Leveraging Linguistic Structures for Named Entity Recognition with Bidirectional Recursive Neural Networks

Peng-Hsuan Li

11/20/2017

Named Entity Recognition (NER)

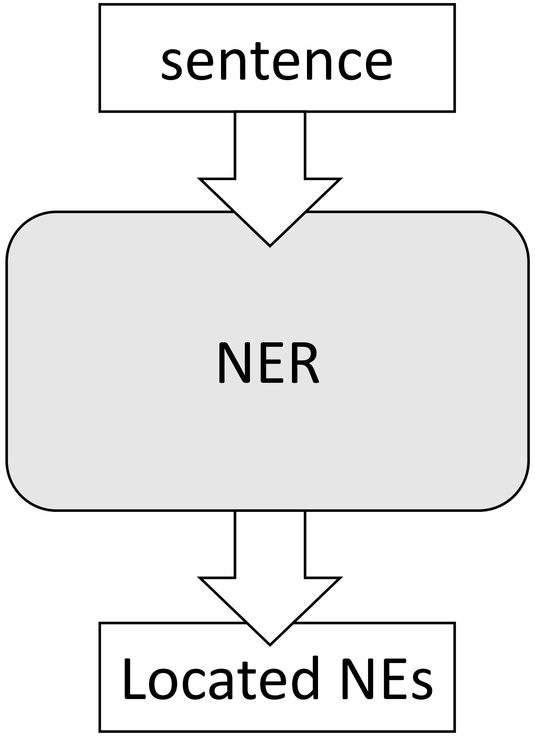
Background

To Locate and Classify Named Entities (NEs)

The $\underbrace{\textit{defense}}_{\textit{ORG}}$ *secretary* $\underbrace{\textit{Donald Rumsfeld}}_{\textit{PER}}$

NE Labels
PER (person) *ORG (organization)*

NER

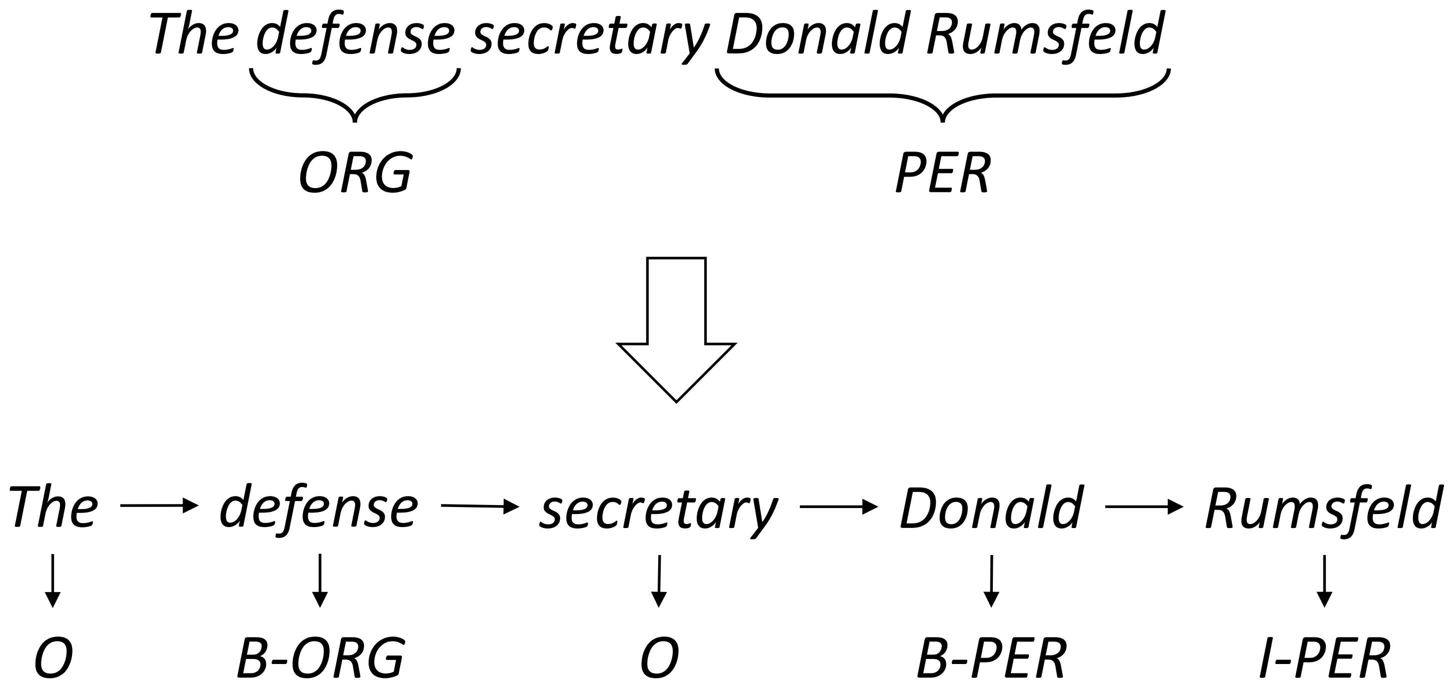


The ***defense*** *secretary* ***Donald Rumsfeld***

1 2 3 4 5

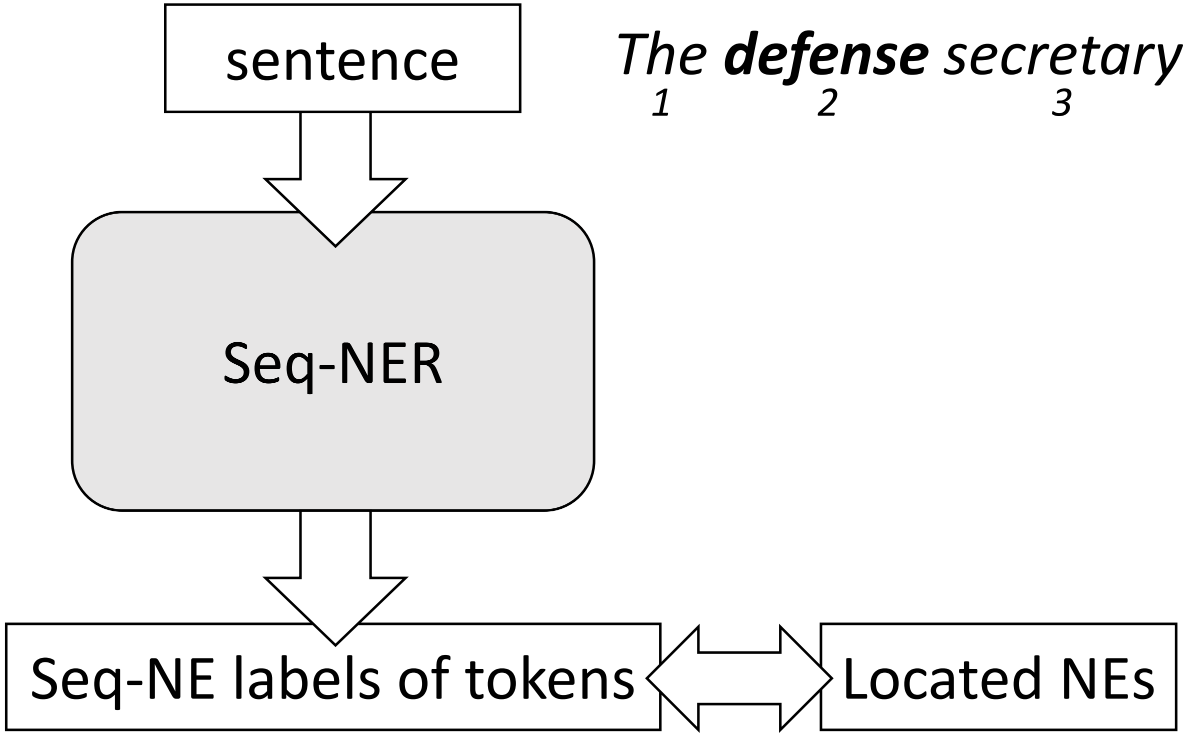
(2,3)-ORG, (4,6)-PER

Sequential Labeling NER (Seq-NER)



Seq Labels
B (begin)
I (inside)
O (outside)

Seq-NER



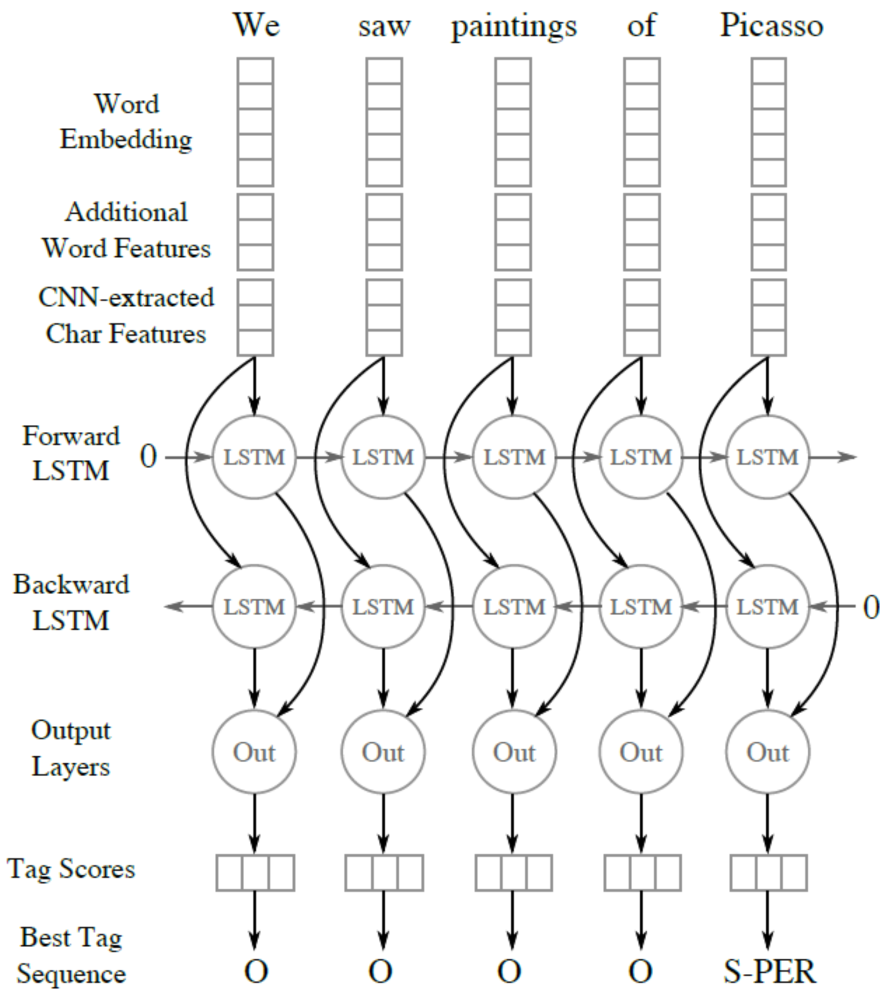
The ***defense*** *secretary* ***Donald Rumsfeld***

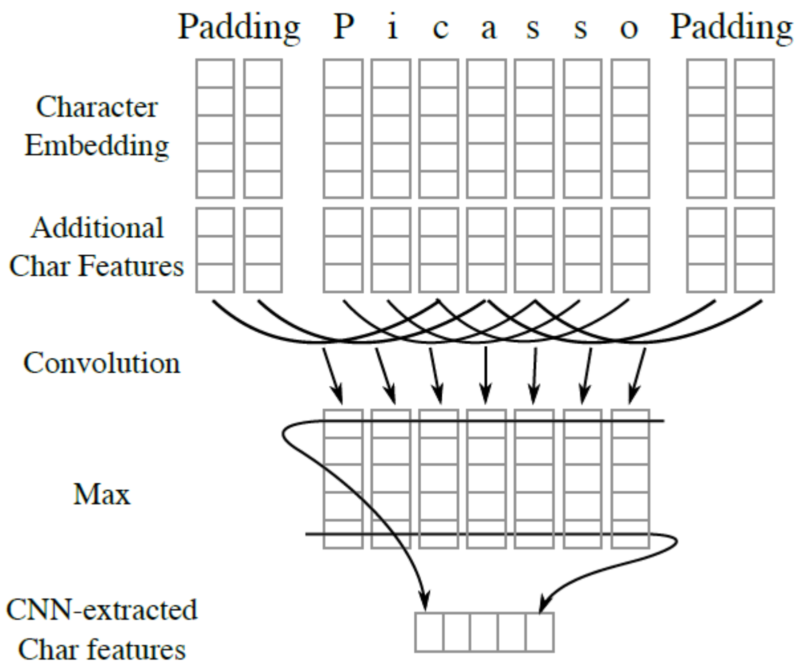
1 2 3 4 5

O B-ORG O B-PER I-PER

(2,3)-ORG, (4,6)-PER

Chiu and Nichols (2016)





J. P. Chiu and E. Nichols. Named Entity Recognition with Bidirectional LSTM-CNNs. *Transactions of the Association for Computational Linguistics*, 4:357–370, 2016.

Seq-NER: the Problem

Motivation

When Predicted (Begin, End) Are Slightly Wrong

B-FAC *I-FAC*

the first couple moves out of the ***White House*** *on January 20th .*

FAC

NE Labels
FAC (facility)

False Positive: Not Even a Linguistic Unit of the Sentence

the White (FAC)

B-FAC I-FAC

the first couple moves out of the ***White House*** *on January 20th .*

The Problem

- Seq-NER neglects the prior knowledge of linguistic structures

Constituency Structures and NER

Pilot Study & Insights

Constituency Structure

```
|--S
    |--S
    |    |--NP
    |    |    |--DT the
    |    |    |--NP
    |    |        |--JJ first
    |    |        |--NN couple
    |    |
    |    |--VP
    |         |--VP
    |         |    |--VBZ moves
    |         |    |--PP
    |         |        |--IN out
    |         |        |--PP
    |         |            |--IN of
    |         |            |--NP
    |         |                |--DT the
    |         |                |--NP
    |         |                    |--NNP White
    |         |                    |--NNP House
    |         |
    |         |--PP
    |              |--IN on
    |              |--NP
    |                  |--NNP January
    |                  |--NN 20th
    |
    |--. .
```

OntoNotes 5.0

Split	Sentences	Tokens	NEs	Consistent NEs
Train	59,924	1,088,503	81,828	76,309
Validate	8,528	147,724	11,066	10,267
Test	8,262	152,728	11,257	10,459
Total	76,714	1,388,955	104,151	97,035

Seq-NER Models

	OntoNotes 5.0		
Model	Precision	Recall	F1
Bi-Recurrent	85.7	86.5	86.10
Chiu and Nichols (2016)	-	-	**86.41**

Removing Non-Constituent Predictions in Post-Processing

		Test		
Model	Const-Only	Precision	Recall	F1
Bi-Recurrent	X	85.7	**86.5**	86.10
Bi-Recurrent	O	**87.2**	85.1	**86.14**

If Constituency Structures and NER Are More Consistent…

Split	Sentences	Tokens	NEs	Consistent NEs
Train	59,924	1,088,503	81,828	76,309
Validate	8,528	147,724	11,066	10,267
Test	8,262	152,728	11,257	10,459

93% → **100%**

More Consistent… Better Recall

		Test		
Model	Const-Only	Precision	Recall	F1
Bi-Recurrent	X	85.7	**86.5**	86.10
Bi-Recurrent	O	**87.2**	85.1	**86.14**

85.1 ↗ 86.5

If Full Structure Information Are Utilized...

```
|--S
     |--S
     |    |--NP
     |    |    |--DT the
     |    |    |--NP
     |    |         |--JJ first
     |    |         |--NN couple
     |    |
     |    |--VP
     |         |--VP
     |         |    |--VBZ moves
     |         |    |--PP
     |         |         |--IN out
     |         |         |--PP
     |         |              |--IN of
     |         |              |--NP
     |         |                   |--DT the
     |         |                   |--NP
     |         |                        |--NNP White
     |         |                        |--NNP House
     |         |
     |         |--PP
     |              |--IN on
     |              |--NP
     |                   |--NNP January
     |                   |--NN 20th
     |
     |--. .
```

NP, *White House*

Structure Information Utilized…

Model	Const-Only	Prediction	False Positive	False Negative
Bi-Recurrent	X	*the White*	Y	Y
Bi-Recurrent	O	-	N	$\boxed{\text{Y}}$
Magic	O	*White House*	N	$\boxed{\text{N}}$

Insights: NER Could Improve If

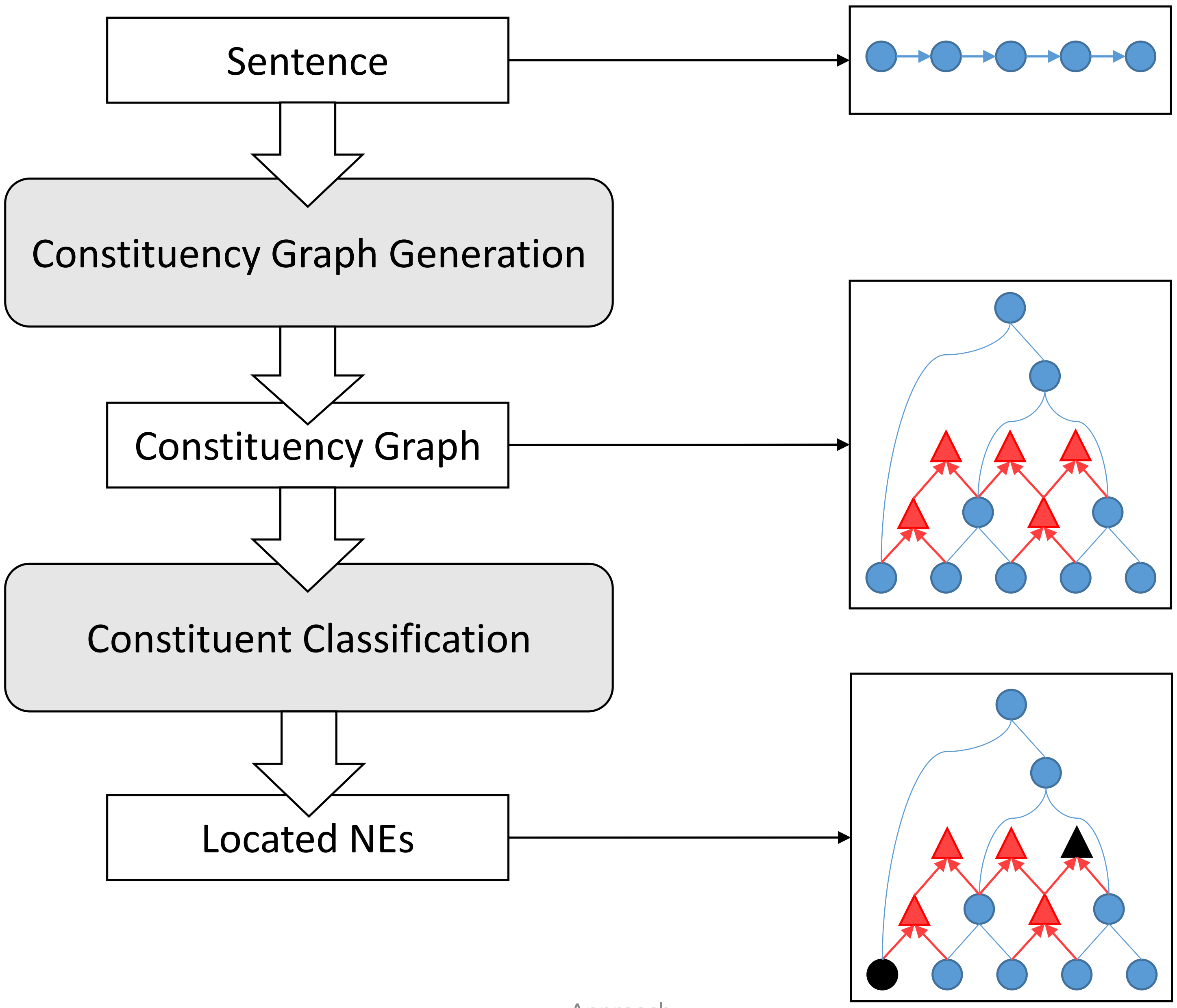
- Inconsistencies between parsing and NER are mitigated
- Prior linguistic structure information is utilized

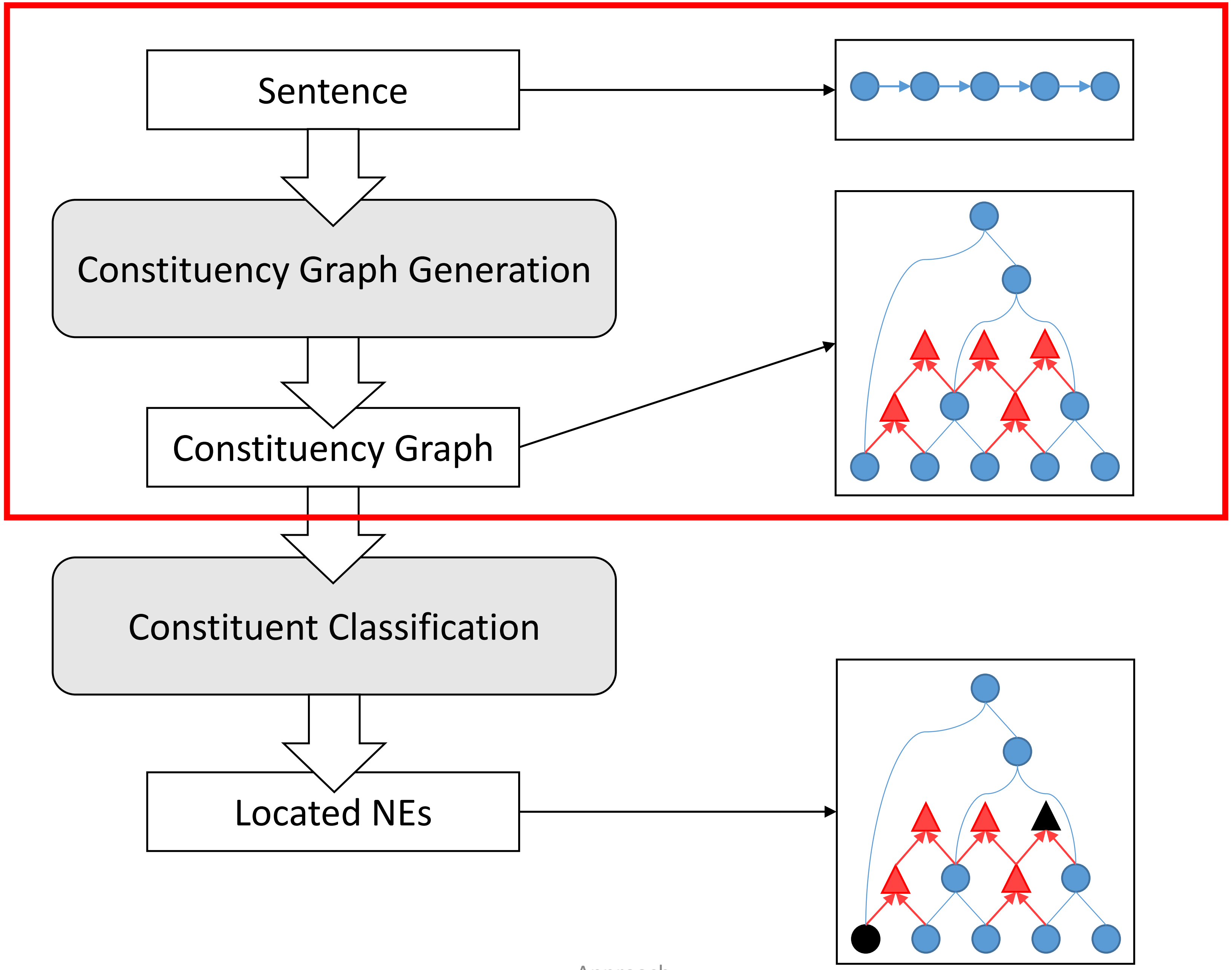
Leveraging Linguistic Structures: the Objectives

- Mitigate the inconsistencies between parsing and NER by restructuring algorithms
- Utilize prior linguistic structure information with constituent-based Bidirectional Recursive Neural Networks (BRNN)

Constituency-Oriented NER

Approach


Sentence
Constituency Graph Generation
Constituency Graph
Constituent Classification
Located NEs


Sentence
Constituency Graph Generation
Constituency Graph
Constituent Classification
Located NEs

Leveraging Linguistic Structures: the Objectives

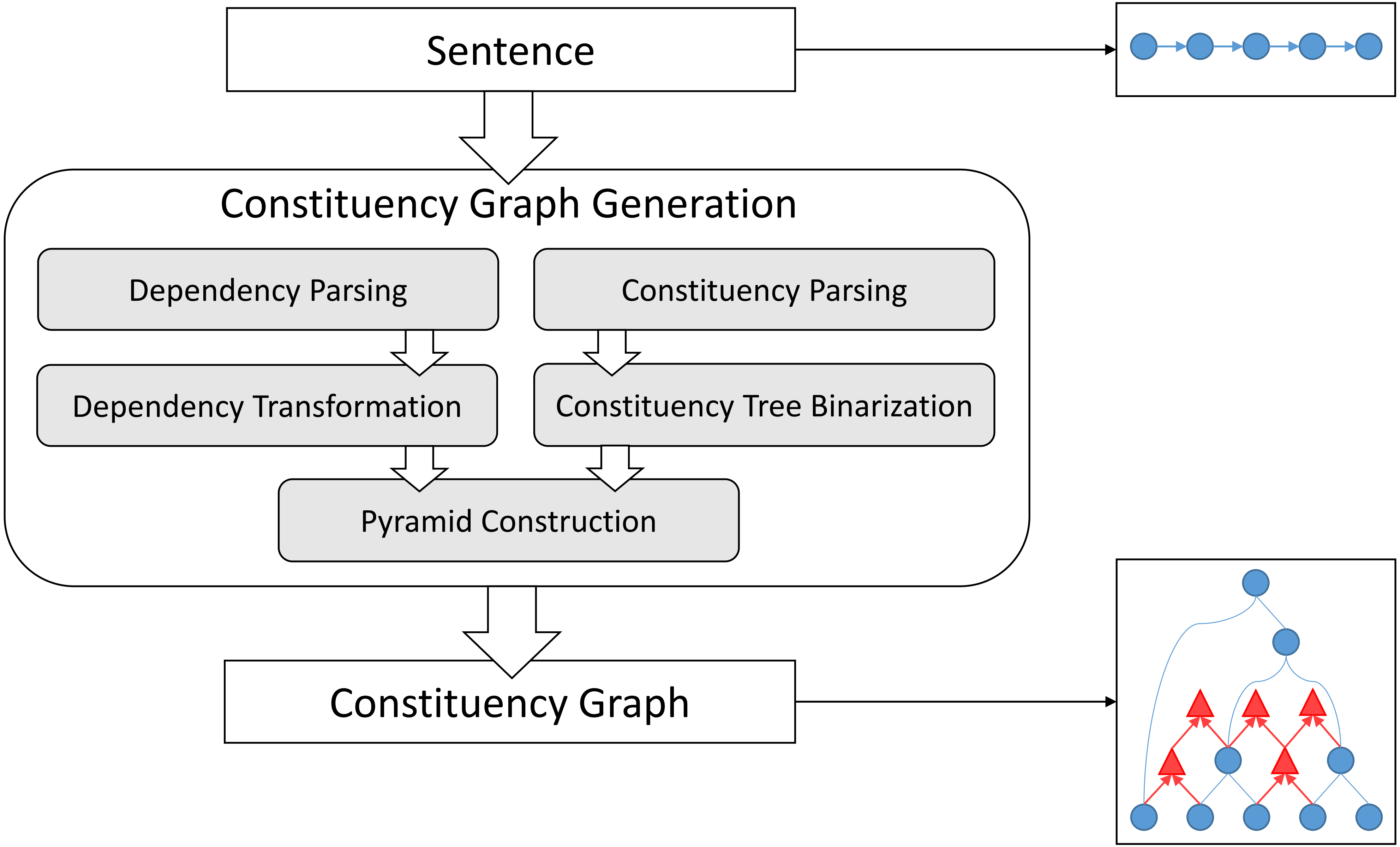
- Mitigate the inconsistencies between parsing and NER by restructuring algorithms

- Utilize prior linguistic structure information with constituent-based Bidirectional Recursive Neural Networks (BRNN)

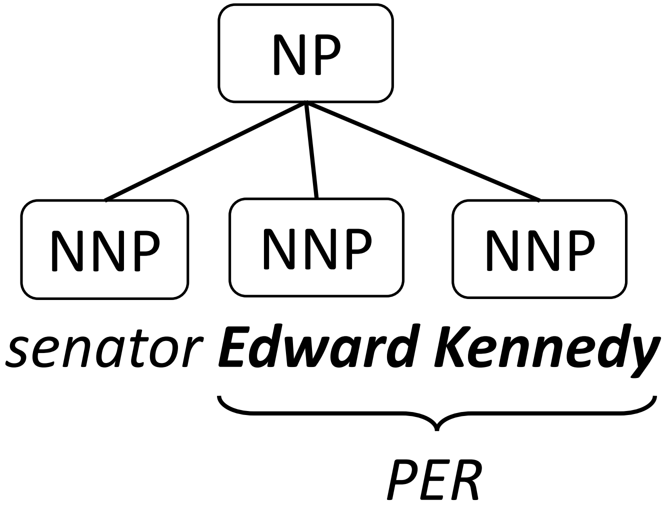
If Constituency Structures and NER Are More Consistent…

Split	Sentences	Tokens	NEs	Consistent NEs
Train	59,924	1,088,503	81,828	76,309
Validate	8,528	147,724	11,066	10,267
Test	8,262	152,728	11,257	10,459

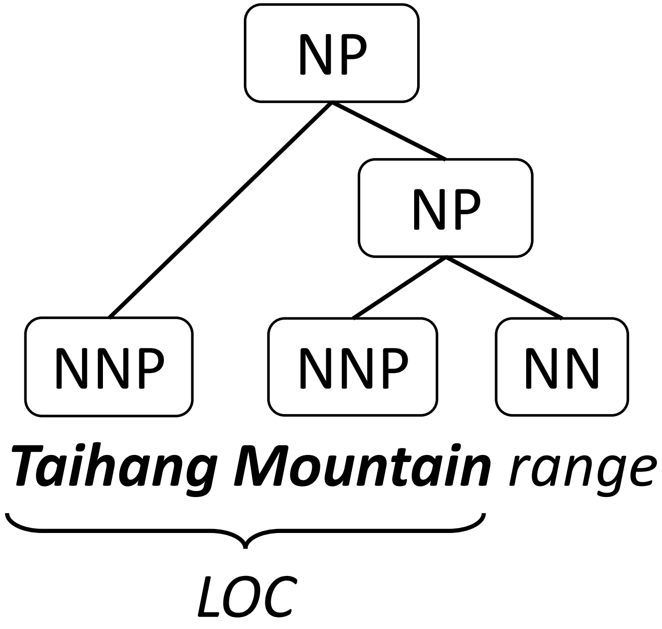
93% ↗ **100%**


Sentence
Constituency Graph Generation
Dependency Parsing
Constituency Parsing
Dependency Transformation
Constituency Tree Binarization
Pyramid Construction
Constituency Graph

Inconsistent NEs

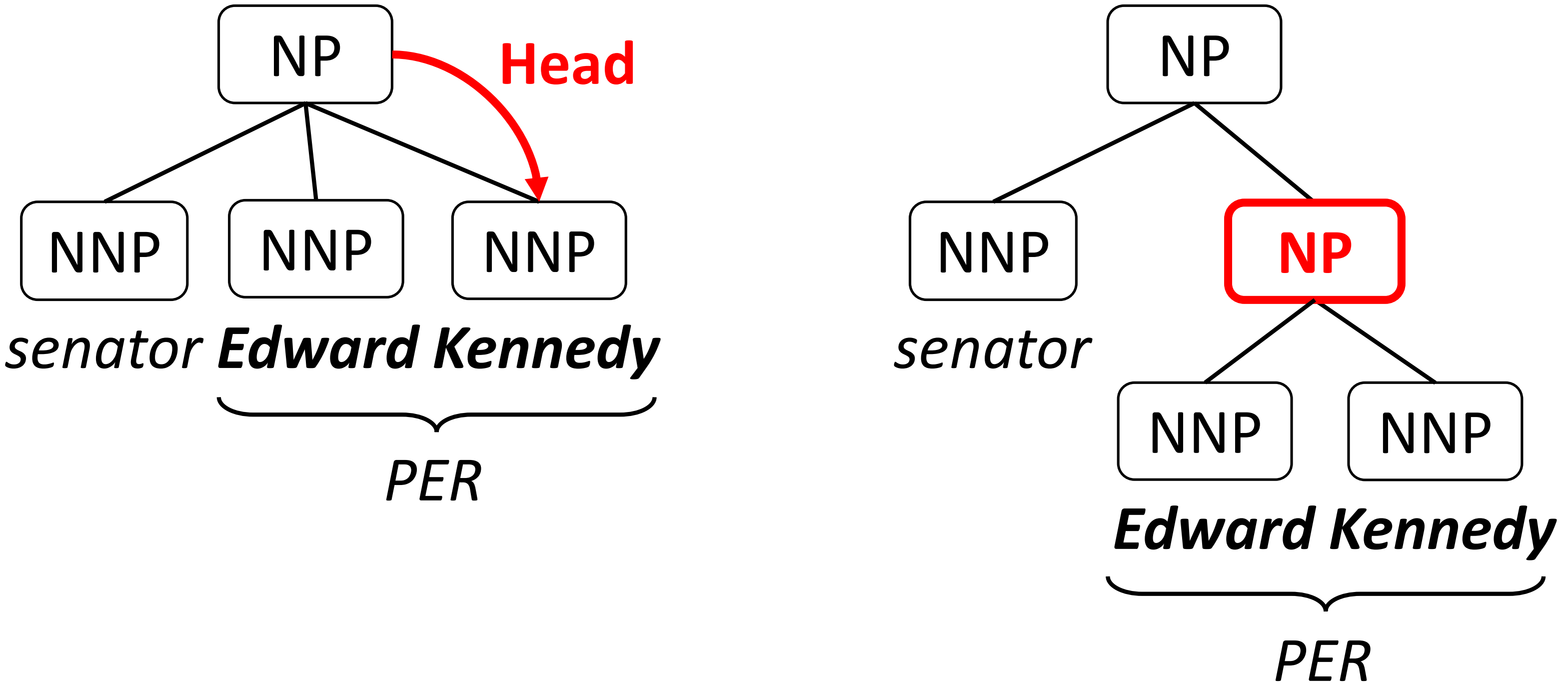


Type-1
Cross Siblings



Type-2
Cross Branches

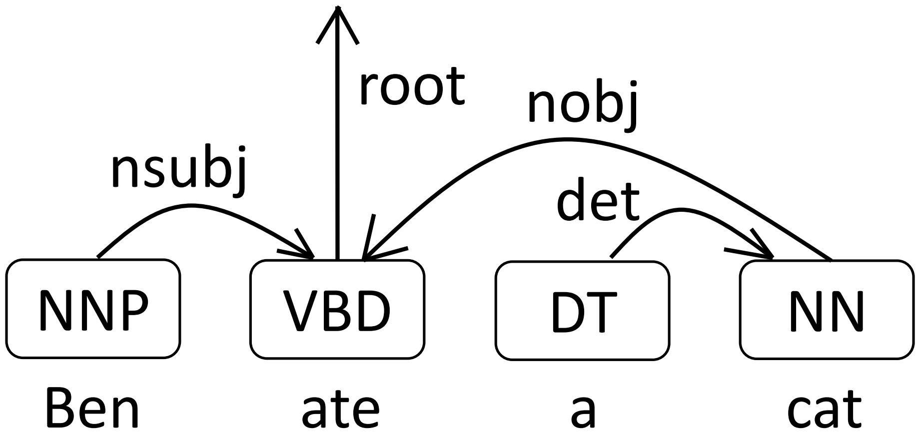
Eliminate Type-1: Constituency Tree Binarization

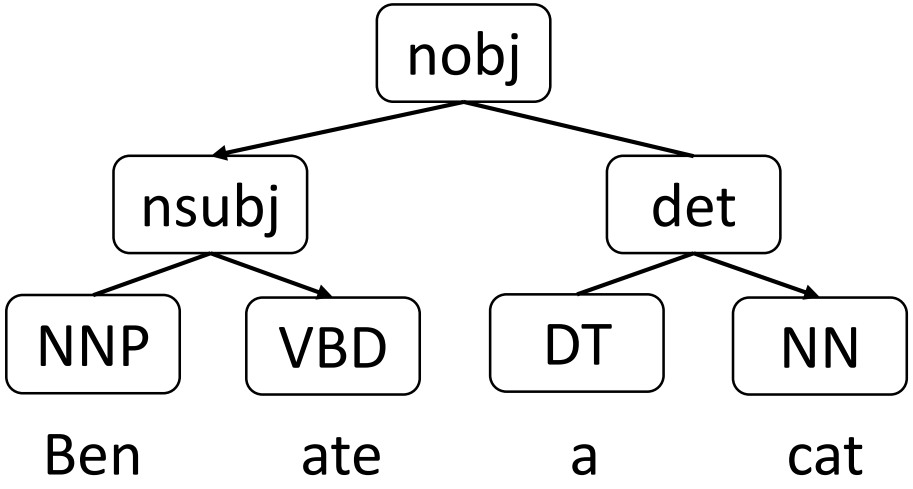


Type-1
Cross Siblings

Consistent

Eliminate Type-1: Dependency Transformation

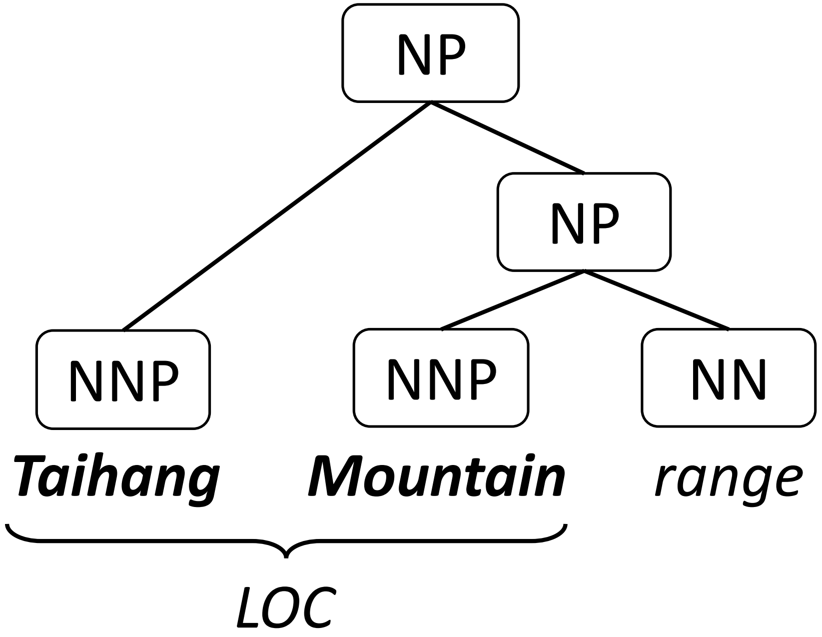


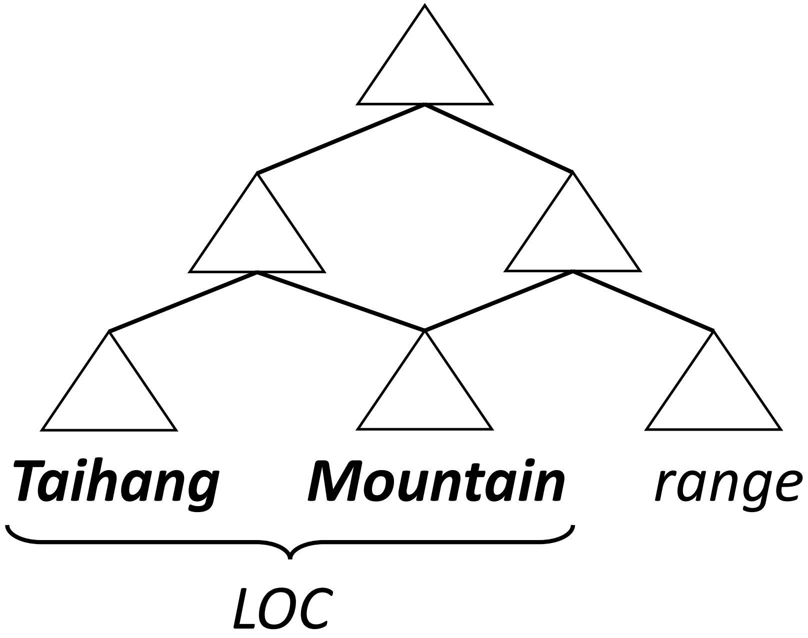


No Constituents

No Type-1 Inconsistencies

Eliminate Type-2: Pyramid Construction

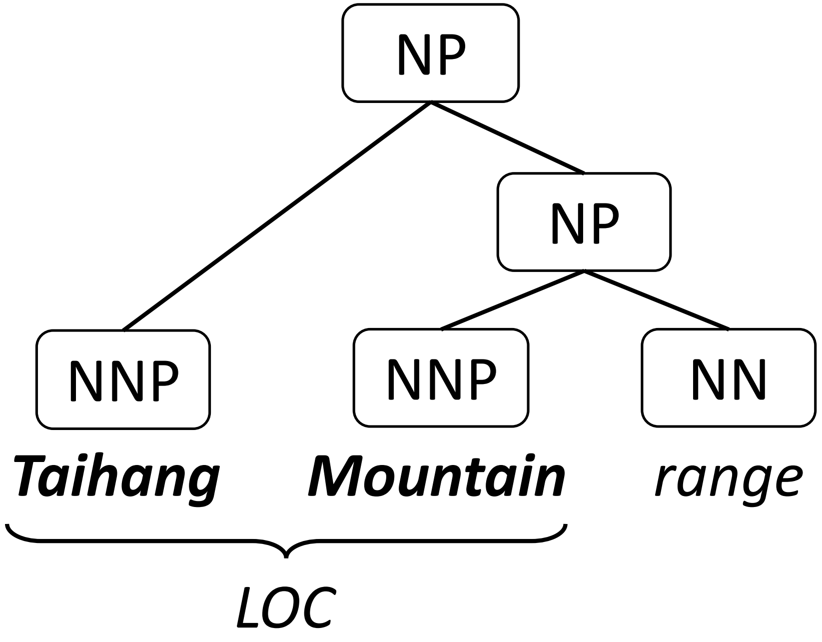




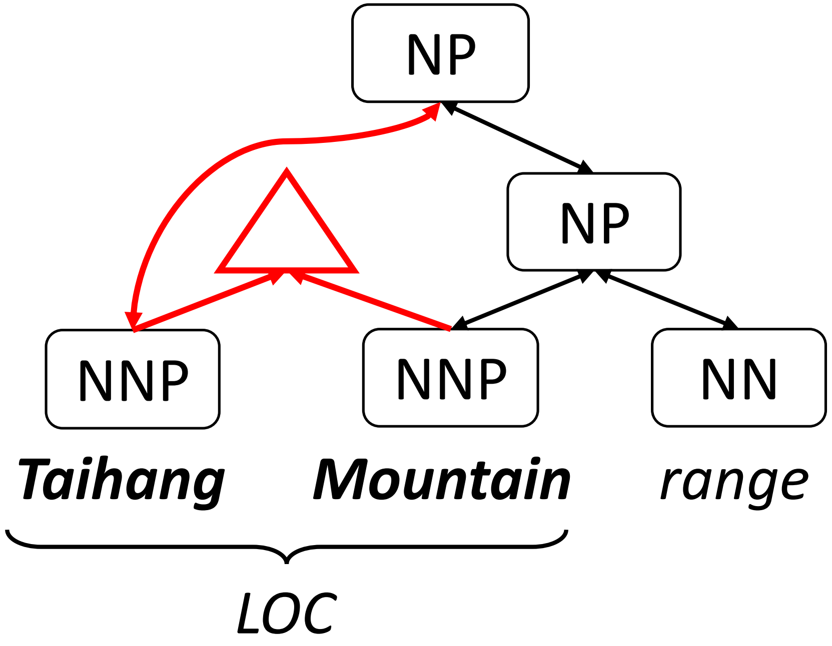
Type-2
Cross Branches

No Liguistic Structures

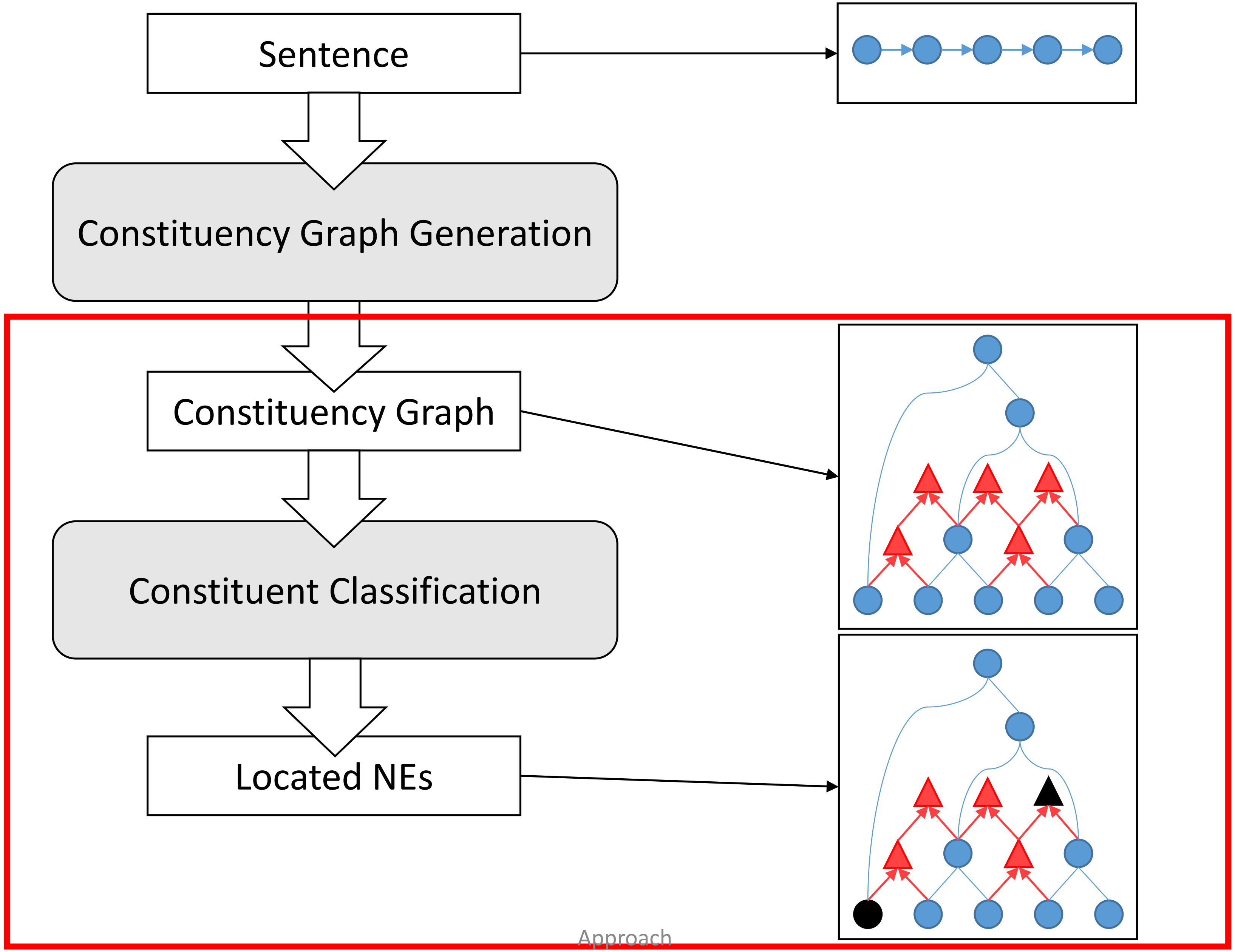
Eliminate Type-2: Pyramid Construction



Type-2
Cross Branches



No Inconsistencies


Sentence
Constituency Graph Generation
Constituency Graph
Constituent Classification
Located NEs

Leveraging Linguistic Structures: the Objectives

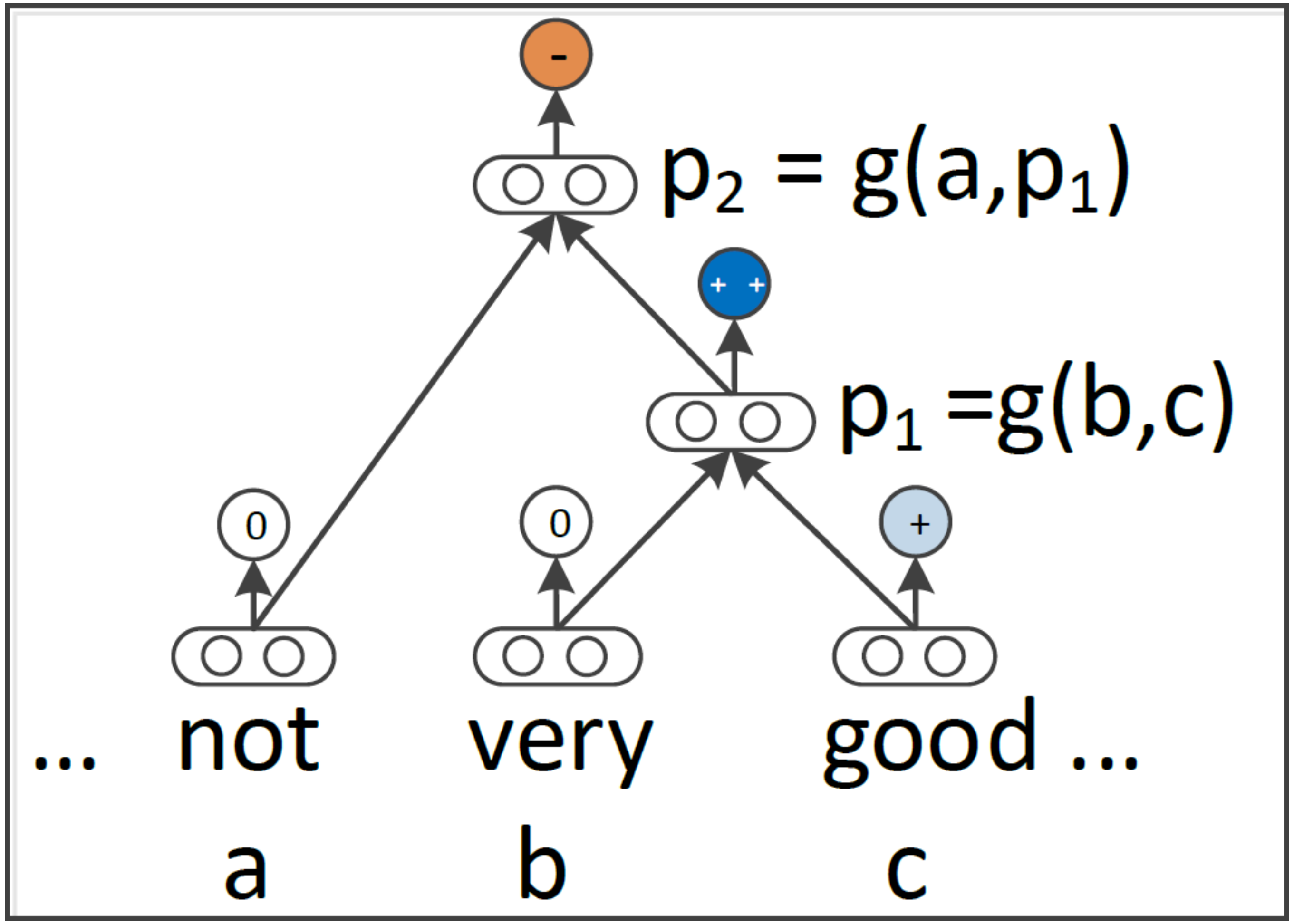
- Mitigate the inconsistencies between parsing and NER by restructuring algorithms
- Utilize prior linguistic structure information with constituent-based Bidirectional Recursive Neural Networks (BRNN)

If Full Structure Information Are Utilized…

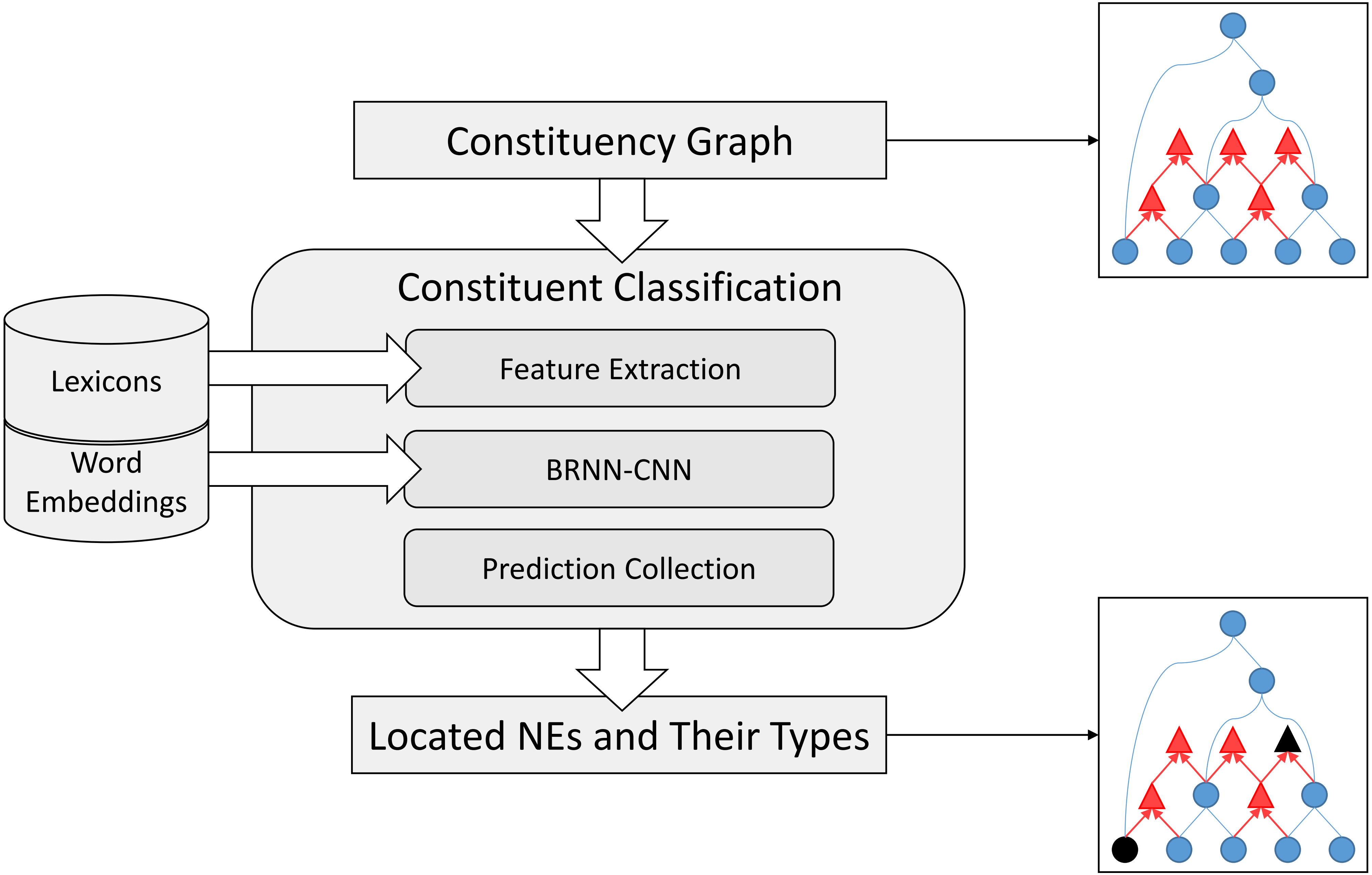
```
|--S
   |--S
   |   |--NP
   |   |   |--DT the
   |   |   |--NP
   |   |       |--JJ first
   |   |       |--NN couple
   |   |
   |   |--VP
   |       |--VP
   |       |   |--VBZ moves
   |       |   |--PP
   |       |       |--IN out
   |       |       |--PP
   |       |           |--IN of
   |       |           |--NP
   |       |              |--DT the
   |       |              |--NP
   |       |                  |--NNP White
   |       |                  |--NNP House
   |       |
   |       |--PP
   |           |--IN on
   |           |--NP
   |               |--NNP January
   |               |--NN 20th
   |
   |--. .
```

NP, *White House*

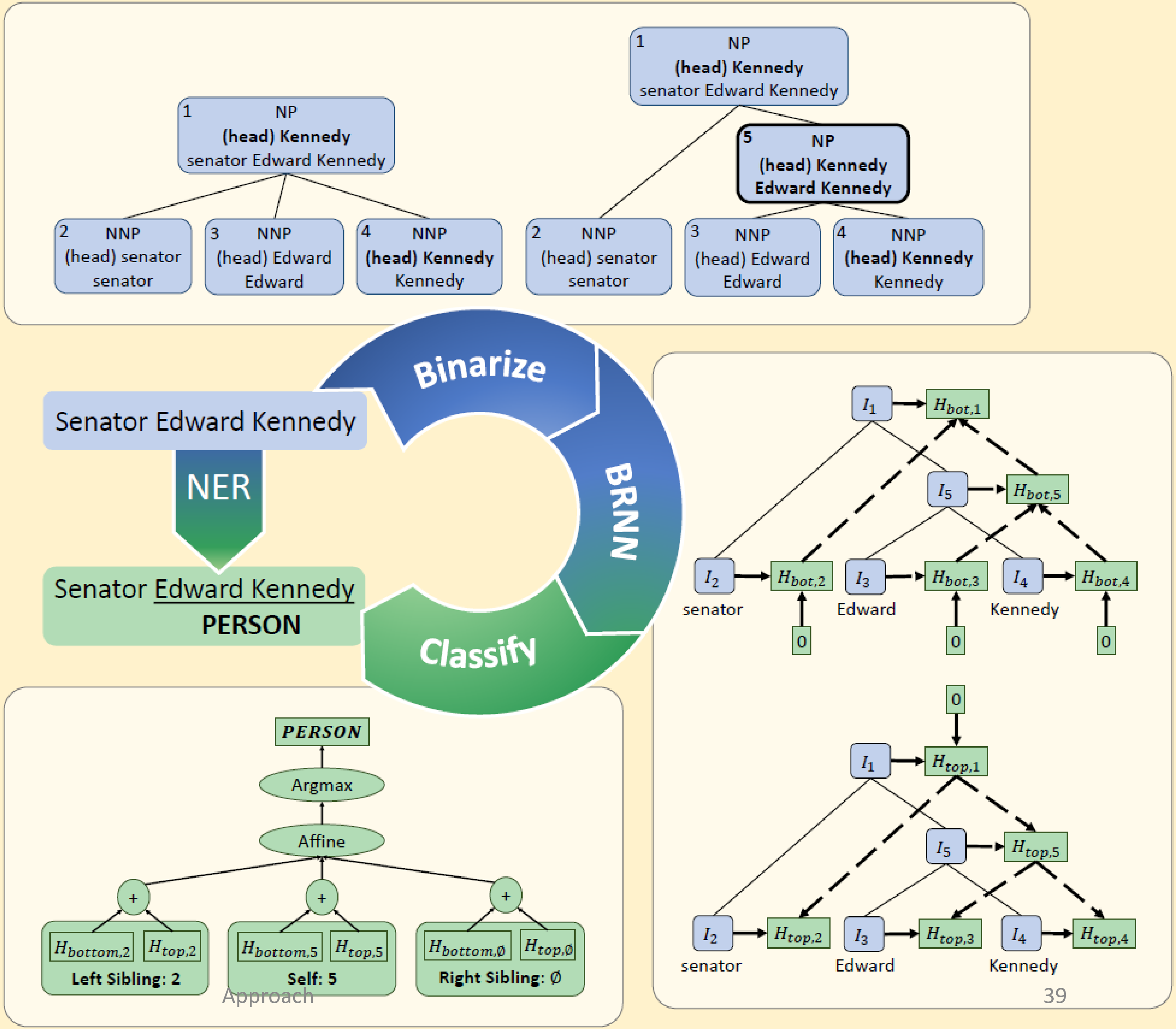
Socher et al.



R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642, 2013.

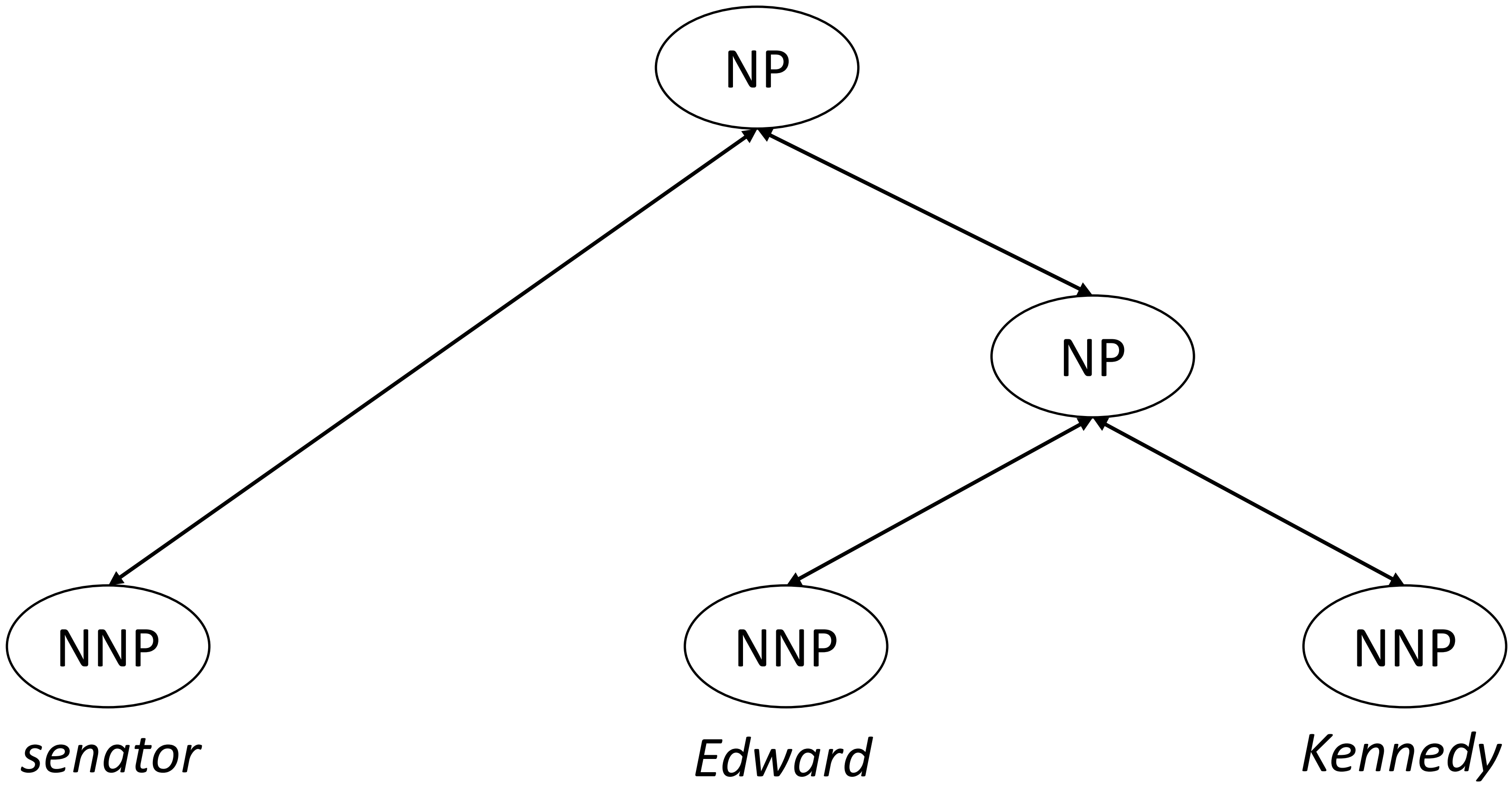

Constituency Graph
Constituent Classification
Lexicons
Word
Embeddings
Feature Extraction
BRNN-CNN
Prediction Collection
Located NEs and Their Types

Li et al. (2017)

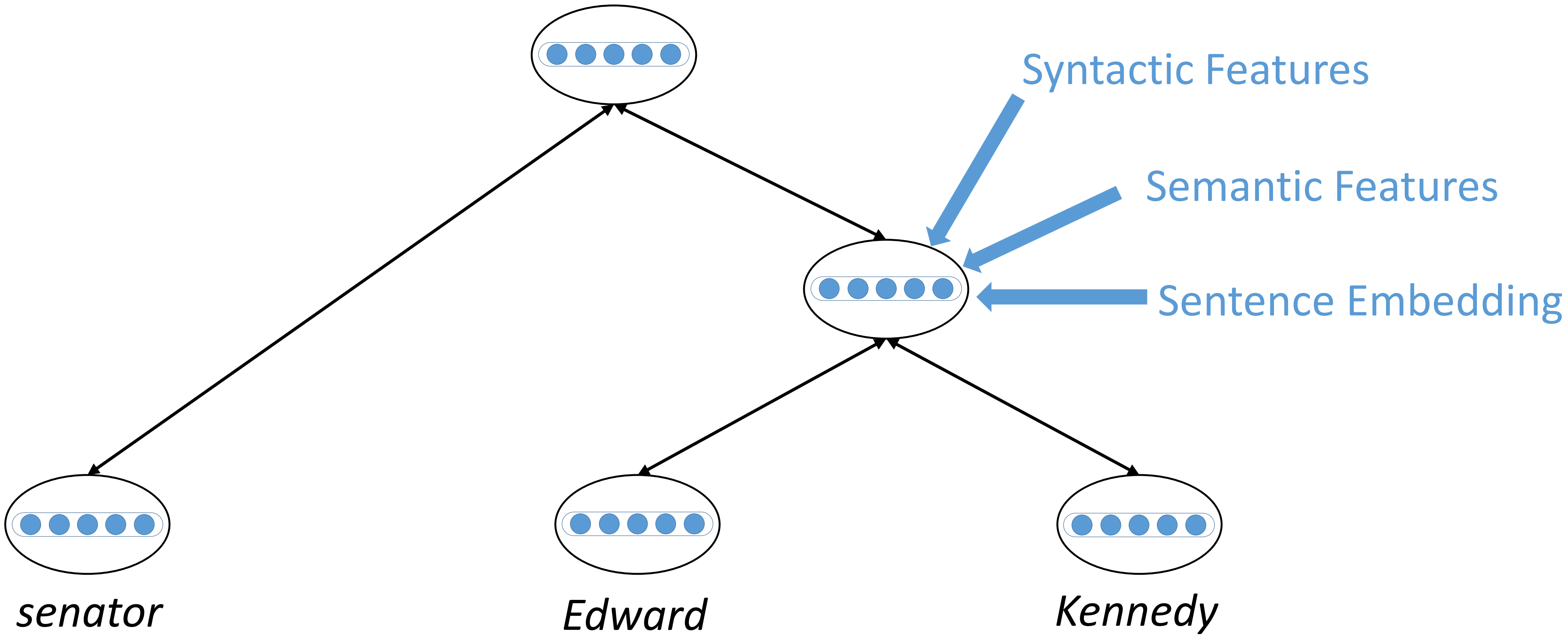


PH Li, RP Dong, YS Wang, JC Chou, and WY Ma. Leveraging Linguistic Structures for Named Entity Recognition with Bidirectional Recursive Neural Networks. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017.

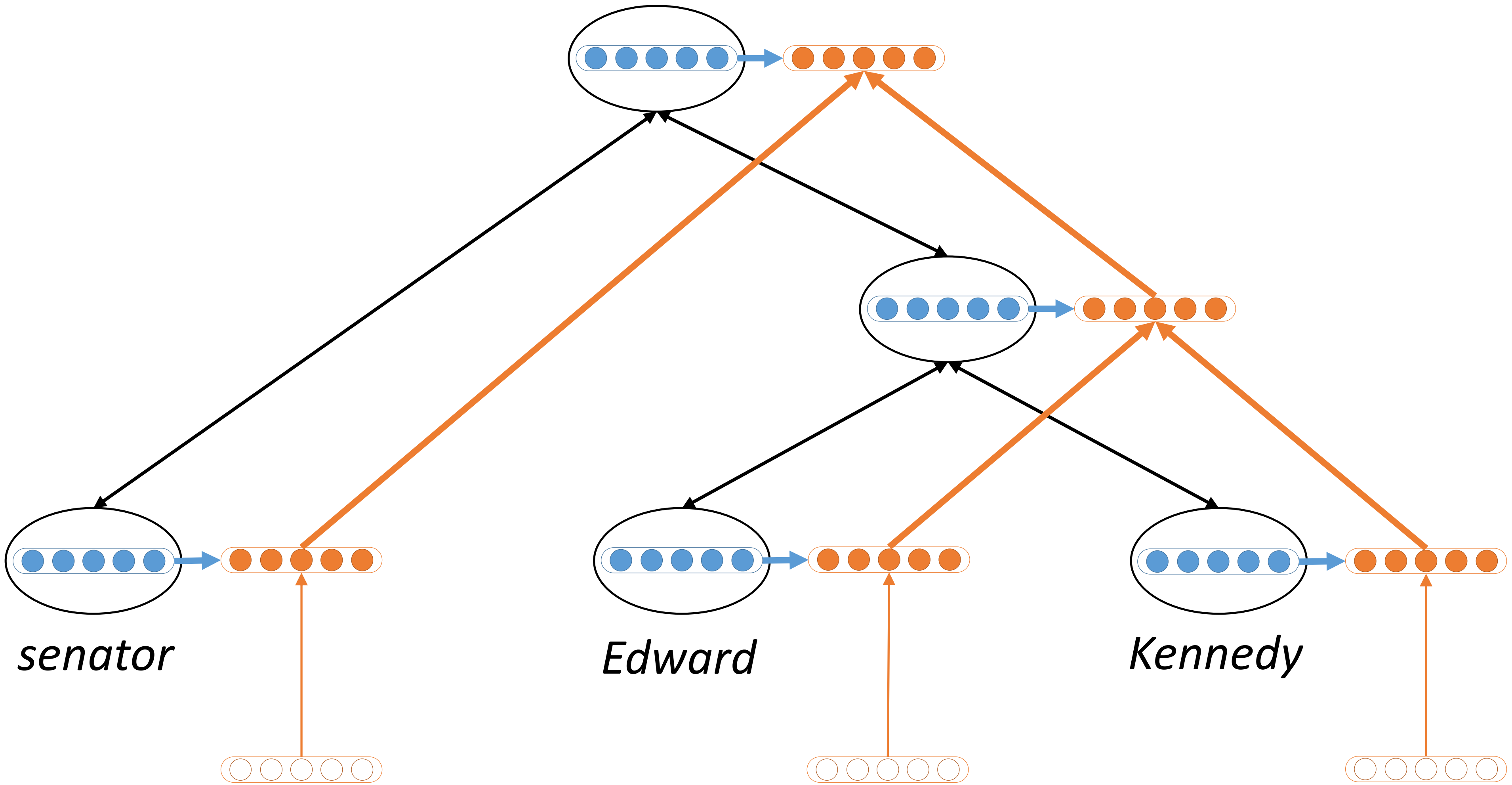
Constituency Graph



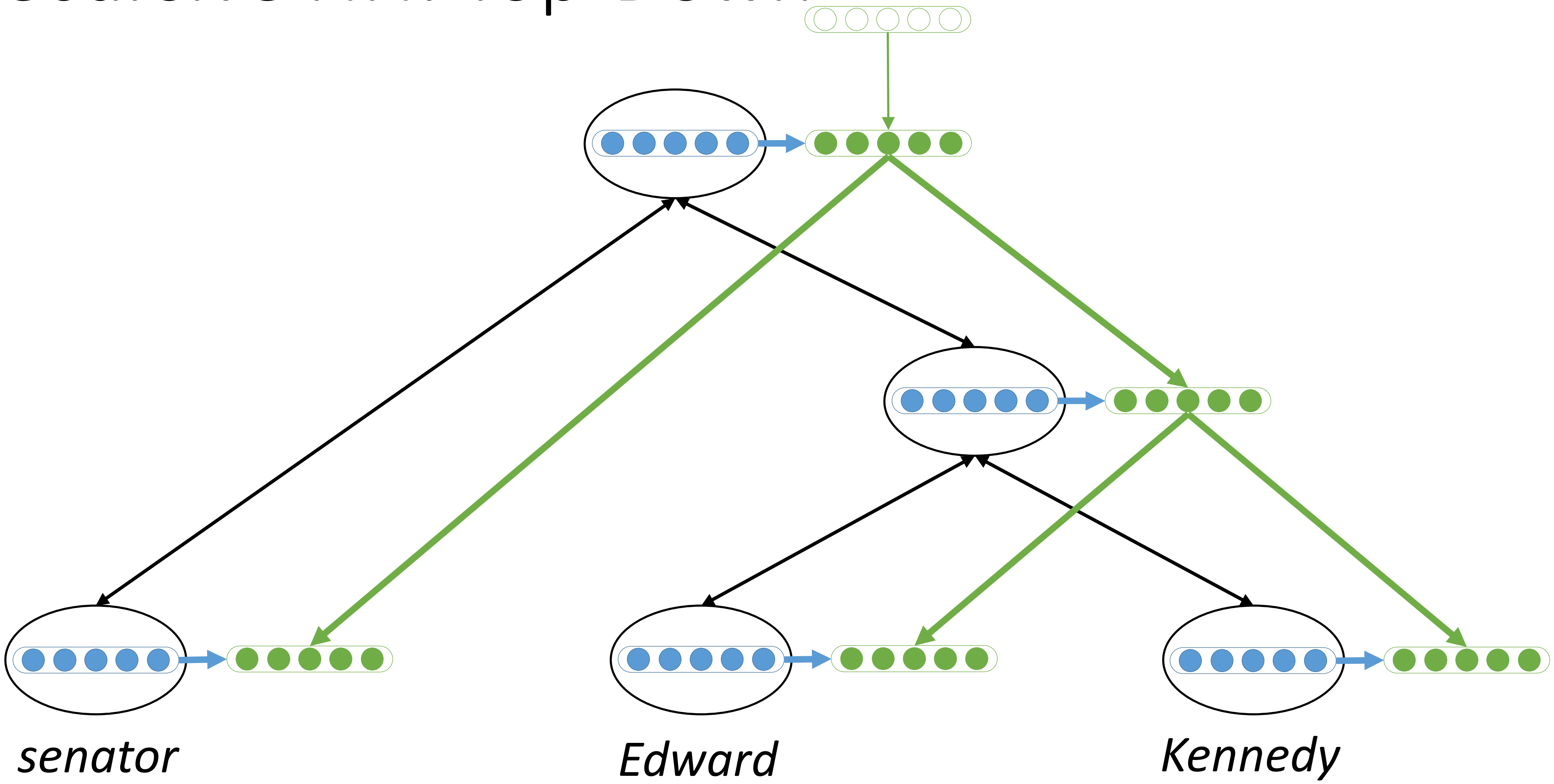
Feature Extraction



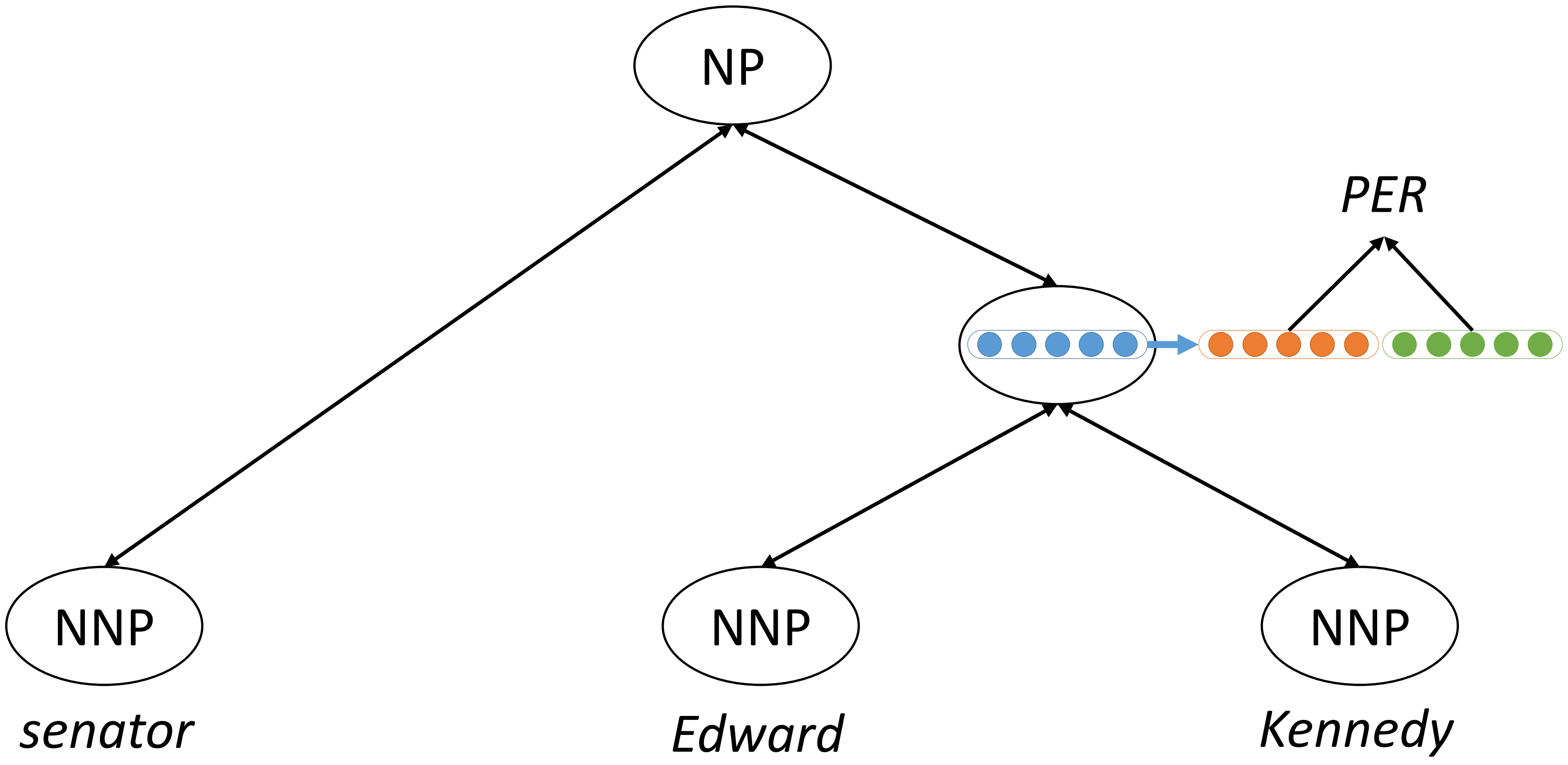
Bi-Recursive NN: Bottom-Up



Bi-Recursive NN: Top-Down



Classification



Discovery

Evaluation & Discussion

Seq-Recurrent vs. Constituency-Oriented BRNN

	93% Consistency			**97% Consistency**		
	CoNLL 2003			OntoNotes 5.0		
Model	Precision	Recall	F1	Precision	Recall	F1
Bi-Recurrent	-	-	-	85.7	86.5	86.10
Chiu and Nichols (2016)	91.4	91.9	**91.62**	-	-	86.41
BRNN(-CNN)	90.2	87.7	88.91	88.0	86.5	**87.21**

Constituent-Based vs. Constituency-Oriented

			OntoNotes 5.0		
Model	Const-Only	Prediction	Precision	Recall	F1
Bi-Recurrent	X	***the White***	85.7	**86.5**	86.10
Bi-Recurrent	O	**-**	87.2	85.1	86.14
BRNN	O	***White House***	**88.0**	86.2	**87.10**

Ablation Study: Constituency Tree Binarization

		OntoNotes 5.0			
Model	Binarize	Consistency	Precision	Recall	F1
BRNN	X	93%	87.3	83.0	85.11
BRNN	O	**97%**	**88.0**	**86.2**	**87.10**

Ablation Study: Dependency Transformation

		CoNLL 2003		
Model	Parser	Precision	Recall	F1
BRNN	StanfordRNN	88.9	86.9	87.91
BRNN	SyntaxNet	**90.2**	**87.7**	**88.91**

Ablation Study: Pyramid Construction

		CoNLL 2003		
Model	Pyramid	Precision	Recall	F1
BRNN	X	89.1	82.9	85.89
BRNN	O	**90.2**	**87.7**	**88.91**

Ablation Study: Bidirectional

		OntoNotes 5.0		
Model	***Koran***	Precision	Recall	F1
Top-Down	-	79.2	69.3	73.93
Bottom-Up	***PERSON***	86.6	86.2	86.41
BRNN	***WORK OF ART***	**88.0**	**86.2**	**87.10**

```
|       |   |--NP
|       |       |--PRP it
|       |
|       |--PP
|           |--IN by
|           |--S
|               |--VP
|                   |--VBG repeating
|                   |--NP
|                       |--NP
|                       |   |--NP
|                       |   |   |--NP
|                       |   |   |   |--NNS verses
|                       |   |   |
|                       |   |   |--PP
|                       |   |       |--IN from
|                       |   |       |--NP
|                       |   |           |--DT the
|                       |   |           |--NP
|                       |   |               |--JJ noble
|                       |   |               |--NNP Koran
|                       |   |
|                       |   |--CC and
|                       |
|                       |--NP
|                           |--DT the
|                           |--NP
|                               |--CD two
|                               |--NNS testimonies
|
|--. .
```

Conclusion

Contribution & Future Work

Recall: the Objectives

- Mitigate the inconsistencies between parsing and NER by restructuring algorithms

- Utilize prior linguistic structure information with constituent-based Bidirectional Recursive Neural Networks (BRNN)

Constituency-Oriented Approach: Contributions

- Elimination of type-1 inconsistencies by constituency binarization and dependency transformation
- Elimination of type-2 inconsistencies by pyramid construction

- Utilization of local structures with bottom-up recursive network
- Utilization of global structures with top-down recursive network

Future Work

- Constituency-oriented approach for nested NEs: parsing biomedical text
- Constituency-oriented approach with one end-to-end model: back-propagation through parsing
- Constituency-oriented approach for identifying text chunks of interest: other than NEs