University of Colorado Boulder

University of California Santa Cruz

Charles University

IBM Research

Apple

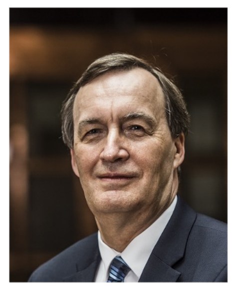
Brandeis University

Tutorial

Meaning Representations for Natural Languages: Design, Models and Applications


Julia Bonn


Jeffrey Flanigan


Jan Hajič


Ishan Jindal


Yunyao Li


Nianwen Xue

IJCAI/2023 MACAO

Tutorial Outline

Morning Session

- Part 1: Introduction – Julia Bonn
- Part 2a: Common Meaning Representations:
  - AMR – Julia Bonn
  - Other Meaning Representations – Jan Hajič
- **Break**
- Part 2b: Common Meaning Representations
  - UMR – Nianwen Xue

Tutorial Outline

Afternoon Session

- Part 3: Modeling Meaning Representation:
    - SRL – Ishan Jindal
    - AMR – Jeff Flanigan
- **Break**
- Part 4: Applying Meaning Representations
  – Yunyao Li, Jeff Flanigan

Part 5: Open Questions and Future Work
– Nianwen Xue

Meaning Representations for Natural Languages Tutorial Part 3a

Modeling Meaning Representation: Semantic Role Labeling (SRL)

Julia Bonn, Jeffrey Flanigan, Jan Hajic, **Ishan Jindal**, Yunyao Li, Nianwen Xue

Semantic Role Labeling (SRL)

Who did what to **whom**, **when**, **where** and **how**?

(Palmar, 1990; Gildea and Jurafsky, 2000; Màrquez et al., 2008)

Semantic Role Labeling (SRL)

Derek **broke** **the window with a hammer to** **escape**

1 Predicate Identification — Identify all predicates in the sentence

Semantic Role Labeling (SRL)

Derek **broke** **the window with a hammer to escape.**

break.01

1. **Predicate Identification** — Identify all predicates in the sentence
2. **Sense Disambiguation** — Classify sense of each predicate

English Propbank	FrameNet Frame	VerbNet
break.01, break	Breaking_apart	Break-45.1
A0: breaker	Pieces	Agent
A1: thing broken	Whole	Patient
A2: instrument	Criterion	Instrument
A3: pieces	Manner	Result
A4: arg1 broken away from what?	Means	
	Place...	

Semantic Role Labeling (SRL)

Derek **broke** **the window** **with a hammer** **to escape**

break.01

1. **Predicate Identification** — Identify all predicates in the sentence
2. **Sense Disambiguation** — Classify sense of each predicate
3. **Argument Identification** — Find all roles of each predicate

Argument identification can either be
- Identification of span, (span SRL) OR
- Identification of head (dependency SRL)

Semantic Role Labeling (SRL)

Derek	broke	the window	with a hammer	to escape
Breaker	break.01	thing broken	instrument	Purpose

1. **Predicate Identification** — Identify all predicates in the sentence
2. **Sense Disambiguation** — Classify sense of each predicate
3. **Argument Identification** — Find all roles of each predicate
4. **Argument Classification** — Assign semantic label to each role

Semantic Role Labeling (SRL)

Derek	broke	the window	with a hammer	to escape
A0: Breaker	break.01	A1: thing broken	A2: instrument	AM-PRP: Purpose

1	Predicate Identification	Identify all predicates in the sentence
2	Sense Disambiguation	Classify sense of each predicate
3	Argument Identification	Find all roles of each predicate
4	Argument Classification	Assign semantic label to each role

If using PropBank

Semantic Role Labeling (SRL)

Derek	broke	the window	with a hammer	to escape
A0: Breaker	break.01	A1: thing broken	A2: instrument	AM-PRP: Purpose

1	Predicate Identification	Identify all predicates in the sentence
2	Sense Disambiguation	Classify sense of each predicate
3	Argument Identification	Find all roles of each predicate
4	Argument Classification	Assign semantic label to each role
5	Global Optimization	Global constraints (predicates and arguments)

Outline

- **Early SRL approaches [< 2017]**
- **Typical neural SRL model components**
  - Performance analysis
- **Syntax-aware neural SRL models**
  - What, When and Where?
  - Performance analysis
  - How to incorporate Syntax?
- **Syntax-agnostic neural SRL models**
  - Performance Analysis
  - Do we really need syntax for SRL?
  - Are high quality contextual embedding enough for SRL task?
- **Practical SRL systems**
  - Should we rely on this pipelined approach?
    - End-to-end SRL systems
  - Can we jointly predict dependency and span?
- **More recent approaches**
  - Handling low-frequency exceptions
  - Incorporate semantic role label definitions
  - SRL as MRC task
- **Practical SRL system evaluations**
  - Are we evaluating SRL systems correctly?
- **Conclusion**

Outline

- ❑ **Early SRL approaches**
- ❑ **Typical neural SRL model components**
  - ❑ Performance analysis
- ❑ **Syntax-aware neural SRL models**
  - ❑ What, When and Where?
  - ❑ Performance analysis
  - ❑ How to incorporate Syntax?
- ❑ **Syntax-agnostic neural SRL models**
  - ❑ Performance Analysis
  - ❑ Do we really need syntax for SRL?
  - ❑ Are high quality contextual embedding enough for SRL task?
- ❑ **Practical SRL systems**
  - ❑ Should we rely on this pipelined approach?
    - ❑ End-to-end SRL systems
  - ❑ Can we jointly predict dependency and span?
- ❑ **More recent approaches**
  - ❑ Handling low-frequency exceptions
  - ❑ Incorporate semantic role label definitions
  - ❑ SRL as MRC task
- ❑ **Practical SRL system evaluations**
  - ❑ Are we evaluating SRL systems correctly?
- ❑ **Conclusion**

Early SRL Approaches

- 2 to 3 steps to obtain complete predicate-argument structure

- Predicate Identification
  - Generally considered as not a task, as all the existing SRL datasets provided Gold predicate location.

- Predicate sense disambiguation
  - Logistic Regression [Roth and Lapata, 2016]

- Argument Identification
  - Binary classifier [Pradhan et al., 2005; Toutanova et al., 2008]

- Role Labeling
  - Labeling is performed using a classifier (SVM, logistic regression)
  - Argmax over roles will result in a local assignment

- Global Optimization
  - 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]

- Requires Feature Engineering
  - Mostly Syntactic [Gildea and Jurafsky, 2002]

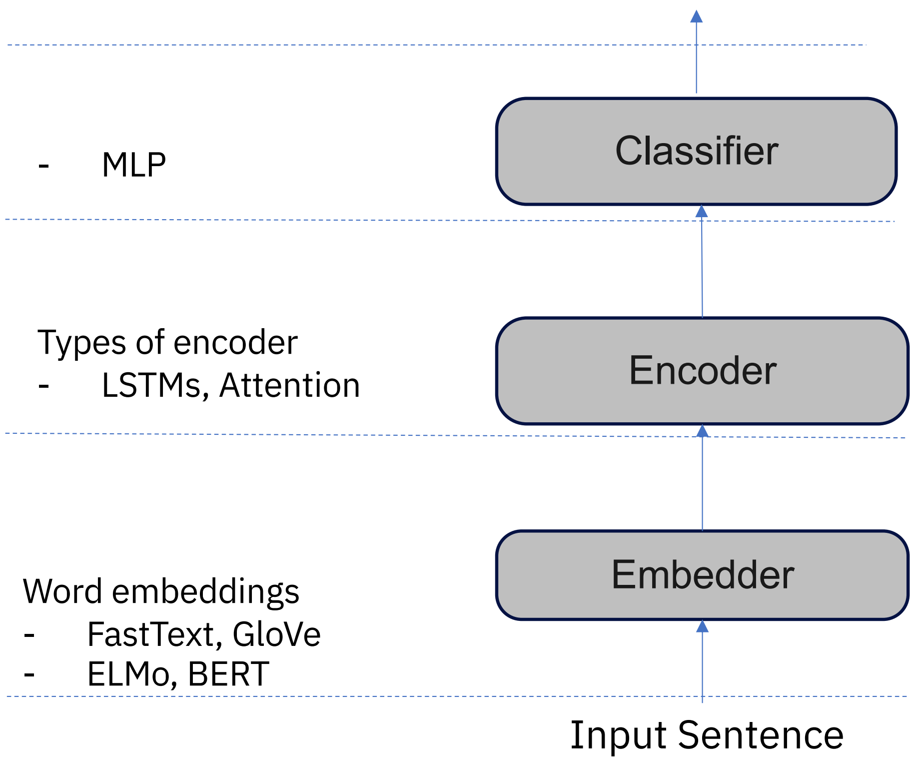
Outline

- ❑ **Early SRL approaches**
- ❑ **Typical neural SRL model components**
  - ❑ Performance analysis
- ❑ **Syntax-aware neural SRL models**
  - ❑ What, When and Where?
  - ❑ Performance analysis
  - ❑ How to incorporate Syntax?
- ❑ **Syntax-agnostic neural SRL models**
  - ❑ Performance Analysis
  - ❑ Do we really need syntax for SRL?
  - ❑ Are high quality contextual embedding enough for SRL task?
- ❑ **Practical SRL systems**
  - ❑ Should we rely on this pipelined approach?
    - ❑ End-to-end SRL systems
  - ❑ Can we jointly predict dependency and span?
- ❑ **More recent approaches**
  - ❑ Handling low-frequency exceptions
  - ❑ Incorporate semantic role label definitions
  - ❑ SRL as MRC task
- ❑ **Practical SRL system evaluations**
  - ❑ Are we evaluating SRL systems correctly?
- ❑ **Conclusion**

Typical Neural SRL Components

A typical neural SRL model contains three components

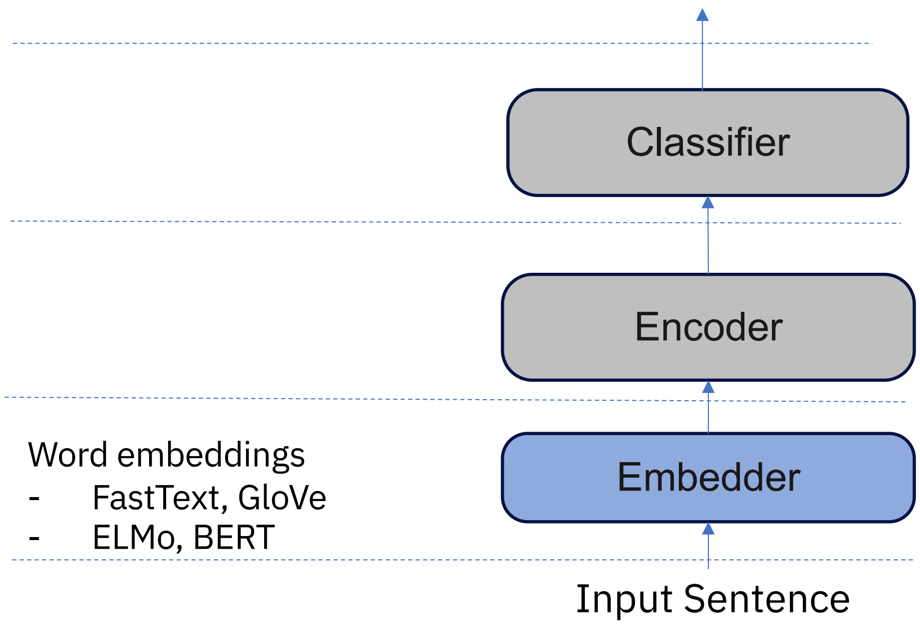
- Classifier
  - Assign a semantic role label to each token in the input sentence. [Local + Global]
- Encoder:
  - Encodes the context information to each token.
- Embedder:
  - Represent input token into continuous vector representation.



Neural SRL Components – Embedder

Sub-task: Argument Classification

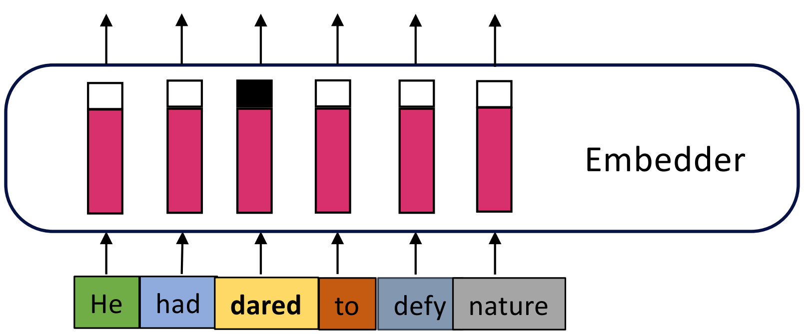
- Embedder:
  - Represent input token into continuous vector representation.



- Could be static or dynamic embeddings
- Could incorporate syntax information

- Usually, a binary flag
- 0 → represents no predicate
- 1 → represent predicate

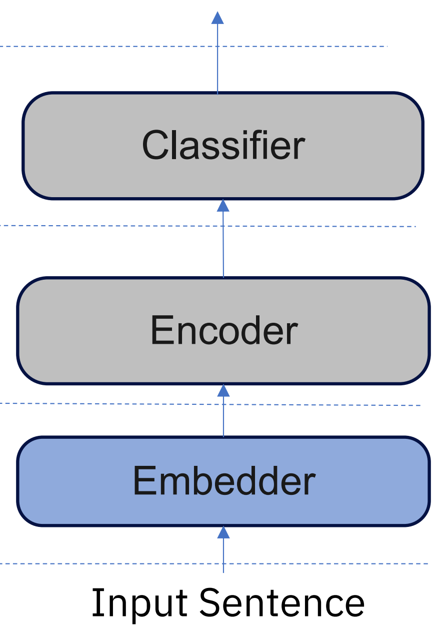
End-to-end systems do not include this flag



Neural SRL Components – Embedder

Sub-task: Argument Classification

- Embedder:
  - Represent input token into continuous vector representation.



Word embeddings
- FastText, GloVe
- ELMo, BERT

GLoVe:
- He et al., 2017
- Strubell et al., 2018

SENNA:
- Ouchi et al., 2018

ELMo:
- Marcheggiani et al., 2017
- Ouchi et al., 2018
- Li et al., 2019
- Lyu et al., 2019
- Jindal et al., 2020
- Li et al., 2020

BERT:
- Shi et al., 2019
- Jindal et al., 2020
- Li et al., 2020

Static Embeddings

BERT:
- Shi et al., 2019
- Conia et al., 2020
- Zhang et al., 2021
- Tian et al., 2022

RoBERTa:
- Conia et al., 2020
- Blloshmi et al., 2021
- Fei et al., 2021
- Wang et al., 2022
- Zhang et al. 2022

XLNet:
- Zhou et al., 2020
- Tian et al., 2022

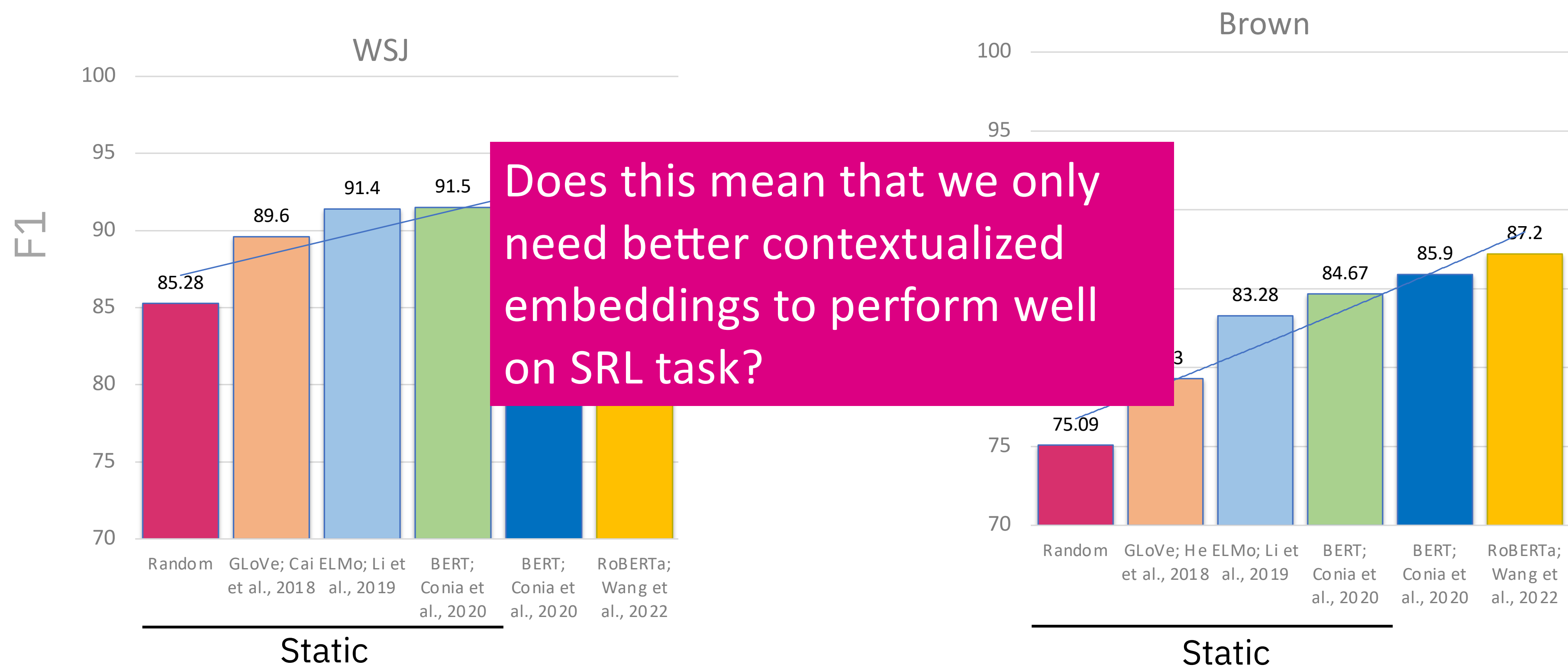
Dynamic Embeddings
Merchant et al., 2020

Performance Analysis

Sub-task: Argument Classification

Dataset: CoNLL09 EN

Best performing model for each word embedding type



Neural SRL Components – Encoder

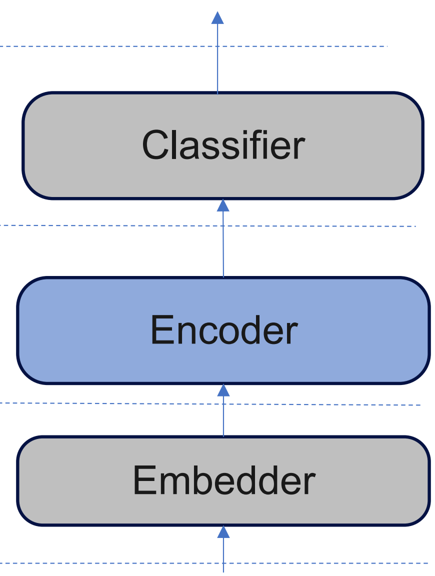
Sub-task: Argument Classification

- Encoder:
  - Encodes the context information to each token.

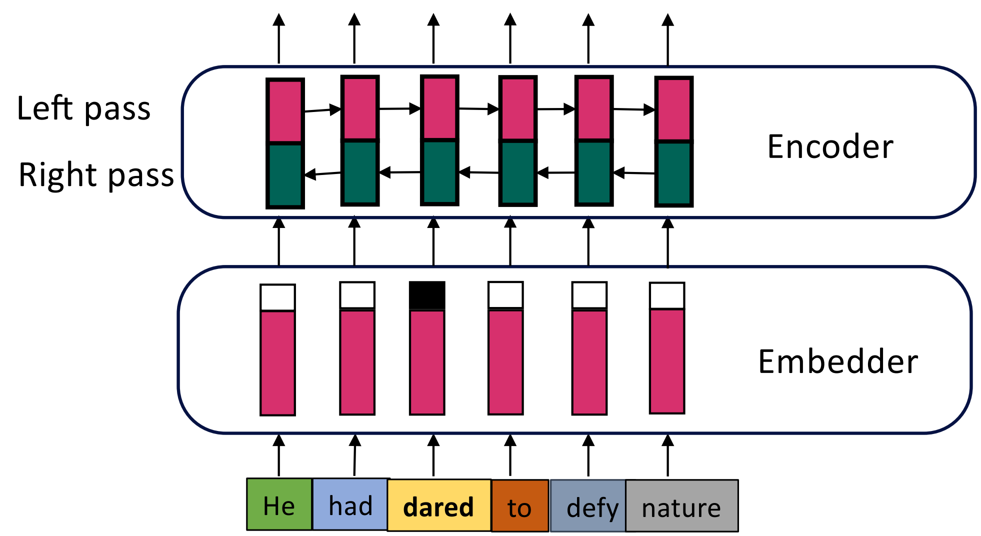
Encoder could be
- Stacked BiLSTMs or some variant of LSTMs
- Attention Network

Incorporates syntax information

Types of encoder
- BiLSTMs
- Attention



Input Sentence

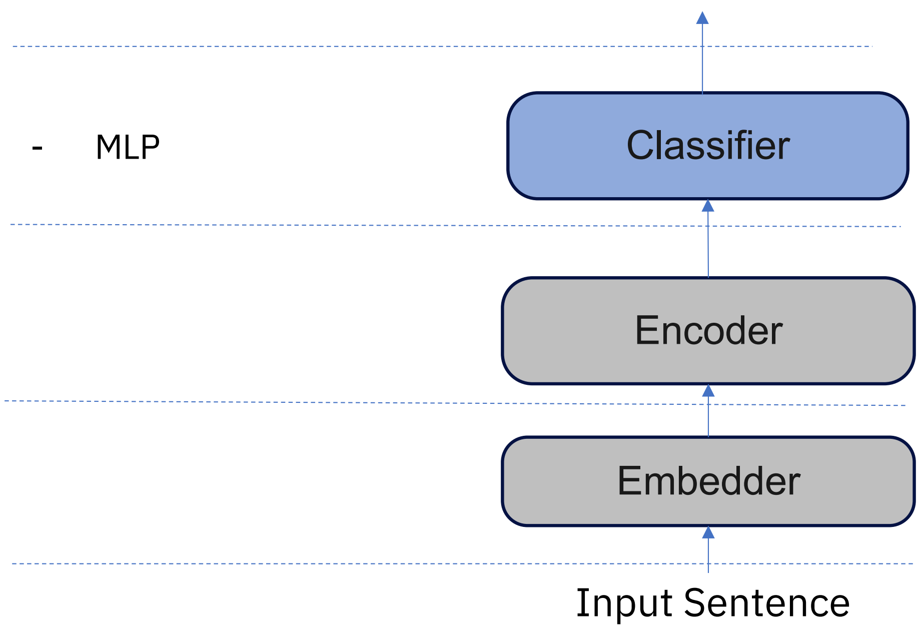


Neural SRL Components – Classifier

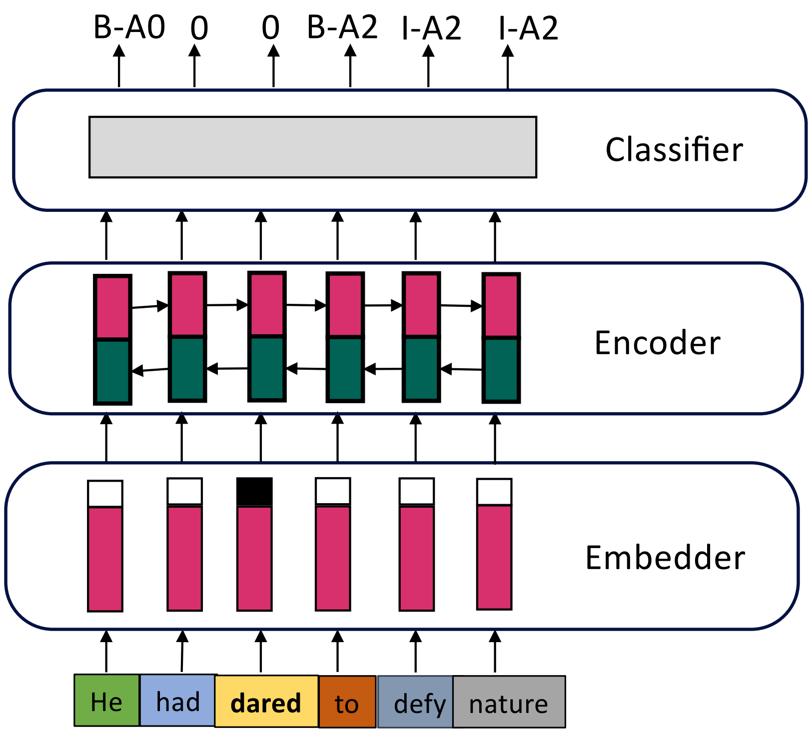
Sub-task: Argument Classification

- Classifier
  - Assign a semantic role label to each token in the input sentence.

- MLP



Usually a FF followed by Softmax



Outline

- **Early SRL approaches**
- **Typical neural SRL model components**
  - Performance analysis
- **Syntax-aware neural SRL models**
  - What, When and Where?
  - Performance analysis
  - How to incorporate Syntax?
- **Syntax-agnostic neural SRL models**
  - Performance Analysis
  - Do we really need syntax for SRL?
  - Are high quality contextual embedding enough for SRL task?
- **Practical SRL systems**
  - Should we rely on this pipelined approach?
    - End-to-end SRL systems
  - Can we jointly predict dependency and span?
- **More recent approaches**
  - Handling low-frequency exceptions
  - Incorporate semantic role label definitions
  - SRL as MRC task
- **Practical SRL system evaluations**
  - Are we evaluating SRL systems correctly?
- **Conclusion**

What and Where Syntax?

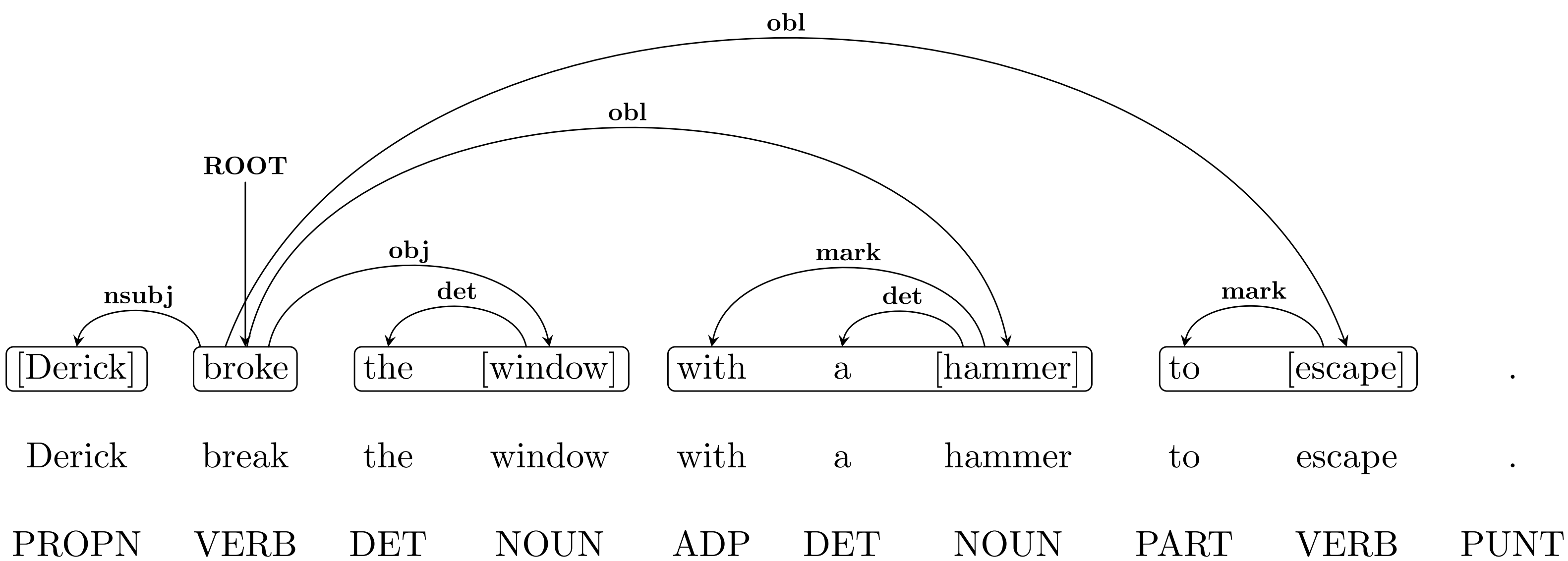
What Syntax for SRL? Everything or anything that explains the syntactic structure of the sentence

Dependency Relation

Surface form

Lemma form

U{X}POS



Parsed with UDPipe Parser: http://lindat.mff.cuni.cz/services/udpipe/

Syntax at Embedder

Where the Syntax is being used?

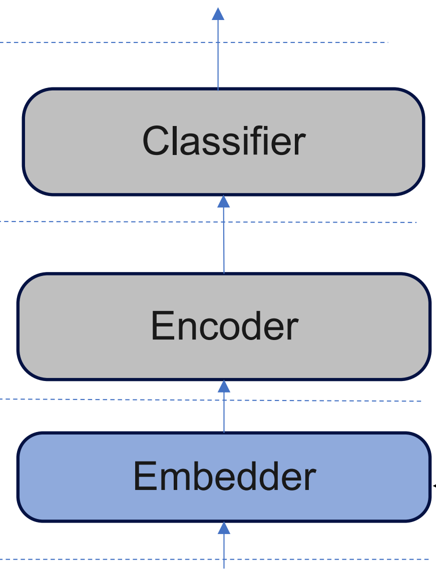
Marcheggiani et al.,2017b

Li et al., 2018
He et al., 2018

Wang et al., 2019
Kasai et al., 2019
HE et al., 2019

Li et al., 2020
Zhou et al., 2020

EMB



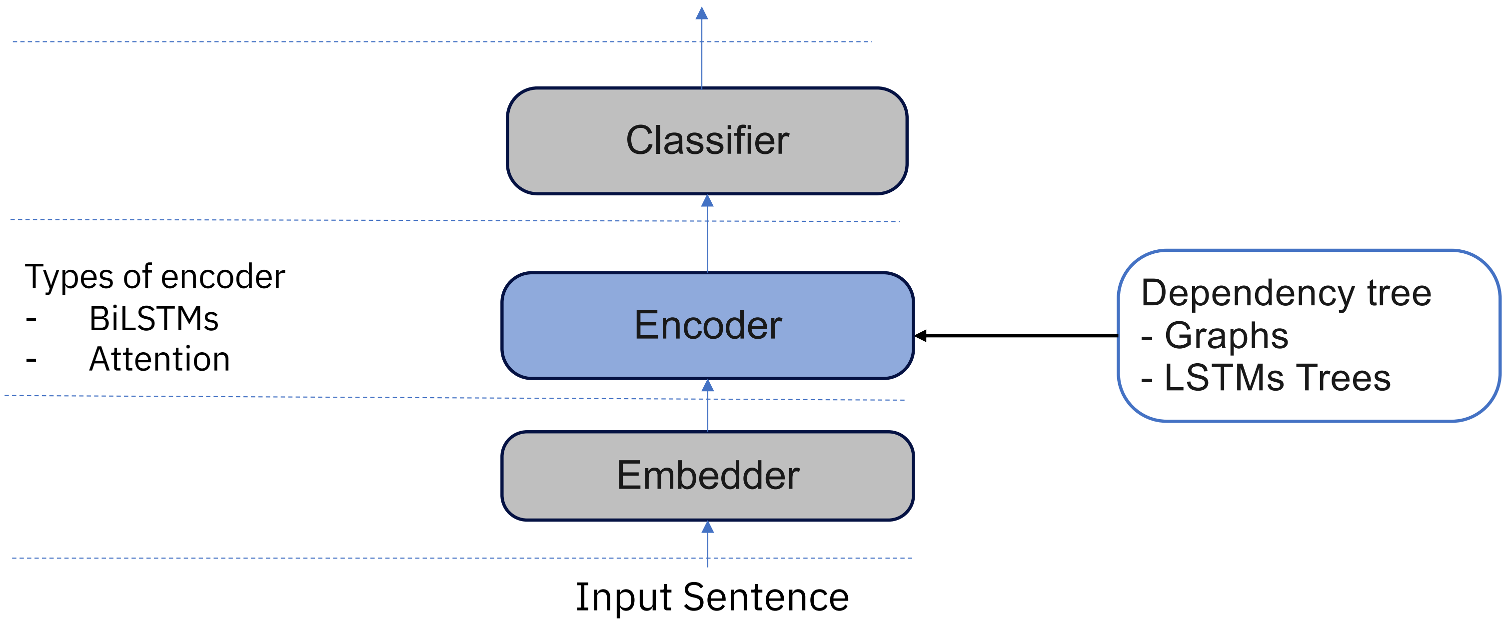
Input Sentence

Word embeddings
- FastText, GloVe
- ELMo, BERT

Concatenate {POS, dependency relation, dependency head and other syntactic information}

Syntax at Encoder

Where the Syntax is being used?



ENC

Marcheggiani et al., 2017

Zhou et al., 2020
Marcheggiani et al., 2020

Zhang et al., 2021

Tian et al., 2022

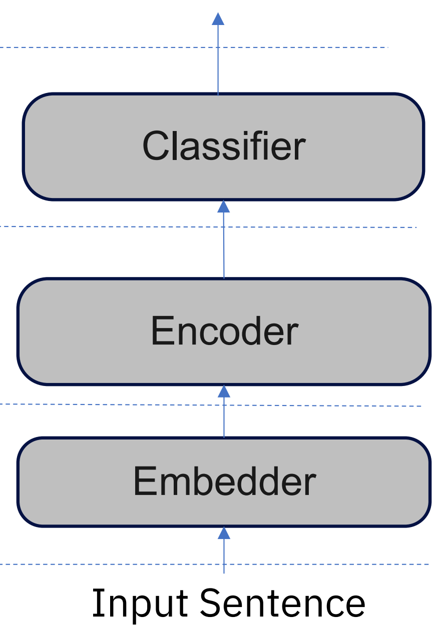
Joint Learning

At what level Syntax is used?

- MLP

Types of encoder
- BiLSTMs
- Attention

Word embeddings
- FastText, GloVe
- ELMo, BERT



Strubell et al., 2018

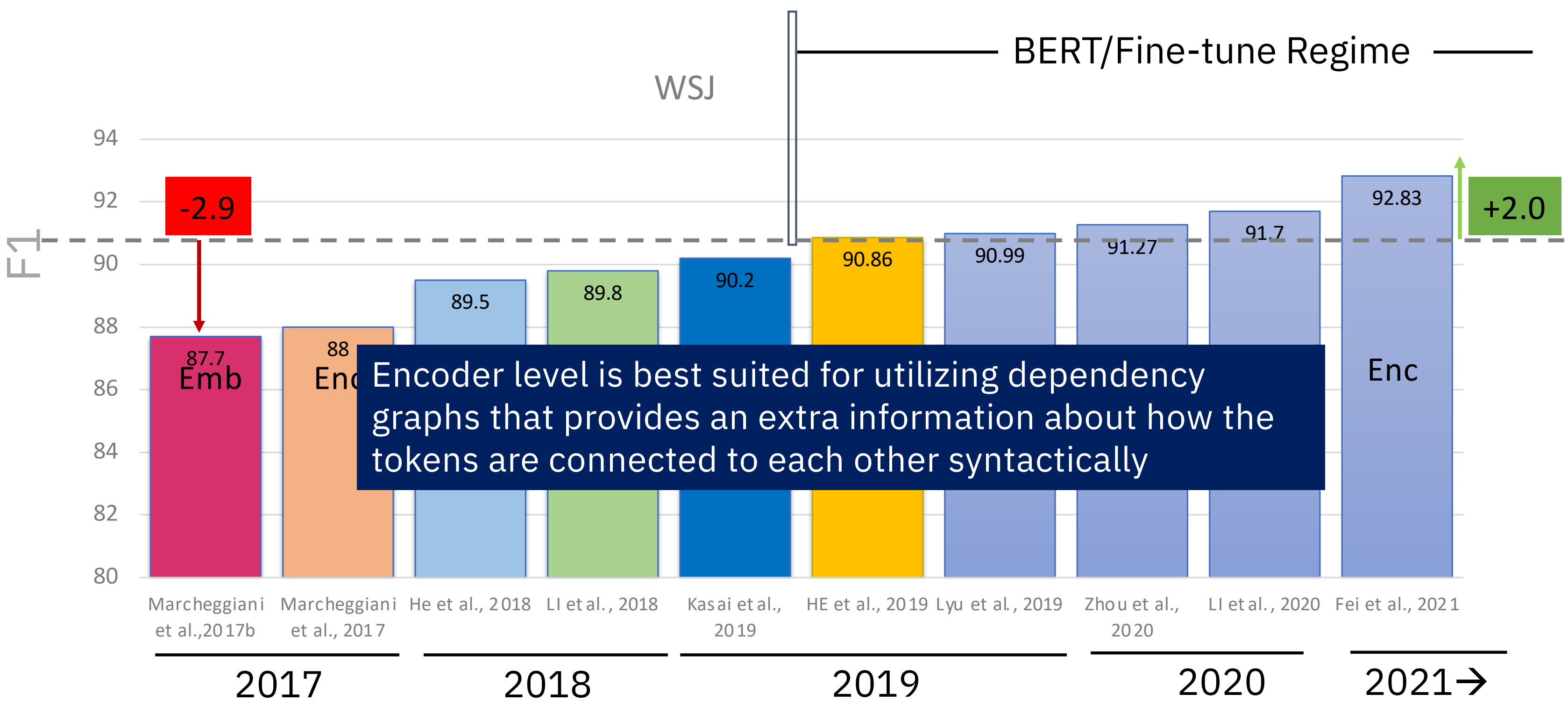
Shi et al., 2020

Multi-task learning

Performance Analysis

Dataset: CoNLL09 EN

Comparing Syntax aware models



Marcheggiani et al., 2017

Syntax at embedder level

A Simple and Accurate Syntax-Agnostic Neural Model for Dependency-based Semantic Role Labeling

- Predict semantic dependency edges between predicates and arguments.
- Use predicate-specific roles (such as make-A0 instead of A0) as opposed to generic sequence labeling task.

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 the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 411–420, Vancouver, Canada. Association for Computational Linguistics.

Marcheggiani et al., 2017

Syntax at embedder level

Embedder OR
Input word representation

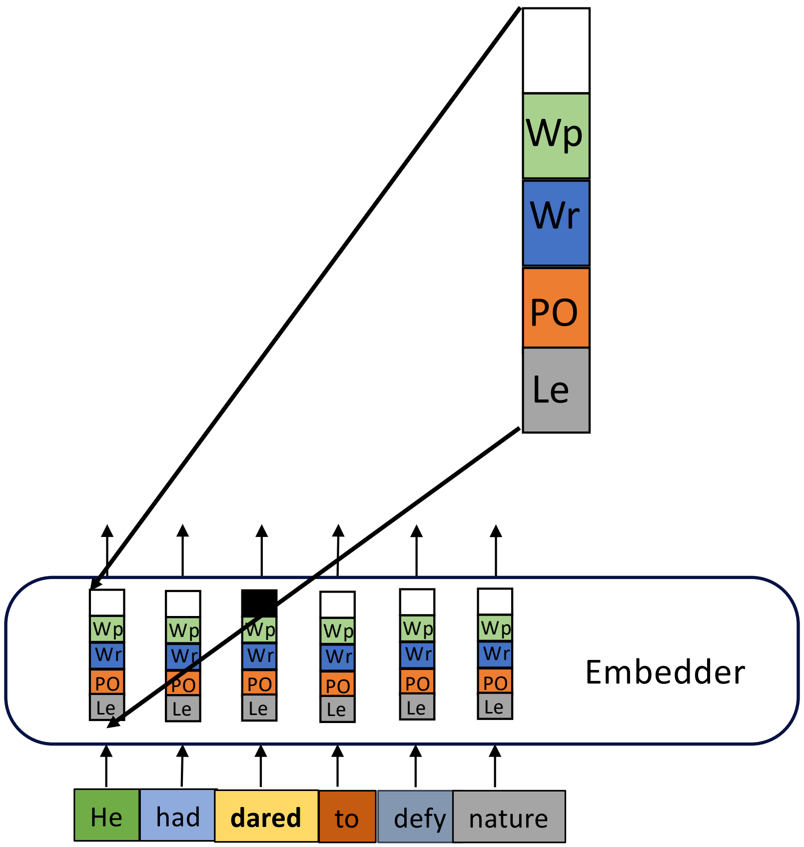
← Randomly initialized word embeddings

← Pre-trained word embeddings

← Randomly initialized POS embeddings

← Randomly initialized Lemma embeddings


← Predicate specific feature [Binary]



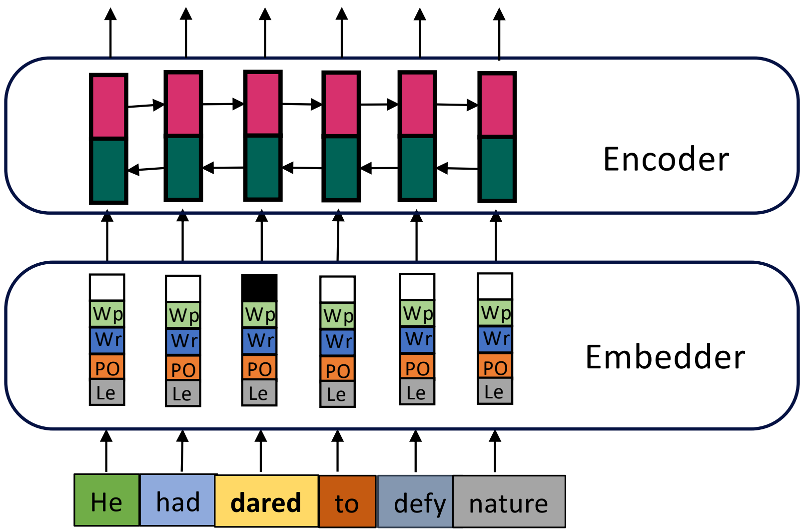
Marcheggiani et al., 2017

Syntax at embedder level

Encoder

Several BiLSTMs layers

- Capturing both the left and the right context
- Each BiLSTM layer takes the lower layer as input



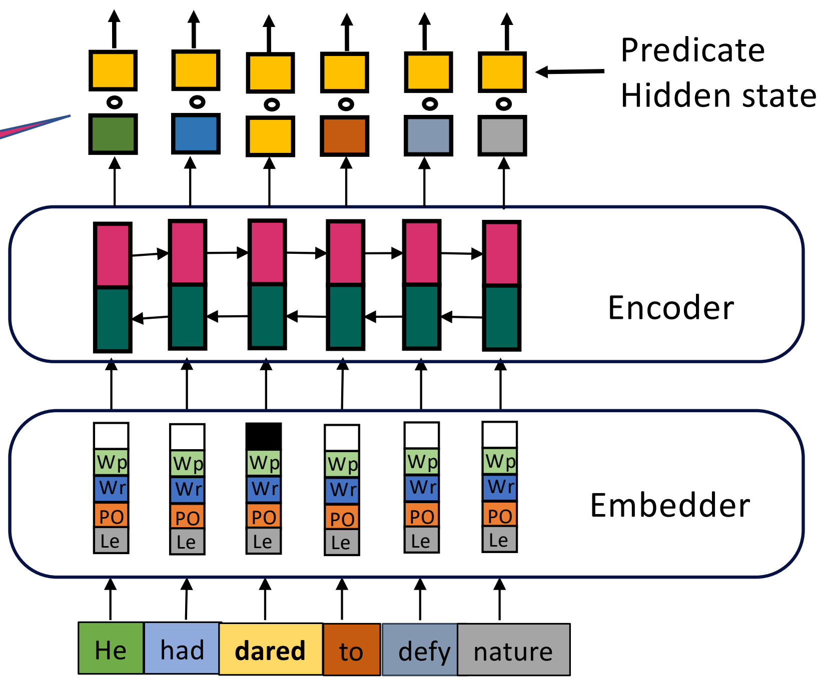
Marcheggiani et al., 2017

Syntax at embedder level

Preparation for classifier

Provide predicate hidden state as another another input to classifier along with each token.

+ ~6% F1 on CoNLL09 EN

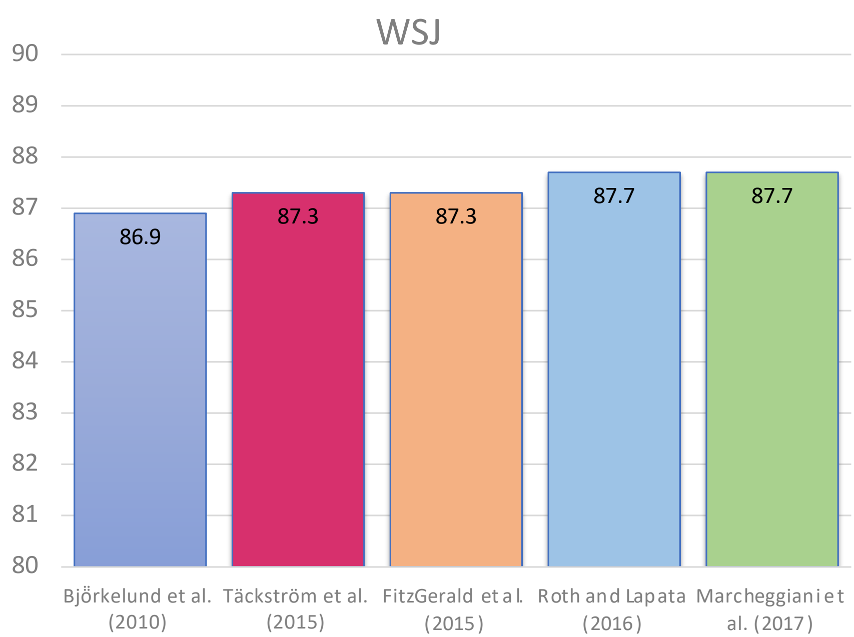


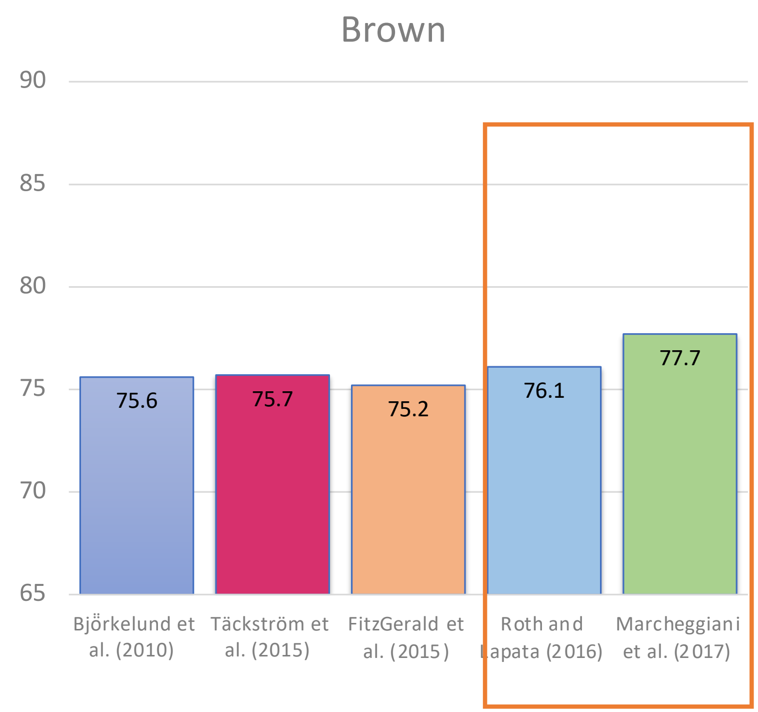
The two ways of encoding predicate information, using predicate-specific flag at embedder level and incorporating the predicate state in the classifier, turn out to be complementary.

Marcheggiani et al., 2017

Syntax at embedder level

Dataset: CoNLL09 EN



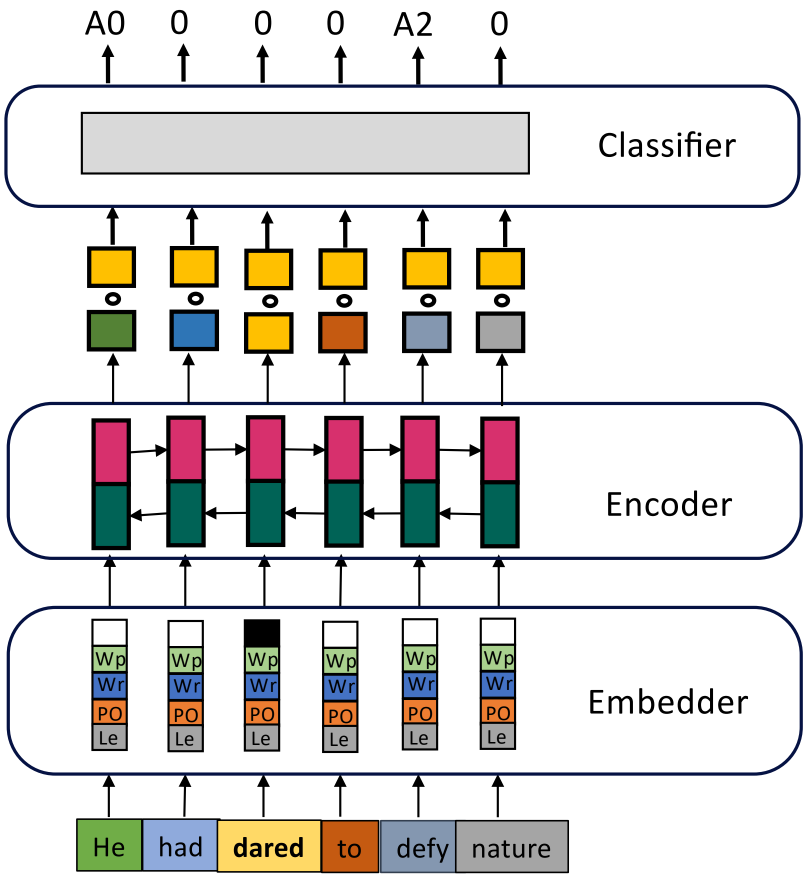


Marcheggiani et al., 2017

Syntax at embedder level

Takeaways

- Appending POS does help → approx. 1 F1 points gain
- Predicate specific encoding does help → approx. 6 F1 point gain
- Quite effective for the classification of arguments which are far from the predicate in terms of word distance.
- **Noted**: Substantial improvement on EN OOD over previous works.

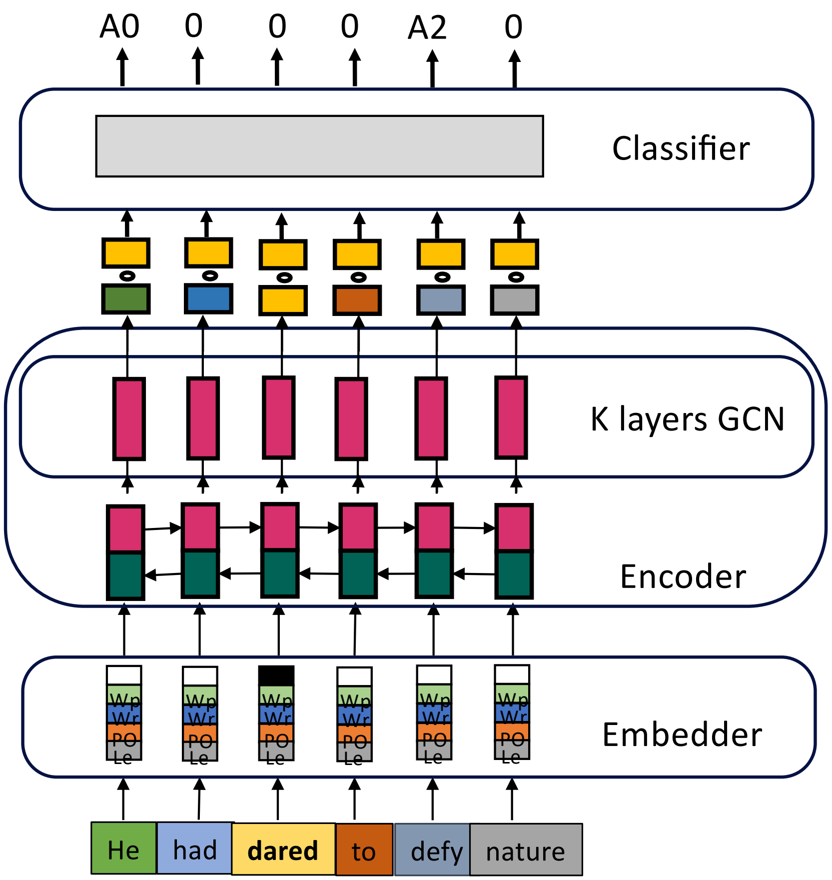


Marcheggiani et al., 2017b

Syntax at encoder level

Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling

- Basic SRL components remains the same as compared to [Marcheggiani et al., 2017]
- GCN layers are inserted between Encoder and Classifier.
  - Re-encoding the encoder representations based on syntactic structure of the sentence.
  - Modeling syntactic dependency structure



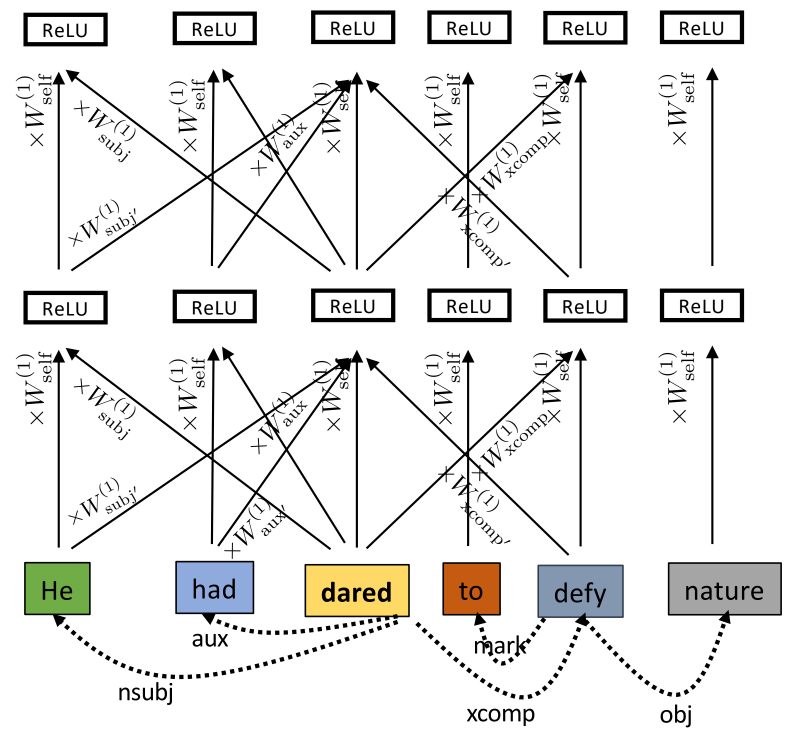
Diego Marcheggiani and Ivan Titov. 2017. [Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling.]() In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1506–1515, Copenhagen, Denmark. Association for Computational Linguistics.

Marcheggiani et al., 2017b

Syntax at encoder level

What is syntactic GCN?

- Want to encode information k nodes away
  - Use k layers to encode k-order neighborhood.
- Helped capture the widened syntactic neighborhood.



Marcheggiani et al., 2017b

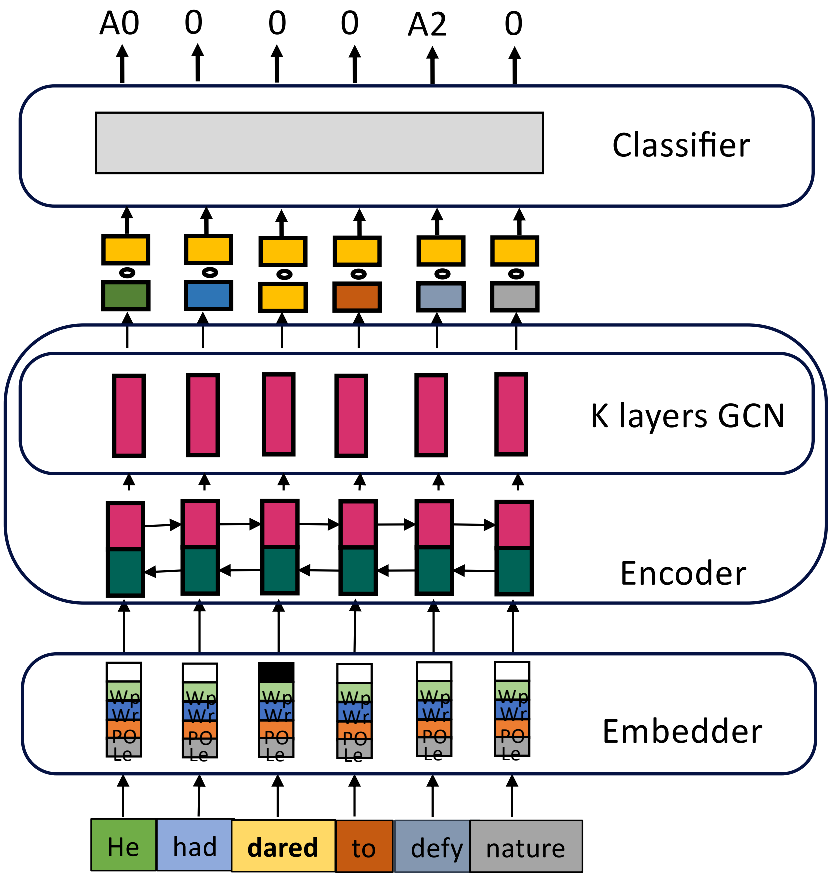
Syntax at encoder level

Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling

- Claim: GCN helps capture long range dependencies.
  - But: encoding k-hope neighborhood seems to hurt the performance. (k = 1 works the best)

System (English)	P	R	F_1
LSTMs	84.3	81.1	82.7
LSTMs + GCNs (K=1)	85.2	81.6	83.3
LSTMs + GCNs (K=2)	84.1	81.4	82.7
LSTMs + GCNs (K=1), no gates	84.7	81.4	83.0
GCNs (no LSTMs), K=1	79.9	70.4	74.9
GCNs (no LSTMs), K=2	83.4	74.6	78.7
GCNs (no LSTMs), K=3	83.6	75.8	79.5
GCNs (no LSTMs), K=4	82.7	76.0	79.2

Table 1: SRL results without predicate disambiguation on the English development set.

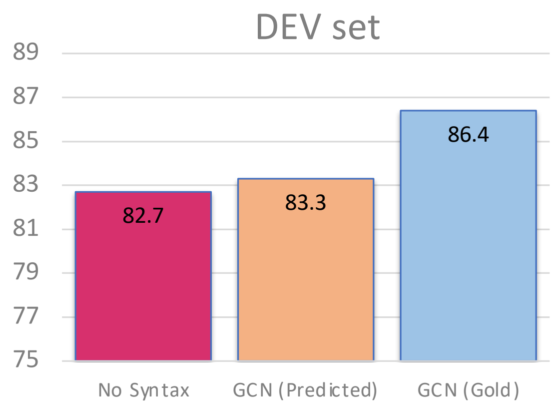


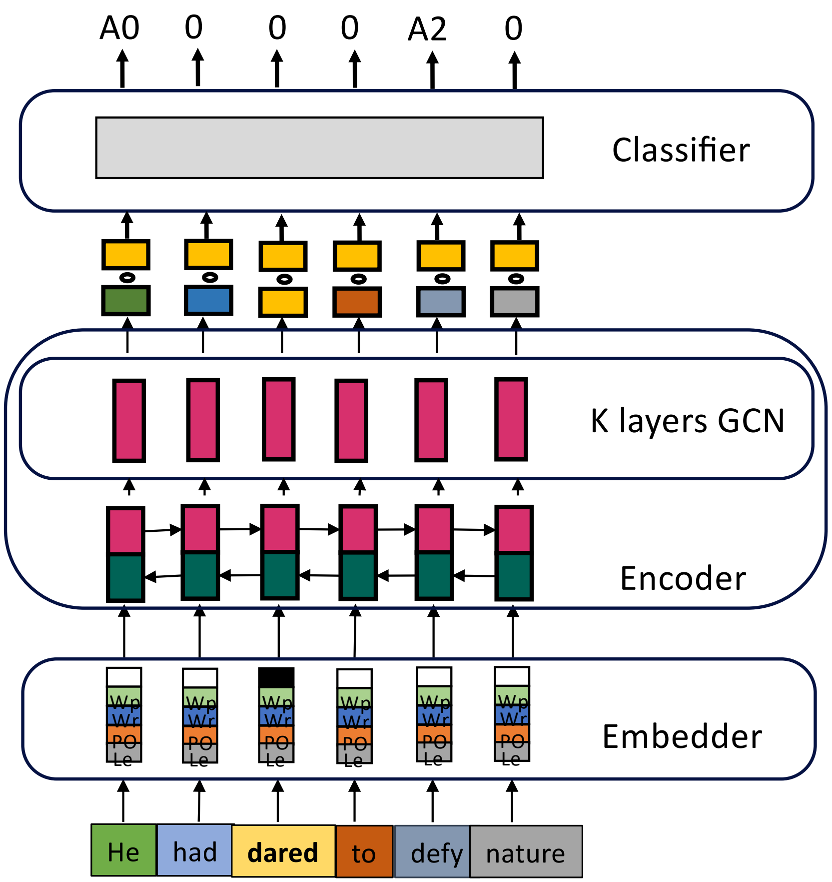
Marcheggiani et al., 2017b

Syntax at encoder level

Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling

- Gold dependency can significantly improve the performance.

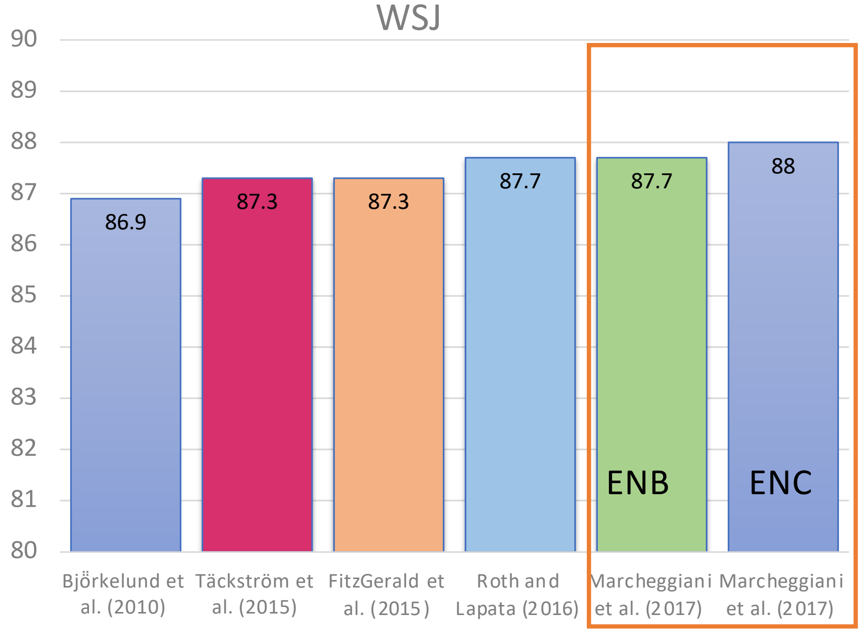


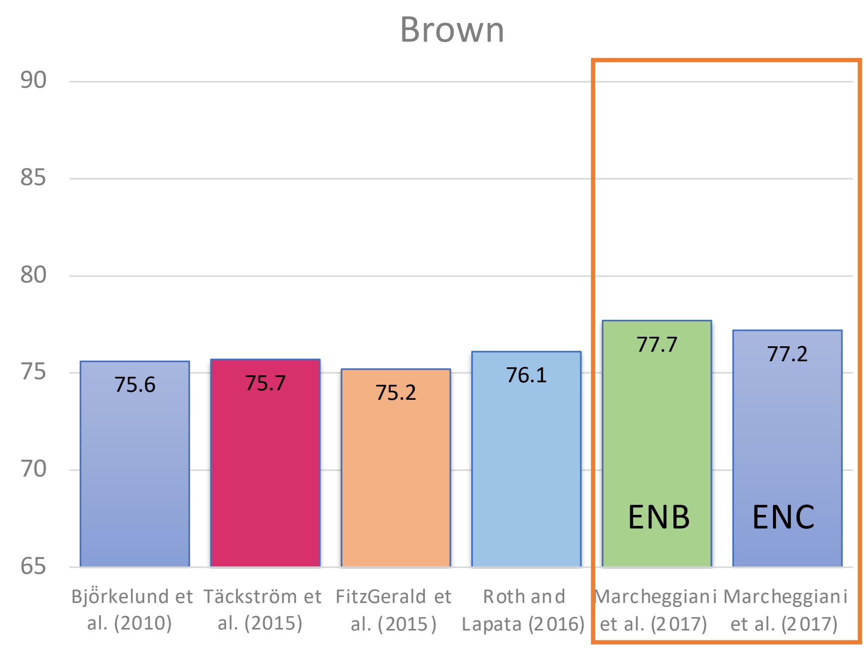


Marcheggiani et al., 2017b

Syntax at encoder level

Dataset: CoNLL09 EN

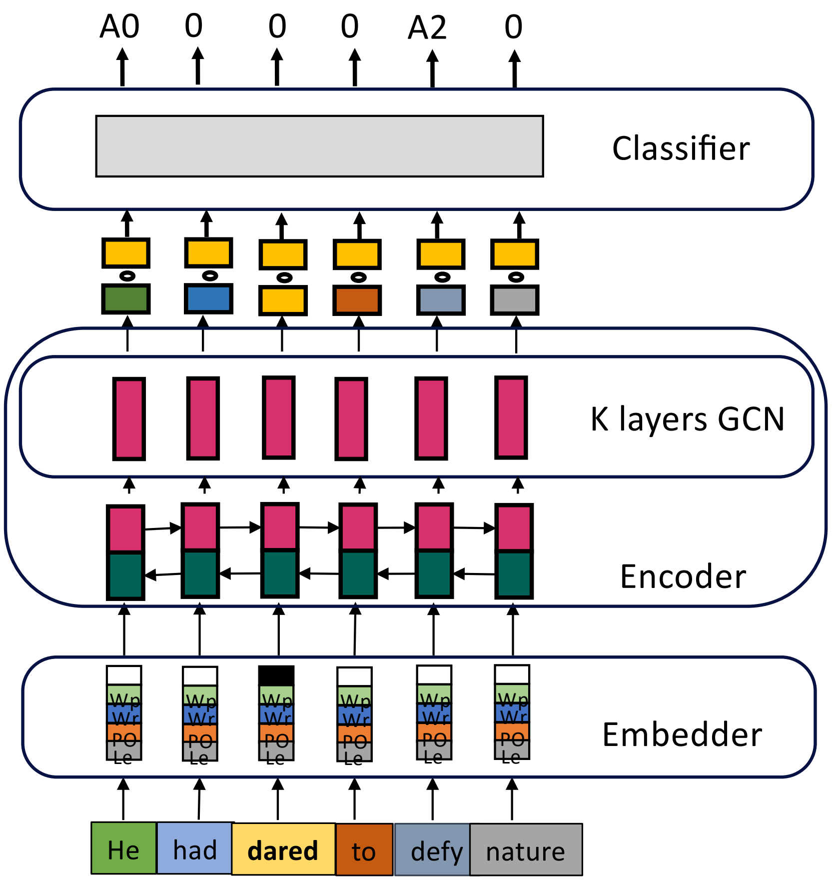




Marcheggiani et al., 2017b

Takeaways

- Appending POS does help → approx. 1 F1 points gain
- Predicate specific encoding does help → approx. 6 F1 point gain
- Model syntactic dependencies via syntactic GCN further improve the SRL performance. NEED High quality syntactic parser
- **Noted**: Improvement only on EN in-domain over previous works.
- However previous work show improvement over OOD set.

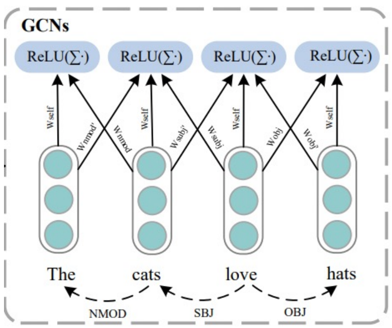


Li et al., 2018

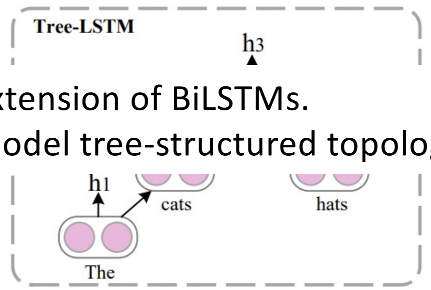
Syntax at encoder level

A unified Syntax-aware Framework for Semantic role labeling

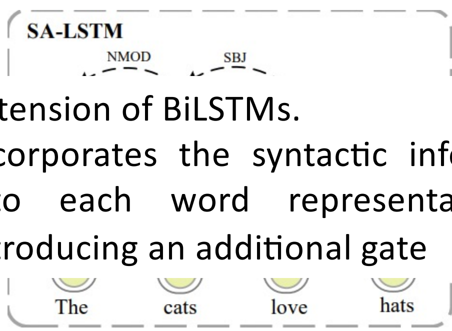
[Marcheggiani et al., 2017b]



[Tai et al., 2015]

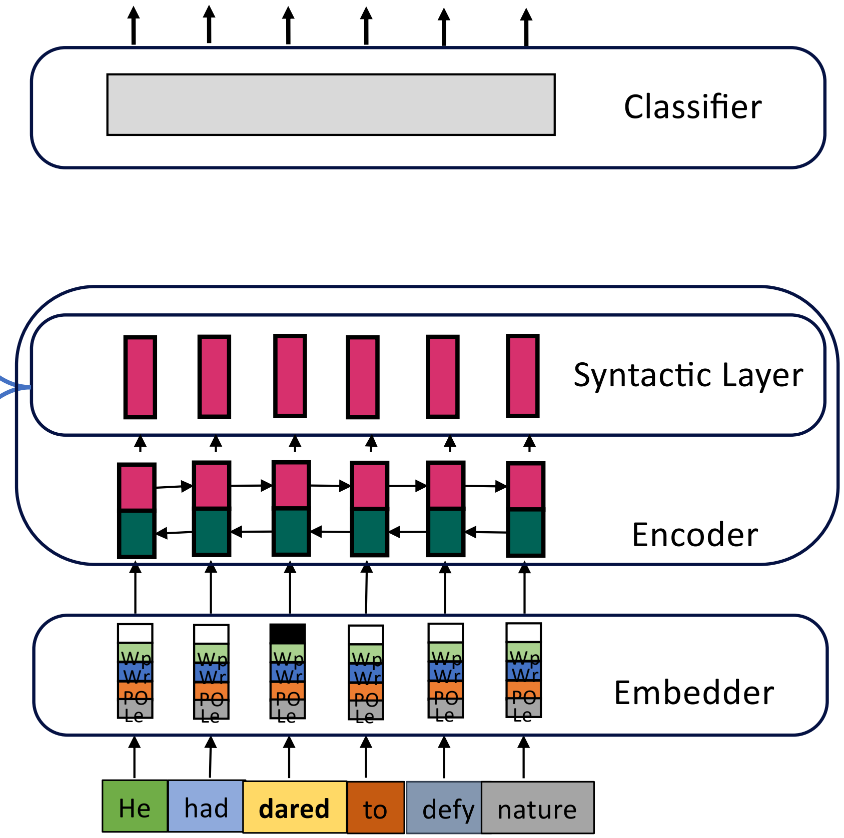


Extension of BiLSTMs.
Model tree-structured topologies



Extension of BiLSTMs.
Incorporates the syntactic information into each word representation by introducing an additional gate

[Qian et al., 2017]

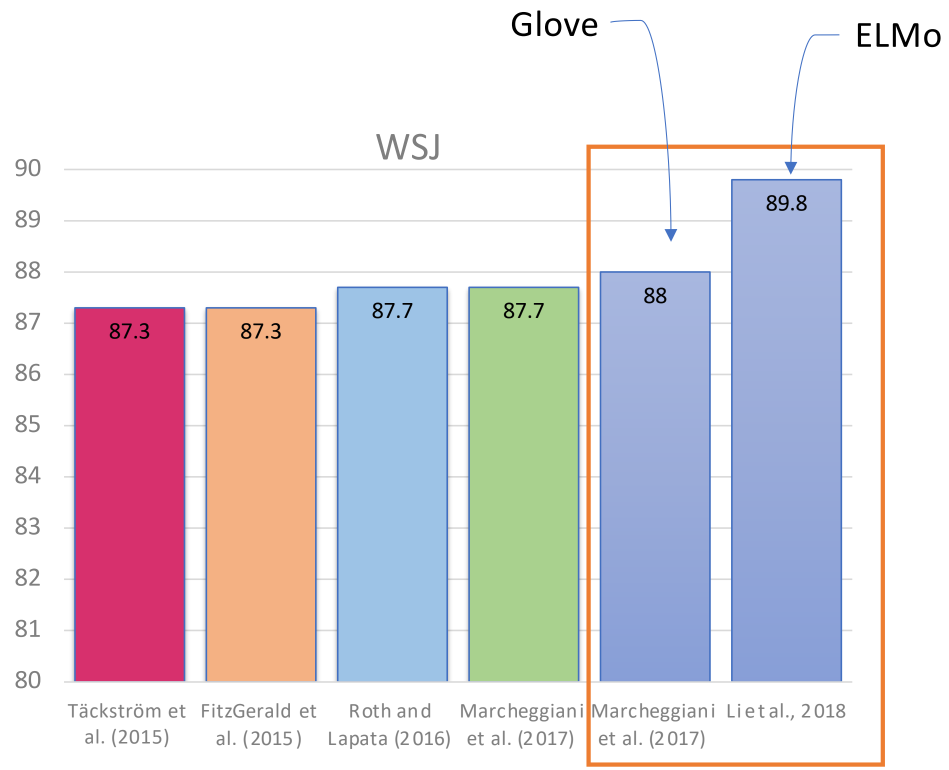


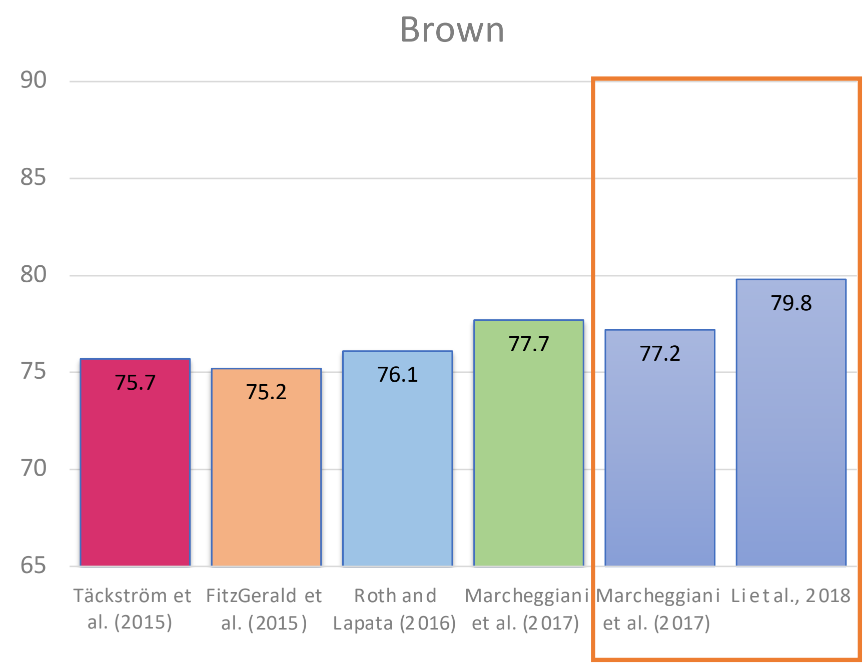
Zuchao Li, Shexia He, Jiaxun Cai, Zhuosheng Zhang, Hai Zhao, Gongshen Liu, Linlin Li, and Luo Si. 2018. A Unified Syntax-aware Framework for Semantic Role Labeling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2401–2411, Brussels, Belgium. Association for Computational Linguistics.

Li et al., 2018

Syntax at encoder level

Dataset: CoNLL09 EN





Outline

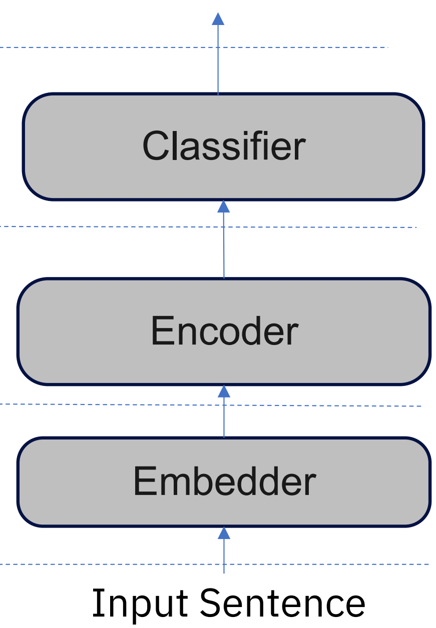
- Early SRL approaches
- **Typical neural SRL model components**
  - Performance analysis
- **Syntax-aware neural SRL models**
  - What, When and Where?
  - Performance analysis
  - How to incorporate Syntax?
- **Syntax-agnostic neural SRL models**
  - Performance Analysis
  - Do we really need syntax for SRL?
  - Are high quality contextual embedding enough for SRL task?
- **Practical SRL systems**
  - Should we rely on this pipelined approach?
    - End-to-end SRL systems
  - Can we jointly predict dependency and span?
- **More recent approaches**
  - Handling low-frequency exceptions
  - Incorporate semantic role label definitions
  - SRL as MRC task
- **Practical SRL system evaluations**
  - Are we evaluating SRL systems correctly?
- **Conclusion**

Syntax-Agnostic Model

- MLP

Types of encoder
- BiLSTMs
- Attention

Word embeddings
- FastText, GloVe
- ELMo, BERT



He et al., 2017

He et al., 2018
Cai et al., 2018
Ouchi et al., 2018

Guan et al., 2019
LI et al., 2019
Shi et al., 2019

Conia et al., 2020
Jindal et al., 2020
Zhou et al., 2020

Conia et al., 2021
Blloshmi et al., 2021

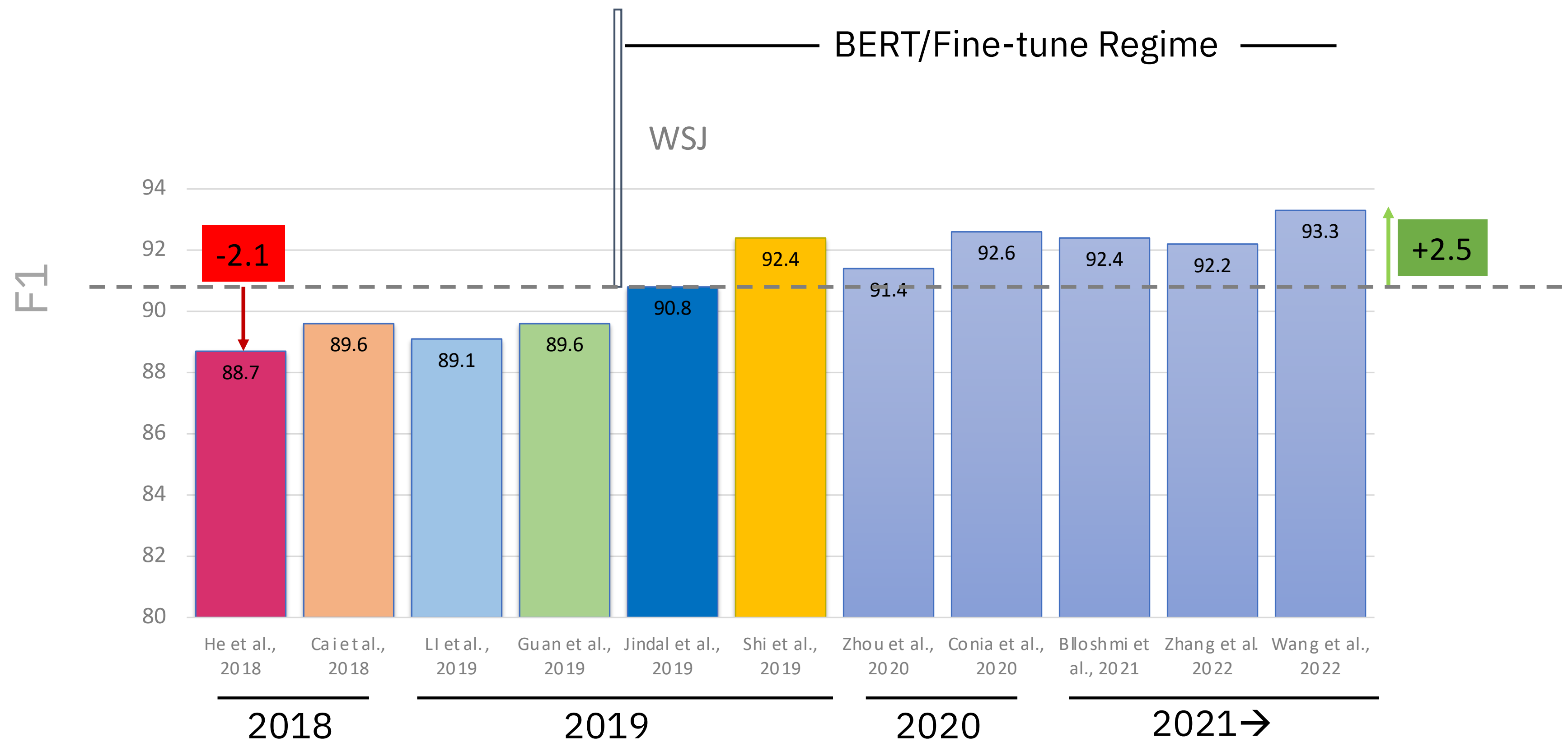
Wang et al., 2022
Zhang et al. 2022

Syntax-Agnostic Models

Performance Analysis

Dataset: CoNLL09 EN

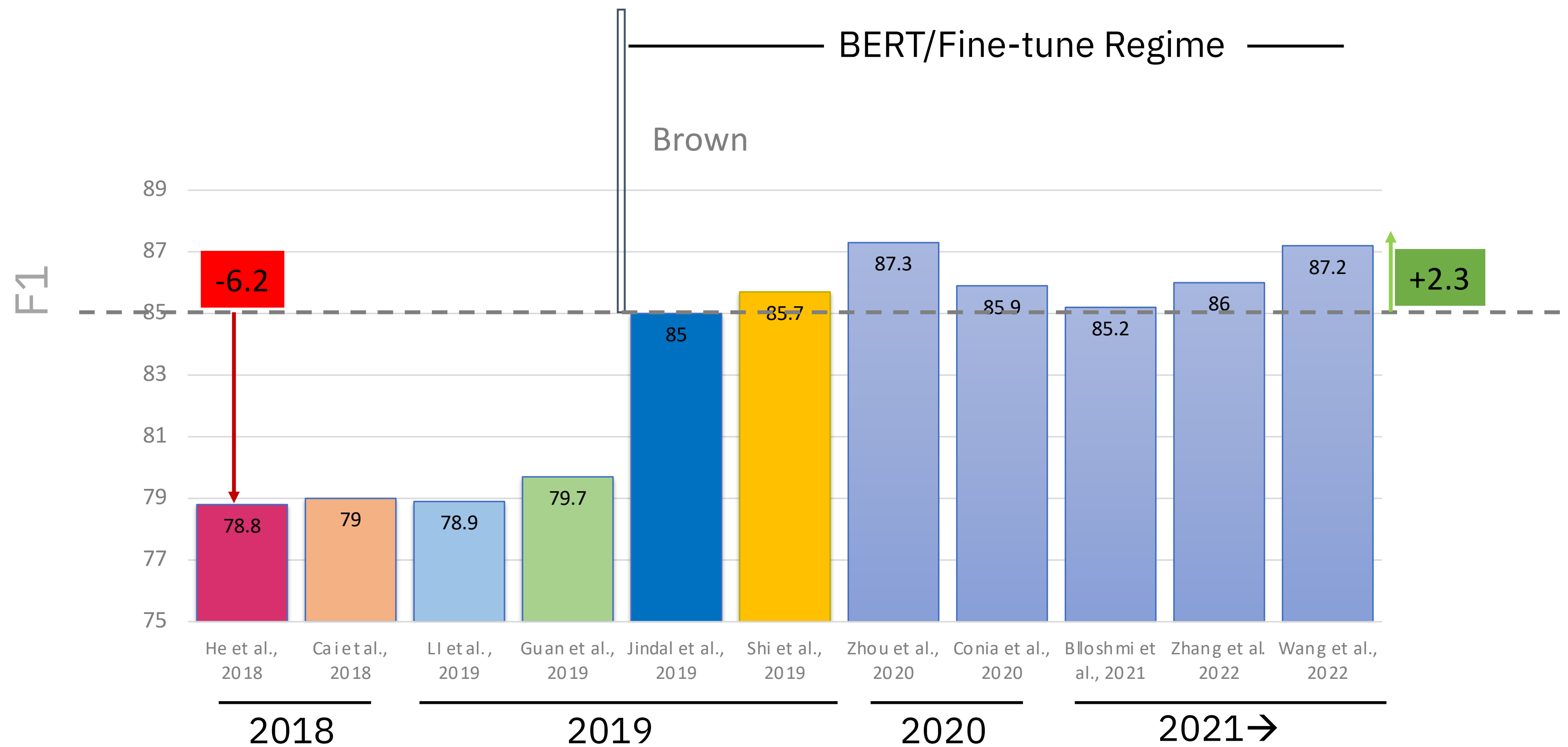
Comparing Syntax agnostic models



Performance Analysis

Dataset: CoNLL09 EN

Comparing Syntax agnostic models



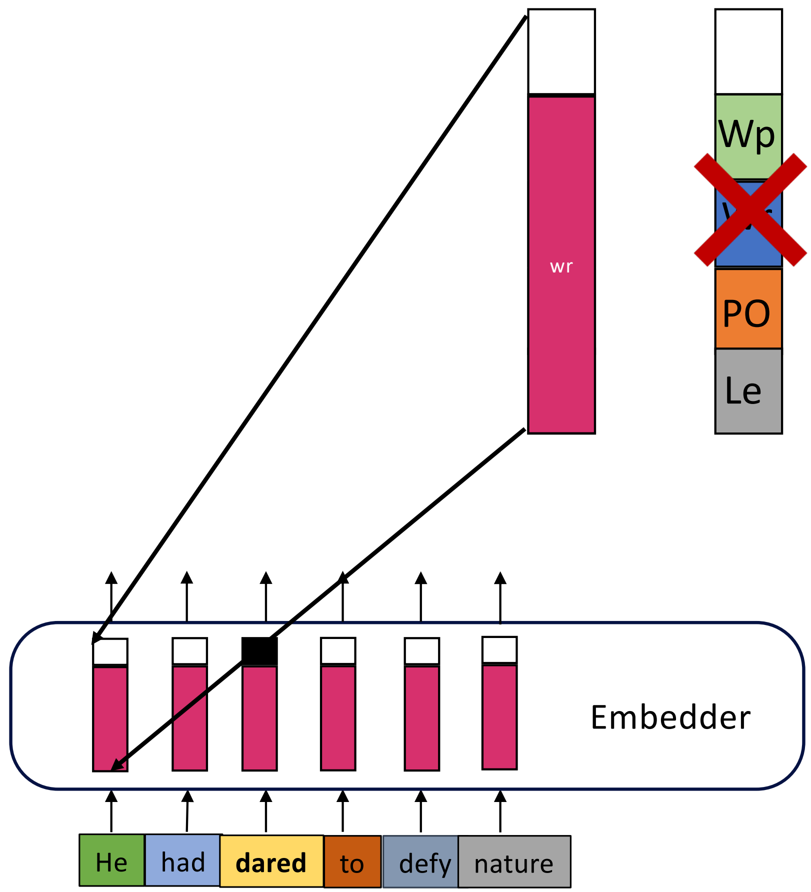
He et al., 2017

Embedder OR Input word representation

- Pre-trained word embeddings
- Use predicate flag

wr ← Pre-trained word embeddings

☐ ← Predicate specific feature [Binary]

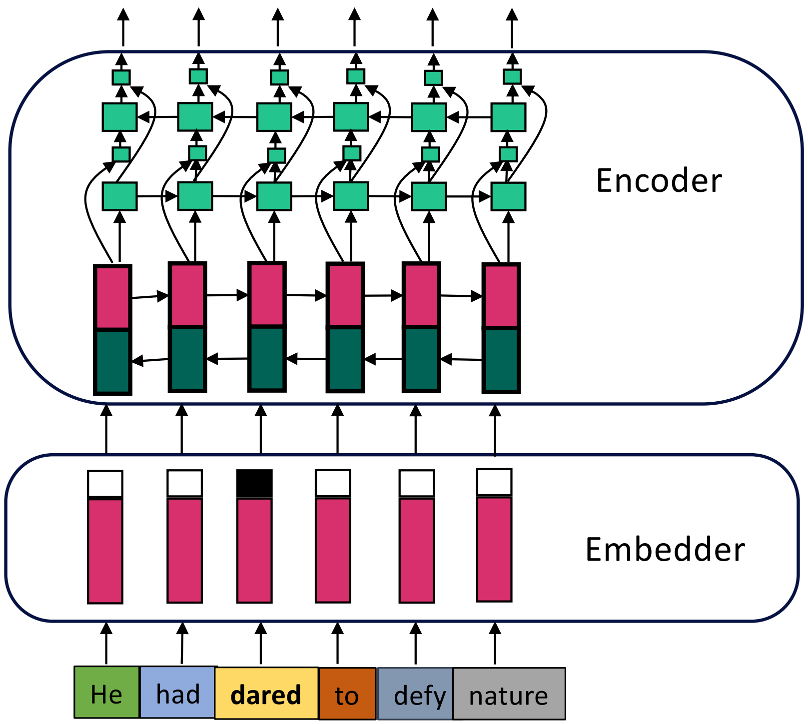


Luheng He, Kenton Lee, Mike Lewis, and Luke Zettlemoyer. 2017. [Deep Semantic Role Labeling: What Works and What's Next](). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 473–483, Vancouver, Canada. Association for Computational Linguistics.

He et al., 2017

Encoder

- Stacked BILSTM
- Highway Connections [Srivastava et al., 2015]
  - To alleviate vanishing gradient problem
- Recurrent Dropout [Gal et al., 2016]
  - To reduce overfitting



He et al., 2017

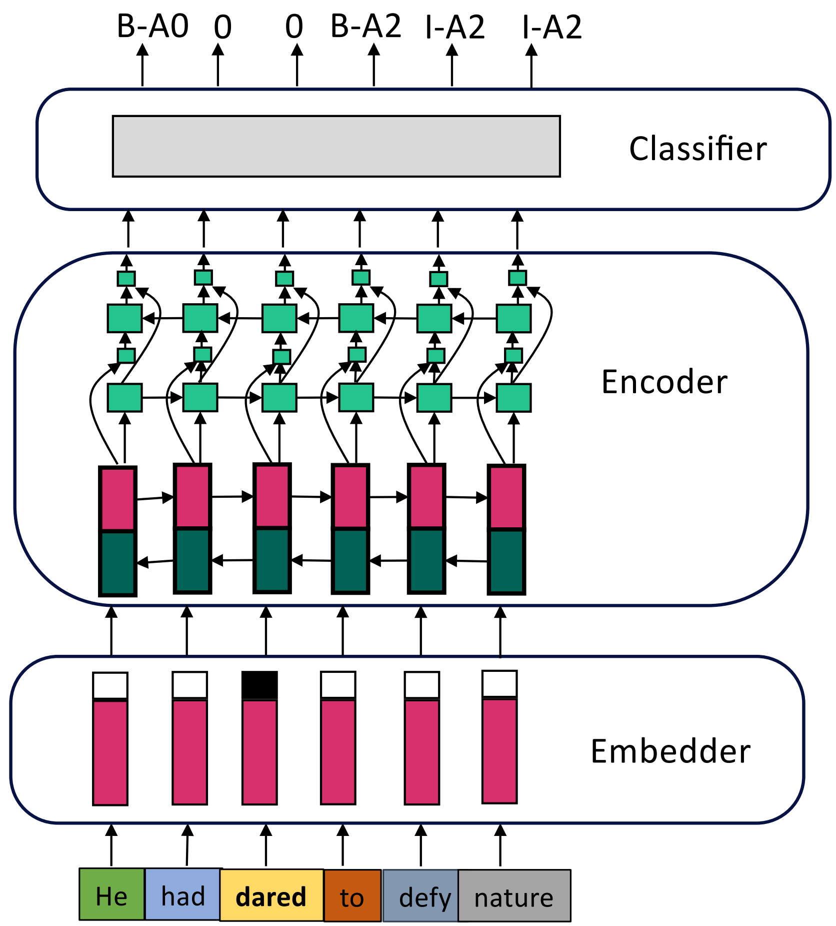
Local classifier

Classifier with MLP layer followed by Softmax

Global optimization

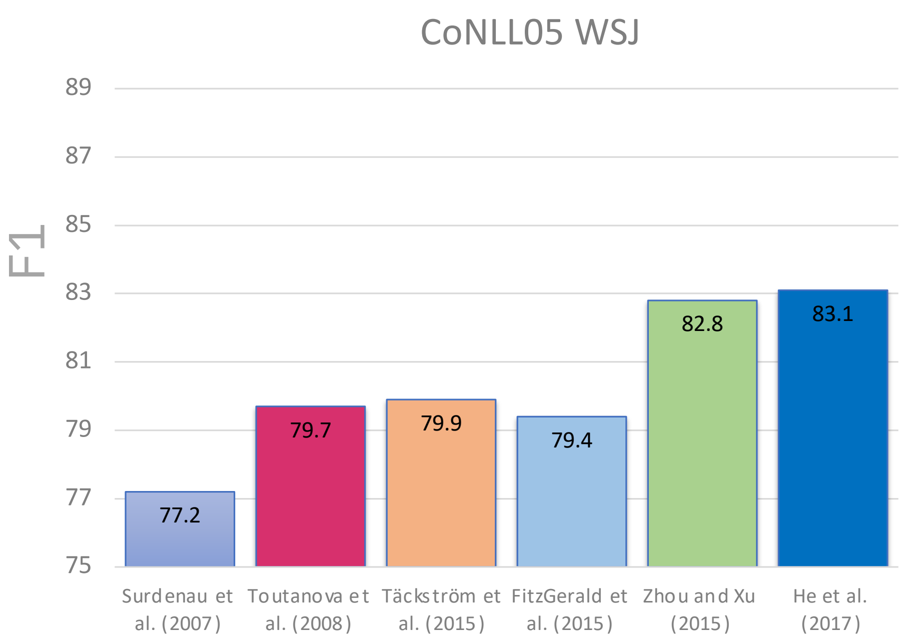
- Constraint A* decoding
  - BIO constraint
  - Unique core roles
  - Continuation Constraint
  - Reference constraint
  - Syntactic constraint

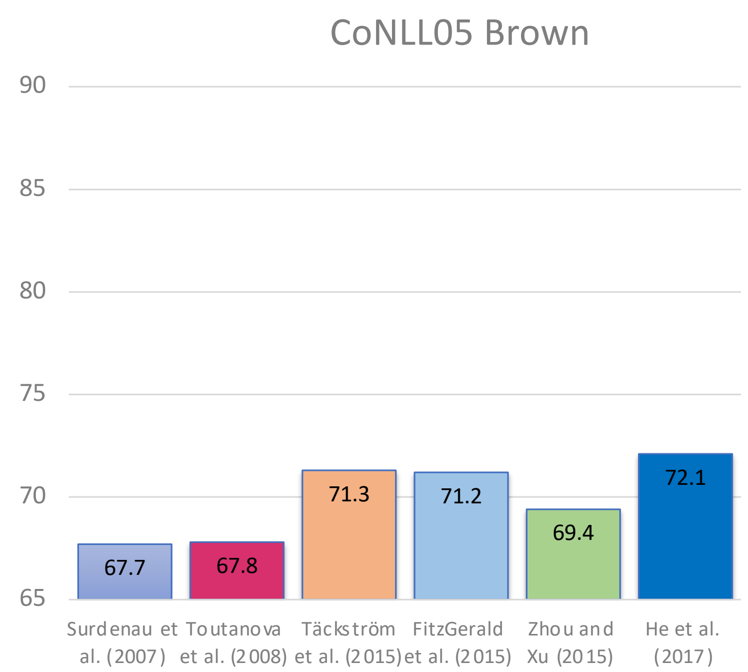
SRL Constraints were previously discussed by Punyakanok et al. (2008) and Tackstrom et al. (2015)



He et al., 2017

Dataset: CoNLL05



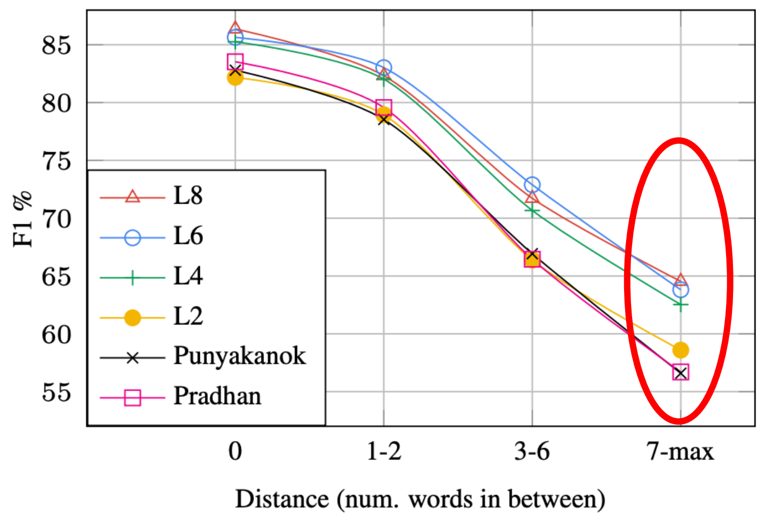


He et al., 2017

How well do LSTMs model global structural consistency, despite conditionally independent tagging decisions?

Long range dependencies:

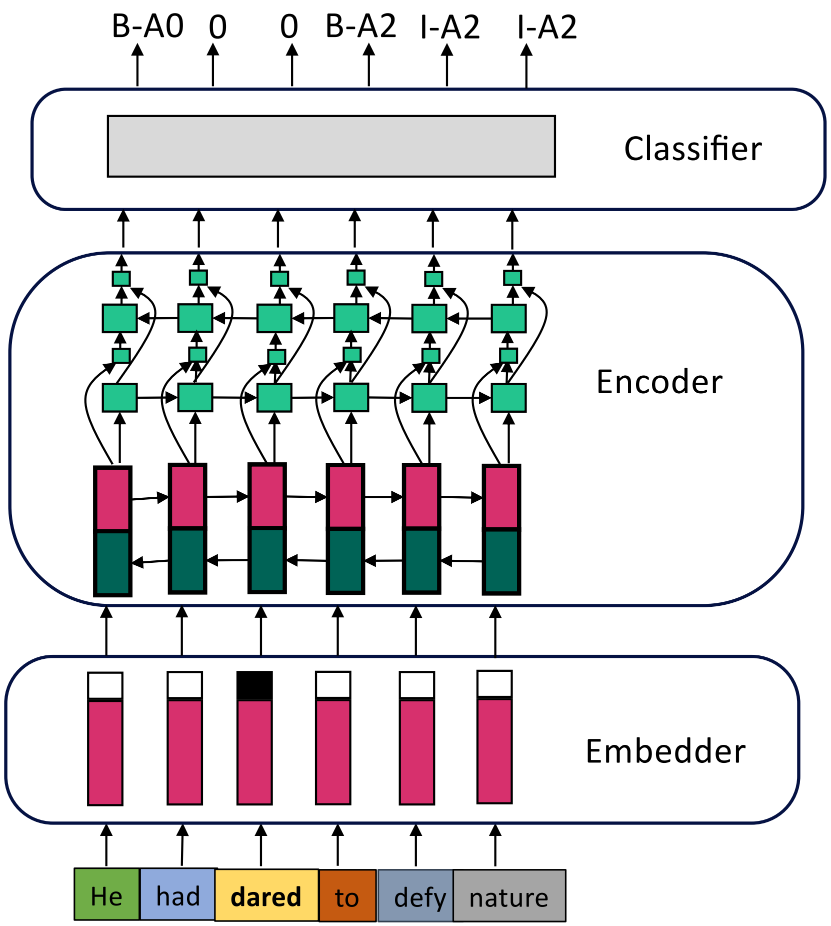
Performance tends to degrade, for all models, for arguments further from the predicate



He et al., 2017

Takeaways

- General label confusion between core arguments and contextual arguments is due to the ambiguous definitions in frame files.
- Layers of BiLSTMs help captures the long–range predicate-argument structures.
- The number of BIO violations decreases when we use a deeper model
- Deeper BiLSTMs are better at enforcing structural consistencies, although not perfectly.

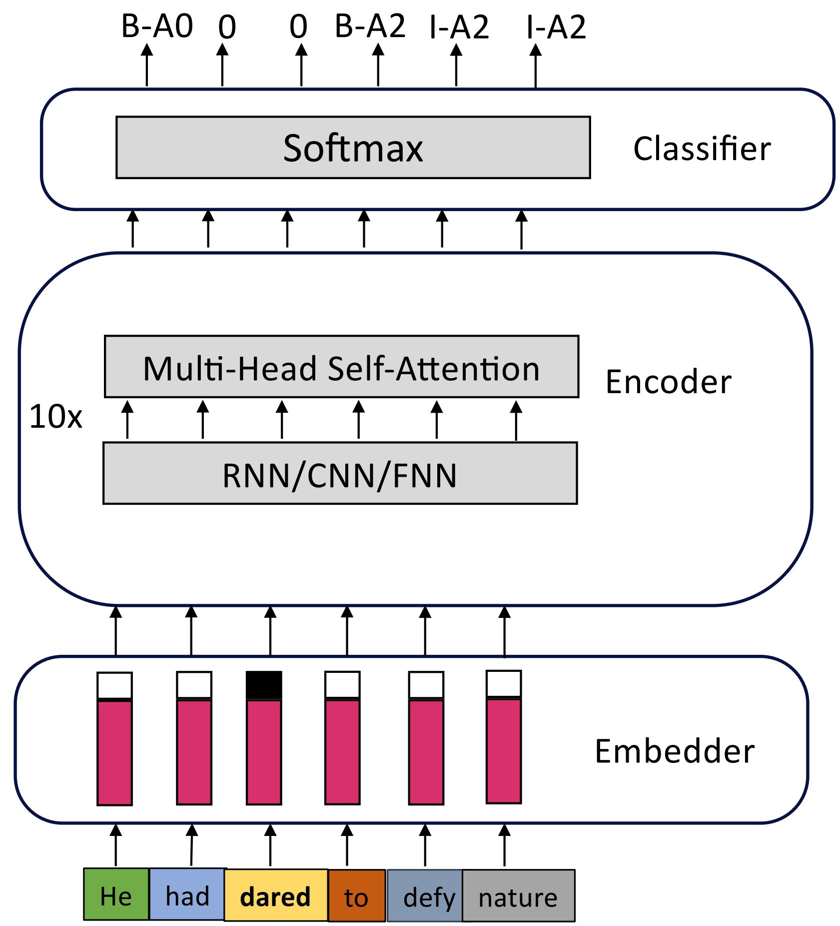


Tan et al., 2018

Do we really need all these hacks!!!! ☺

Let's break Recurrence and allow every position in the sentence to attend over all positions in the input sequence

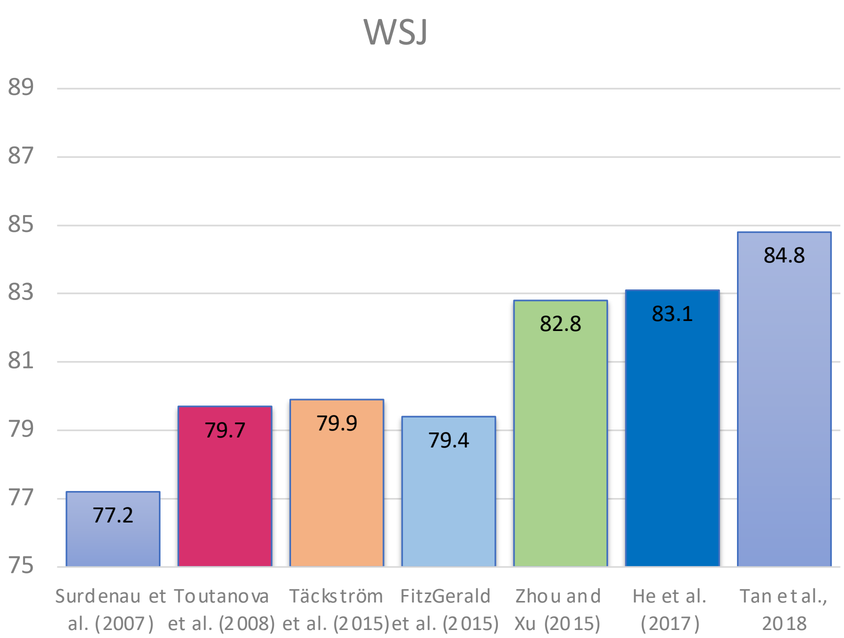
No Syntax
Use predicate specific flag
Use Multi-head self attention
Use Glove Embeddings

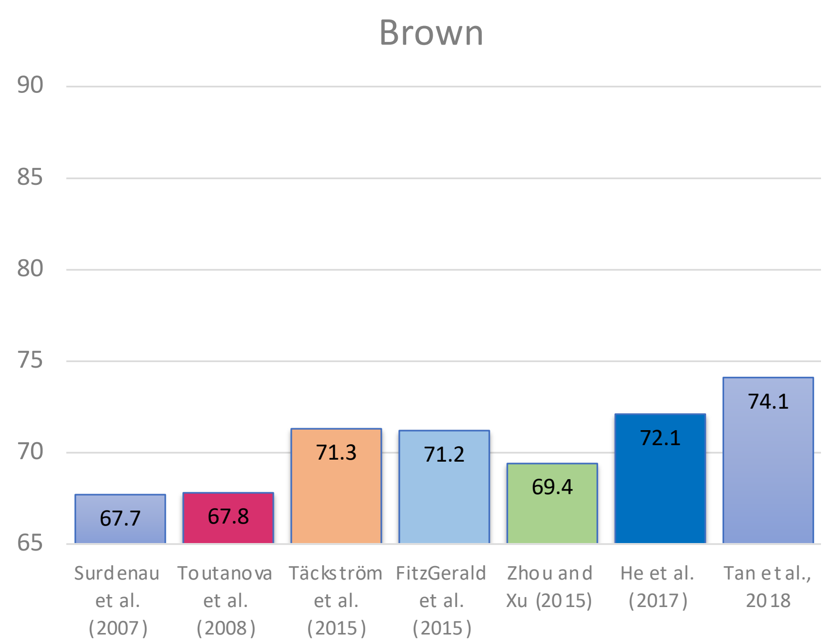


Tan, Zhixing, Mingxuan Wang, Jun Xie, Yidong Chen, and Xiaodong Shi. "Deep semantic role labeling with self-attention." In *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1. 2018.

Tan et al., 2018

Dataset: CoNLL05

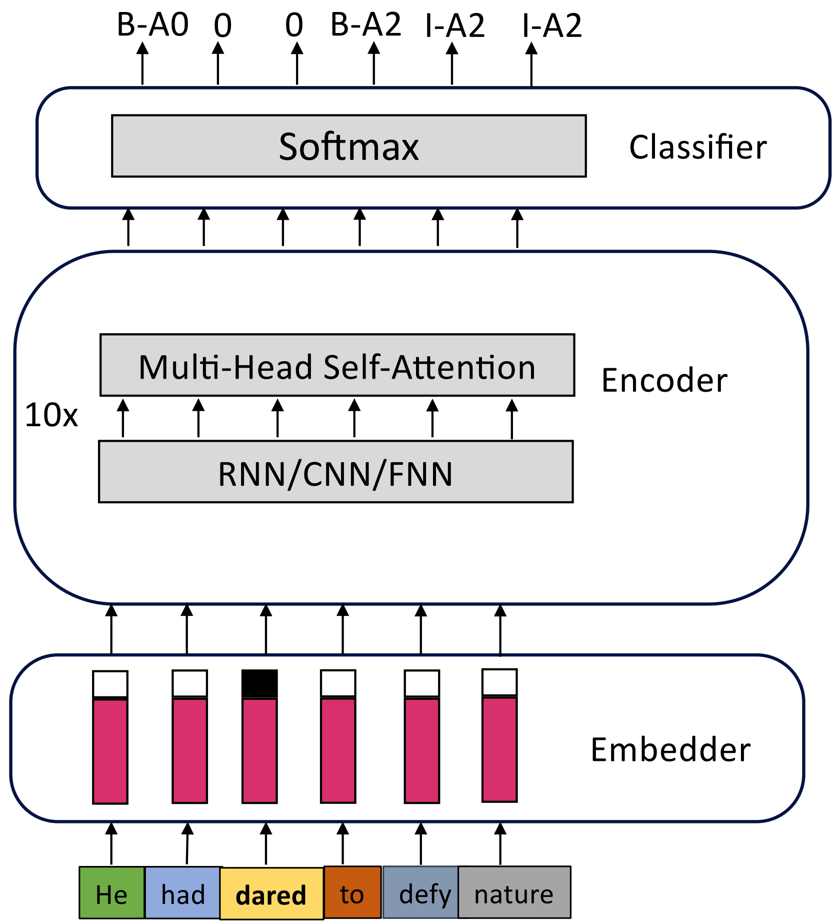




Tan et al., 2018

Takeaways

- Substantial improvements on CoNLL05 WSJ as compared to [He et al., 2017]
- No need of CONSTRAINED Decoding (slows down) Just use Argmax decoding. 83.1 → 83.0 [Token classification]
- As reported earlier, Model depth is the key as compared against model width
- FNN seems better choice over CNN and RNN when attention is used as encoder
- Positional embeddings are necessary to gain actual performance

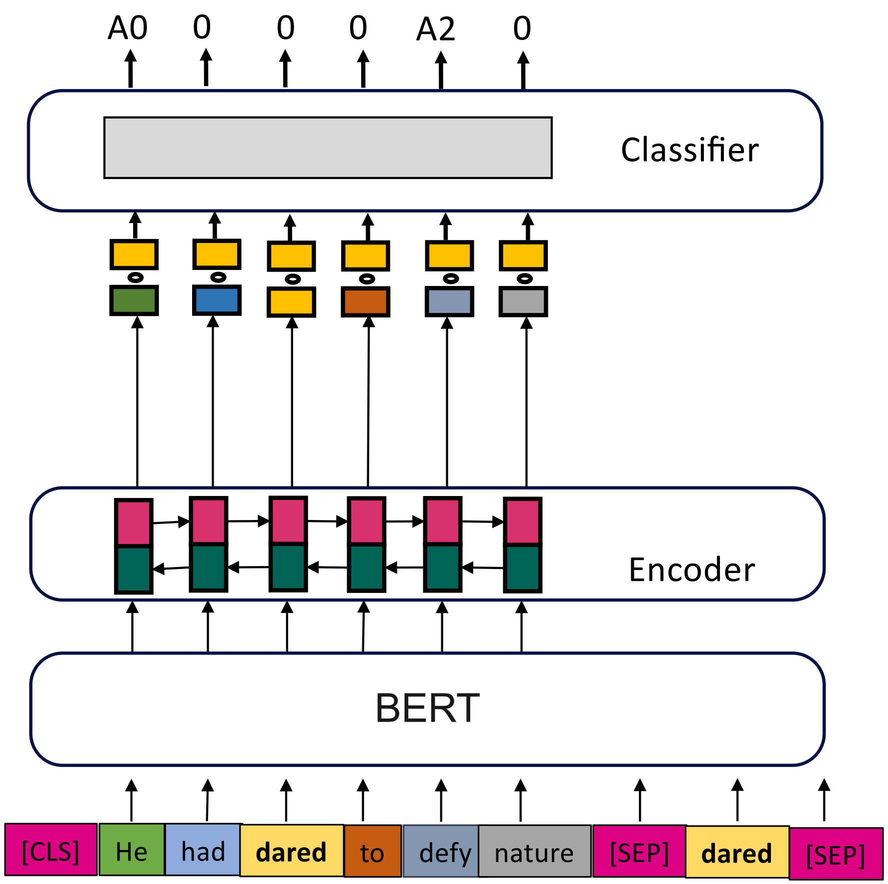


Shi et al., 2019

Are high quality contextual embedding enough for SRL task?

Simple BERT model for relation extraction and SRL

- Use BERT LM to obtain predicate-aware contextualized embeddings for encoder.
- BiLSTMs are encoder layer (1x)
- Concatenate predicate hidden state to the hidden state of the rest of the tokes similar to [Marcheggiani et al., 2017] and then fed into one-layer MLP classification.

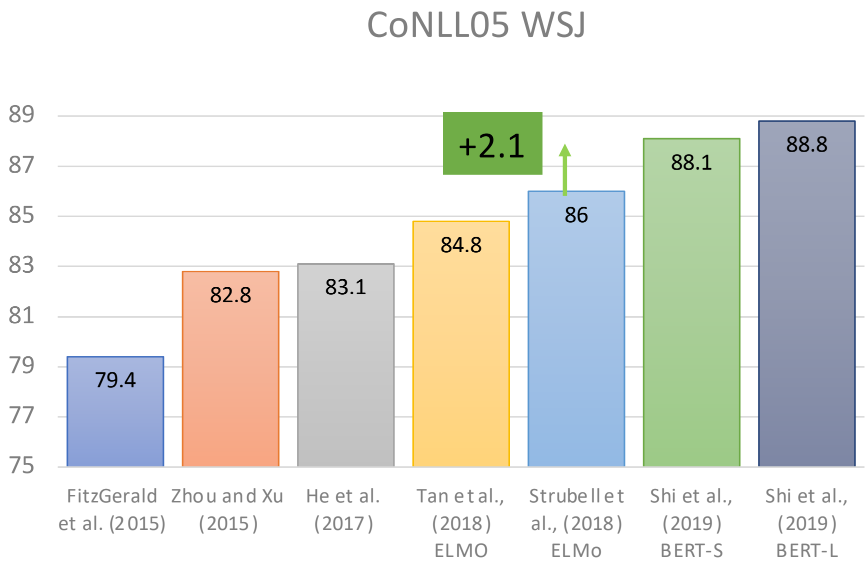


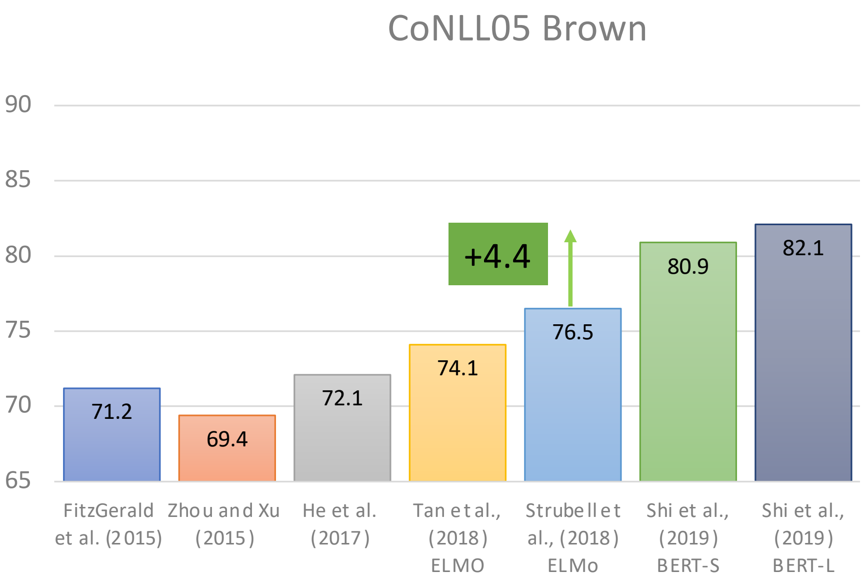
Shi, Peng, and Jimmy Lin. "Simple bert models for relation extraction and semantic role labeling." *arXiv preprint arXiv:1904.05255* (2019).

Shi et al., 2019

Are high quality contextual embedding enough for SRL task?

Dataset: CoNLL05

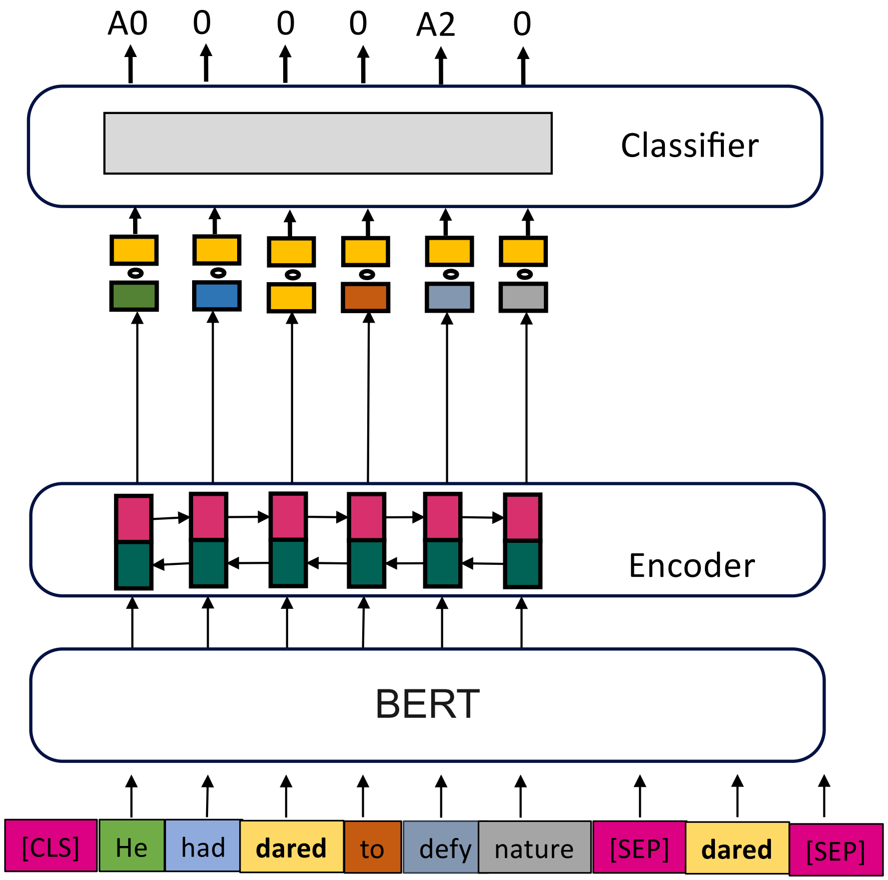




Shi et al., 2019

Are high quality contextual embedding enough for SRL task?

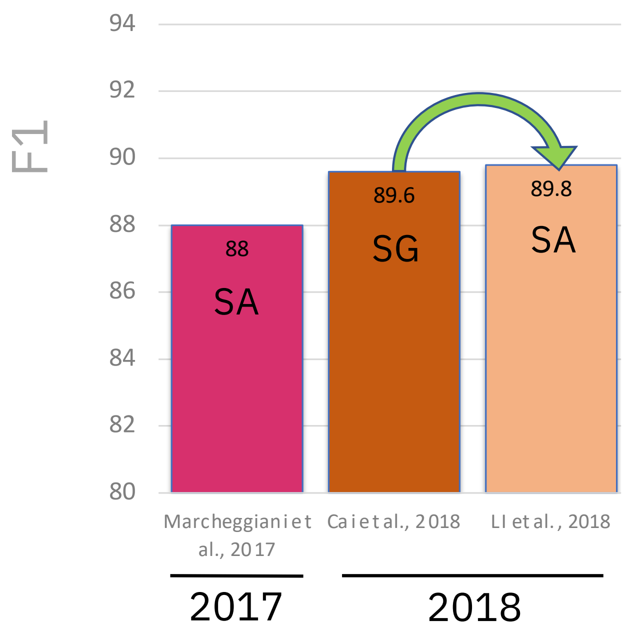
- Powerful Contextualized embeddings is all we need for SRL??
- We do not need syntax to perform better on SRL??
- Do we know if BERT embeddings encodes syntax implicitly??
  - Yes [Jawaher et al., 2019]
  - Explicit syntax information shown to further improve the SoTA SRL performance.



Comparison

Dataset: CoNLL09 EN

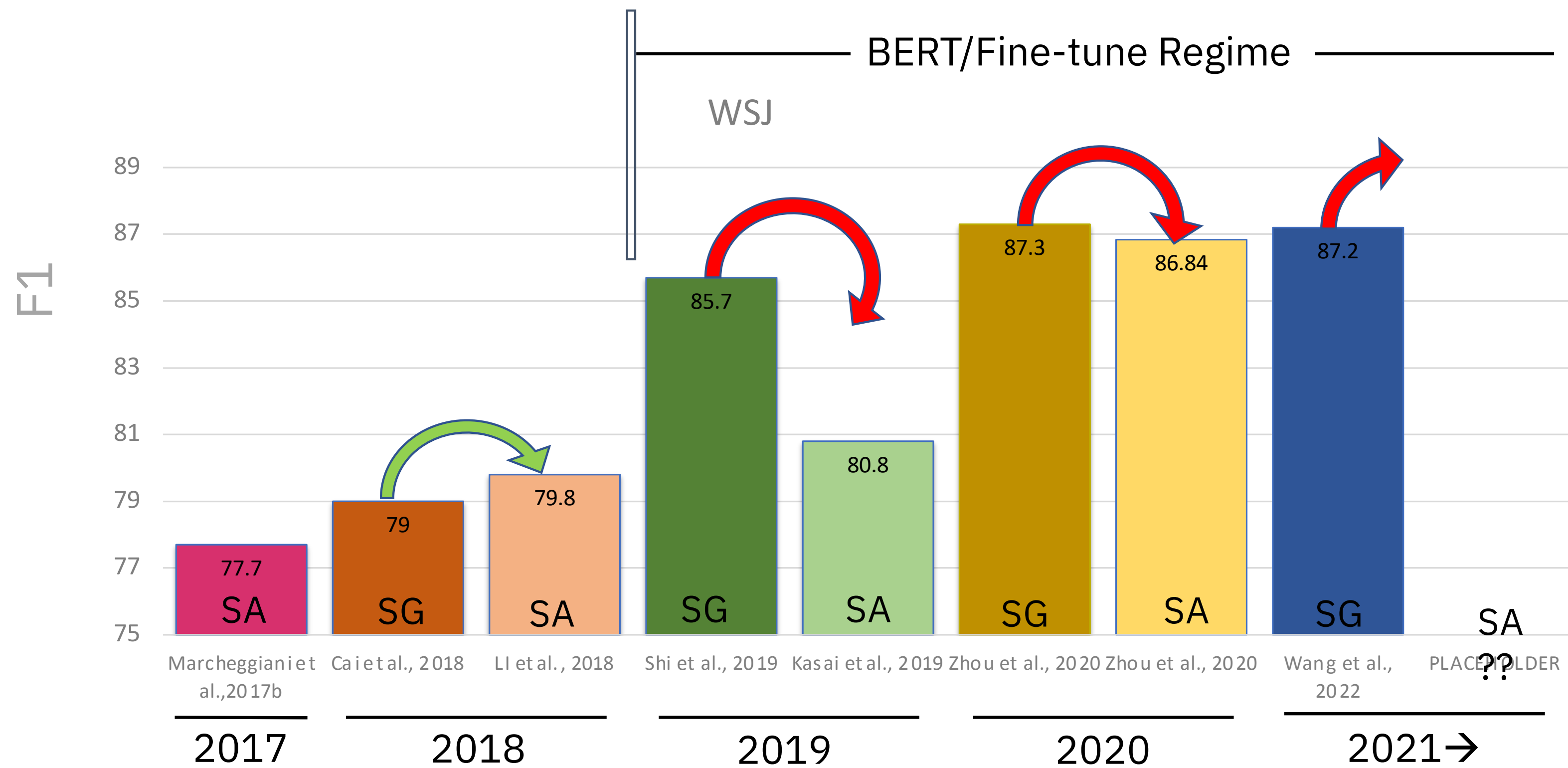
Syntax-agnostic (SG) Vs. Syntax-aware(SA) models



Comparison

Dataset: CoNLL09 EN

Syntax-agnostic (SG) Vs. Syntax-aware(SA) models



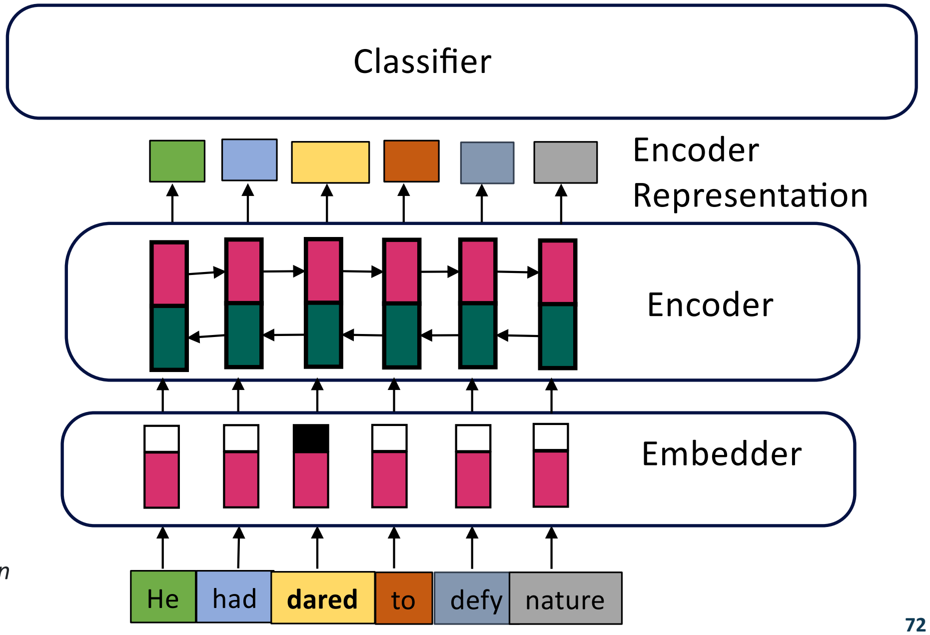
Outline

- **Early SRL approaches**
- **Typical neural SRL model components**
  - Performance analysis
- **Syntax-aware neural SRL models**
  - What, When and Where?
  - Performance analysis
  - How to incorporate Syntax?
- **Syntax-agnostic neural SRL models**
  - Performance Analysis
  - Do we really need syntax for SRL?
  - Are high quality contextual embedding enough for SRL task?
- **Practical SRL systems**
  - Should we rely on this pipelined approach?
    - End-to-end SRL systems
  - Can we jointly predict dependency and span?
- **More recent approaches**
  - Handling low-frequency exceptions
  - Incorporate semantic role label definitions
  - SRL as MRC task
- **Practical SRL system evaluations**
  - Are we evaluating SRL systems correctly?
- **Conclusion**

He et al., 2018

Syntax-agnostic end-to-end SRL system

- Jointly predicting all predicates, arguments spans and the relation between them
- Build upon coreference resolution model [Lee et al., 2017].
- Embedder:
  - No predicate location specified instead concatenate word embeddings with the output of charCNN.
- Each edge is identified by independently predicting which role, if any, holds between every possible pair of text spans, while using aggressive beam pruning for efficiency. The final graph is simply the union of predicted SRL roles (edges) and their associated text spans (nodes)



Luheng He, Kenton Lee, Omer Levy, and Luke Zettlemoyer. 2018. Jointly Predicting Predicates and Arguments in Neural Semantic Role Labeling. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 364–369, Melbourne, Australia. Association for Computational Linguistics.

He et al., 2018

Syntax-agnostic end-to-end SRL system

Task: Predict a set of labeled predicate argument relations

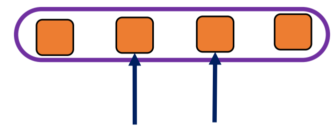
$$Y \subset \mathcal{P} \times \mathcal{A} \times \mathcal{L}$$

Set of all SRL labels

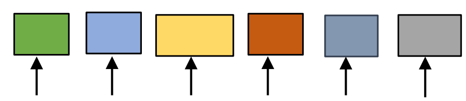
Set of all the possible spans

Set of all tokens

Set of all predicate-argument relations



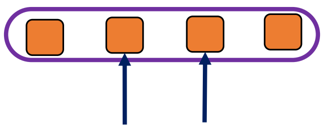
$$P(y_{p,a} = l|X)$$



Encoder Representation

He et al., 2018

Syntax-agnostic end-to-end SRL system



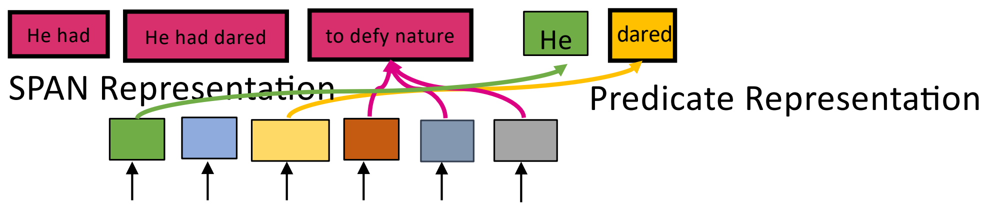
$$P(y_{p,a} = l|X)$$

To obtain predicate and argument representations

Predicate representation is simply the BiLSTM output at the position index p

Argument Representation contains the following:

- End points from BiLSTM ouput
- A soft head word
- Embedded span width feature



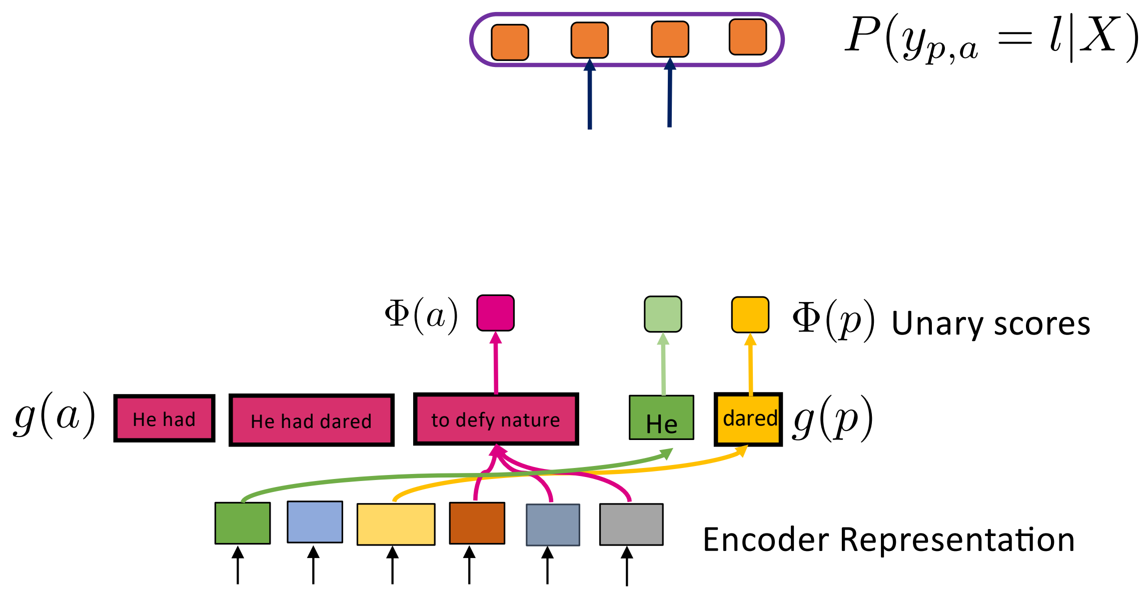
He et al., 2018

Syntax-agnostic end-to-end SRL system

Jointly predicting predicates and Arguments in Neural SRL

Compute Unary score for predicates and arguments

$$\Phi_{\mathrm{a}}(a) = \mathbf{w}_{\mathrm{a}}^{\intercal}\mathrm{MLP}_{\mathrm{a}}(\mathbf{g}(a))$$
$$\Phi_{\mathrm{p}}(p) = \mathbf{w}_{\mathrm{p}}^{\intercal}\mathrm{MLP}_{\mathrm{p}}(\mathbf{g}(p))$$



He et al., 2018

Syntax-agnostic end-to-end SRL system

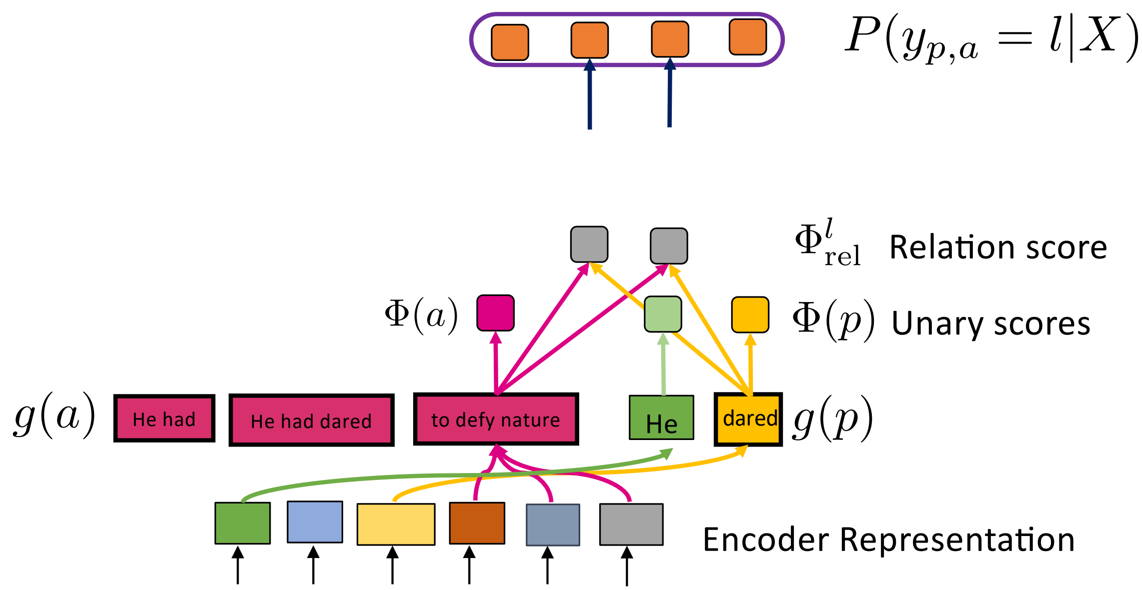
Jointly predicting predicates and Arguments in Neural SRL

Compute Unary score for predicates and arguments

$$\Phi_{\mathrm{a}}(a) = \mathbf{w}_{\mathrm{a}}^{\intercal}\mathrm{MLP}_{\mathrm{a}}(\mathbf{g}(a))$$
$$\Phi_{\mathrm{p}}(p) = \mathbf{w}_{\mathrm{p}}^{\intercal}\mathrm{MLP}_{\mathrm{p}}(\mathbf{g}(p))$$

Compute Relation score between predicates and arguments

$$\Phi_{\mathrm{rel}}^{(l)}(a,p) = \mathbf{w}_{\mathrm{r}}^{(l)\intercal}\mathrm{MLP}_{\mathrm{r}}([\mathbf{g}(a);\mathbf{g}(p)])$$



Number of possible relations

$$\mathcal{O}(n^3|\mathcal{L}|)$$

He et al., 2018

Syntax-agnostic end-to-end SRL system

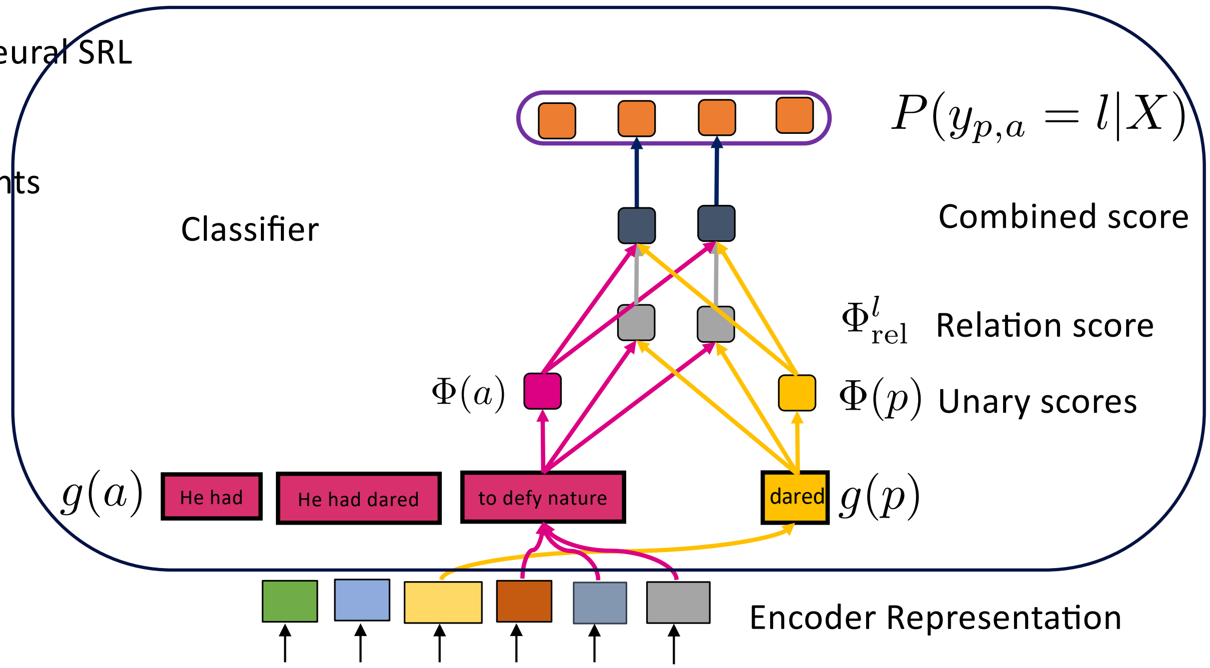
Jointly predicting predicates and Arguments in Neural SRL

Compute Unary score for predicates and arguments

$$\Phi_{\mathrm{a}}(a) = \mathbf{w}_{\mathrm{a}}^{\intercal}\mathrm{MLP}_{\mathrm{a}}(\mathbf{g}(a))$$
$$\Phi_{\mathrm{p}}(p) = \mathbf{w}_{\mathrm{p}}^{\intercal}\mathrm{MLP}_{\mathrm{p}}(\mathbf{g}(p))$$

Compute Relation score between predicates and arguments

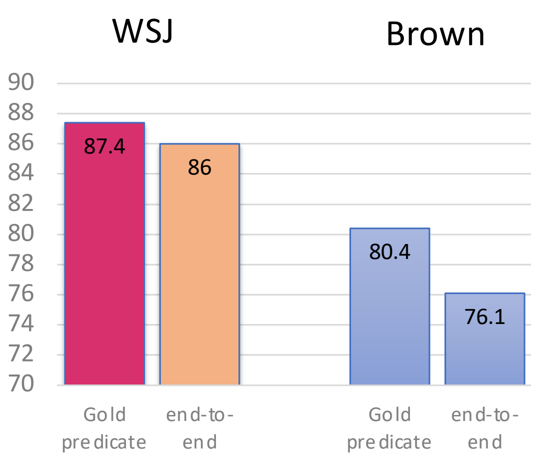
$$\Phi_{\mathrm{rel}}^{(l)}(a,p) = \mathbf{w}_{\mathrm{r}}^{(l)\intercal}\mathrm{MLP}_{\mathrm{r}}([\mathbf{g}(a);\mathbf{g}(p)])$$



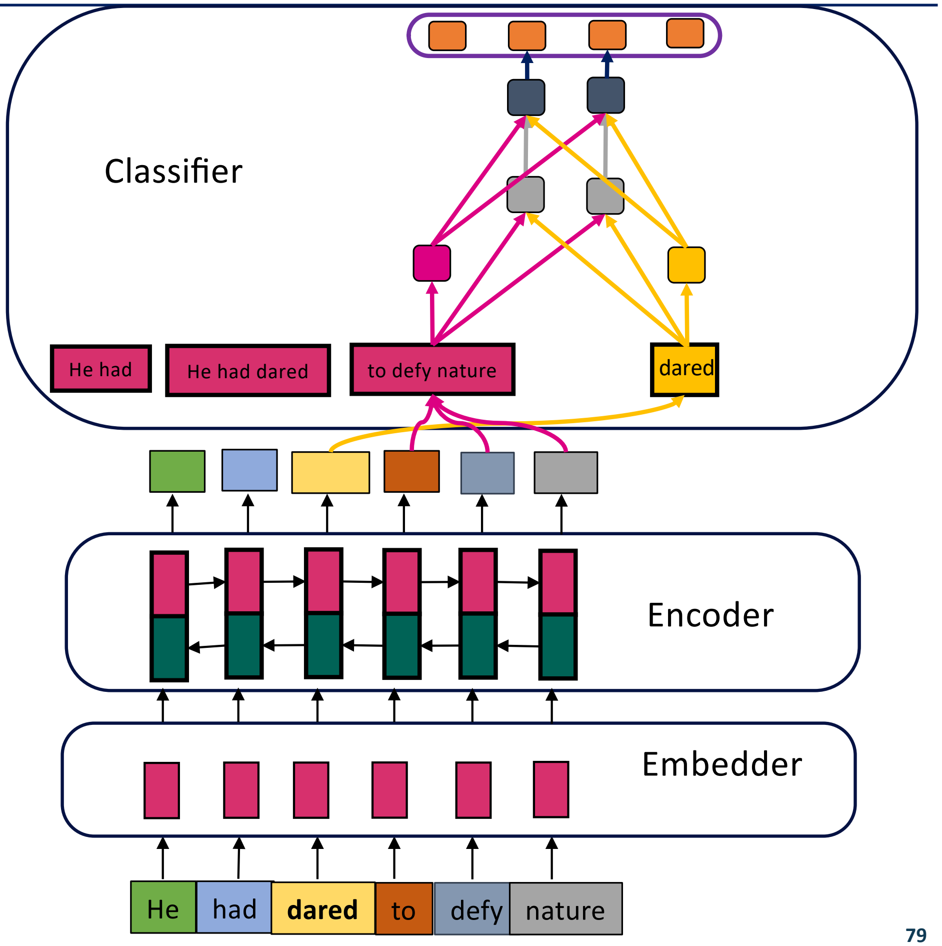
He et al., 2018

- An end-to-end Neural SRL Model

Argument classification results on CoNLL05



Syntax-agnostic end-to-end SRL system

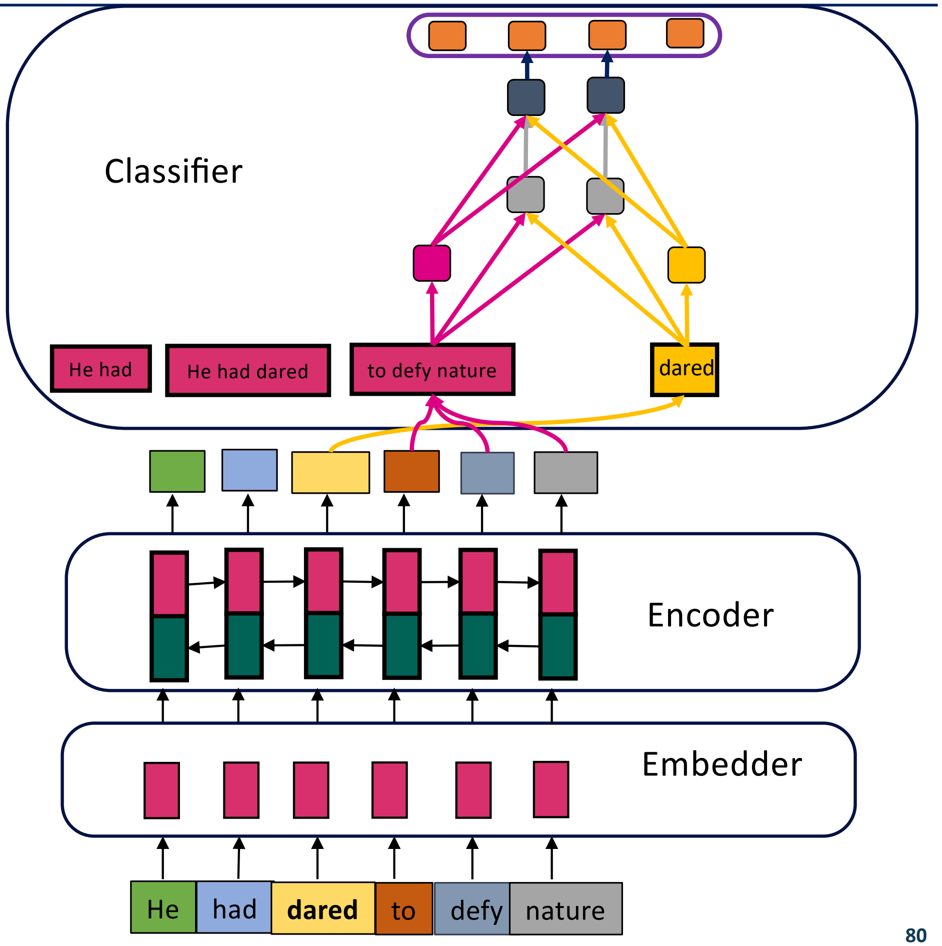


He et al., 2018

Syntax-agnostic end-to-end SRL system

Takeaways

- First end-to-end neural SRL model.
- Strong performance against models with gold predicates.
- Empirically, the model does better at long range dependencies and agreement with syntactic boundaries, but is weaker at global consistency, due to our strong independence assumption



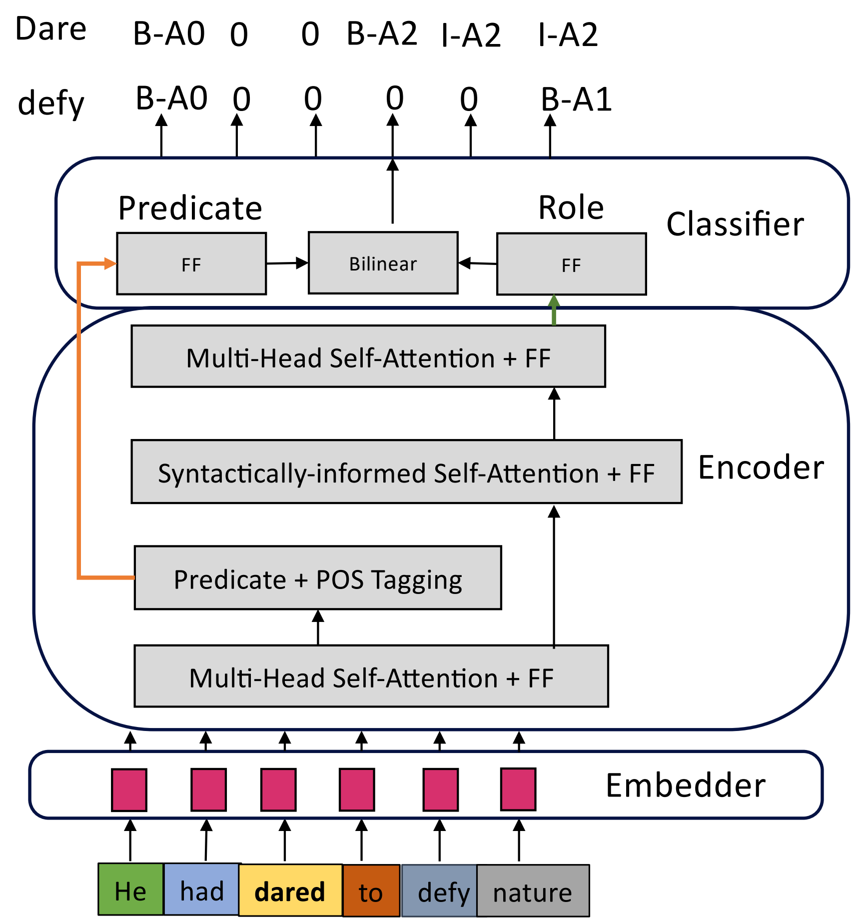
Strubell et al., 2018

Linguistically-Informed Self-Attention for Semantic Role Labeling

Syntax strikes back

- A multi-task learning framework with stacked multi-head self-attention
- Jointly predicts POS and predicates
- Perform parsing
- Attend to syntactic parse parent while assigning semantic role label.

Syntax-aware end-to-end SRL system

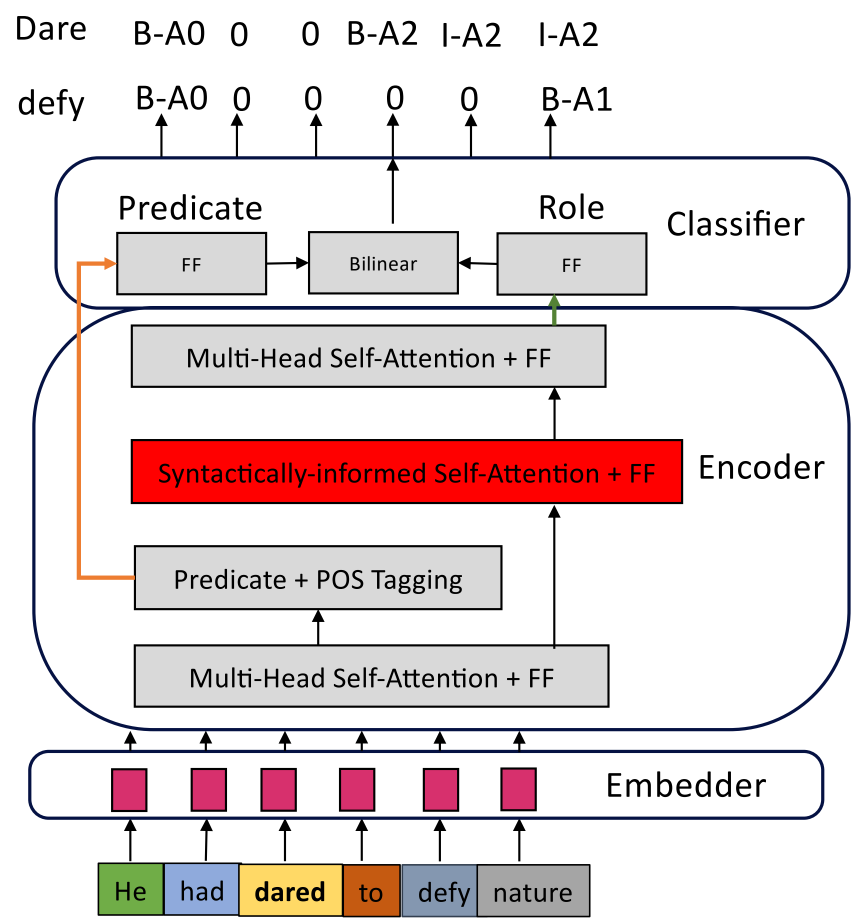


Emma Strubell, Patrick Verga, Daniel Andor, David Weiss, and Andrew McCallum. 2018. [Linguistically-Informed Self-Attention for Semantic Role Labeling](). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5027–5038, Brussels, Belgium. Association for Computational Linguistics.

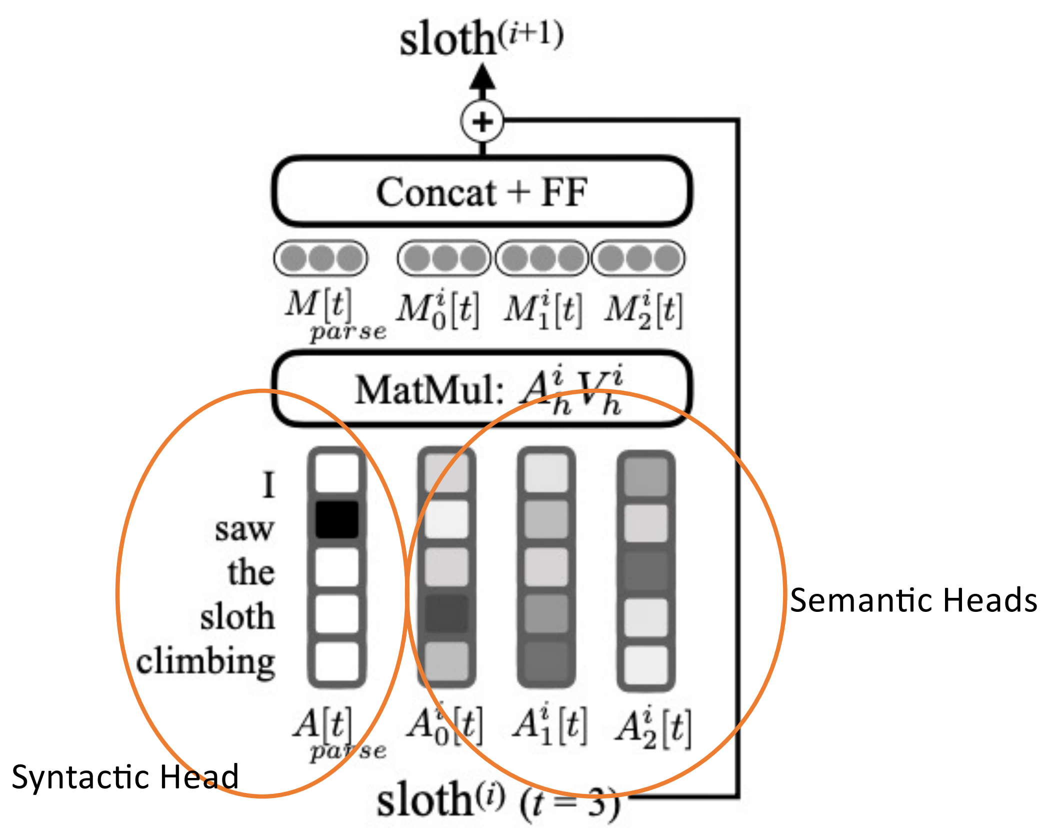
Strubell et al., 2018

- ❑ Replace one attention head with the deep bi-affine model of Dozat and Manning (2017).
- ❑ Use a bi-affine operator U to obtain attention weights for that single head.
- ❑ Encode both the dependency and the dependency label

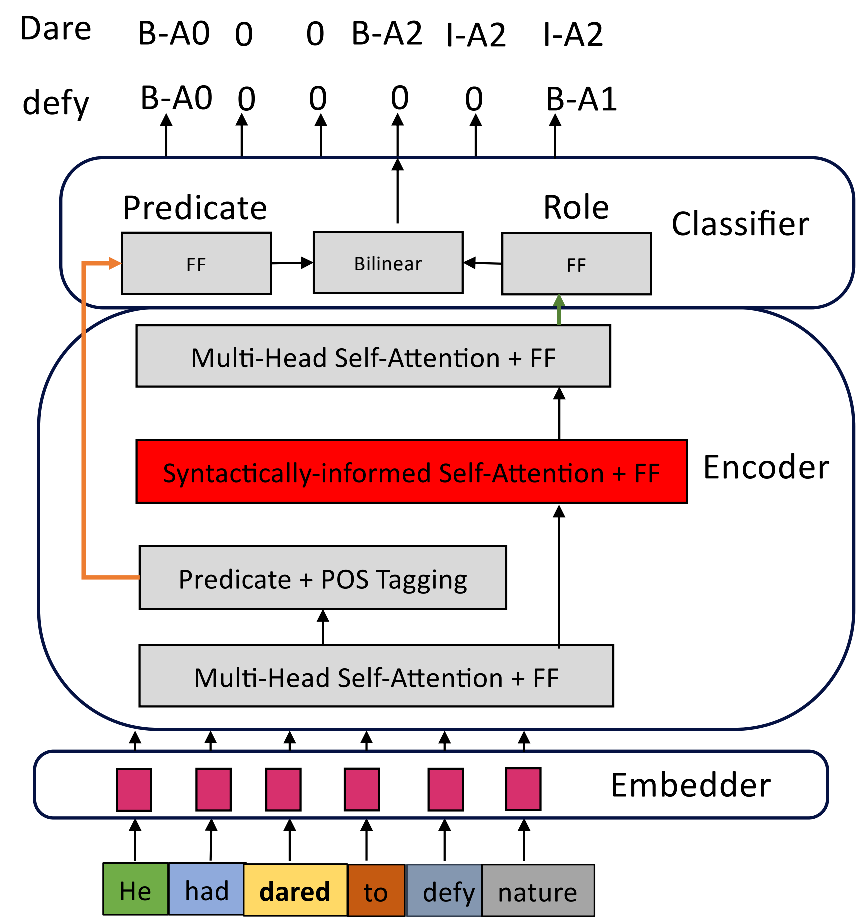
Syntax-aware end-to-end SRL system



Strubell et al., 2018



Syntax-aware end-to-end SRL system



Strubell et al., 2018

Linguistically-Informed Self-Attention for Semantic Role Labeling

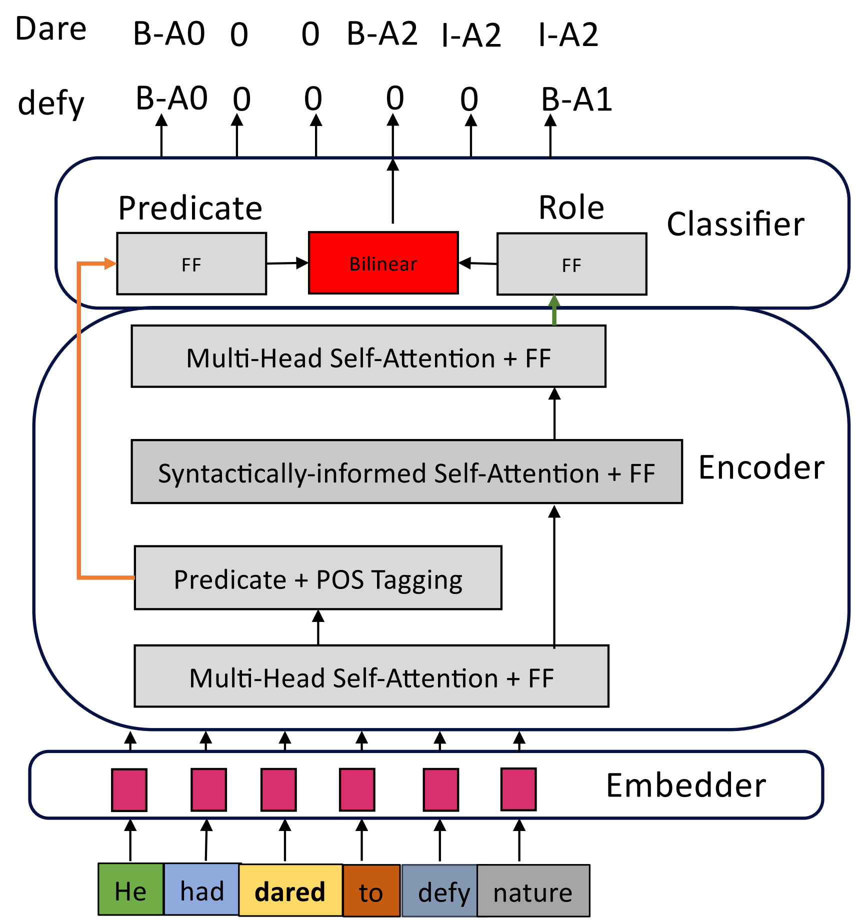
$$s_{ft} = (s_f^{pred})^T U s_t^{role}$$

Predicate-specific representation

Bilinear Transformation operator

Argument-specific representation

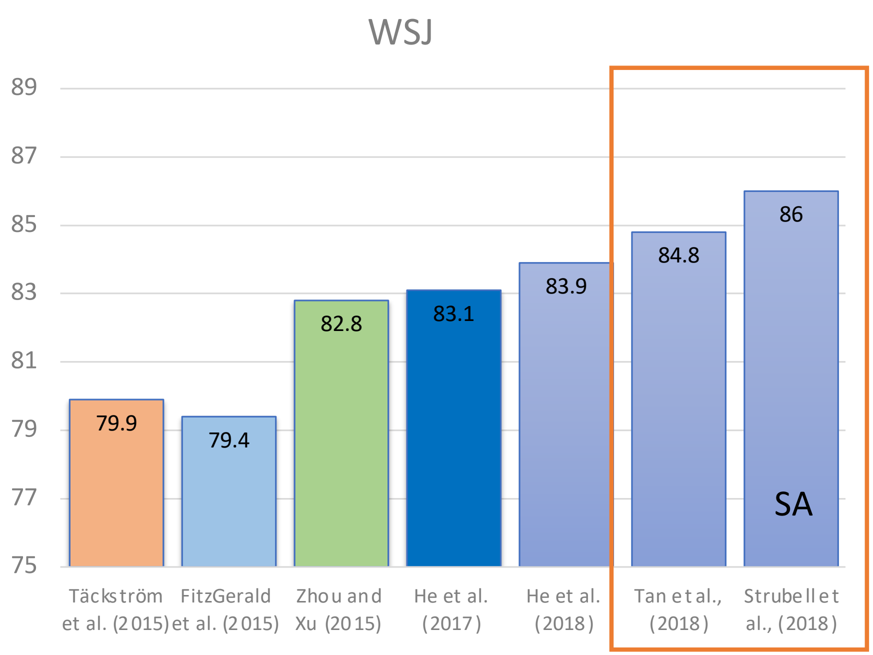
Syntax-aware end-to-end SRL system

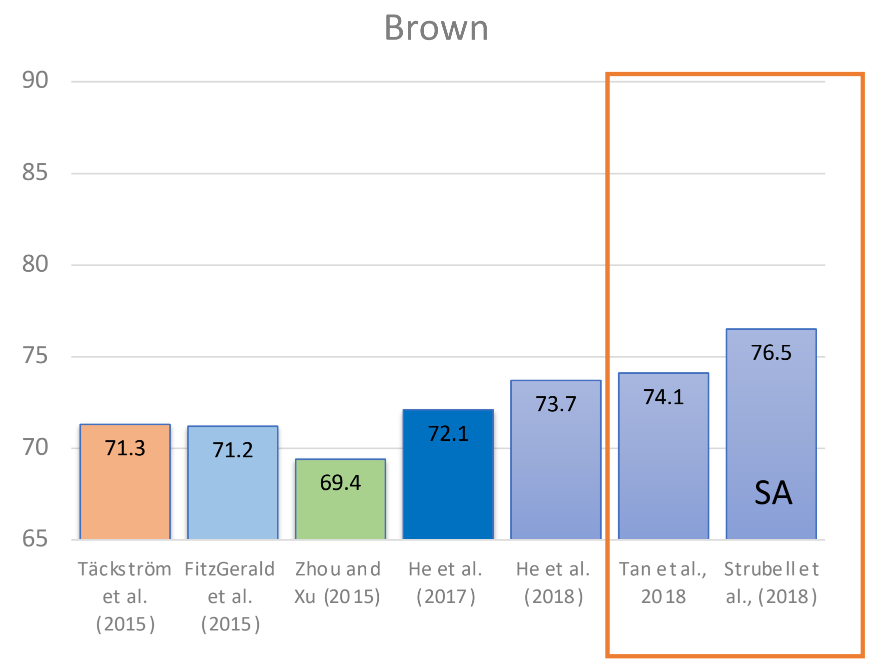


Strubell et al., 2018

Syntax-aware end-to-end SRL system

Dataset: CoNLL05



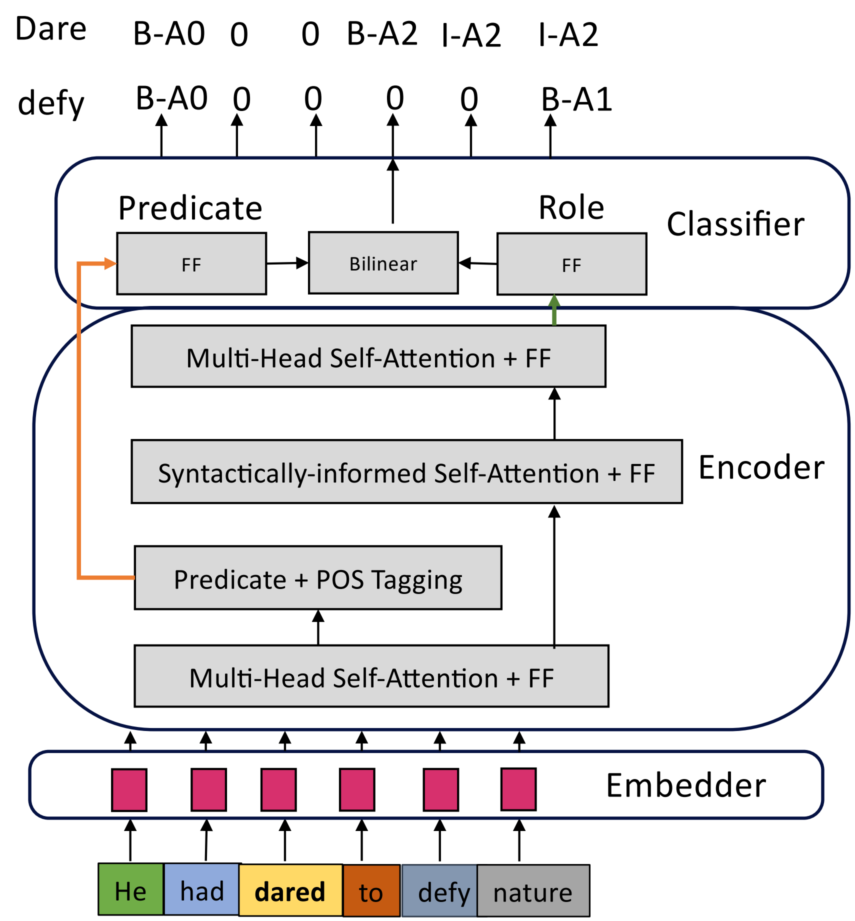


Strubell et al., 2018

Takeaways

- Shows strong performance gain over other methods with and w/o gold predicate location
- Incorporating parse information helpful for resolving span boundary errors (Merge spans, split spans etc.)

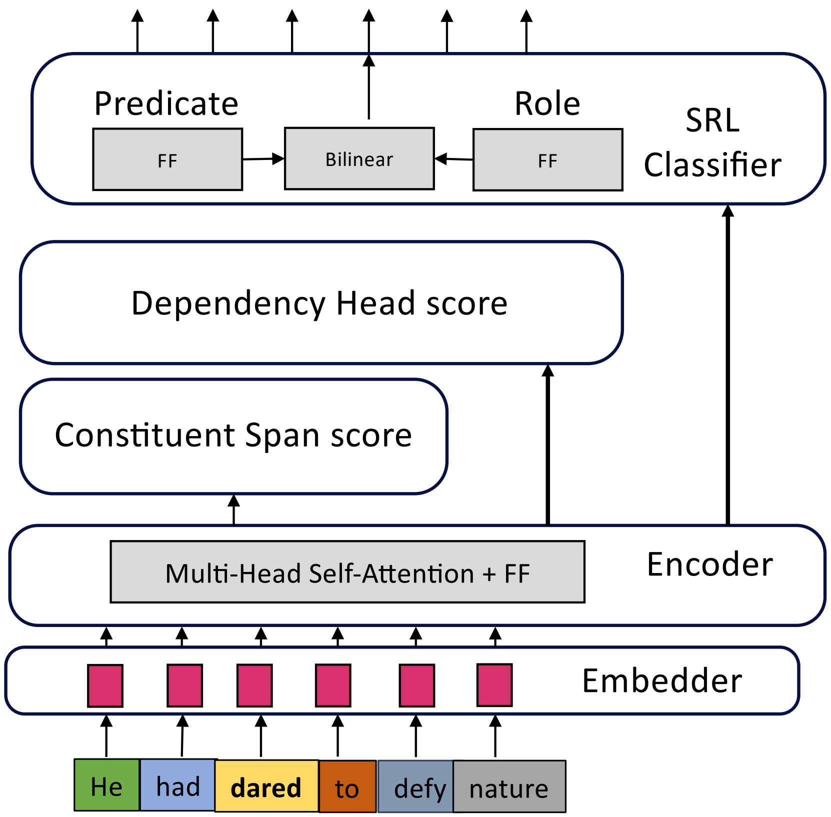
Syntax-aware end-to-end SRL system



Zhou et al., 2019

- ❑ Semantics is usually considered as a higher layer of linguistics over syntax, most previous studies focus on how the latter helps the former
- ❑ Semantics benefit from syntax, but syntax may also benefit from semantics.
- ❑ Joint training of (Multi-task learning) following 5 tasks
  - ❑ Semantic
    - ❑ Dependency
    - ❑ Span
    - ❑ Predicate
  - ❑ Syntax
    - ❑ Constituent
    - ❑ Dependency

Syntax-aware end-to-end SRL system



Junru Zhou, Zuchao Li, and Hai Zhao. 2020. Parsing All: Syntax and Semantics, Dependencies and Spans. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4438–4449, Online. Association for Computational Linguistics.

Zhou et al., 2019

Can we jointly predict dependency and span?

Interesting Insights

SEMANTICS

- ❑ Joint training of dependency and span for SRL helps improve both. Further strengthened by Fei et al. (2021).
- ❑ Further improve for both is observed when combined with syntactic constituent.
- ❑ Not so when combined with syntactic dependency

SYNTAX

- ❑ Though marginal, semantic do improve syntax

System	SEM_{span}	SEM_{dep}	SYN_{con}	SYN_{dep}	
	F_1	F_1	F_1	UAS	LAS
End-to-end					
SEM_{span}	82.27	–	–	–	–
SEM_{dep}	–	84.90	–	–	–
$SEM_{span,dep}$	83.50	84.92	–	–	–
$SEM_{span,dep}$, SYN_{con}	**83.81**	**84.95**	**93.98**	–	–
$SEM_{span,dep}$, SYN_{dep}	83.13	84.24	–	95.80	94.40
$SYN_{con,dep}$	–	–	93.78	95.92	94.49
$SEM_{span,dep}$, $SYN_{con,dep}$	83.12	83.90	**93.98**	**95.95**	**94.51**
Given predicate					
SEM_{span}	83.16	–	–	–	–
SEM_{dep}	–	88.23	–	–	–
$SEM_{span,dep}$	84.74	88.32	–	–	–
$SEM_{span,dep}$, SYN_{con}	84.46	**88.40**	93.78	–	–
$SEM_{span,dep}$, SYN_{dep}	**84.76**	87.58	–	95.94	94.54
$SEM_{span,dep}$, $SYN_{con,dep}$	84.43	87.58	**94.07**	**96.03**	**94.65**

Table 2 from the paper: Joint learning analysis on CoNLL-2005, CoNLL-2009, and PTB dev sets

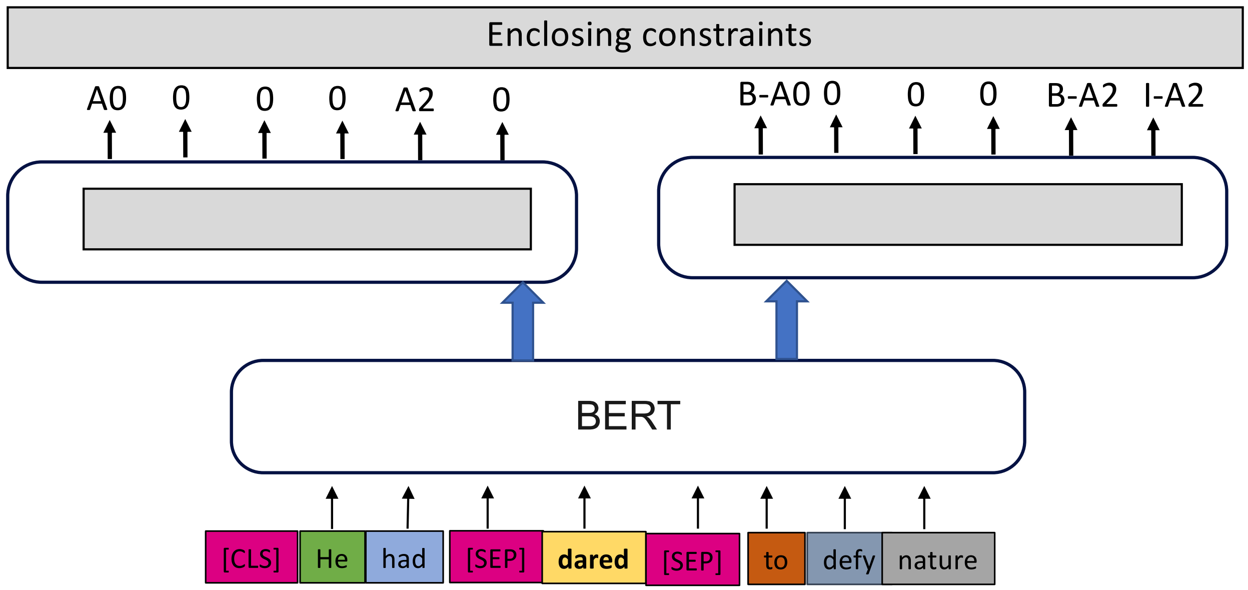
Hao Fei, Shengqiong Wu, Yafeng Ren, Fei Li, and Donghong Ji. 2021. Better Combine Them Together! Integrating Syntactic Constituency and Dependency Representations for Semantic Role Labeling. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 549–559, Online. Association for Computational Linguistics.

Jindal et al., 2022

Can we jointly predict dependency and span?

SPADE: **SPA**n and **DE**pendency SRL model

A multi-task learning framework
- Train simultaneously on argument heads and the argument spans.



Observations:
- ❑ Slight drop on argument head performance.
- ❑ Gain on argument span performance.

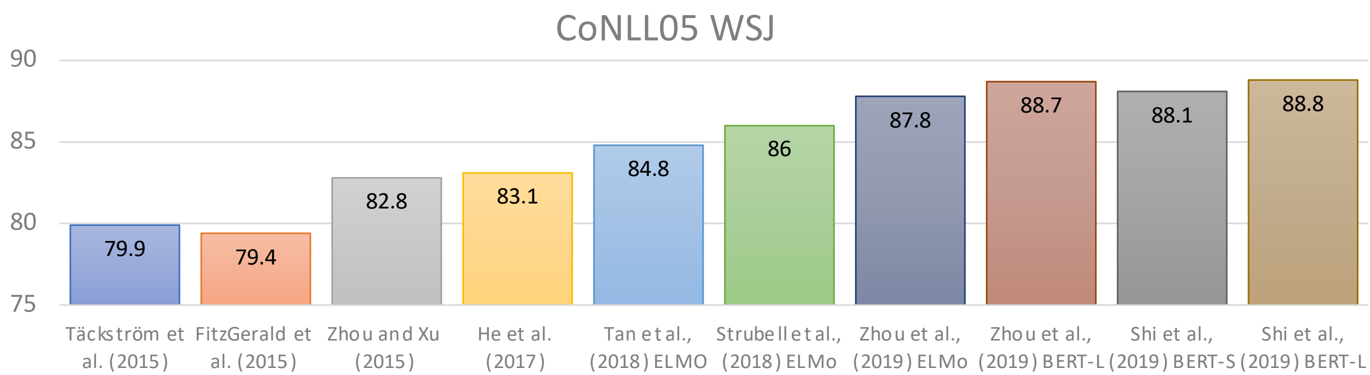
These observations are consistent with Zhou et at., 2019

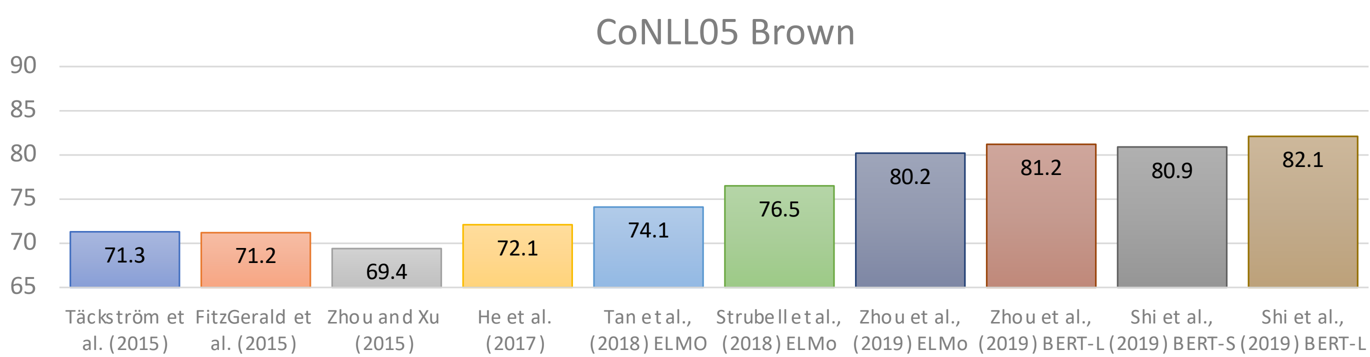
Ishan Jindal, Alexandre Rademaker, Michał Ulewicz, Ha Linh, Huyen Nguyen, Khoi-Nguyen Tran, Huaiyu Zhu, and Yunyao Li. 2022. Universal Proposition Bank 2.0. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1700–1711, Marseille, France. European Language Resources Association.

Zhou et al., 2019

Can we jointly predict dependency and span?

Parsing All: Syntax and Semantics, Dependencies and Spans





Outline

- **Early SRL approaches**
- **Typical neural SRL model components**
  - Performance analysis
- **Syntax-aware neural SRL models**
  - What, When and Where?
  - Performance analysis
  - How to incorporate Syntax?
- **Syntax-agnostic neural SRL models**
  - Performance Analysis
  - Do we really need syntax for SRL?
  - Are high quality contextual embedding enough for SRL task?
- **Practical SRL systems**
  - Should we rely on this pipelined approach?
    - End-to-end SRL systems
  - Can we jointly predict dependency and span?
- **More recent approaches**
  - Learn low-frequency exceptions
  - Incorporate semantic role label definitions
  - SRL as MRC task
- **Practical SRL system evaluations**
  - Are we evaluating SRL systems correctly?
- **Conclusion**

Low-frequency Exceptions

Argument labeling task:

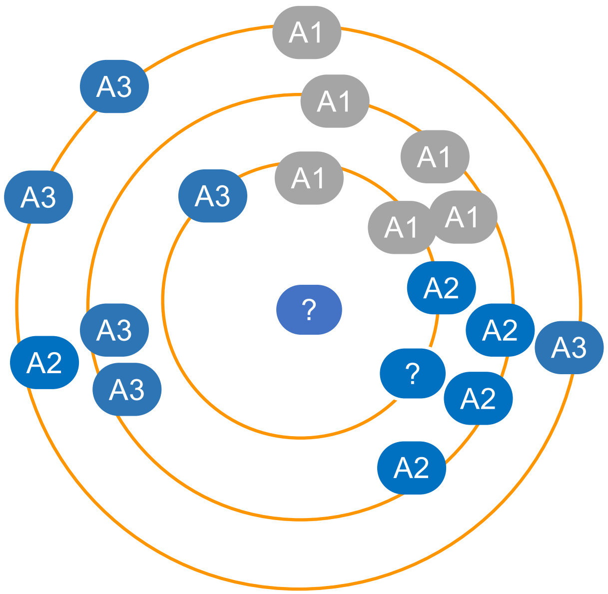
- Arguments that are syntactically realized as passive subjects are typically labeled Arg1
  - However, there exist numerous low-frequency exceptions to this rule.
    - Passive subjects of certain frames (such as the frame TELL.01) are most commonly labeled as Arg2

Observations based on CoNLL09 training data [Akbik and Li, 2015]:

- 57% of all subjects are labeled A0
- 33% of all subjects are labeled A1

- 74% of active subjects are labeled A0
- 86% of passive subjects are labeled A1
  - 100% of passive subjects of SELL.01 are labeled A1
  - 88% of passive subjects of TELL.01 are labeled A2

Low-frequency Exceptions



Instance-based learning

- ❑ Extrapolates predictions from the most similar instances in the training data. [Akbik and Li, 2016, Jindal et al., 2020]
- ❑ Generally, staged approaches where base model is trained first to get the word/span representations. [Guan et al., 2019, Jindal et al., 2020]

[Akbik and Li, 2016] Alan Akbik and Yunyao Li. 2016. K-SRL: Instance-based Learning for Semantic Role Labeling. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 599–608, Osaka, Japan. The COLING 2016 Organizing Committee.

[Guan et al., 2019] Chaoyu Guan, Yuhao Cheng, and Hai Zhao. 2019. Semantic Role Labeling with Associated Memory Network. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3361–3371, Minneapolis, Minnesota. Association for Computational Linguistics.

[Jindal et al., 2020] Jindal, Ishan, Ranit Aharonov, Siddhartha Brahma, Huaiyu Zhu, and Yunyao Li. "Improved Semantic Role Labeling using Parameterized Neighborhood Memory Adaptation." *arXiv preprint arXiv:2011.14459* (2020).

Understand BERT for SRL

Understanding BERT based model better for better SRL performance.

BERT "rediscovers" the classical NLP pipeline [Tenney et al., 2019]

- ❑ Lower layers tend to encode mostly lexical level information, while
- ❑ Upper layers seem to favor sentence-level information.

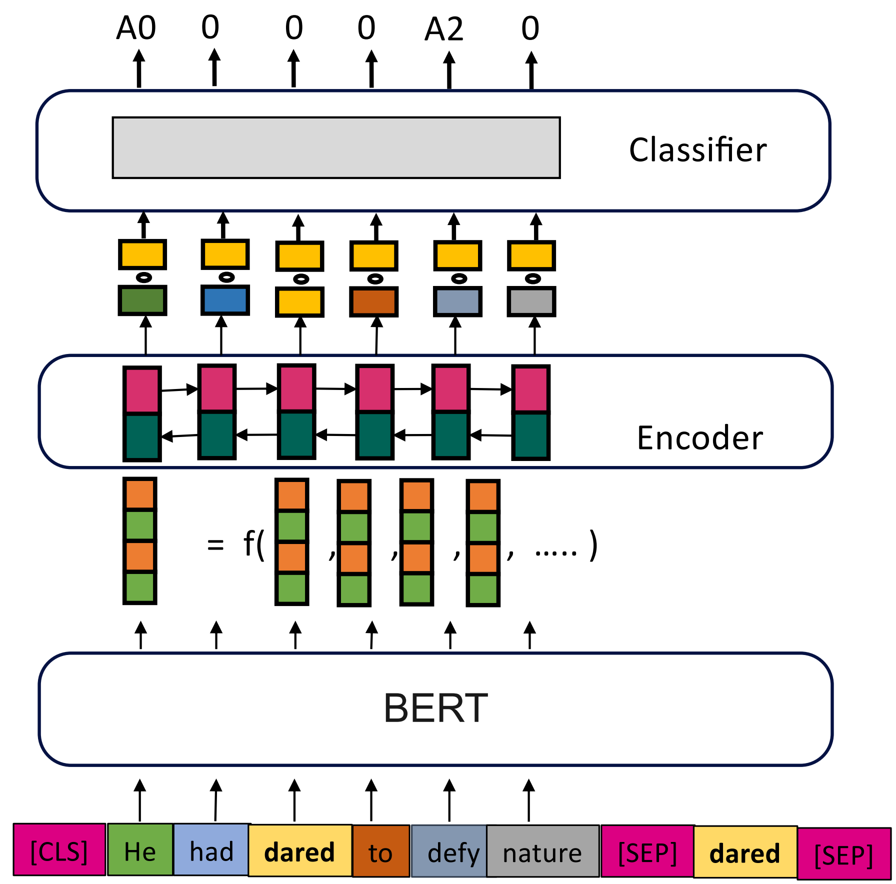
Ilia Kuznetsov and Iryna Gurevych. 2020. A matter of framing: The impact of linguistic formalism on probing results. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 171–182, Online. Association for Computational Linguistics.

Understand BERT for SRL

Understanding BERT based model better for better SRL performance.

		BERT	RoBERTa	m-BERT	XLM-R
Linear	Random	72.8	72.8	–	–
	Static	75.1	75.3	–	–
	Top-4	85.3	85.3	–	–
	W-Avg	**85.7**	**86.1**	–	–
Non-Linear	Random	75.9	75.9	75.8	75.7
	Static	76.3	76.5	76.2	76.3
	Top-4	89.2	88.8	88.0	88.9
	W-Avg	**89.4**	**89.3**	**88.8**	**89.1**

Static: Last layer activations as static embeddings
Top-4: Concatenate top 4 layers activations
W-avg: Parametric sum of all layer activations



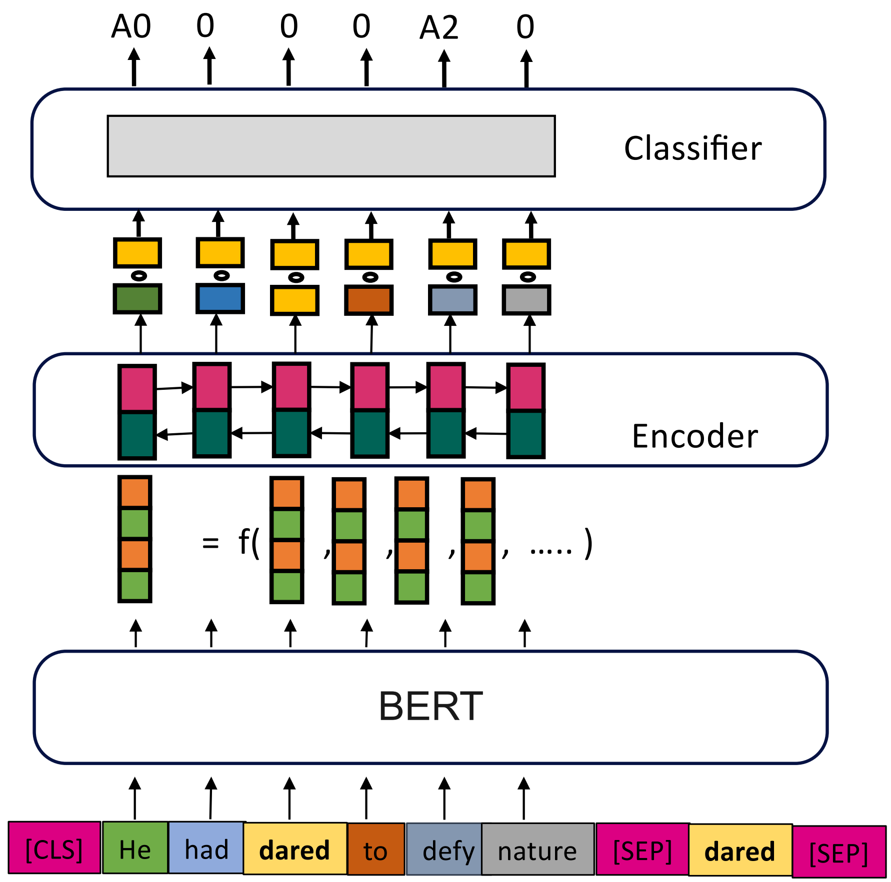
Simone Conia and Roberto Navigli. 2022. Probing for Predicate Argument Structures in Pretrained Language Models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4622–4632, Dublin, Ireland. Association for Computational Linguistics.

Understand BERT for SRL

Understanding BERT based model better for better SRL performance.

Interesting Insights

- ❑ Predicate senses and argument structures are encoded at different layers in LMs
- ❑ Verbal and nominal predicate-argument structures are represented differently across the layers of a LM;
  - ❑ SRL system benefits from treating them separately;



Incorporating Role Definitions

Label-aware NLP

- Model is given the definitions of labels, and can effectively leverage them in many tasks
  - Sentiment/entailment: (Schick and Schutze, 2021)
  - Event extraction: (Du and Cardie, 2020; Hongming et al., 2021)
  - Word sense disambiguation: (Kumar et al., 2019)
- Strong even with few-shot

- Many more, but NOT for SRL (why?)
  - Semantic roles are specific to predicates
  - There are many predicates, thus many roles; very sparse
  - 8500 Predicate senses in CoNLL09 data
  - ~8500*3 argument labels ~ 25K

Incorporating Role Definitions

Label-aware NLP for SRL

[Zhang et al., 2022]

- ❑ Make n+1 copies of the sentence where n is number of core arguments defined for frame.
  - ❑ N is number of core arguments
  - ❑ +1 is for contextual arguments
- ❑ Append label definition at the end of the sentence.
- ❑ Convert K class classification problem into binary class classification.
  - ❑ That is to determine whether a token is worker or not in this example.

He	[SEP]	works	[SEP]	at	ACL	now	.
A0	O	O	O	O	A2	TMP	O

Predicate "work" has sense *work.01*
Frame of *work.01*

```
<roleset sense="work.01" lemma="work">
<roles>
  <role descr="worker" n="A0"/>
  <role descr="job, project" n="A1"/>
  <role descr="employer" n="A2"/>
  ...
</roles>
```

He	[SEP]	works	[SEP]	at	ACL	now	.	[SEP]	worker
A	O	O	O	O	O	O	O	O	O

He	[SEP]	works	[SEP]	at	ACL	now	.	[SEP]	job
O	O	O	O	O	O	O	O	O	O

He	[SEP]	works	[SEP]	at	ACL	now	.	[SEP]	employer
O	O	O	O	O	A	O	O	O	O

He	[SEP]	works	[SEP]	at	ACL	now	.	[SEP]	context.
O	O	O	O	O	O	TMP	O	O	O

Li Zhang, Ishan Jindal, and Yunyao Li. 2022. Label Definitions Improve Semantic Role Labeling. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5613–5620, Seattle, United States. Association for Computational Linguistics.

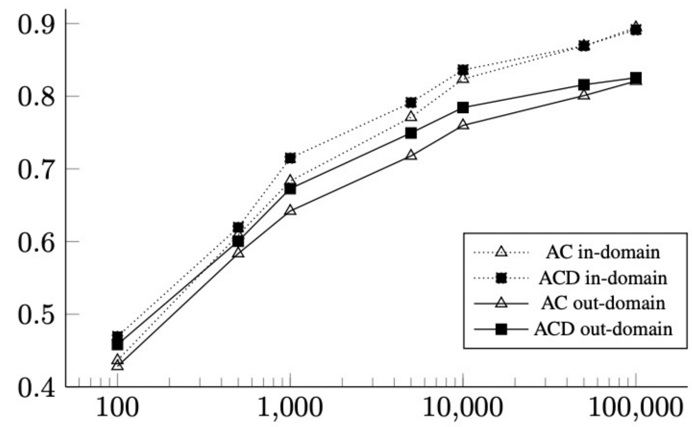
Incorporating Role Definitions

Interesting Insights

Low-Frequency Predicates.
- SRL suffers from the long-tail phenomenon.
- LD outperforms base by up to 4.4 argument F1 for unseen predicates, notably helping with low-frequency predicates.

Few-Shot Learning.
- LD outperforms base by up to 3.2 F1 in- and out-domain.
- The performance gap diminishes as training size approaches 100, 000.

Distant Domain Adaptation
- evaluate models trained on CoNLL09 (news articles) on the Biology PropBank.
- LD model achieves 55.5 argument F1, outperforming base which achieves 54.6..



SRL as MRC Task

SRL as extractive machine Reading Comprehension task [Wang et al., 2022]

Input Sentence
The stock has been < p> beaten </p> down for two days.

Multiple-Choice MRC for Predicate Disambiguation
Question: What is the sense of predicate "beaten"?
A. (Cause) pulsating motion that often makes sound
B. push, cause motion
C. win over some competitor
Answer: B

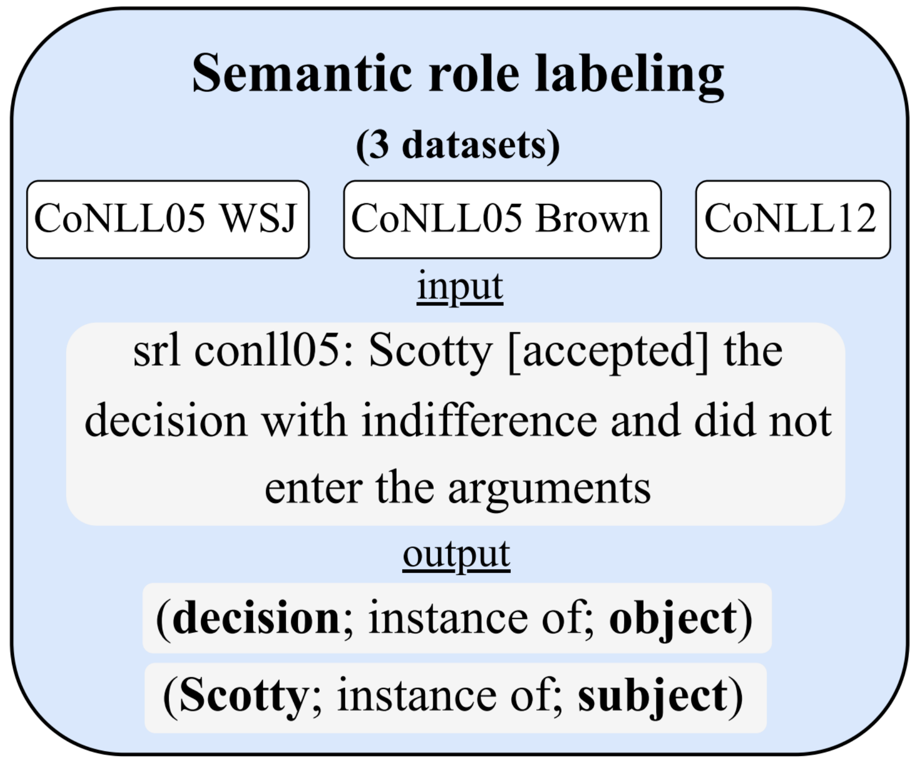
Extractive MRC for Argument Labeling
Question for A0: What are the arguments with meaning "causer of motion"?
Answer: No Answer
Question for A1: What are the arguments with meaning "thing moving"?
Answer: the stock
Question for A2: What are the arguments with meaning "direction, destination"?
Answer: down
Question for TMP: What are the time modifiers of predicate "beaten"?
Answer: for two days

Nan Wang, Jiwei Li, Yuxian Meng, Xiaofei Sun, Han Qiu, Ziyao Wang, Guoyin Wang, and Jun He. 2022. An MRC Framework for Semantic Role Labeling. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 2188–2198, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

SRL Generation

DEEPSTRUCT: Pretraining of Language Models for Structure Prediction [Wang et al., 2022]

- SRL as structure prediction task
- Reformulating structure prediction as a series of unit tasks–triple prediction tasks
- Showed significant performance gain over autoencoders models.



Wang, Chenguang, Xiao Liu, Zui Chen, Haoyun Hong, Jie Tang, and Dawn Song. "DeepStruct: Pretraining of Language Models for Structure Prediction." In *Findings of the Association for Computational Linguistics: ACL 2022*, pp. 803-823. 2022.

Outline

- ❑ **Early SRL approaches**
- ❑ **Typical neural SRL model components**
  - ❑ Performance analysis
- ❑ **Syntax-aware neural SRL models**
  - ❑ What, When and Where?
  - ❑ Performance analysis
  - ❑ How to incorporate Syntax?
- ❑ **Syntax-agnostic neural SRL models**
  - ❑ Performance Analysis
  - ❑ Do we really need syntax for SRL?
  - ❑ Are high quality contextual embedding enough for SRL task?
- ❑ **Practical SRL systems**
  - ❑ Should we rely on this pipelined approach?
    - ❑ End-to-end SRL systems
  - ❑ Can we jointly predict dependency and span?
- ❑ **More recent approaches**
  - ❑ Handling low-frequency exceptions
  - ❑ Incorporate semantic role label definitions
  - ❑ SRL as MRC task
- ❑ **Practical SRL system evaluations**
  - ❑ Are we evaluating SRL systems correctly?
- ❑ **Conclusion**

SRL Evaluation – Issues with Evaluation Metrics

Two official evaluation scripts

- ❑ Evaluation script from CoNLL05 Shared task (eval05.pl)
- ❑ Evaluation script from CoNLL09 Shared task (eval09.pl)

ERROR PROPOGATATION

	Eval05.pl	Eval09.pl
Predicate identification	✗ Assume gold predicate location	✗ Assume gold predicate location
Predicate sense disambiguation	✗	✓
Argument identification	✓ Span only	✓ Head only
Argument classification	✓	✓

All tasks are evaluated independently

SRL Evaluation – Predicate Error Types

Example:

text = Yesterday, John bought a car.

ID	FORM	FLAG	PRED SENSE	Predicate-argument prediction			
				Gold	P1	P2	P3
1	Yesterday	_	_	TMP	TMP	TMP	TMP
2	,	_	_	_	_	_	_
3	John	_	_	A0	A0	A0	A0
4	bought	Y	buy.01	buy.01	*buy_out.03*	*buy.05*	*sell.01*
5	a	_	_	_	_	_	_
6	car	_	_	A1	A1	A1	A1
7	.	_	_	_	_	_	_

Predicate sense evaluation			Gold	P1	P2	P3
	R	Eval05.pl	Do not evaluate			
		Eval09.pl	1/1	0/1	0/1	1/1
	P	Eval05.pl	Do not evaluate			
		Eval09.pl	1/1	0/1	0/1	1/1

SRL Evaluation – Error Examples

Real errors from a SoTA SRL model.

All of these predicate senses are marked correct by the CoNLL09 evaluation script.

Id	Sentence	Gold	Predicted
1	He was able, now, to sit for hours in a chair in the living room and stare out at the bleak yard without moving.	stare.01	look.01
2	She greeted her husband's colleagues with smiling politeness , offering nothing.	politeness.01	minimalism.01
3	It was a Negro section of peeling row houses, store-front churches and ragged children.	peel.01	peer.01
4	He was calm, drugged , and lazy.	drug.01	dropper.01
5	The walk and his fears had served to overheat him and his sweaty armpits cooled at the touch of the night air.	overheat.01	soothe.01
6	He did not resent their supervision or Virginia's sometimes tiring sympathy.	tire.01	hiring.01

SRL Evaluation – Argument Error Types

Example:

text = Yesterday, John bought a car.

ID	FORM	FLAG	PRED SENSE	Predicate-argument prediction			
				Gold	P1	P2	P3
1	Yesterday	_	_	TMP	TMP	TMP	TMP
2	,	_	_	_	_	_	_
3	John	_	_	A0	A0	A0	A0
4	bought	Y	buy.01	buy.01	*buy_out.03*	*buy.05*	*sell.01*
5	a	_	_	_	_	_	_
6	car	_	_	A1	A1	A1	A1
7	.	_	_	_	_	_	_

Argument evaluation	R	Eval05.pl	3/3	0/3	3/3	0/3
		Eval09.pl	3/3	3/3	3/3	3/3
	P	Eval05.pl	3/3	0/3	3/3	0/3
		Eval09.pl	3/3	3/3	3/3	3/3

An Improved Evaluation Scheme

Summary of issues with existing SRL evaluation metric:

- ❑ Proper evaluation of predicate sense disambiguation task;
- ❑ Argument label evaluation in conjunction with predicate sense;
- ❑ Proper evaluation for discontinuous arguments and reference arguments; and
- ❑ Unified evaluation of argument head and span.

PriMeSRL-Eval: A Practical Quality Metric for Semantic Role Labeling Systems Evaluation [Jindal et al., 2022]

Jindal, Ishan, Alexandre Rademaker, Khoi-Nguyen Tran, Huaiyu Zhu, Hiroshi Kanayama, Marina Danilevsky, and Yunyao Li. "PriMeSRL-Eval: A Practical Quality Metric for Semantic Role Labeling Systems Evaluation." In *Findings of the Association for Computational Linguistics: EACL 2023*, pp. 1761-1773. 2023.

An Improved Evaluation Scheme

With PriMeSRL-Eval we made the following observations:

- ❑ Current evaluation scripts exaggerate the SRL model quality.
- ❑ A clear drop on ~7F1 points on OOD set is observed.
- ❑ The relative ranking of the SoTA SRL models changes.

Model	Evaluation script	In-domain PSD	In-domain Argument Classification				Out-of-domain PSD	Out-of-domain Argument Classification			
		F1	P	R	F1	(r)	F1	P	R	F1	(r)
(Conia et al., 2021)	CoNLL09	96.9	89.5	89.5	89.5	(3)	87.8	82.0	81.9	81.9	(3)
	PriMeSRL	95.5(↓1.4)	86.6	86.6	86.6(↓2.9)	(2)	80.9(↓6.9)	72.4	72.6	72.5(↓9.4)	(4)
(Blloshmi et al., 2021)$_{nested}$	CoNLL09	97.1	89.3	81.9	85.4	(4)	89.7	82.8	75.7	79.1	(4)
	PriMeSRL	96.4(↓0.7)	86.8	79.8	83.1(↓2.3)	(4)	86.7(↓3.0)	75.7	69.9	72.7(↓6.4)	(3)
(Blloshmi et al., 2021)$_{flat}$	CoNLL09	97.4	90.9	89.6	**90.2**	(1)	90.1	83.9	82.1	83.0	(2)
	PriMeSRL	96.9(↓0.5)	88.6	87.4	**88.0**(↓2.2)	(1)	87.8(↓2.3)	77.6	76.3	**76.9**(↓6.1)	(1)
(Jindal et al., 2022)	CoNLL09	96.8	89.9	89.3	89.6	(2)	89.8	82.9	83.1	**83.02**	(1)
	PriMeSRL	95.5(↓1.3)	86.8	86.3	86.55(↓3.0)	(3)	83.4(↓6.4)	73.9	74.3	74.1(↓8.9)	(2)

Conclusion

Observations

- **Syntax matters**
  - Yes, at least for argument spans.
  - Not for dependency SRL.
    - Eventually, you need syntax to compute span.
  - SRL can help syntax

- **Contextualized embeddings**
  - Carry major chunk of performance gain in SRL.
  - Fine-tunning LM for SRL further raised the bar.

- **End-to-End Systems**
  - More practical, but computationally expensive
  - Predicate and arguments task shown to improve each other.

Opportunities

- **SRL in few shot setting**
  - Probe SRL information from large LMs.
    - Given the sparsity of the SRL label space finding a right prompt is quite challenging.

- **Multilingual SRL**
  - Multilingual SRL Resources
  - Universal PropBanks for SRL
    - A long way to go

- **Datasets**
  - Dataset without predicate sense annotations
  - Ethical issues

- **SRL Model Re-Evaluations**

References

1. Merchant, A., Rahimtoroghi, E., Pavlick, E., & Tenney, I. (2020, November). What Happens To BERT Embeddings During Fine-tuning?. In *Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP* (pp. 33-44).
2. Tan, Z., Wang, M., Xie, J., Chen, Y., & Shi, X. (2018, April). Deep semantic role labeling with self-attention. In Proceedings of the AAAI conference on artificial intelligence (Vol. 32, No. 1).
3. Marcheggiani, D., Frolov, A., & Titov, I. (2017, August). A Simple and Accurate Syntax-Agnostic Neural Model for Dependency-based Semantic Role Labeling. In Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017) (pp. 411-420).
4. A Unified Syntax-aware Framework for Semantic Role Labeling
5. Tian, Y., Qin, H., Xia, F., & Song, Y. (2022, June). Syntax-driven Approach for Semantic Role Labeling. In Proceedings of the Thirteenth Language Resources and Evaluation Conference (pp. 7129-7139).
6. Zhang, Z., Strubell, E., & Hovy, E. (2021, August). Comparing span extraction methods for semantic role labeling. In Proceedings of the 5th Workshop on Structured Prediction for NLP (SPNLP 2021) (pp. 67-77).
7. Fei, H., Wu, S., Ren, Y., Li, F., & Ji, D. (2021, August). Better combine them together! integrating syntactic constituency and dependency representations for semantic role labeling. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021 (pp. 549-559).
8. Wang, N., Li, J., Meng, Y., Sun, X., & He, J. (2021). An mrc framework for semantic role labeling. arXiv preprint arXiv:2109.06660.
9. Blloshmi, R., Conia, S., Tripodi, R., & Navigli, R. (2021). Generating Senses and RoLes: An End-to-End Model for Dependency-and Span-based Semantic Role Labeling. In IJCAI (pp. 3786-3793).
10. Zhang, L., Jindal, I., & Li, Y. (2022, July). Label Definitions Improve Semantic Role Labeling. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 5613-5620).
11. Cai, J., He, S., Li, Z., & Zhao, H. (2018, August). A full end-to-end semantic role labeler, syntactic-agnostic over syntactic-aware?. In Proceedings of the 27th International Conference on Computational Linguistics (pp. 2753-2765).
12. He, S., Li, Z., & Zhao, H. (2019, November). Syntax-aware Multilingual Semantic Role Labeling. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 5350-5359).
13. Conia, S., Bacciu, A., & Navigli, R. (2021, June). Unifying cross-lingual Semantic Role Labeling with heterogeneous linguistic resources. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 338-351).

References

14. Conia, S., & Navigli, R. (2020, December). Bridging the gap in multilingual semantic role labeling: a language-agnostic approach. In Proceedings of the 28th International Conference on Computational Linguistics (pp. 1396-1410).
15. Kasai, J., Friedman, D., Frank, R., Radev, D., & Rambow, O. (2019, June). Syntax-aware Neural Semantic Role Labeling with Supertags. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 701-709).
16. He, L., Lee, K., Levy, O., & Zettlemoyer, L. (2018, July). Jointly Predicting Predicates and Arguments in Neural Semantic Role Labeling. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 364-369).
17. Shi, T., Malioutov, I., & İrsoy, O. (2020, November). Semantic Role Labeling as Syntactic Dependency Parsing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 7551-7571).
18. Zhou, J., Li, Z., & Zhao, H. (2020, November). Parsing All: Syntax and Semantics, Dependencies and Spans. In Findings of the Association for Computational Linguistics: EMNLP 2020 (pp. 4438-4449).
19. Zhou, J., Li, Z., & Zhao, H. (2020, November). Parsing All: Syntax and Semantics, Dependencies and Spans. In Findings of the Association for Computational Linguistics: EMNLP 2020 (pp. 4438-4449).
20. Wang, Y., Johnson, M., Wan, S., Sun, Y., & Wang, W. (2019, July). How to best use syntax in semantic role labelling. In Annual Meeting of the Association for Computational Linguistics (57th: 2019) (pp. 5338-5343). Association for Computational Linguistics.
21. He, S., Li, Z., Zhao, H., & Bai, H. (2018, July). Syntax for semantic role labeling, to be, or not to be. In Proceedings of the 56th annual meeting of the association for computational linguistics (Volume 1: Long papers) (pp. 2061-2071).
22. Marcheggiani, D., & Titov, I. (2020, November). Graph Convolutions over Constituent Trees for Syntax-Aware Semantic Role Labeling. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 3915-3928).
23. Marcheggiani, D., & Titov, I. (2017, September). Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (pp. 1506-1515).
24. Marcheggiani, D., & Titov, I. (2017, September). Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (pp. 1506-1515).
25. Li, Z., Zhao, H., Wang, R., & Parnow, K. (2020, November). High-order Semantic Role Labeling. In Findings of the Association for Computational Linguistics: EMNLP 2020 (pp. 1134-1151).

References

26. Lyu, C., Cohen, S. B., & Titov, I. (2019, November). Semantic Role Labeling with Iterative Structure Refinement. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 1071-1082).
27. Li, Z., He, S., Zhao, H., Zhang, Y., Zhang, Z., Zhou, X., & Zhou, X. (2019, July). Dependency or span, end-to-end uniform semantic role labeling. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 01, pp. 6730-6737).
28. Ouchi, H., Shindo, H., & Matsumoto, Y. (2018). A Span Selection Model for Semantic Role Labeling. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (pp. 1630-1642).
29. Strubell, E., Verga, P., Andor, D., Weiss, D., & McCallum, A. (2018). Linguistically-Informed Self-Attention for Semantic Role Labeling. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (pp. 5027-5038).
30. He, L., Lee, K., Lewis, M., & Zettlemoyer, L. (2017, July). Deep semantic role labeling: What works and what's next. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 473-483).
31. FitzGerald, N., Täckström, O., Ganchev, K., & Das, D. (2015, September). Semantic role labeling with neural network factors. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 960-970).
32. Guan, C., Cheng, Y., & Zhao, H. (2019, June). Semantic Role Labeling with Associated Memory Network. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 3361-3371).
33. Jindal, I., Aharonov, R., Brahma, S., Zhu, H., & Li, Y. (2020). Improved Semantic Role Labeling using Parameterized Neighborhood Memory Adaptation. arXiv preprint arXiv:2011.14459.
34. Jindal, Ishan, Alexandre Rademaker, Khoi-Nguyen Tran, Huaiyu Zhu, Hiroshi Kanayama, Marina Danilevsky, and Yunyao Li. "PriMeSRL-Eval: A Practical Quality Metric for Semantic Role Labeling Systems Evaluation." In *Findings of the Association for Computational Linguistics: EACL 2023*, pp. 1761-1773. 2023.
35. Zhang, Li, Ishan Jindal, and Yunyao Li. "Label definitions improve semantic role labeling." In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 5613-5620. 2022.