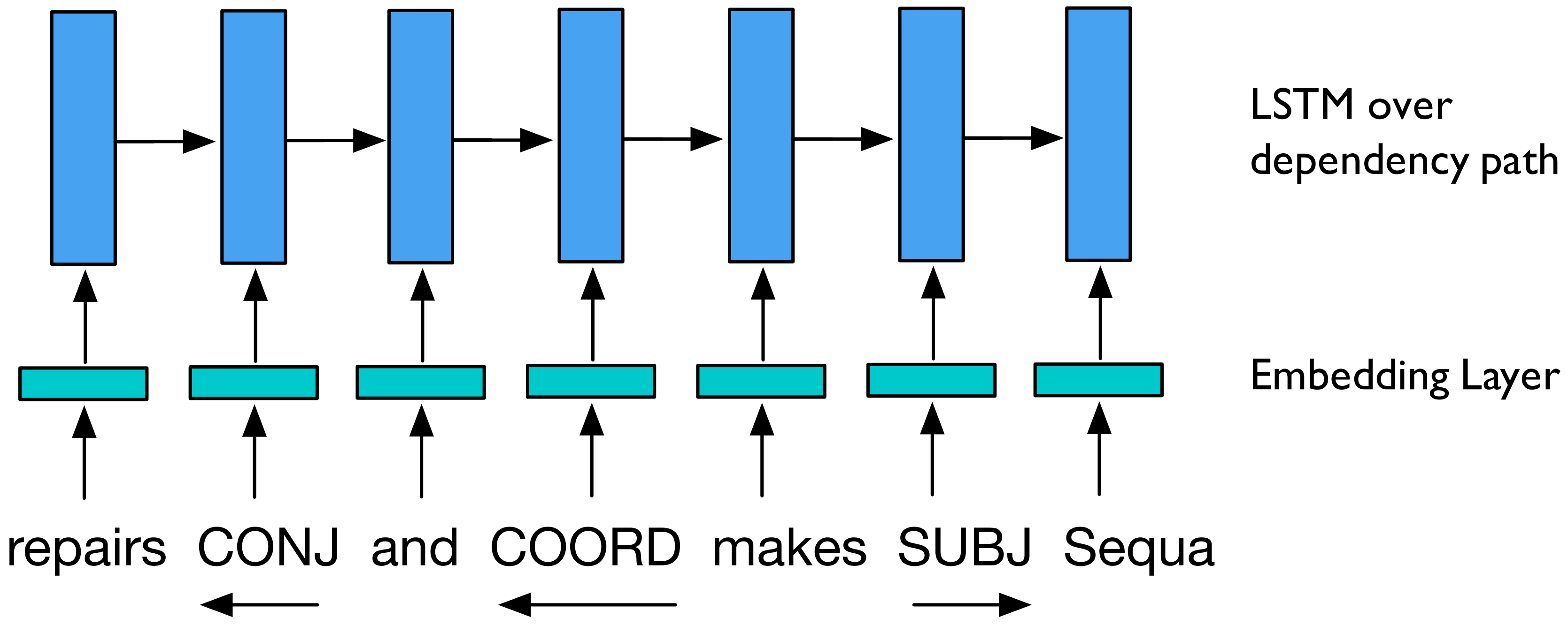
- Syntactic paths between predicates and arguments are an important feature
- It may be extremely sparse
- Creating a distributed representation can solve the problem
- Use LSTM [Hochreiter and Schmidhuber, 1995] to encode paths

Roth and Lapata, 2016: Example

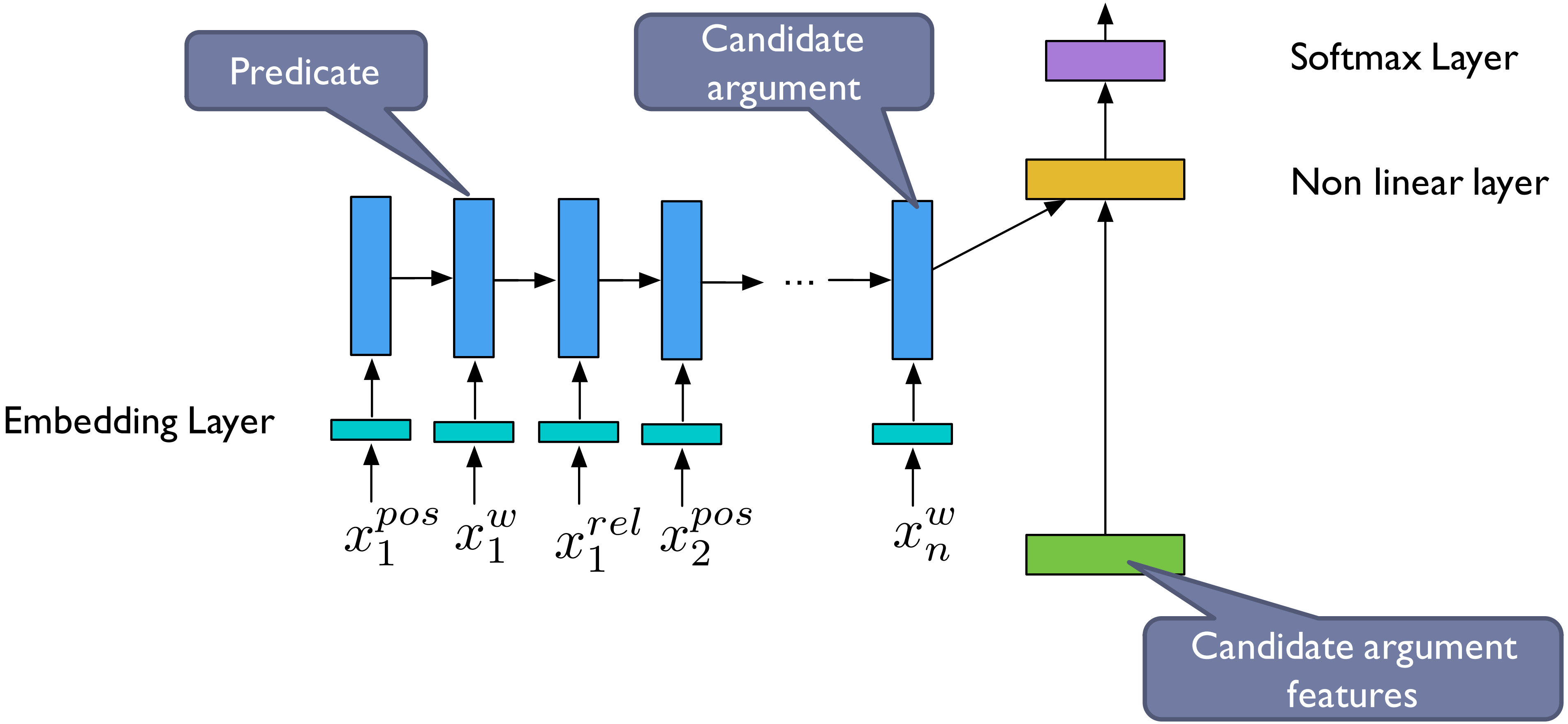


repairs CONJ and COORD makes SUBJ Sequa

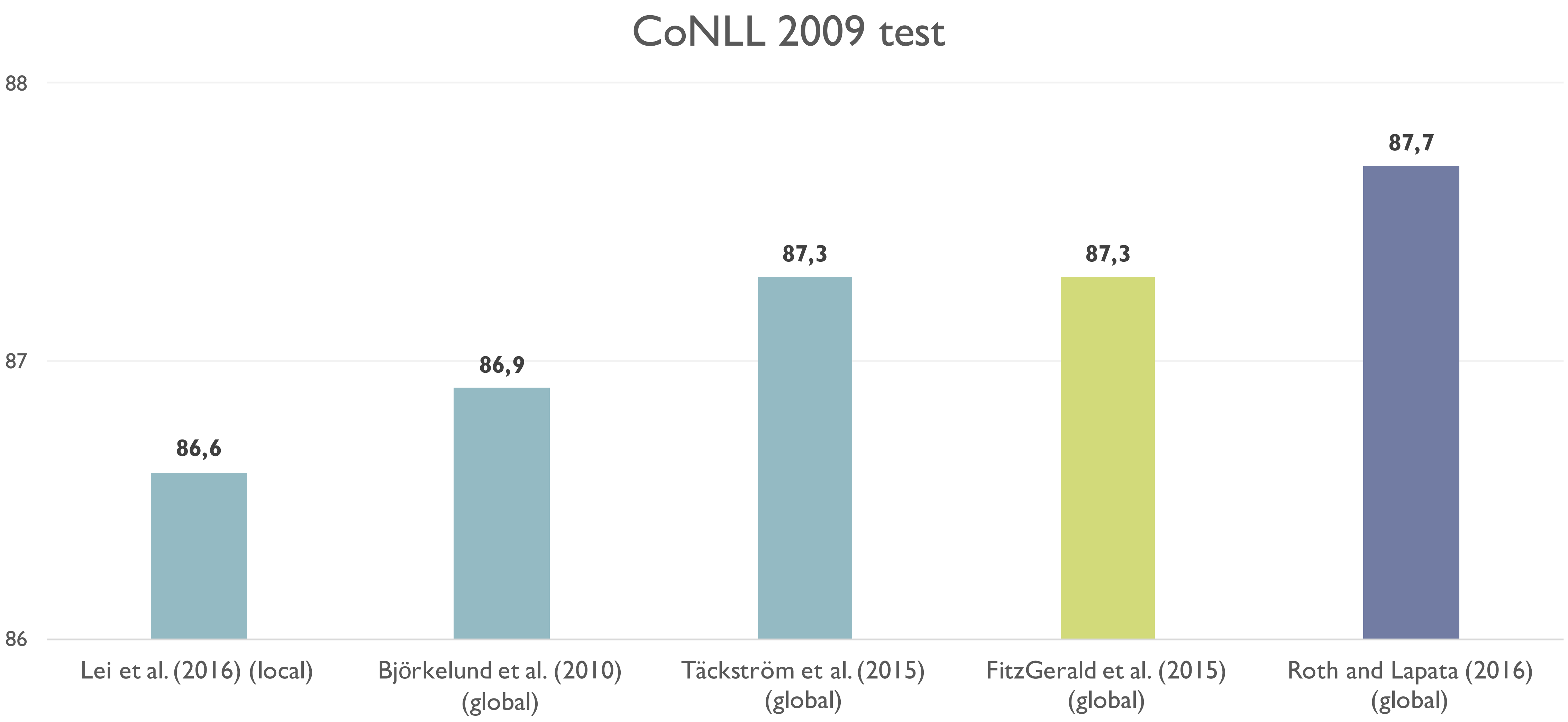
Roth and Lapata, 2016: Dependency path embeddings example



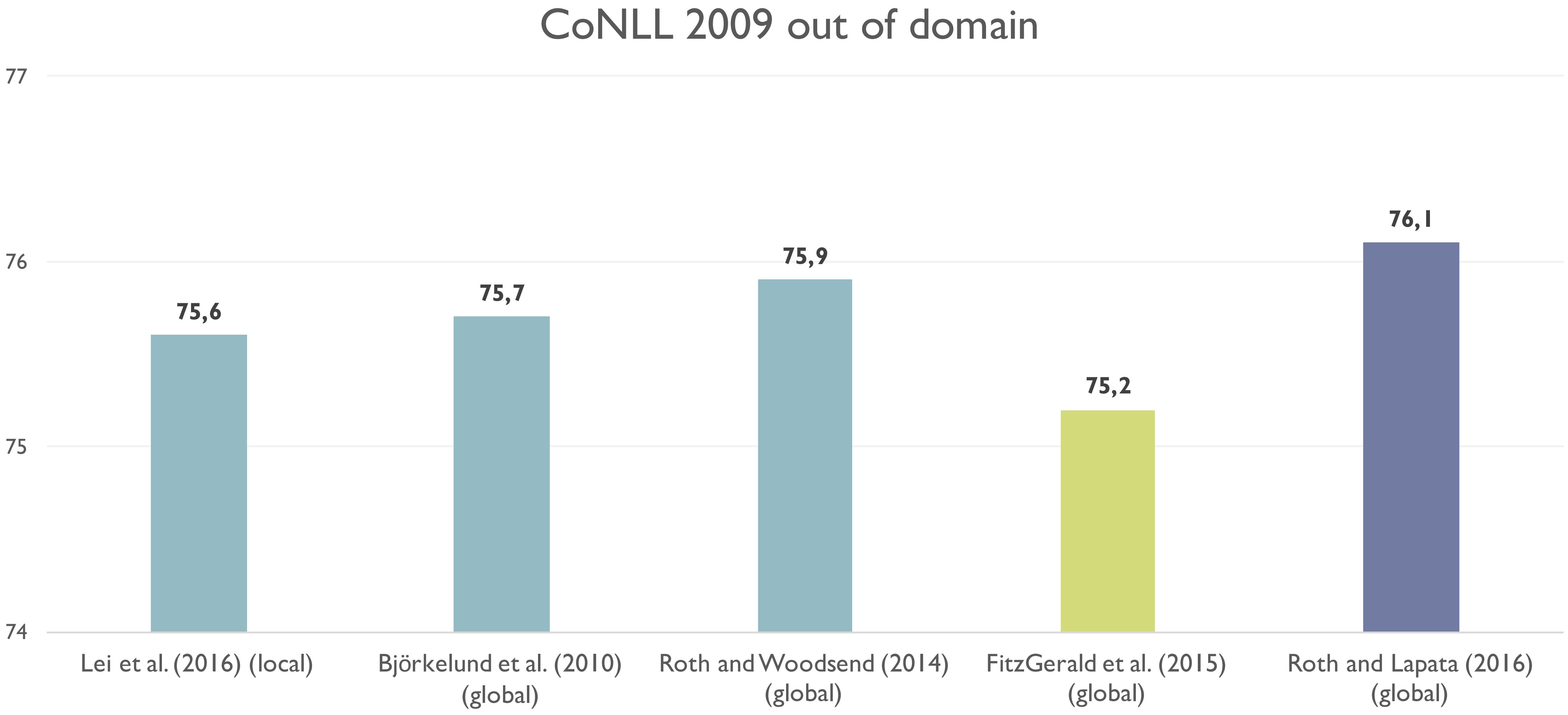
Roth and Lapata, 2016: Architecture



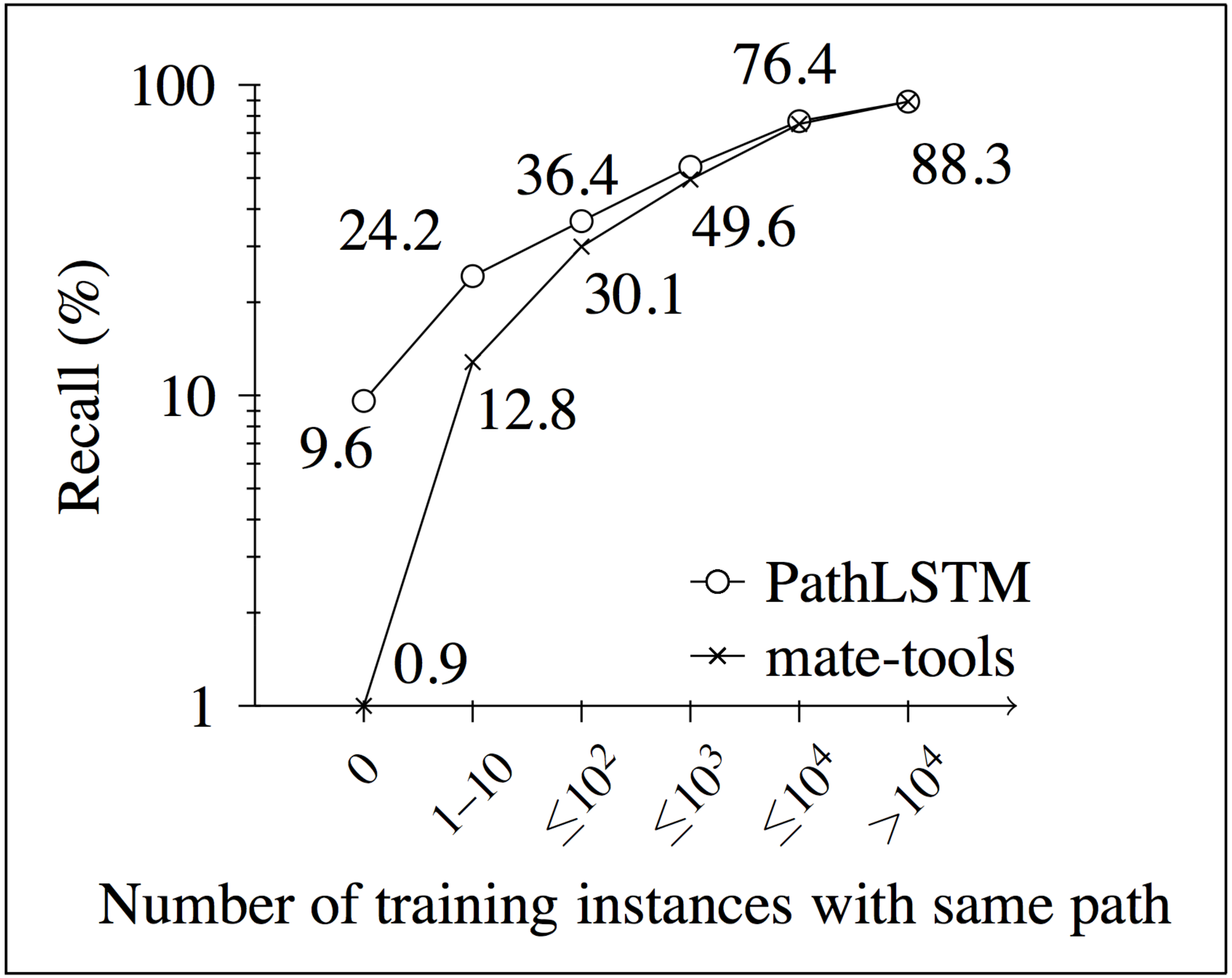
Roth and Lapata, 2016: Dependency-based SRL results



Roth and Lapata, 2016: Dependency-based SRL results



Roth and Lapata, 2016: Analysis



Roth and Lapata, 2016

- Encode syntactic paths with LSTMs
  - Overcome sparsity
- Combination of symbolic features and continuous syntactic paths

Outline: the fall and rise of syntax in SRL

- Early SRL framework
- Symbolic approaches + Neural networks
- **Syntax-agnostic neural methods (the fall)**
- Syntax-aware neural methods

Syntax-agnostic neural methods

- SRL as a sequence labeling task

ARG 0			repair.01	ARG 1
Sequa	makes	and	repairs	jet engines

Syntax-agnostic neural methods

- SRL as a sequence labeling task
  - Argument identification and role labeling in one step

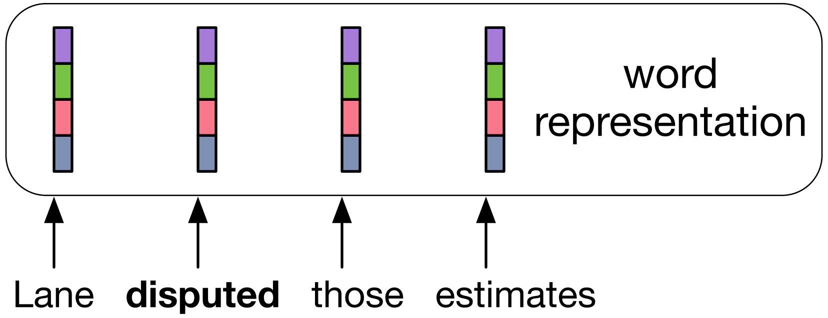
ARG 0			repair.01	ARG 1	
Sequa	makes	and	repairs	jet	engines
B-A0	O	O	O	B-A1	I-A1

Syntax-agnostic neural methods

- General architecture
  - Word encoding
  - Sentence encoding (via LSTM)
  - Decoding
- No use of any kind of treebank syntax (not trivial to encode it)
- Differentiable end-to-end
  - [Collobert et al., (2011)]

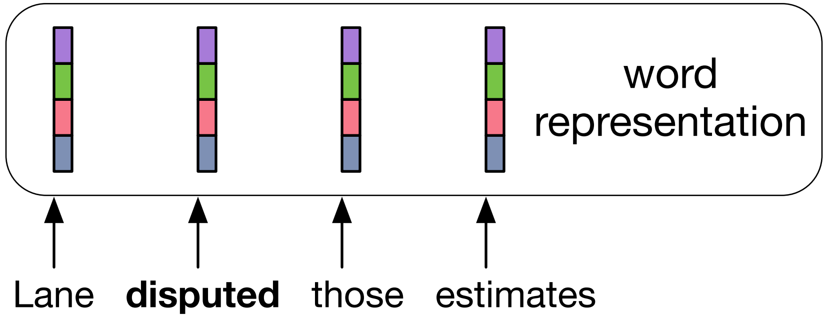
Zhou and Xu, 2015: Word encoding

- Pretrained word embedding



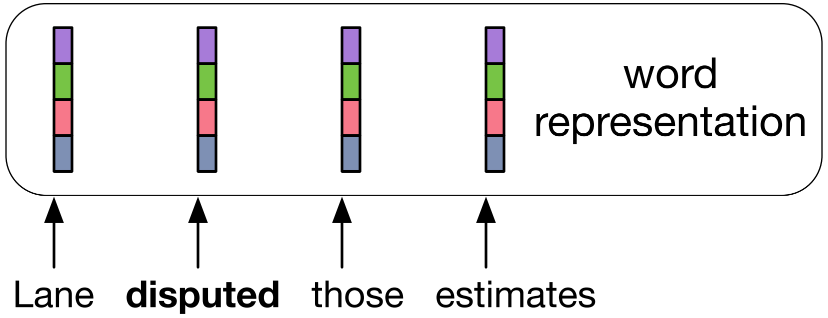
Zhou and Xu, 2015: Word encoding

- Pretrained word embedding
- Distance from the predicate



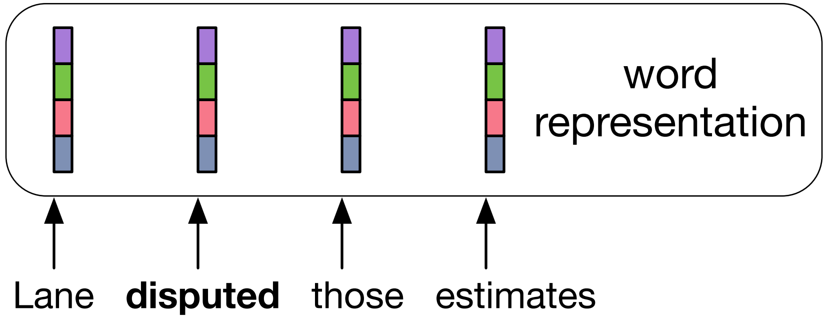
Zhou and Xu, 2015: Word encoding

- Pretrained word embedding
- Distance from the predicate
- Predicate context (for disambiguation)



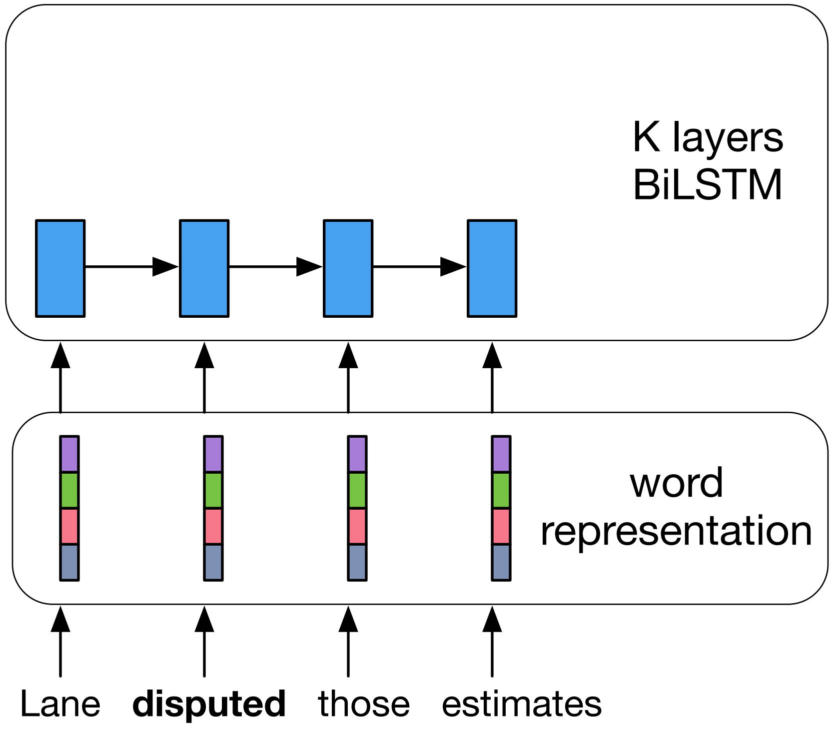
Zhou and Xu, 2015: Word encoding

- Pretrained word embedding
- Distance from the predicate
- Predicate context (for disambiguation)
- Predicate region mark



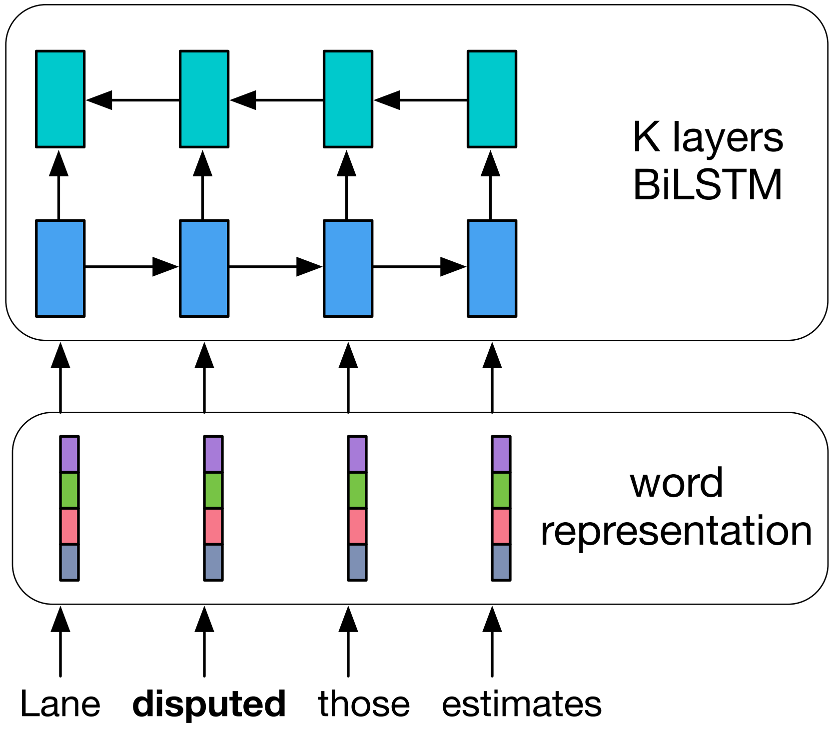
Zhou and Xu, 2015: Sentence encoding

- Bidirectional LSTM
  - Forward (left context)



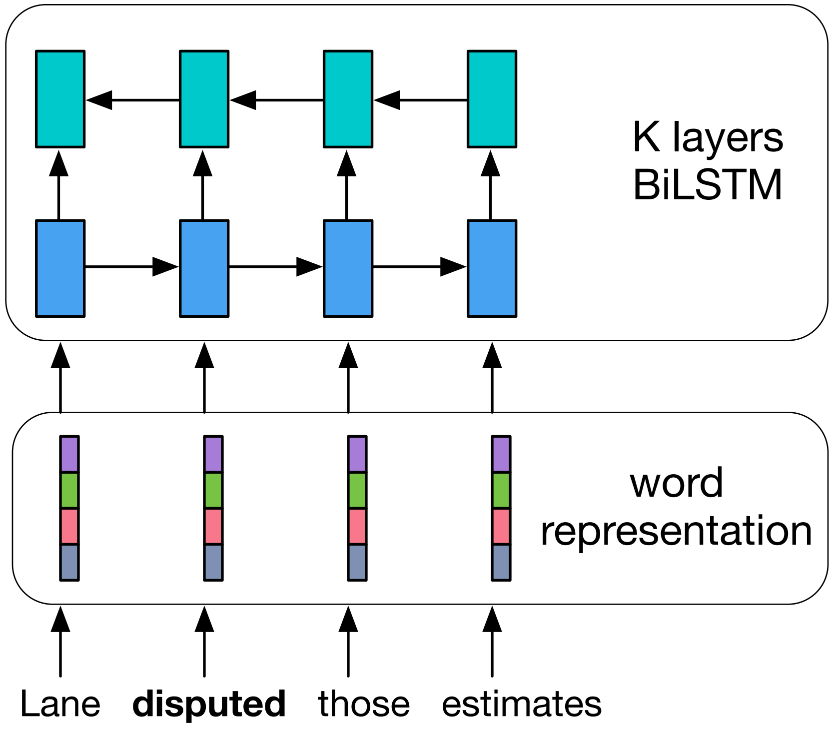
Zhou and Xu, 2015: Sentence encoding

- Bidirectional LSTM
  - Forward (left context)
  - Backward (right context)



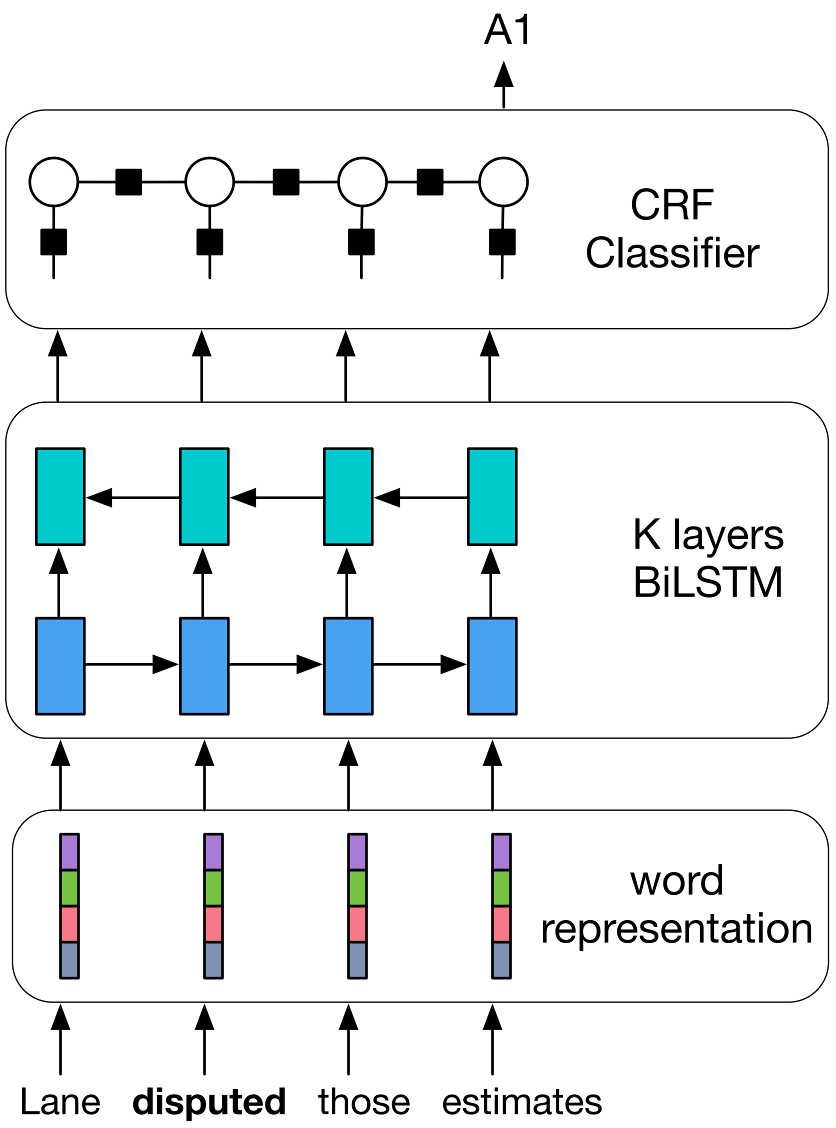
Zhou and Xu, 2015: Sentence encoding

- Bidirectional LSTM
  - Forward (left context)
  - Backward (right context)
  - Snake BiLSTM

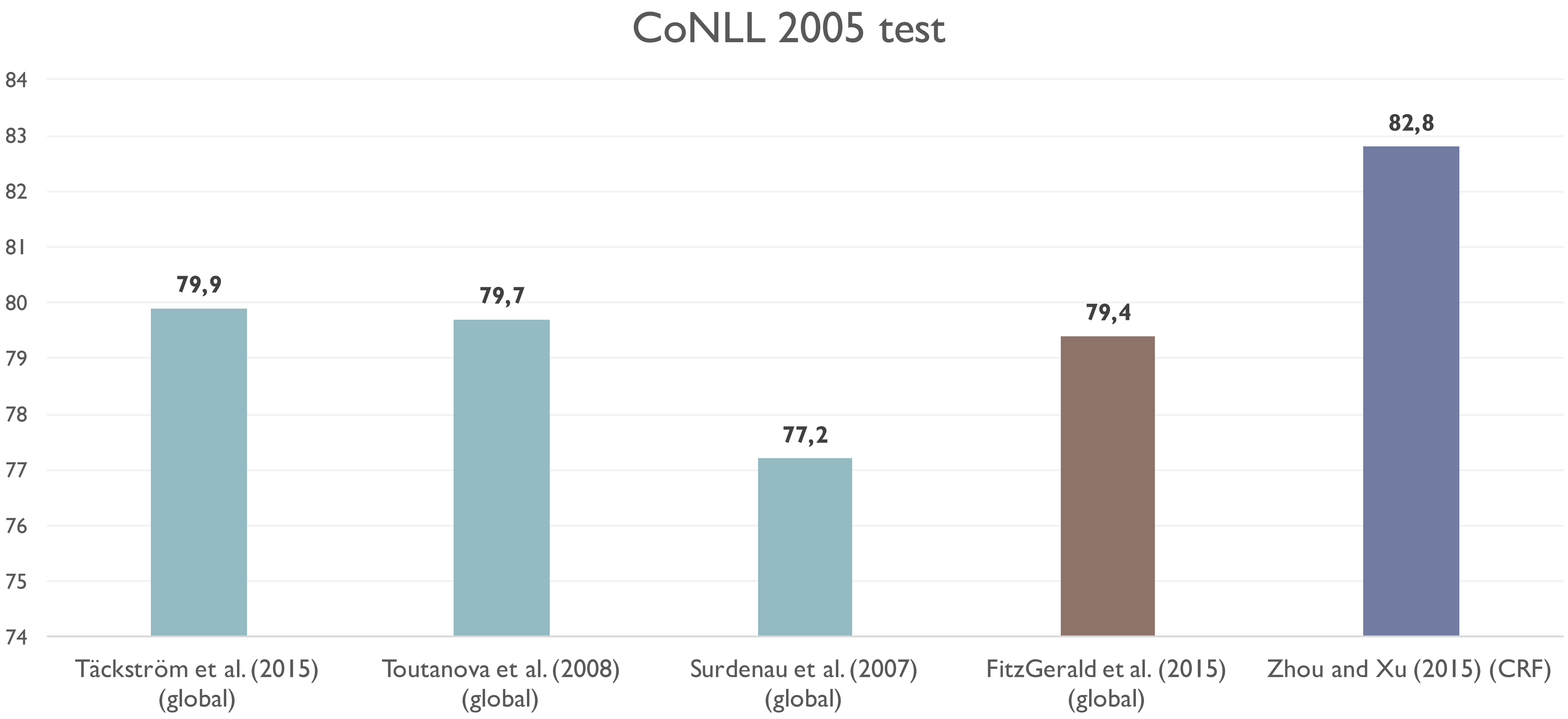


Zhou and Xu, 2015: Decoder

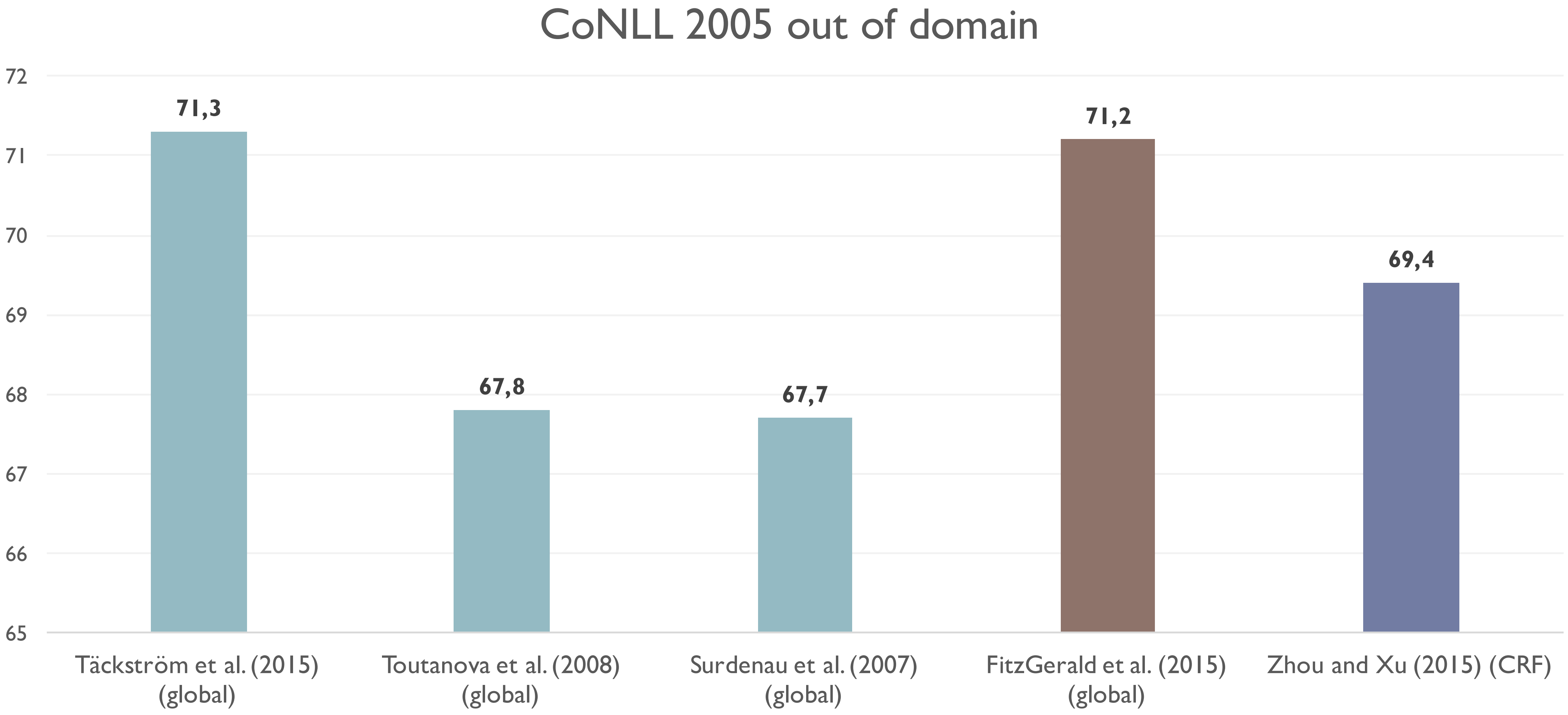
- Conditional Random Field
  - [Lafferty et al., 2001]
  - Markov assumption between role labels



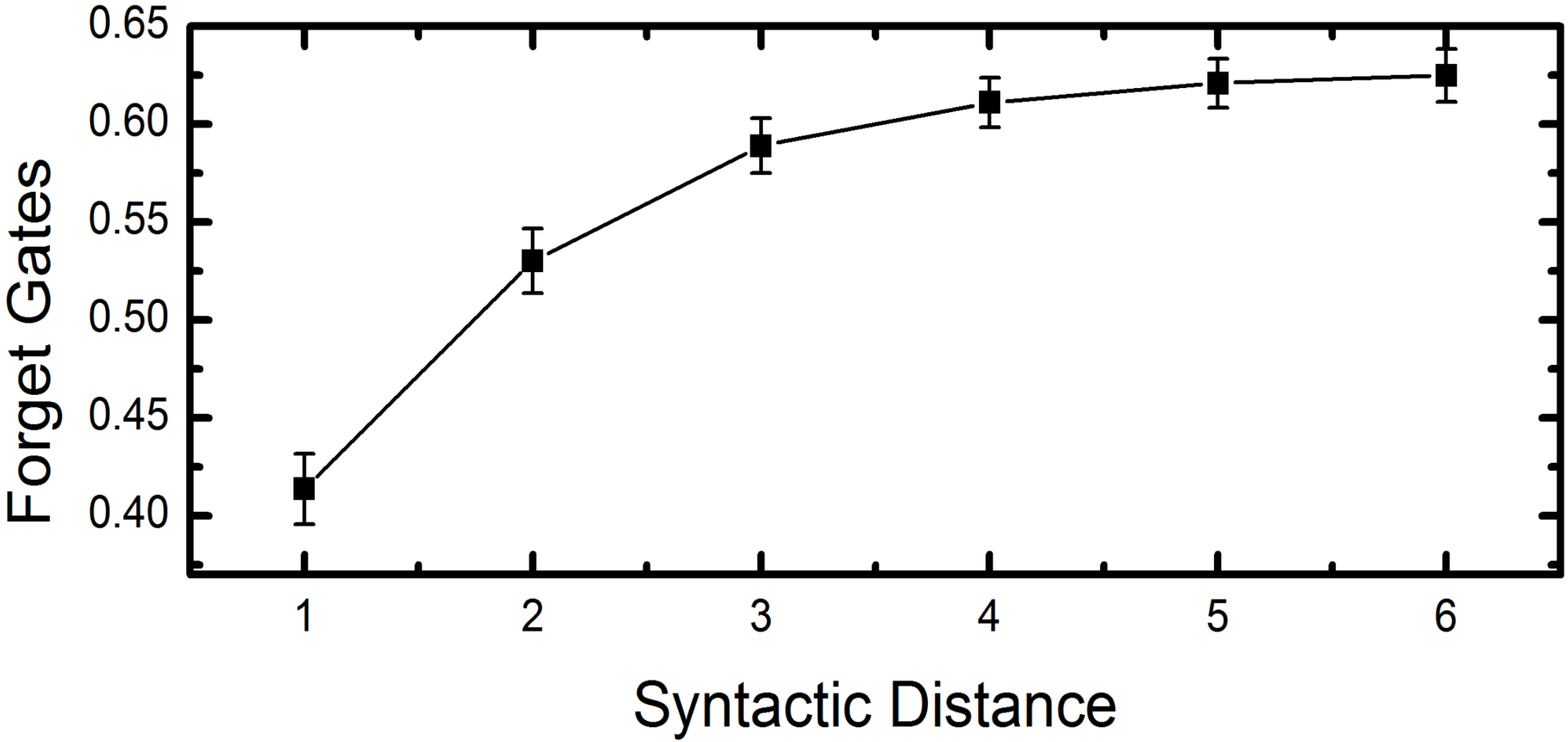
Zhou and Xu, 2015: Results



Zhou and Xu, 2015: Results



Zhou and Xu, 2015: Analysis

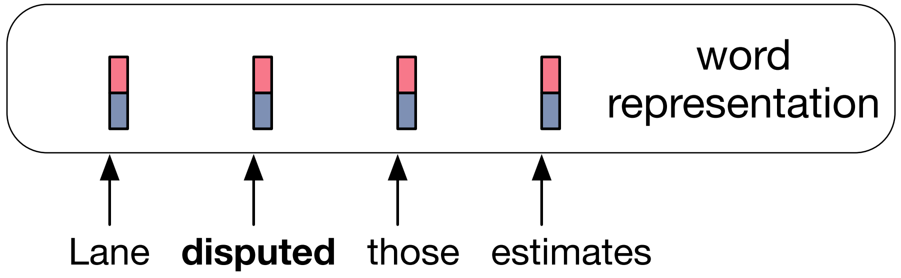


Zhou and Xu, 2015

- No syntax
- Minimal word representation
- Sentence encoding with "Snake" BiLSTM

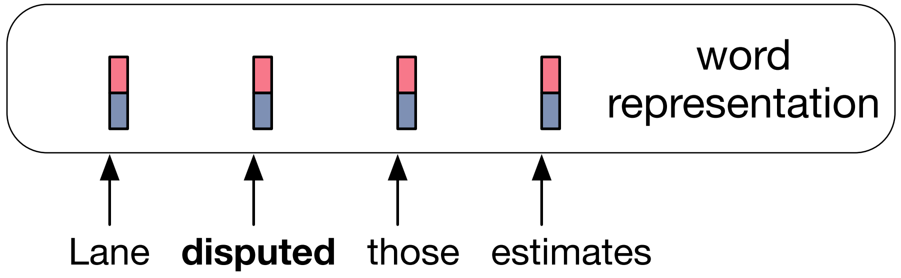
He et al., 2017: Word encoding

- Pretrained word embedding
- Predicate flag

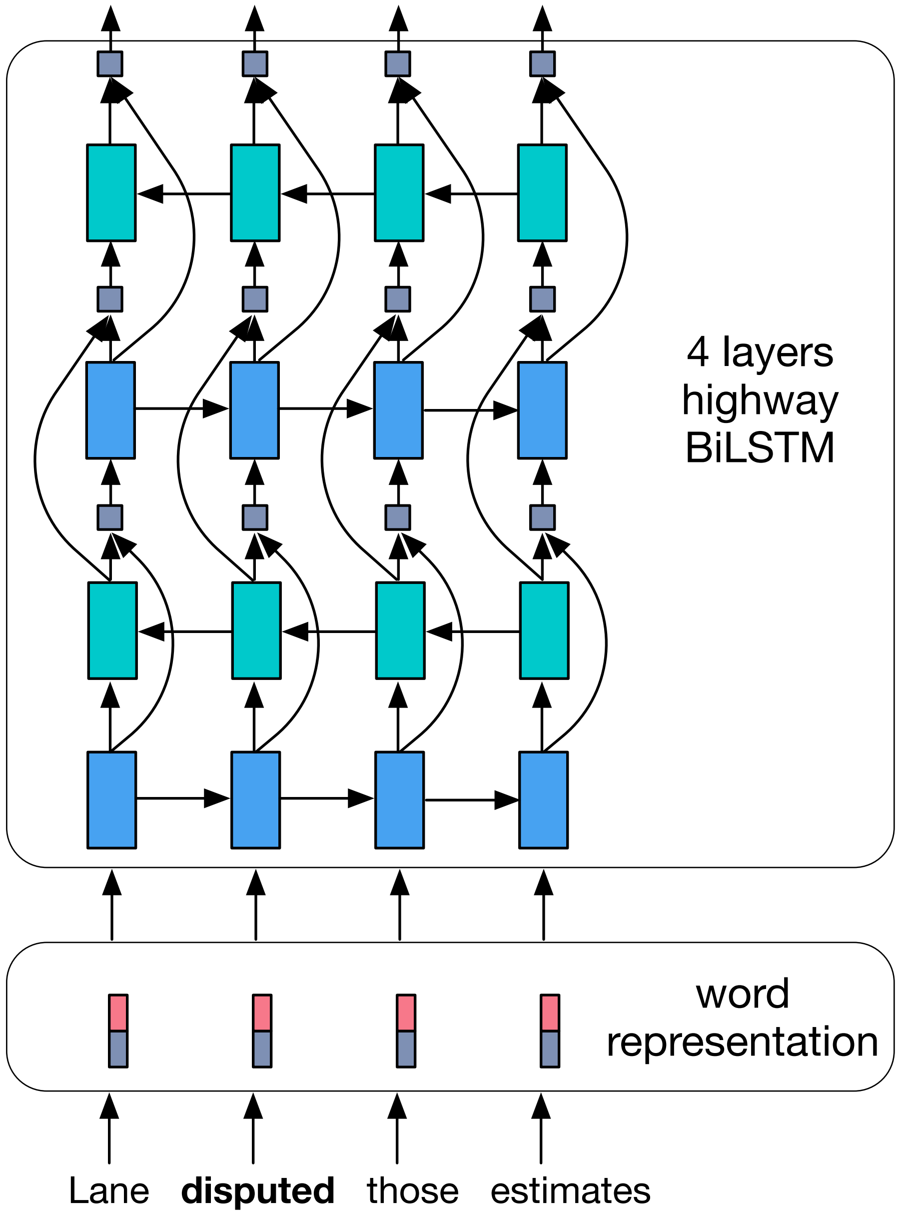


He et al., 2017: Sentence encoding

- “Snake” Bi-LSTM
- Highway connections [Srivastava et al., 2015]
- Recurrent dropout [Gal and Ghahramani, 2016]



He et al., 2017: Highway connections [Srivastava et al., 2015]



He et al., 2017: Highway connections [Srivastava et al., 2015]

Transform gate

$$r_{l,t} = \sigma(W^l(h_{l,t-1} \circ h_{l-1,t}))$$

He et al., 2017: Highway connections [Srivastava et al., 2015]

Transform gate

$$r_{l,t} = \sigma(W^l(h_{l,t-1} \circ h_{l-1,t}))$$

Gated hidden state

Previous layer hidden state

$$h_{l,t} = \underbrace{r_{l,t} \odot h'_{l,t}}_{\text{Current hidden state}} + \overbrace{(1 - r_{l,t}) \odot V h_{l-1,t}}^{\text{Previous layer hidden state}}$$

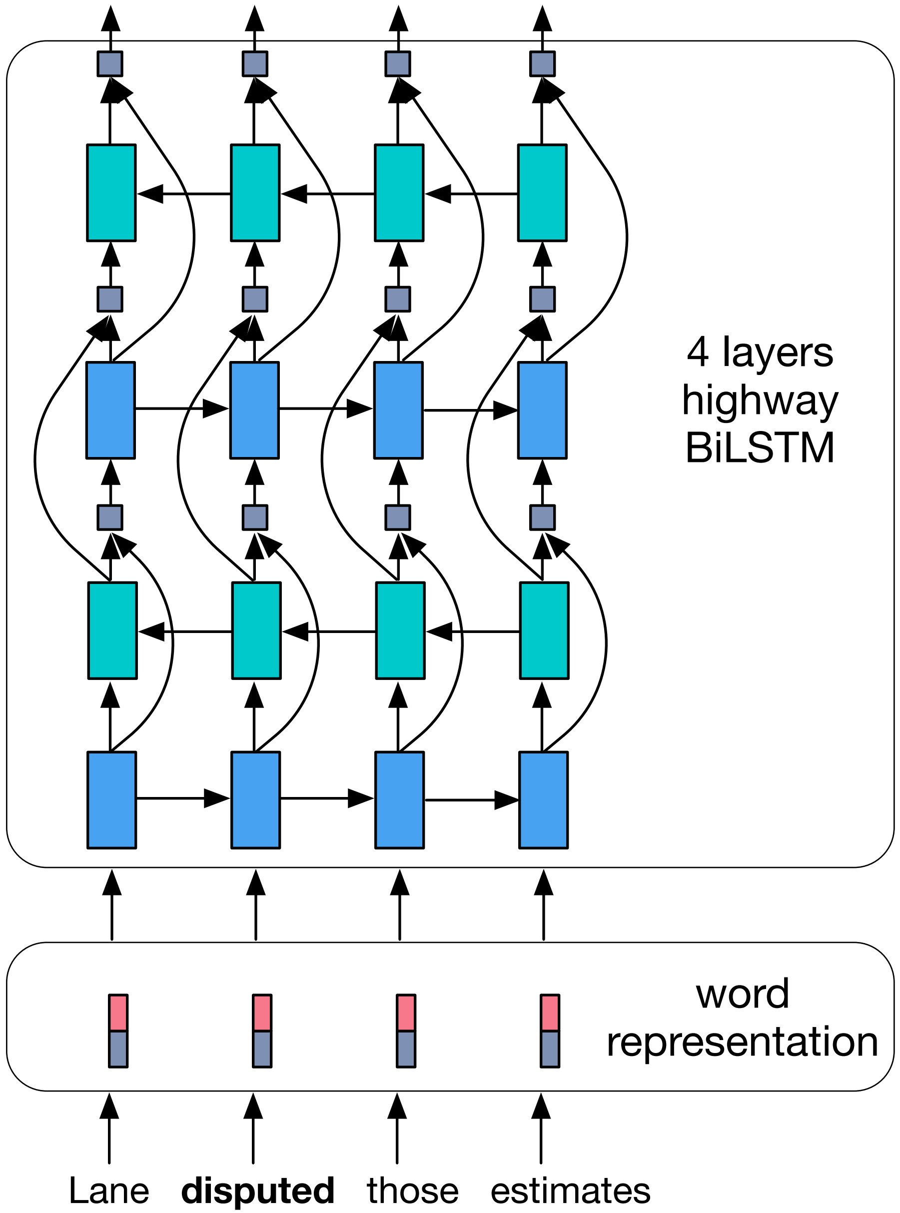
Current hidden state

He et al., 2017: Recurrent dropout [Gal and Ghahramani, 2016]

Gated hidden state

$$\tilde{h}_{l,t} = z_l \odot h_{l,t}$$

Random binary mask

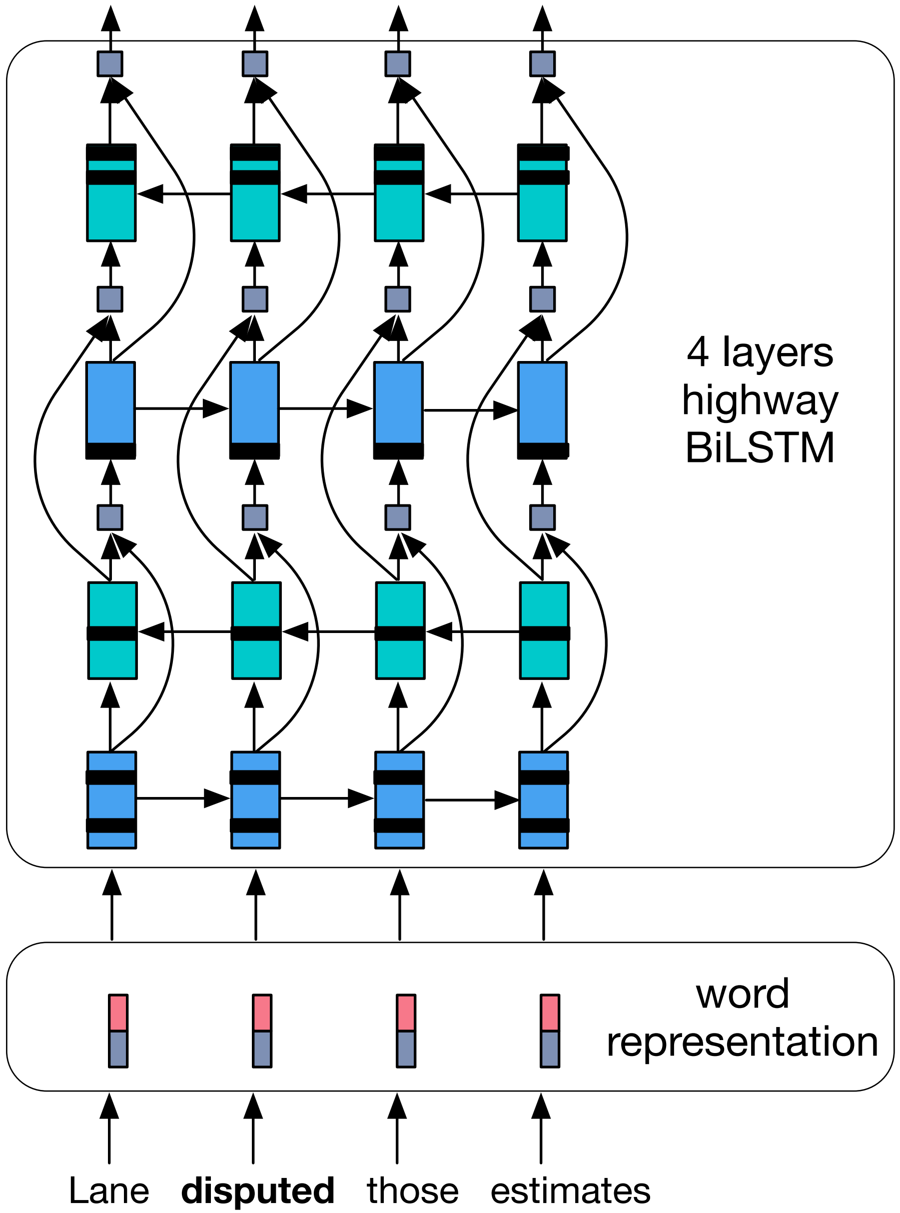


He et al., 2017: Recurrent dropout [Gal and Ghahramani, 2016]

Gated hidden state

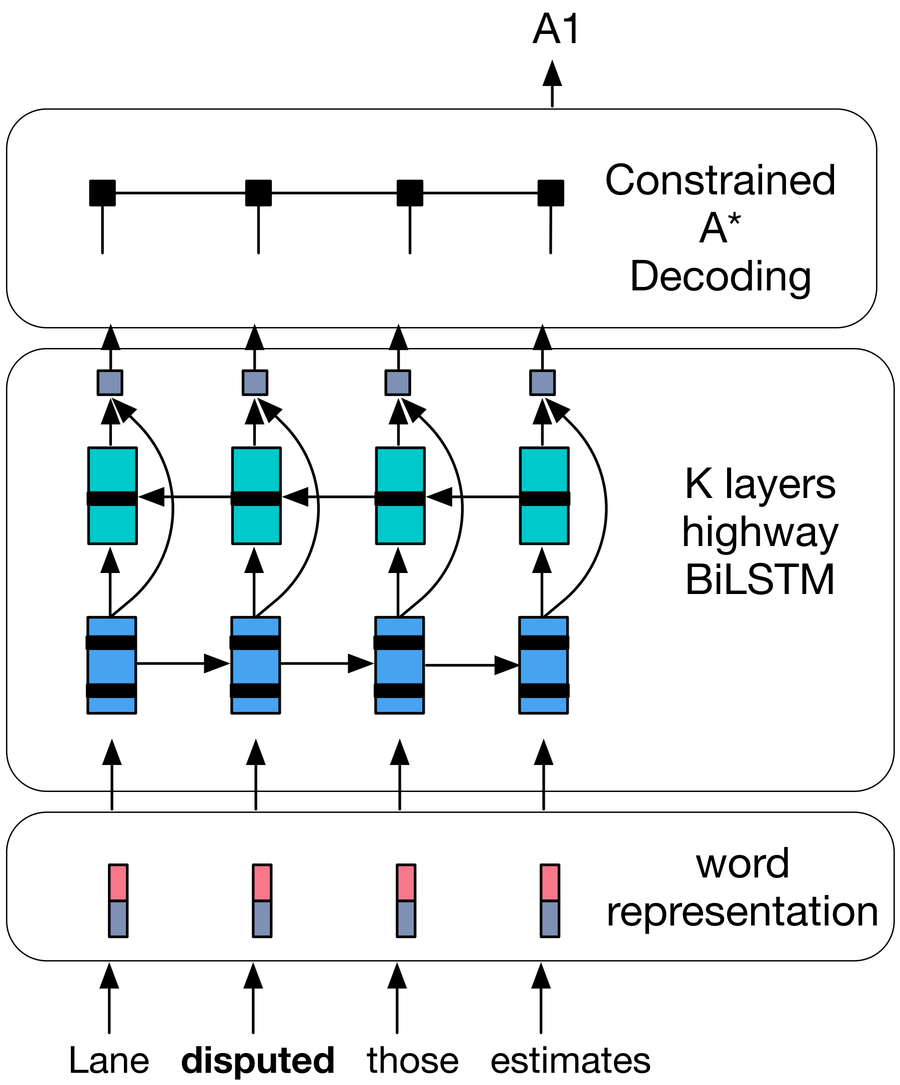
$$\tilde{h}_{l,t} = z_l \odot h_{l,t}$$

Random binary mask

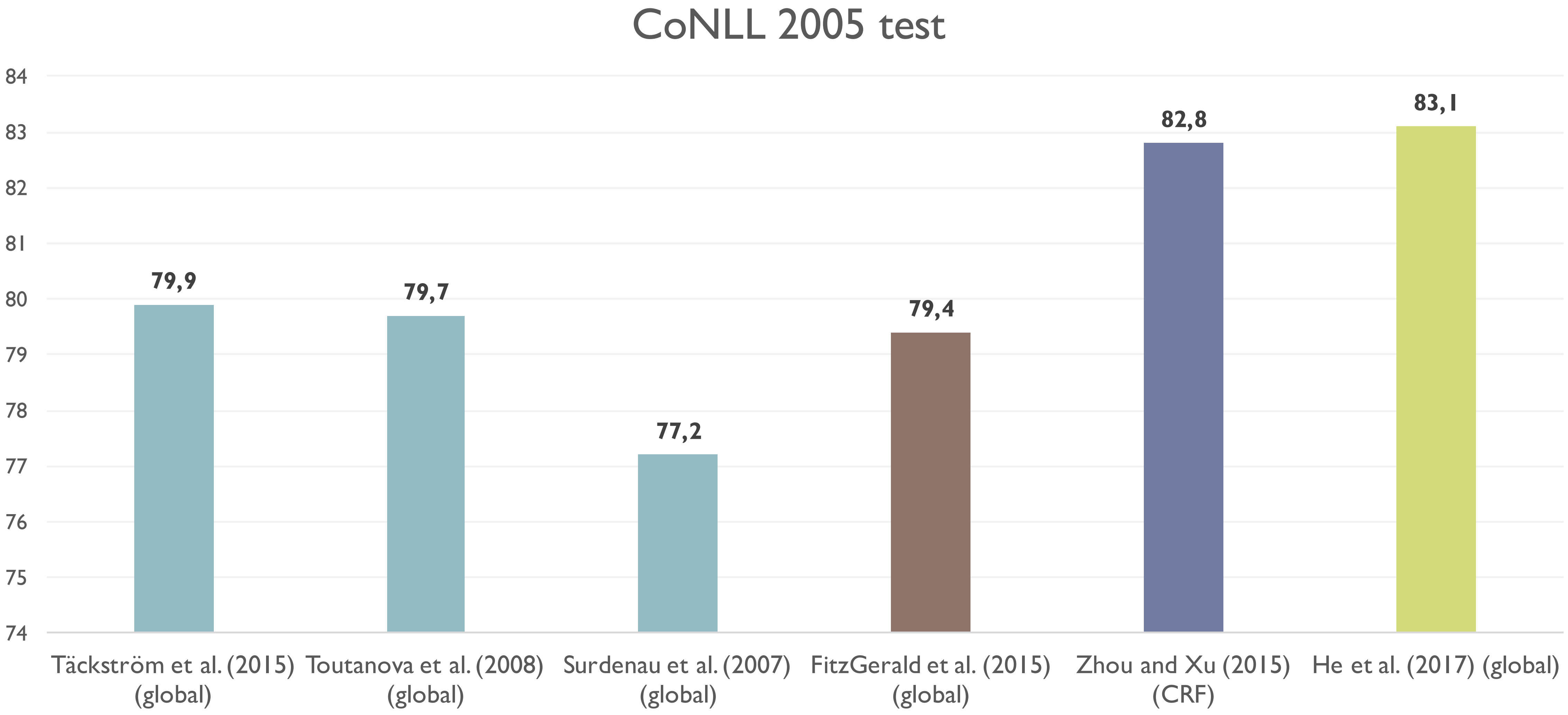


He et al., 2017: Decoding

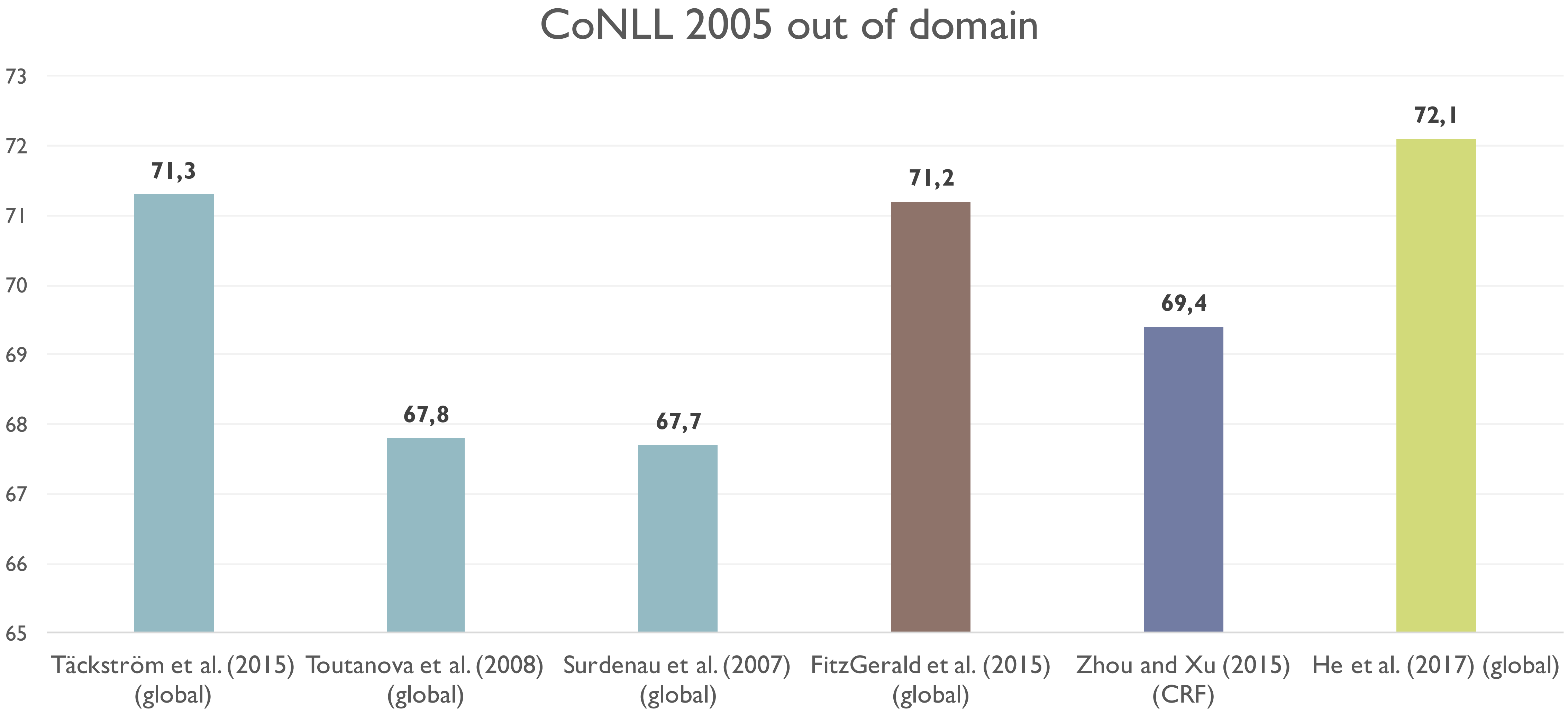
- A* decoding algorithm
  - BIO constraint
  - Continuation constraint
  - Uniqueness core roles
  - Reference constraint
  - Syntactic constraint



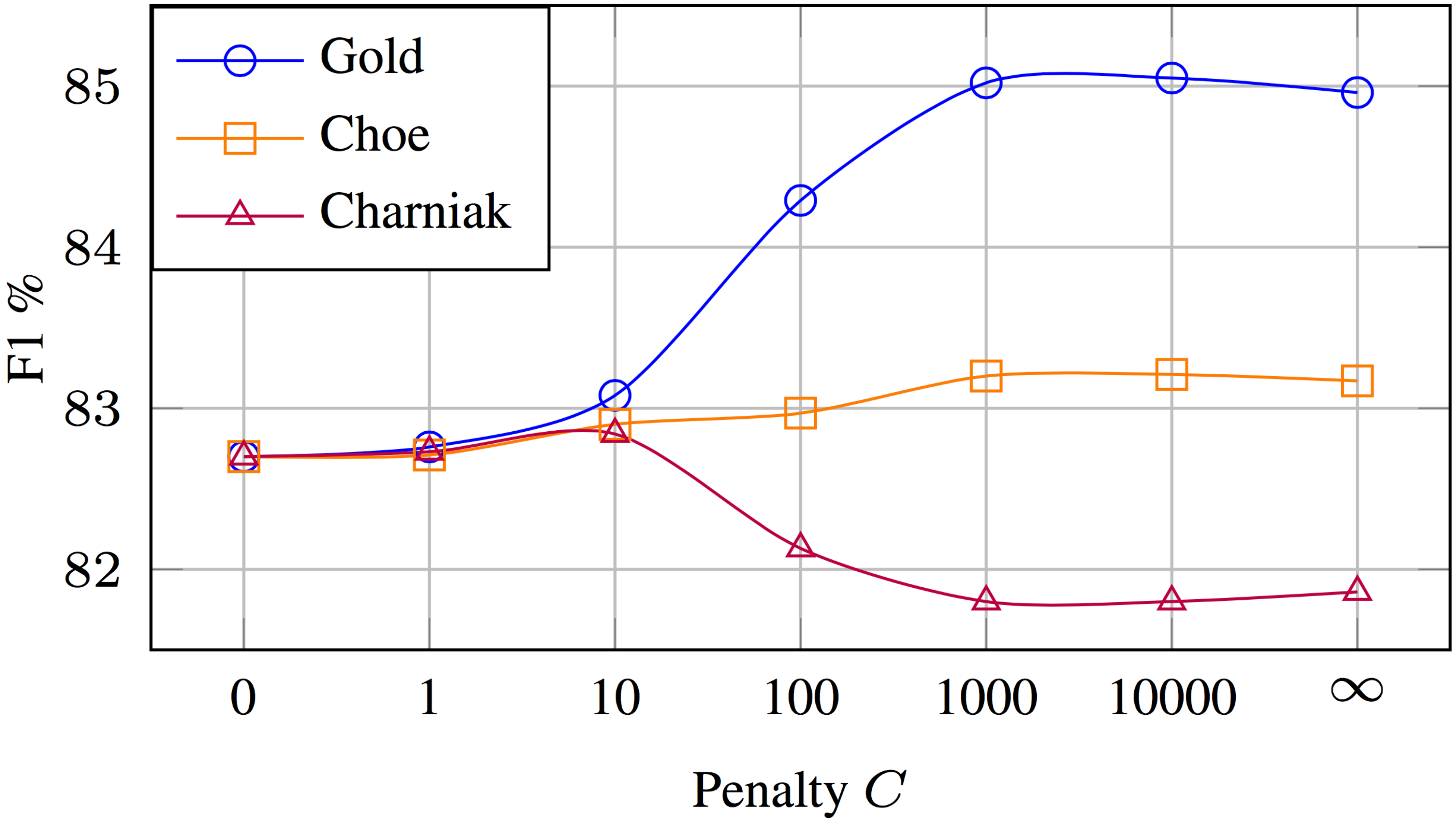
He et al., 2017: Results



He et al., 2017: Results



He et al., 2017: Analysis syntactic constraints



He et al., 2017

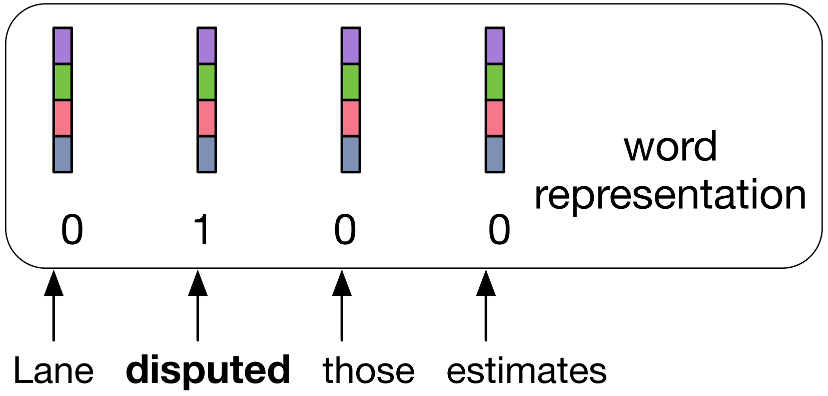
- No syntax
- Super minimal word representation
- Exploit at best the representational power of NN
  - Highway networks
  - Recurrent dropout

Marcheggiani et al., 2017

- Dependency-based SRL
- Shallow syntactic information (POS tags)
- Intuitions from syntactic dependency parsing
- Local classifier

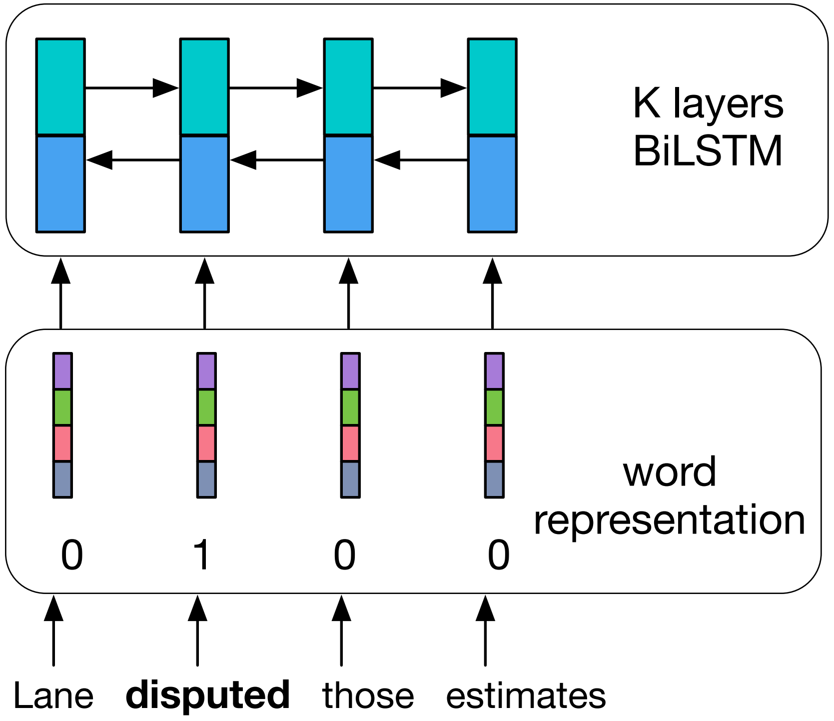
Marcheggiani et al., 2017: Word encoding

- Pretrained word embedding
- Randomly initialized embedding
- Randomly initialized embedding of POS tags
- Embeddings of the predicate lemmas
- Predicate flag



Marcheggiani et al., 2017: Sentence encoding

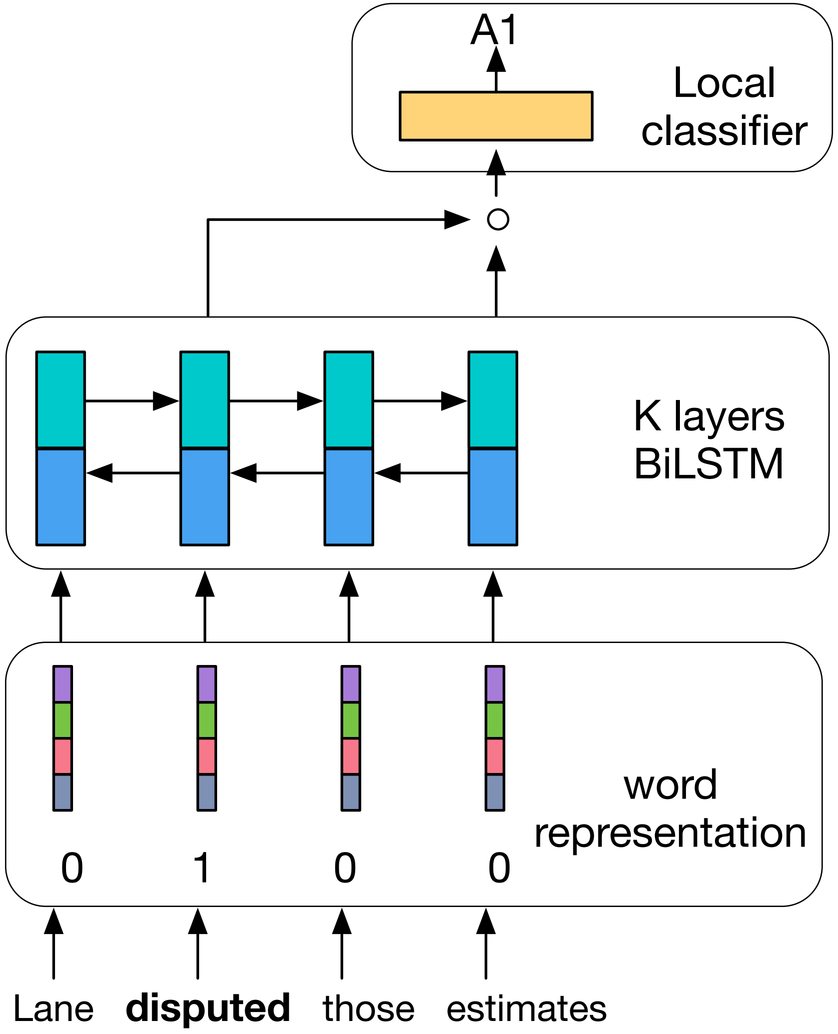
- Standard (non-snake) BI-LSTM
  - Forward LSTM encode left context
  - Backward LSTM encode right context
  - Forw. and Backw. states are concatenated



Marcheggiani et al., 2017: Decoding

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

Concatenation of argument and predicate states
[Kiperwasser and Goldberg, 2016]



Marcheggiani et al., 2017: Decoding

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

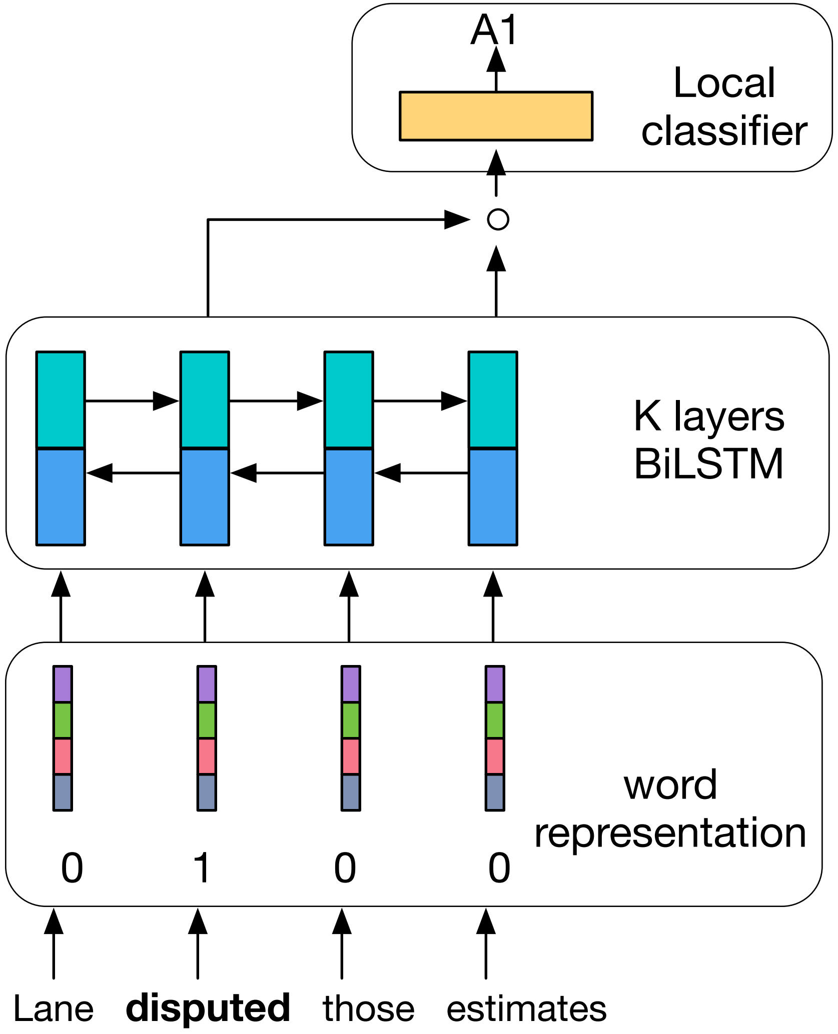
Concatenation of argument and predicate states [Kiperwasser and Goldberg, 2016]

$$W_{l,r} = ReLU(U(q_l \circ q_r))$$

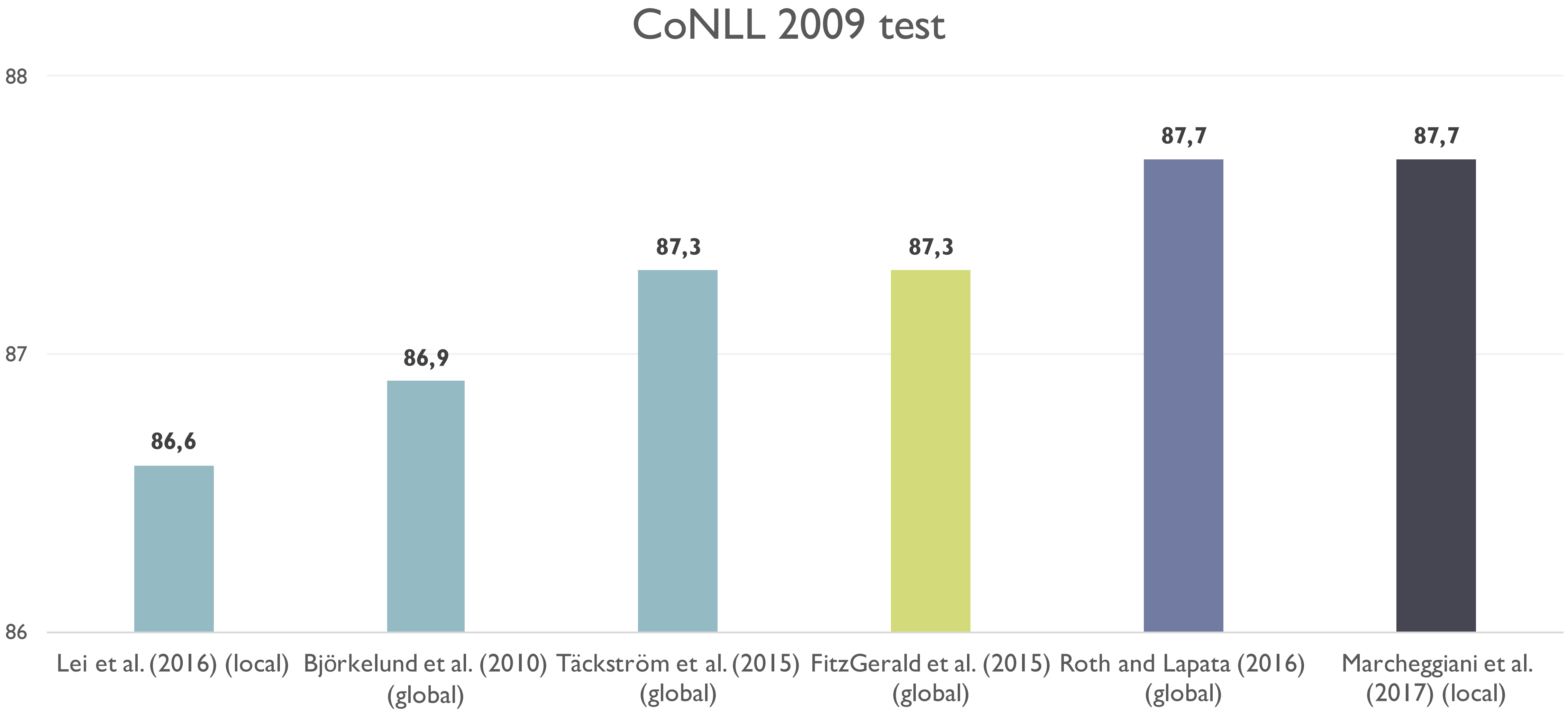
Predicate lemma embedding

Role embedding

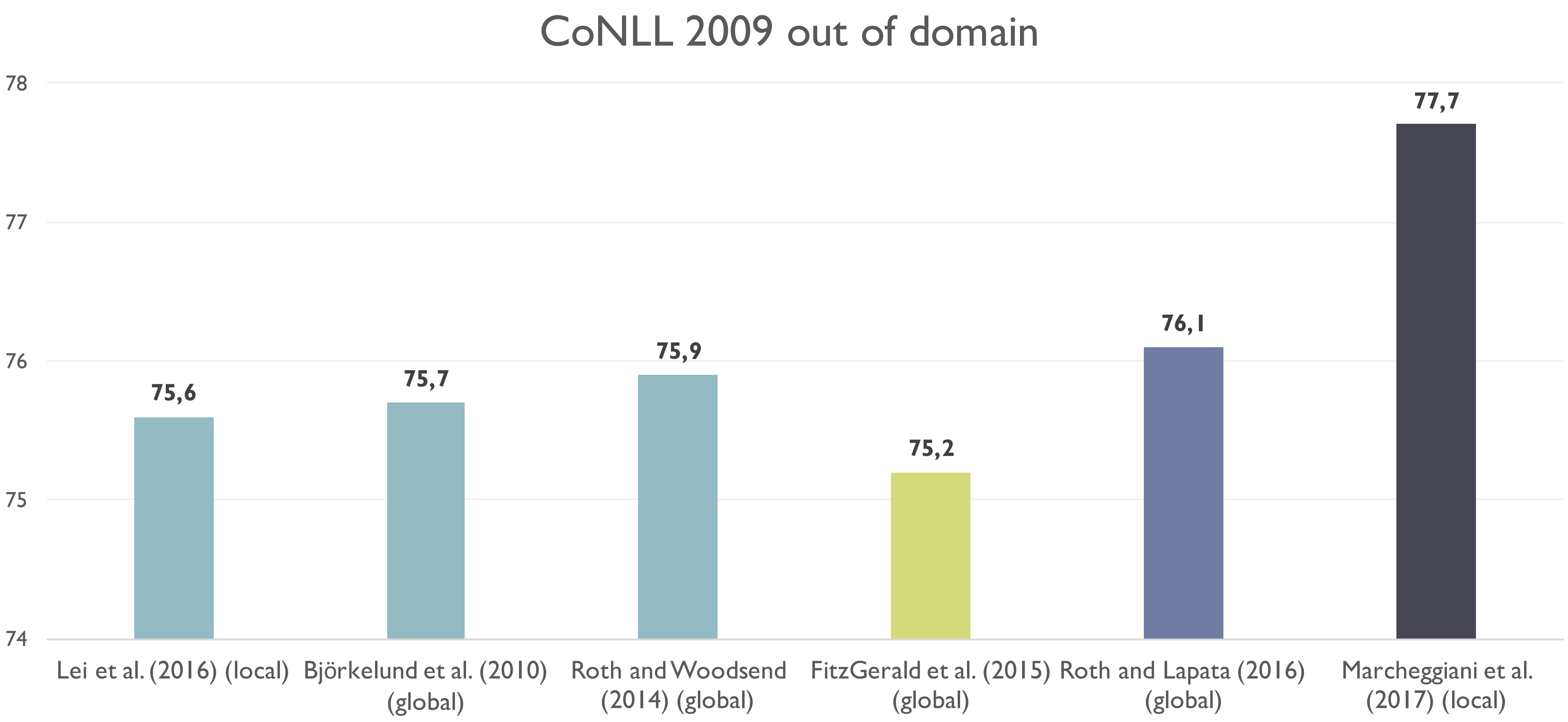
Fitzgerald et al. 2015



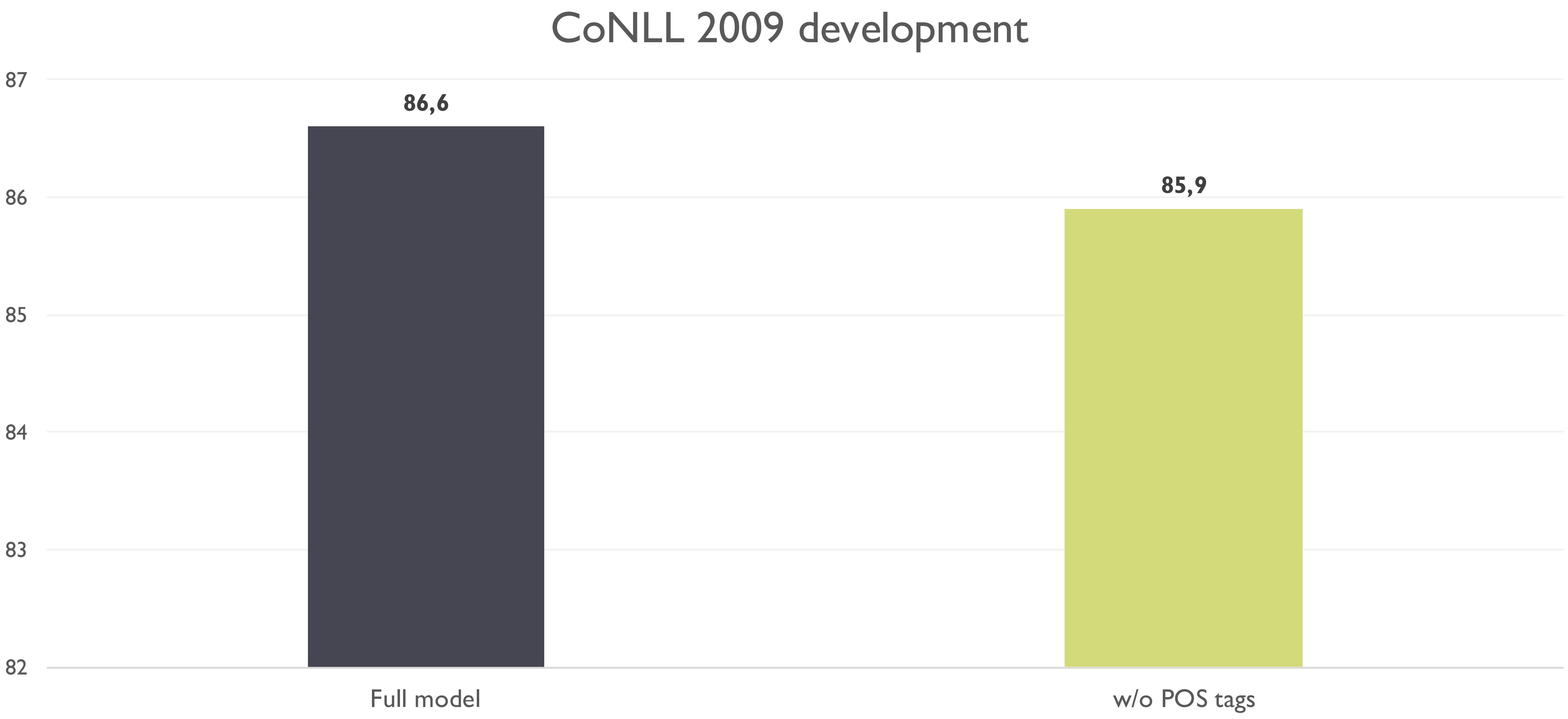
Marcheggiani et al., 2017: Results



Marcheggiani et al., 2017: Results



Marcheggiani et al., 2017: Ablation study



Marcheggiani et al., 2017

- Little bit of syntax (POS tags)
- More sophisticated word representation
- Fast local classifier conditioned on predicate representation

Outline: the fall and rise of syntax in SRL

- Early SRL framework
- Symbolic approaches + Neural networks
- Syntax-agnostic neural methods
- **Syntax-aware neural methods (syntax strikes back!)**

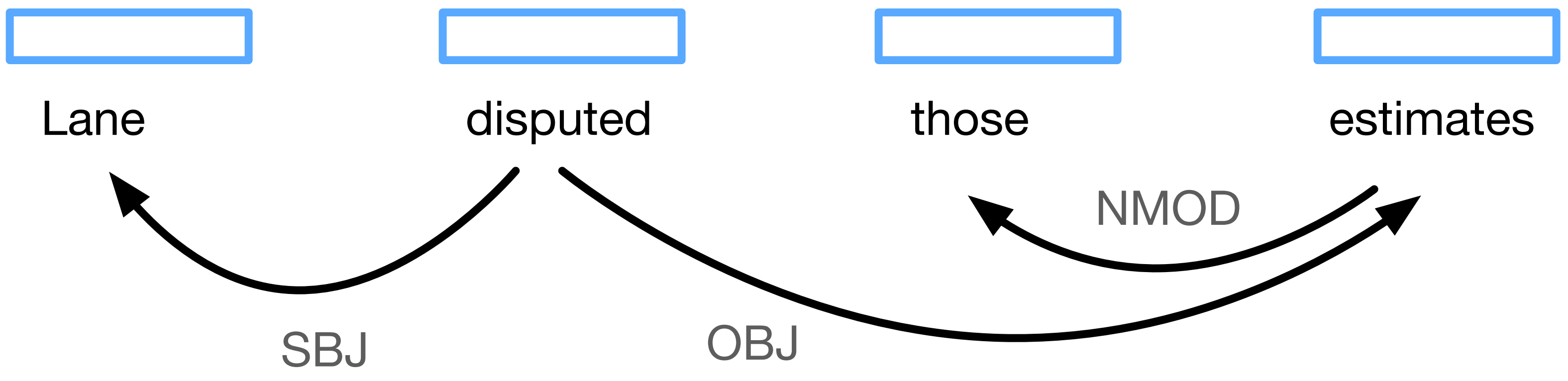
Is syntax important for semantics?

- POS tags are beneficial [Marcheggiani et al., 2017]
- Gold syntax is beneficial (but hard to encode) [He at al., 2017]
- Encoding syntax with Graph Convolutional Networks
  - [Marcheggiani and Titov, 2017]

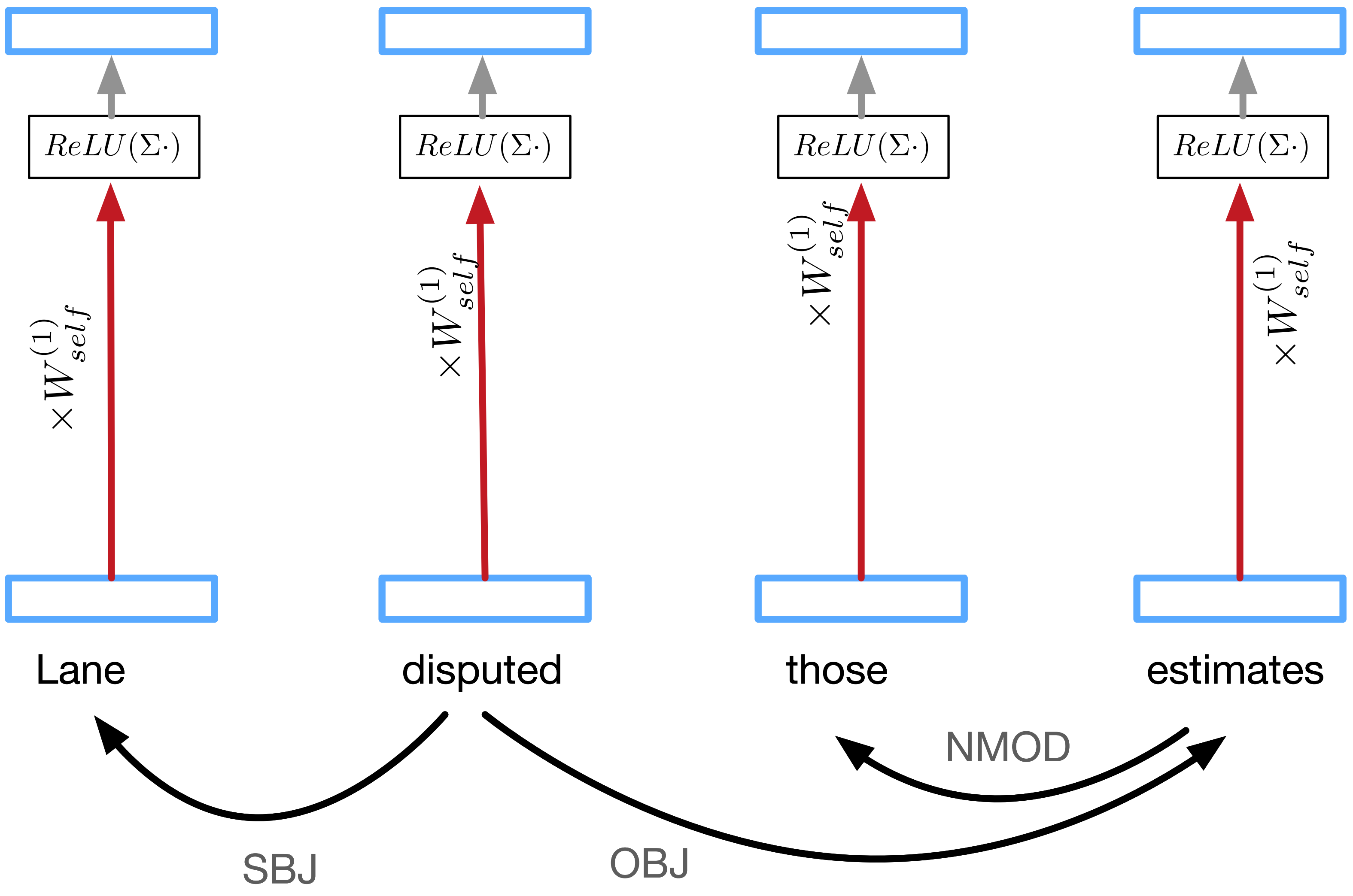
Marcheggiani and Titov, 2017

- Word encoding [Marcheggiani et. al, 2017]
- Sentence encoding with BiLSTM [Marcheggiani et. al, 2017]
- Syntax encoding with Graph Convolutional Networks (GCN)
  - [Kipf and Welling, 2016]
  - Each word is enriched with the representation of its syntactic neighborhood
- Local classifier [Marcheggiani et. al, 2017]

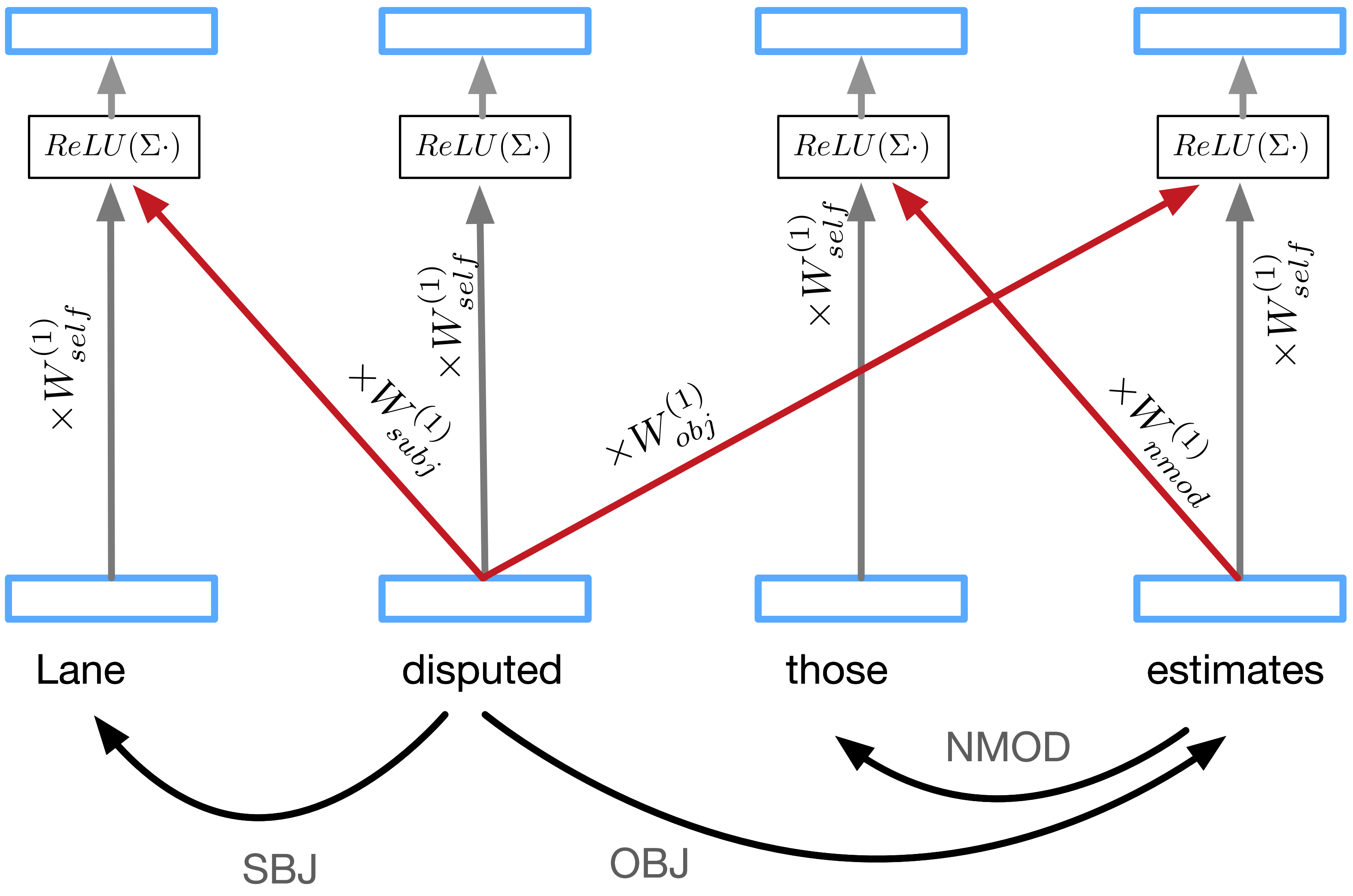
Marcheggiani and Titov, 2017: Syntactic GCN example



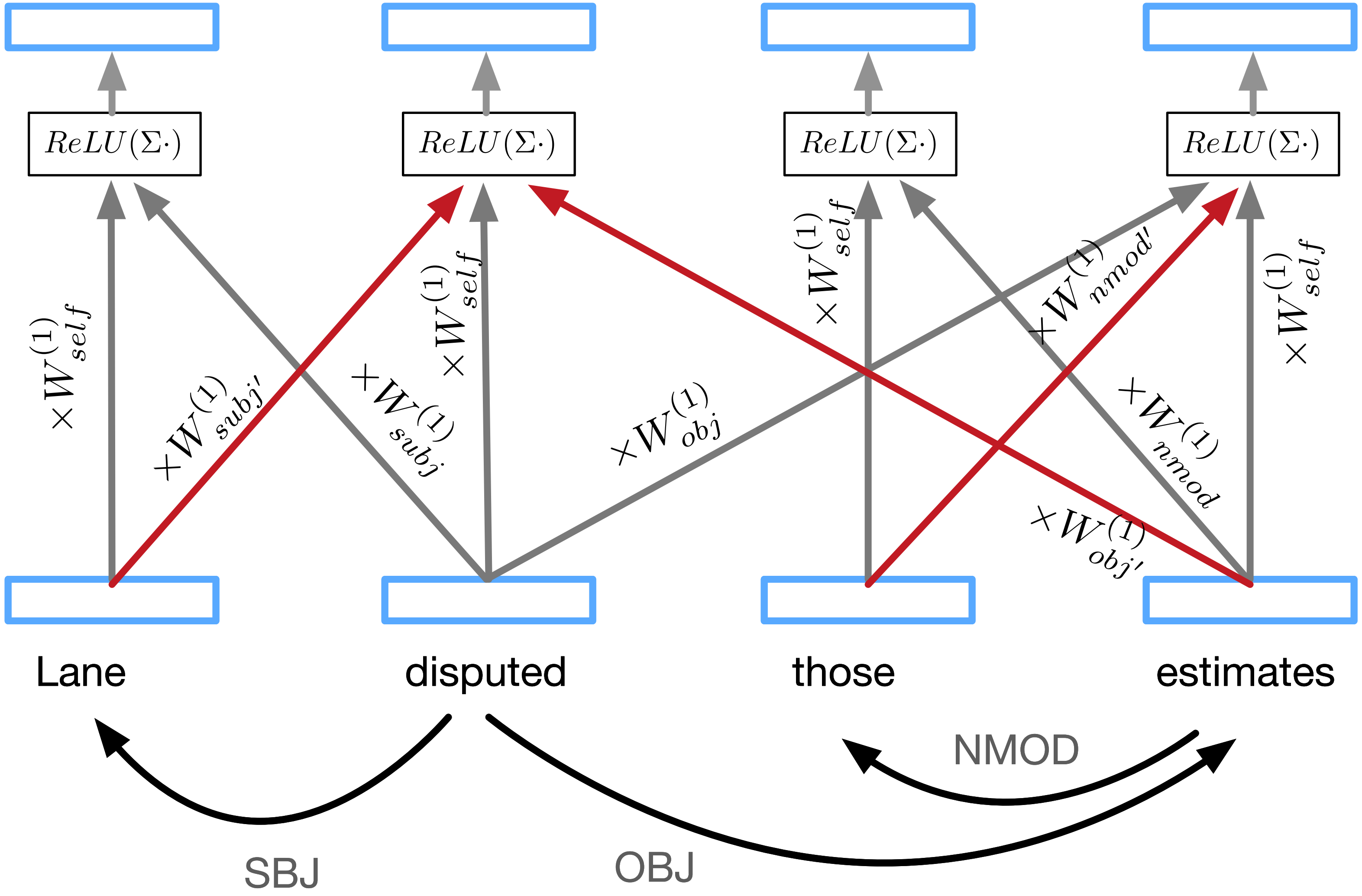
Marcheggiani and Titov, 2017: Syntactic GCN example



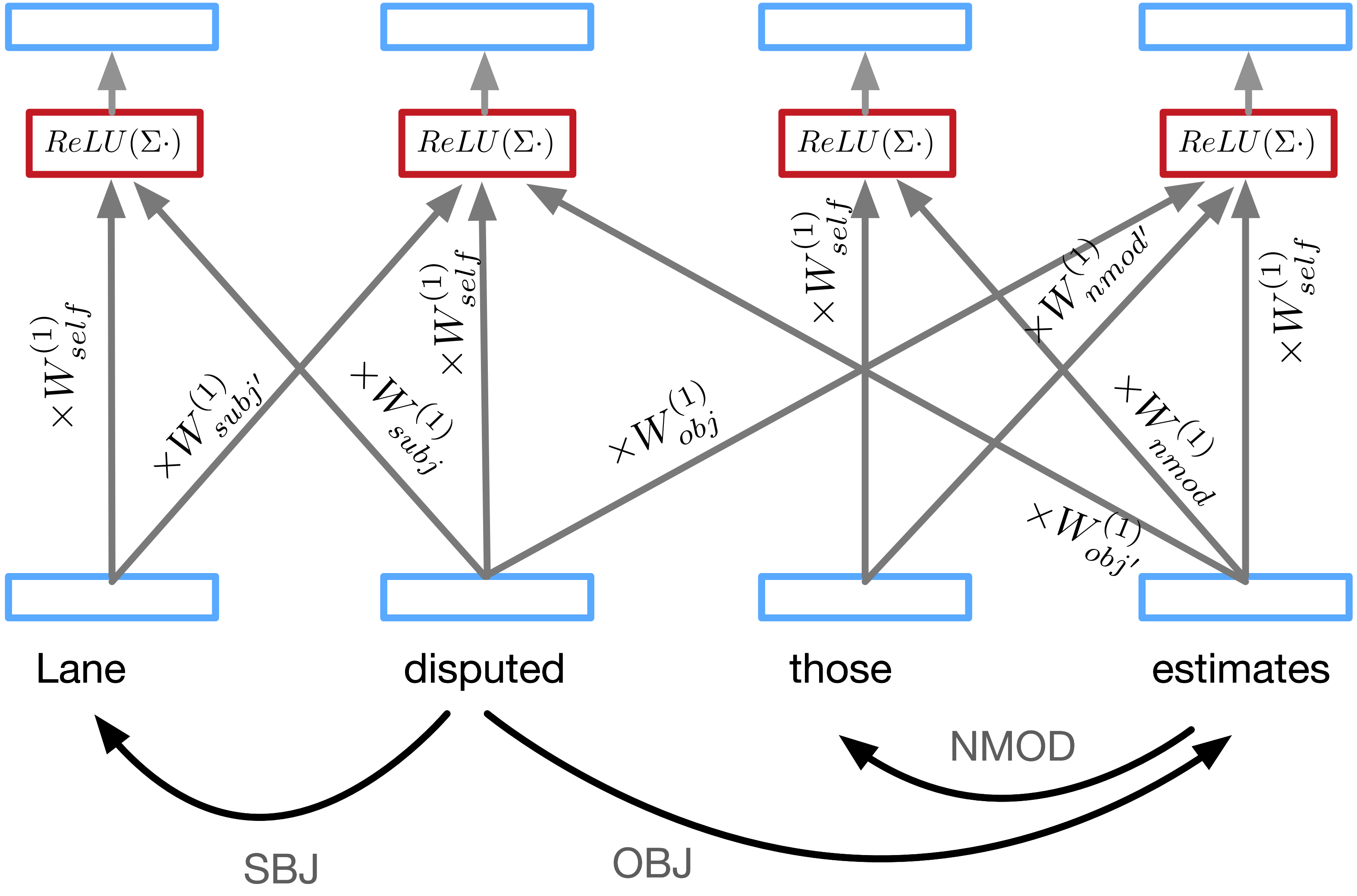
Marcheggiani and Titov, 2017: Syntactic GCN example



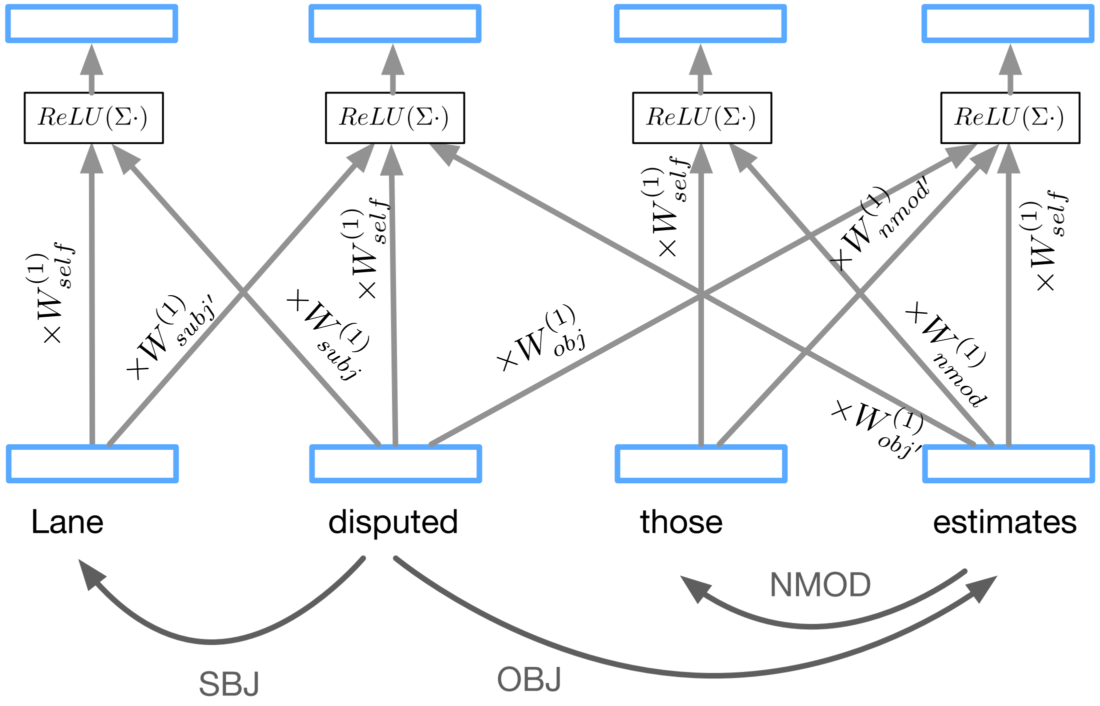
Marcheggiani and Titov, 2017: Syntactic GCN example



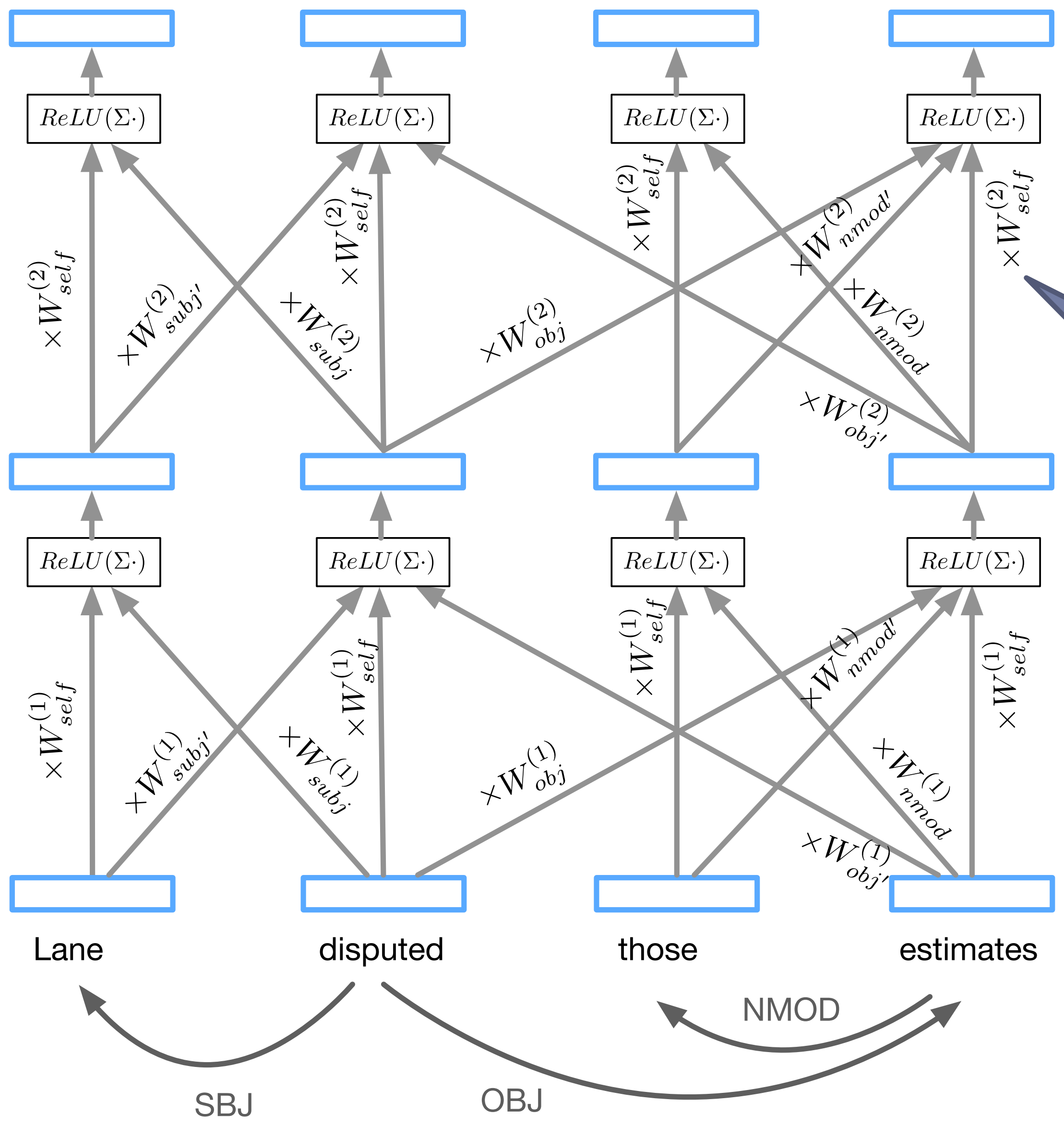
Marcheggiani and Titov, 2017: Syntactic GCN example



Marcheggiani and Titov, 2017: Syntactic GCN example



Marcheggiani and Titov, 2017: Syntactic GCN example



Stacking GCNs widens the syntactic neighborhood

Marcheggiani and Titov, 2017: Syntactic GCN

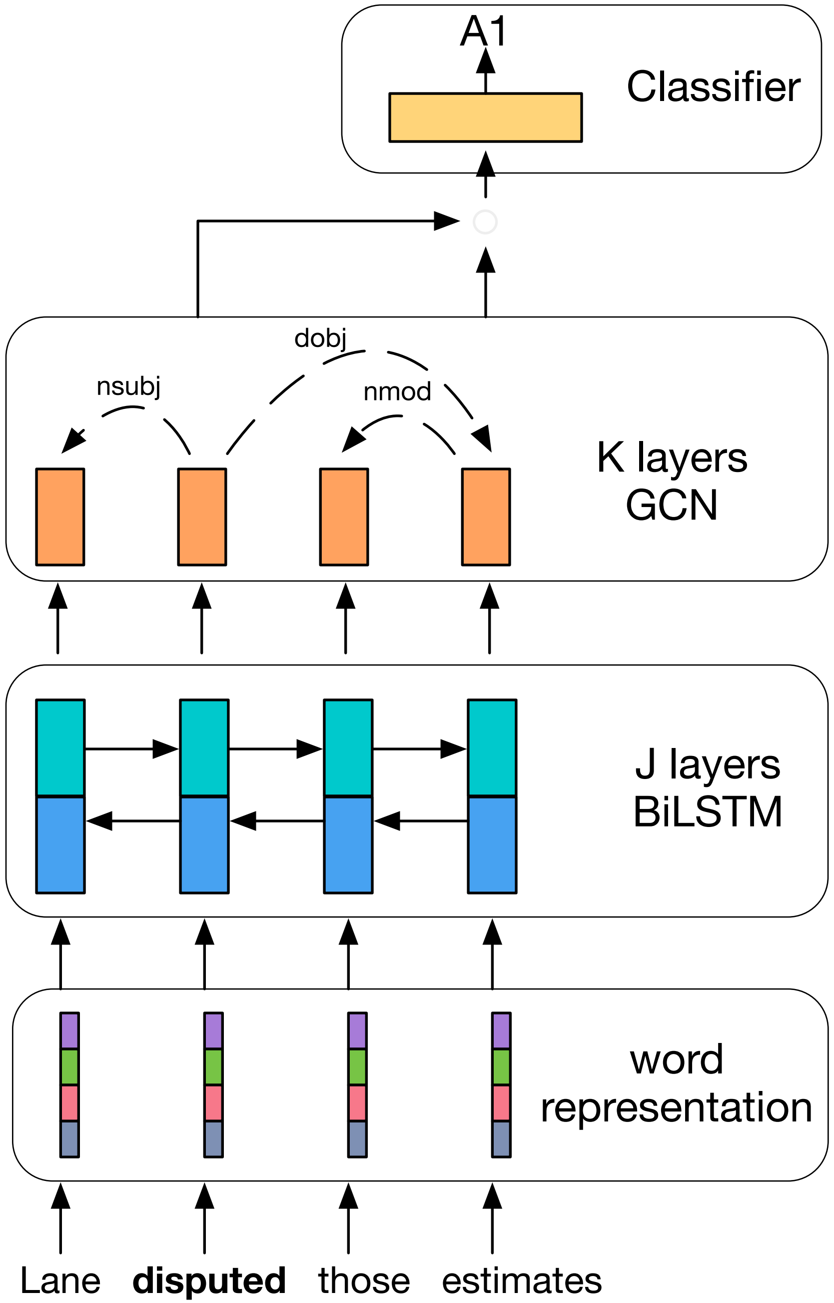
Sum over the syntactic neighborhood

Each node is transformed according to label and direction

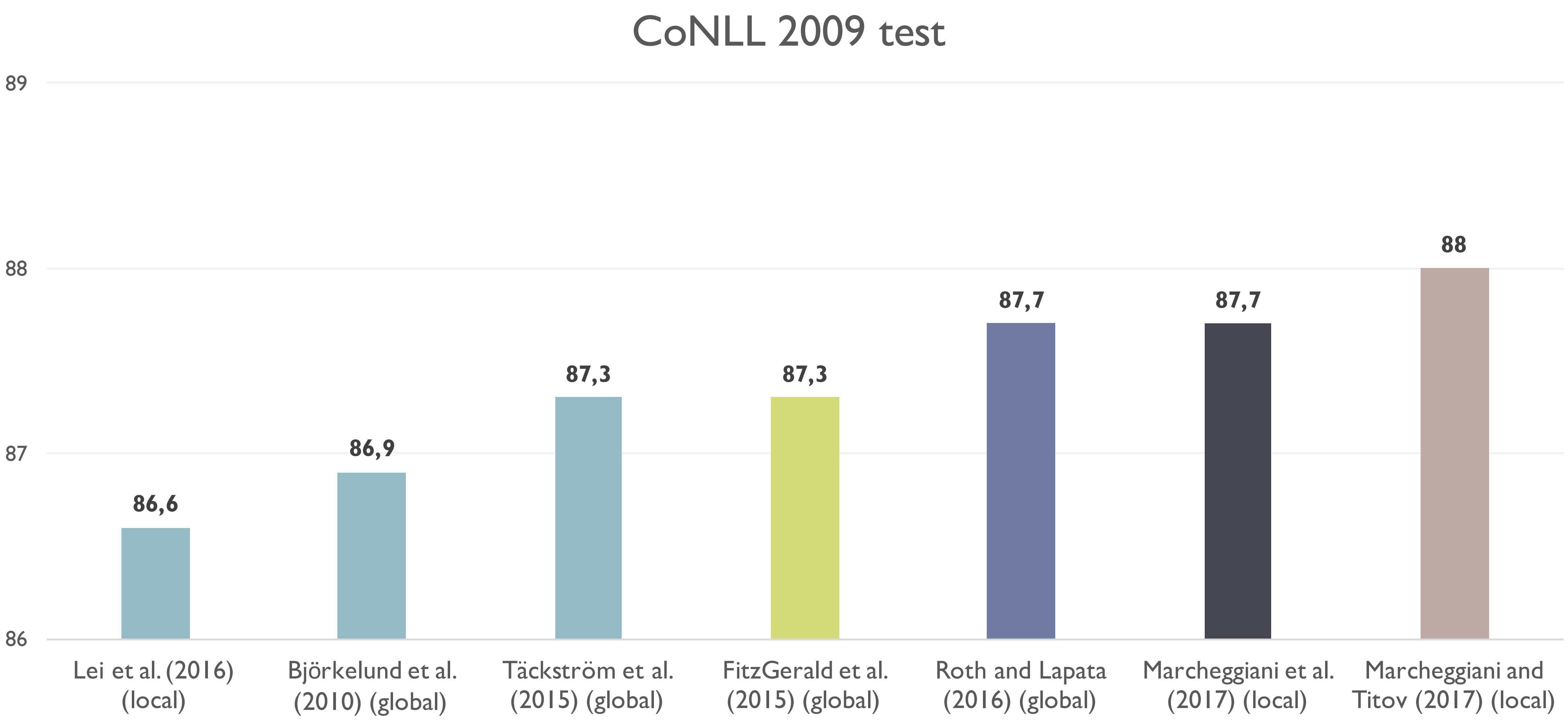
$$h_v^{(k+1)} = ReLU\left(\sum_{u \in \mathcal{N}(v)} \boxed{W_{L(u,v)}^{(k)} h_u^{(k)} + b_{L(u,v)}^{(k)}}\right)$$

Marcheggiani and Titov, 2017: Architecture

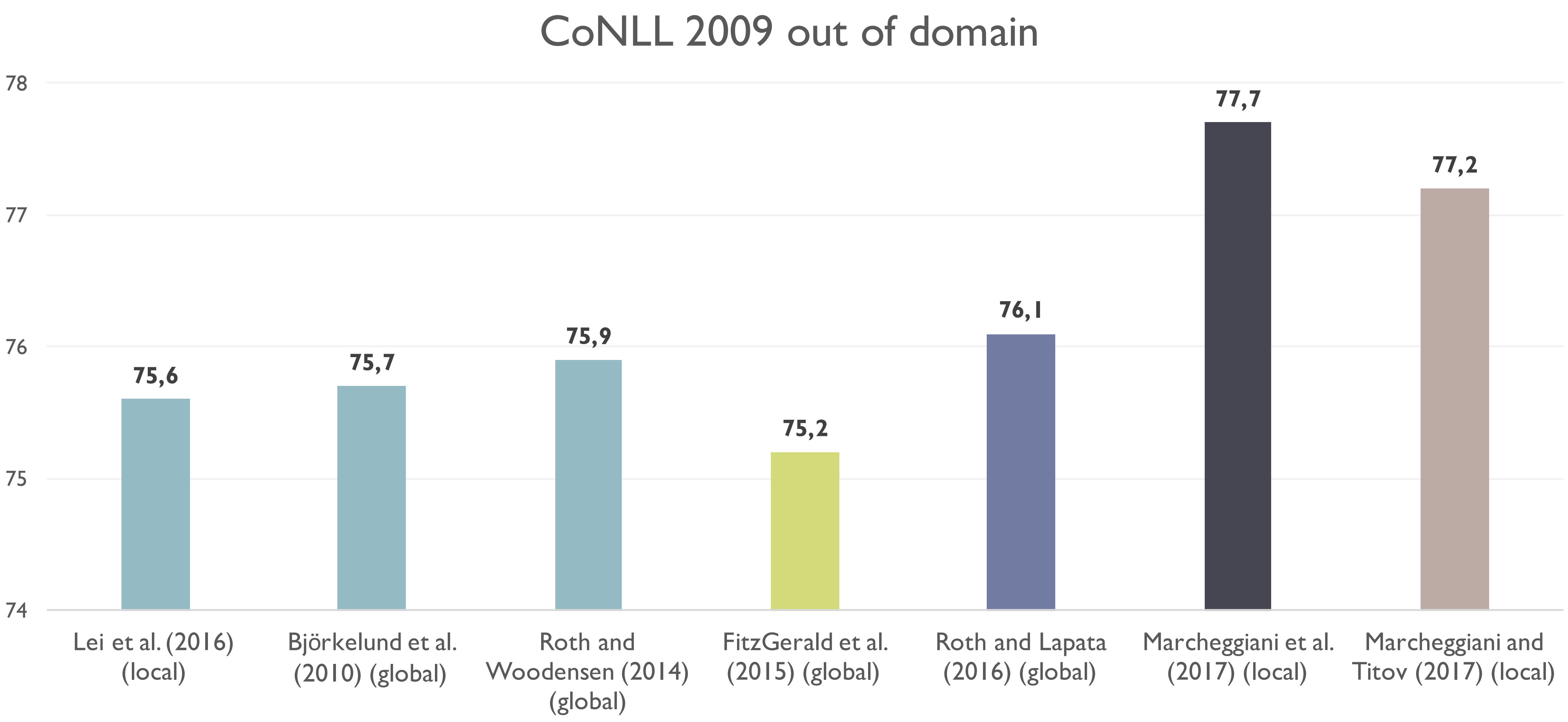
- Same architecture of [Marcheggiani et al., 2017]
- Syntactic GCN after BiLSTM encoder
  - Skip connections
  - Longer dependencies are captured



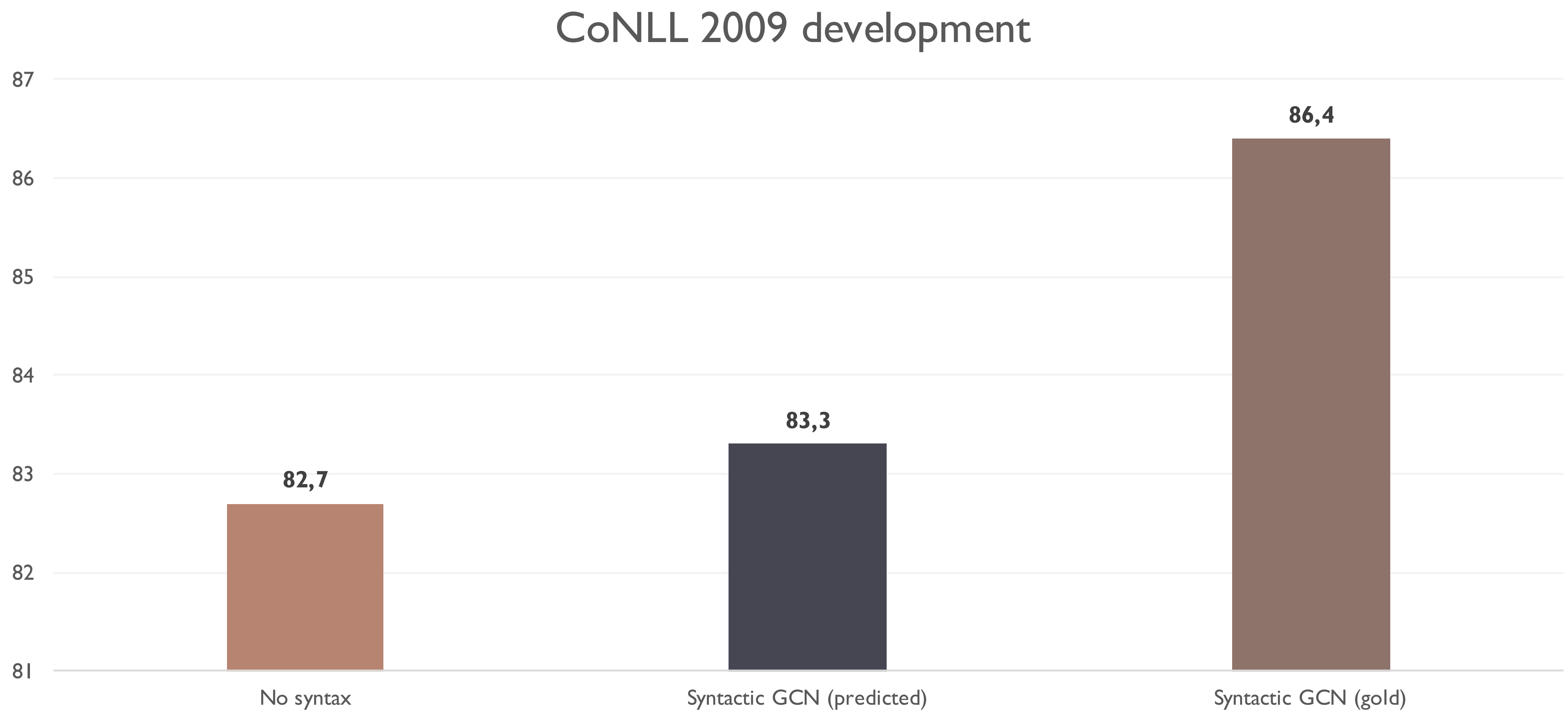
Marcheggiani and Titov, 2017: Results



Marcheggiani and Titov, 2017: Results



Marcheggiani and Titov, 2017: Analysis



Marcheggiani and Titov, 2017

- Encoding structured prior linguistic knowledge in NN
  - Syntax
  - Semantics
  - Coreference
  - Discourse
- Complement LSTM with skip connections for long dependencies

Conclusion

- We can live without syntax (out of domain)

Conclusion

- We can live without syntax (out of domain)
- But life with syntax is better

Conclusion

- We can live without syntax (out of domain)
- But life with syntax is better
  - and the better the syntax (parsers) the better our semantic role labeler

Conclusion

- We can live without syntax (out of domain)
- But life with syntax is better
  - and the better the syntax (parsers) the better our semantic role labeler
- What's the (present) future?

Conclusion

- We can live without syntax (out of domain)
- But life with syntax is better
  - and the better the syntax (parsers) the better our semantic role labeler

- What's the (present) future?
  - Multi-task learning
  - Swayamdipta et al. (2017) frame-semantic parsing + syntax
  - Peng et al. (2017) multi-task on different semantic formalisms

Conclusion

- We can live without syntax (out of domain)
- But life with syntax is better
  - and the better the syntax (parsers) the better our semantic role labeler

- What's the (present) future?
  - Multi-task learning
  - Swayamdipta et al. (2017) frame-semantic parsing + syntax
  - Peng et al. (2017) multi-task on different semantic formalisms

- Neural networks work (I kid you not) …

Conclusion

- We can live without syntax (out of domain)
- But life with syntax is better
  - and the better the syntax (parsers) the better our semantic role labeler

- What's the (present) future?
  - Multi-task learning
  - Swayamdipta et al. (2017) frame-semantic parsing + syntax
  - Peng et al. (2017) multi-task on different semantic formalisms

- Neural networks work (I kid you not) …
- … but we do have (a lot of) linguistic prior knowledge…

Conclusion

- We can live without (treebank) syntax (out of domain)
- But life with syntax is better
  - and the better the syntax (parsers) the better our semantic role labeler

- What's the (present) future?
  - Multi-task learning
  - Swayamdipta et al. (2017) frame-semantic parsing + syntax
  - Peng et al. (2017) multi-task on different semantic formalisms

- Neural networks work (I kid you not) …
- … but we do have (a lot of) linguistic prior knowledge…
- … and it is time to use it again.

References

- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106.
- Nianwen Xue and Martha Palmer. 2004. Calibrating features for semantic role labeling. In *Proceedings of EMNLP*.
- Sameer Pradhan, Kadri Hacioglu, Valerie Krugler, Wayne Ward, James H Martin, and Daniel Jurafsky. 2005. Support vector learning for semantic argument classification. *Machine Learning*, 60(1-3):11– 39.
- Kristina Toutanova, Aria Haghighi, and Christopher D Manning. 2008. A global joint model for semantic role labeling. *Computational Linguistics*, 34(2):161–191.
- Vasin Punyakanok, Dan Roth, and Wen-tau Yih. 2008. The importance of syntactic parsing and inference in semantic role labeling. *Computational Linguistics*, 34(2):257–287.

References

- Oscar Täckström, Kuzman Ganchev, and Dipanjan Das. 2015. Efficient inference and structured learning for semantic role labeling. *Transactions of the Association for Computational Linguistics*, 3:29–41.
- Anders Björkelund, Bernd Bohnet, Love Hafdell, and Pierre Nugues. 2010. A high-performance syntactic and semantic dependency parser. In *Proceedings of ICCL: Demonstrations*.
- Anders Björkelund, Love Hafdell, and Pierre Nugues. 2009. Multilingual semantic role labeling. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*.
- Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.
- Nicholas FitzGerald, Oscar Täckström, Kuzman Ganchev, and Dipanjan Das. 2015. Semantic role labeling with neural network factors. In *Proceedings EMNLP*.

References

- Michael Roth and Mirella Lapata. 2016. Neural semantic role labeling with dependency path embeddings. In *Proceedings of ACL*.
- Tao Lei, Yuan Zhang, Lluîs Márquez, Alessandro Moschitti, and Regina Barzilay. 2015. High-order low-rank tensors for semantic role labeling. In *Proceedings of NAACL*.
- Mihai Surdeanu, Lluîs Màrquez, Xavier Carreras, and Pere Comas. 2007. Combination strategies for semantic role labeling. *Journal of Artificial Intelligence Research*, 29:105–151.
- Ronan Collobert, Jason Weston, Le´on Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *The Journal of Machine Learning Re- search*, 12:2493–2537.
- Michael Roth and Kristian Woodsend. 2014. Composition of word representations improves semantic role labelling. In *Proceedings of EMNLP*.

References

- Jie Zhou and Wei Xu. 2015. End-to-end learning of semantic role labeling using recurrent neural networks. In *Proceedings of ACL.*
- John Lafferty, Andrew McCallum, and Fernando Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of ICML.*
- Luheng He, Kenton Lee, Mike Lewis, and Luke Zettlemoyer. 2017. Deep Semantic Role Labeling: What Works and What's Next. In *Proceedings of ACL.*
- Yarin Gal and Zoubin Ghahramani. 2016. A theoretically grounded application of dropout in recurrent neural networks. In *Proceedings of NIPS.*
- Rupesh K Srivastava, Klaus Greff, and Jürgen Schmidhuber. 2015. Training very deep networks. In *Proceedings of NIPS.*

References

- Diego Marcheggiani, Anton Frolov, and Ivan Titov. 2017. A simple and accurate syntax-agnostic neural model for dependency-based semantic role labeling. In *Proceedings of CoNLL.*
- Eliyahu Kiperwasser and Yoav Goldberg. 2016. Simple and accurate dependency parsing using bidirectional LSTM feature representations. In *Transactions of the Association for Computational Linguistics*.
- Thomas N. Kipf and Max Welling. 2017. Semi- supervised classification with graph convolutional networks. In *Proceedings of ICLR.*
- Diego Marcheggiani, and Ivan Titov. 2017. Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling. In *Proceedings of EMNLP.*
- Swabha Swayamdipta, Sam Thomson, Chris Dyer, and Noah A. Smith. 2017. Frame-Semantic Parsing with Softmax-Margin Segmental RNNs and a Syntactic Scaffold. arXiv preprint.

References

- Hao Peng, Sam Thomson, and Noah A. Smith. 2017. Deep multitask learning for semantic dependency parsing. In *Proceedings of ACL*.