Learning Challenges in Natural Language Processing

Swabha Swayamdipta

April 08, 2019

UWNLP

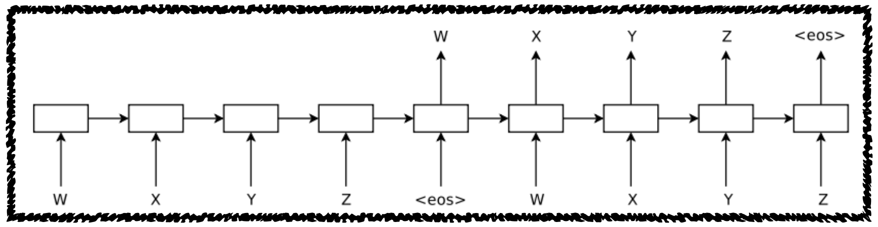
Carnegie Mellon University
Language Technologies Institute

NLP today

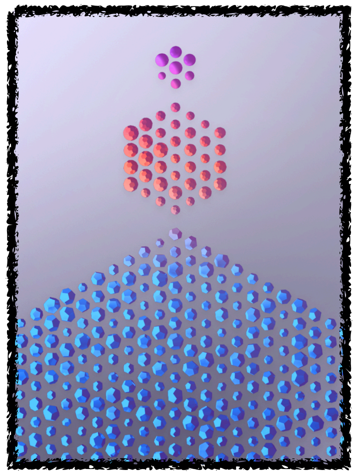
NLP today

Contextualized Representations

NLP today



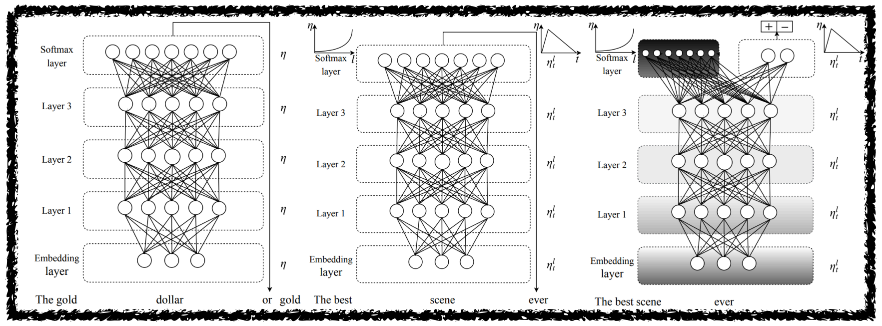
[Dai & Le, 2015]



[Radford et. al., 2018]

[Peters et. al., 2018]

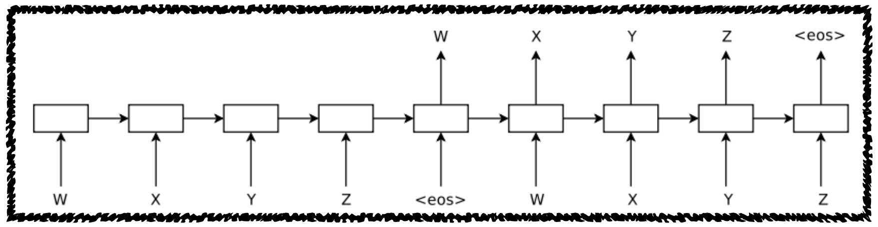
Contextualized Representations



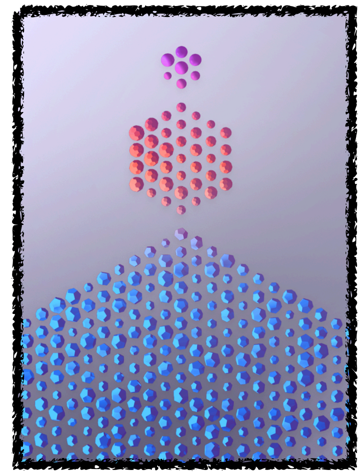
[Howard & Ruder, 2018]

[Devlin et. al., 2018]

NLP today



[Dai & Le, 2015]

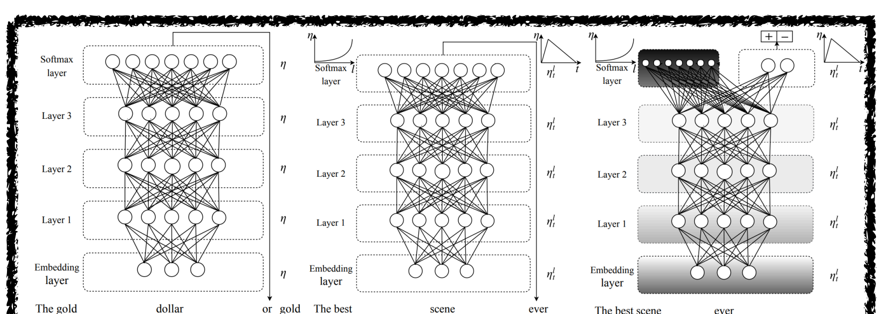


[Radford et. al., 2018]

Large Language Model

[Peters et. al., 2018]

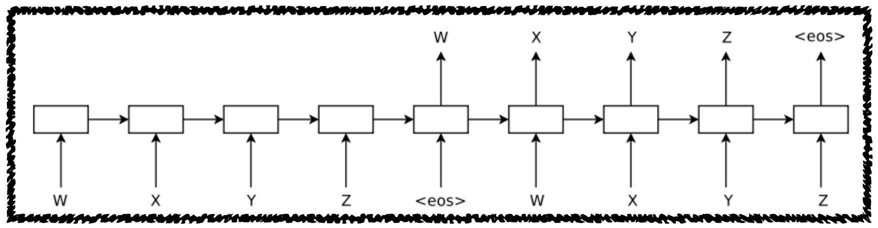
Contextualized Representations



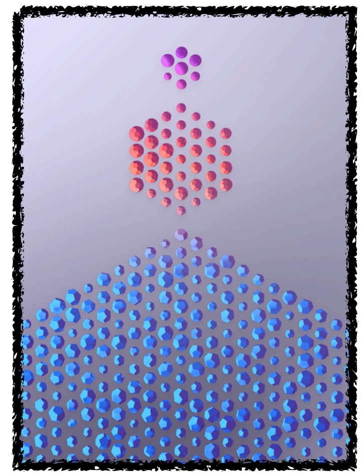
[Howard & Ruder, 2018]

[Devlin et. al., 2018]

NLP today



[Dai & Le, 2015]



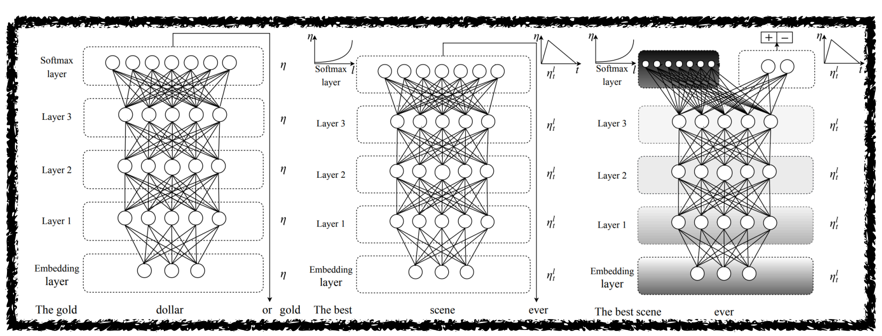
[Radford et. al., 2018]

Large Language Model

[Peters et. al., 2018]

Contextualized Representations

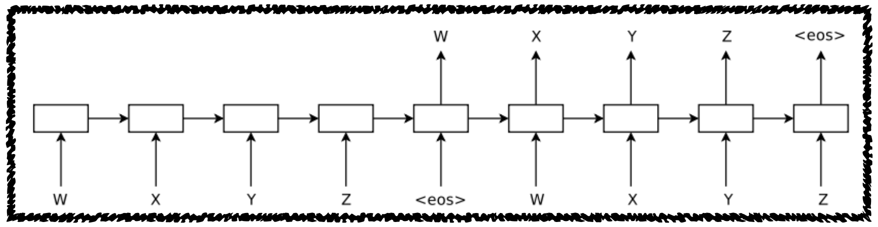
Downstream Tasks



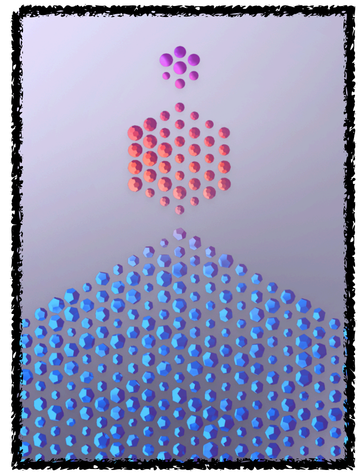
[Howard & Ruder, 2018]

[Devlin et. al., 2018]

NLP today



[Dai & Le, 2015]

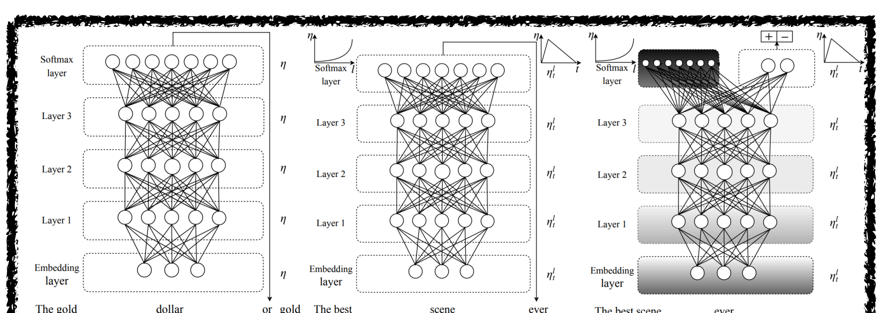

[Radford et. al., 2018]

Large Language Model


[Peters et. al., 2018]

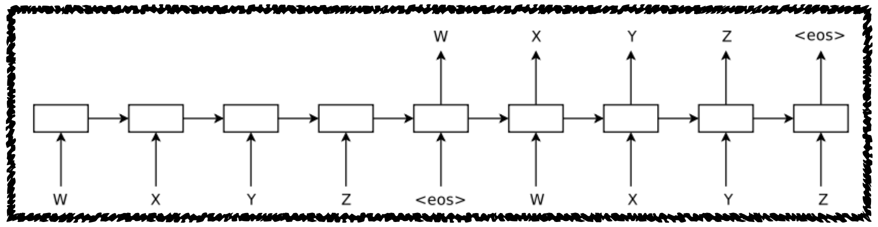

Contextualized Representations

Downstream Tasks

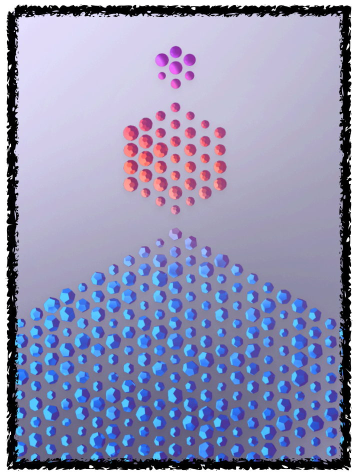

[Howard & Ruder, 2018]


[Devlin et. al., 2018]

NLP today



[Dai & Le, 2015]



[Radford et. al., 2018]

Large Language Model

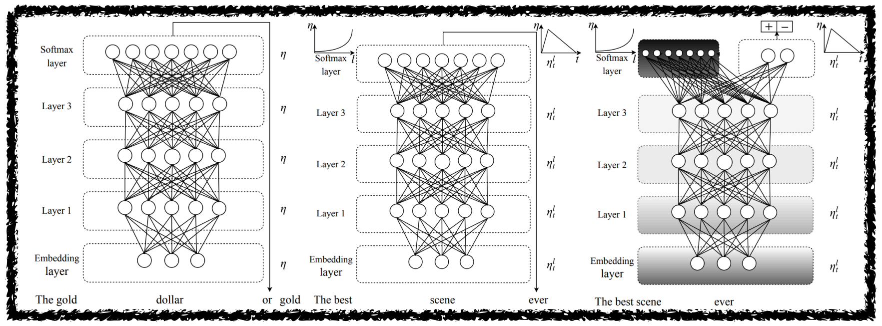
Unsupervised

[Peters et. al., 2018]

Contextualized Representations

Downstream Tasks

Supervised



[Howard & Ruder, 2018]

[Devlin et. al., 2018]

A closer look...

On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of new Huguenot colonists declined after what year?

A closer look...

On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of new Huguenot colonists declined after what year?

A closer look...

On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of old Acadian colonists declined after the year 1675.

The number of new Huguenot colonists declined after what year?

[Jia & Liang, 2017]

A closer look...

On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of old Acadian colonists declined after the year 1675.

The number of new Huguenot colonists declined after what year?

[Jia & Liang, 2017]

A closer look...

On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of old Acadian colonists declined after the year 1675.

The number of new Huguenot colonists declined after what year?

[Jia & Liang, 2017]

A closer look...

On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of old Acadian colonists declined after the year 16~~7~~5.

The number of new Huguenot colonists declined after what year?

[Jia & Liang, 2017]

Learning Challenges

Learning Challenges

Part I

Can we incorporate some priors about language to improve our models?

- [ ] Syntactic Scaffolds for Semantic Structures (EMNLP 2018)

Learning Challenges

Part I

Can we incorporate some priors about language to improve our models?

- [ ] Syntactic Scaffolds for Semantic Structures (EMNLP 2018)

Part II

What in our data is causing models to achieve high performance?

- [ ] Annotation Artifacts in Natural Language Inference Data (NAACL 2018)

Learning Challenge #1

- Can we incorporate some priors about language?

On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of new Huguenot colonists declined after what year?

Learning Challenge #1

- Can we incorporate some priors about language?
- One kind of prior - Linguistic Structure

On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of new Huguenot colonists declined after what year?

Learning Challenge #1

- Can we incorporate some priors about language?
- One kind of prior - Linguistic Structure
- Can linguistic structure act as an informative prior?

On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of new Huguenot colonists declined after what year?

Linguistic Structure: Semantics

Linguistic Structure: Semantics

- Who did what to whom?

Linguistic Structure: Semantics

- Who did what to whom?

After encouraging them , he told them goodbye and left for Macedonia

Linguistic Structure: Semantics

- Who did what to whom?

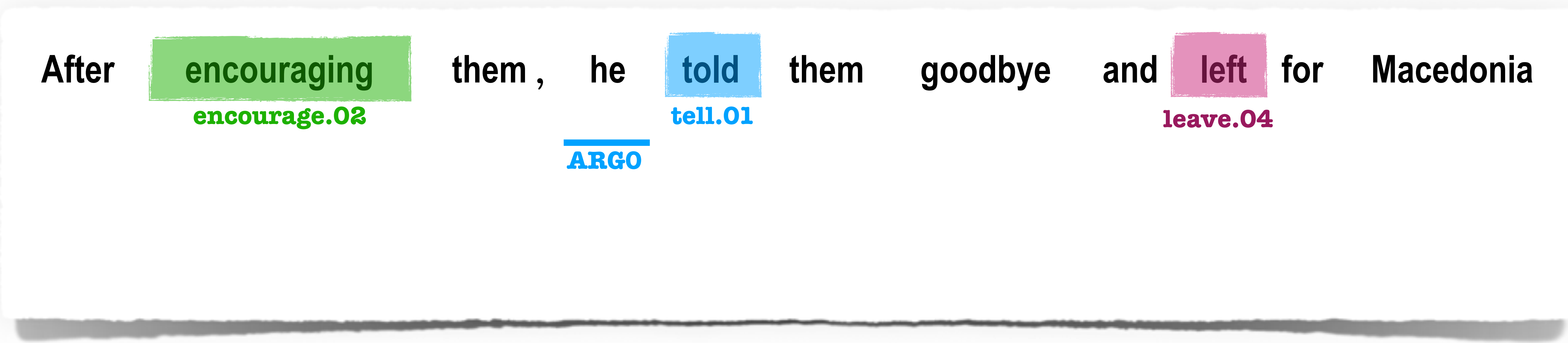
After encouraging them , he told them goodbye and left for Macedonia

encourage.02 tell.01 leave.04

PropBank [Palmer et. al., 05]

Linguistic Structure: Semantics

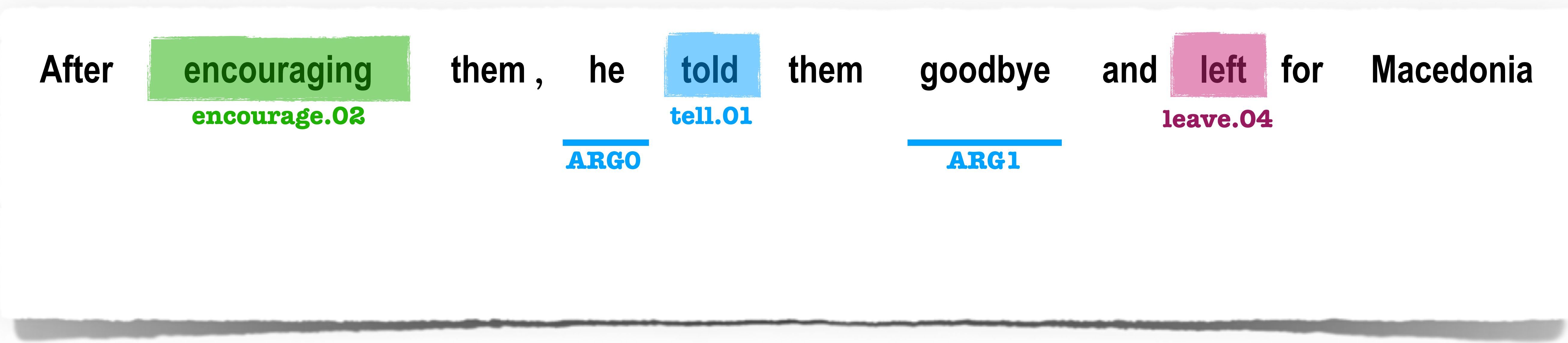
- Who did what to whom?



PropBank [Palmer et. al., 05]

Linguistic Structure: Semantics

- Who did what to whom?



PropBank [Palmer et. al., 05]

Linguistic Structure: Semantics

- Who did what to whom?

After encouraging them , he told them goodbye and left for Macedonia

encourage.02 tell.01 leave.04

ARG0 ARG2 ARG1

PropBank [Palmer et. al., 05]

Linguistic Structure: Semantics

- Who did what to whom?

PropBank [Palmer et. al., 05]

Linguistic Structure: Semantics

- Who did what to whom?

PropBank [Palmer et. al., 05]

Linguistic Structure: Semantics

- Who did what to whom?

PropBank [Palmer et. al., 05]

Linguistic Structure: Semantics

- Who did what to whom?
- This talk: **Span**-based semantics.

PropBank [Palmer et. al., 05]

Linguistic Structure: Semantics

- Who did what to whom?
- This talk: **Span**-based semantics.
- Can span-based semantics serve as a linguistic prior?

PropBank [Palmer et. al., 05]

Linguistic Structure: Semantics

- Who did what to whom?
- This talk: **Span**-based semantics.
- ~~Can span-based semantics serve as a linguistic prior?~~

PropBank [Palmer et. al., 05]

A Prior for Semantics

A Prior for Semantics

- **Syntax** - a foundation for sentence meaning / semantics

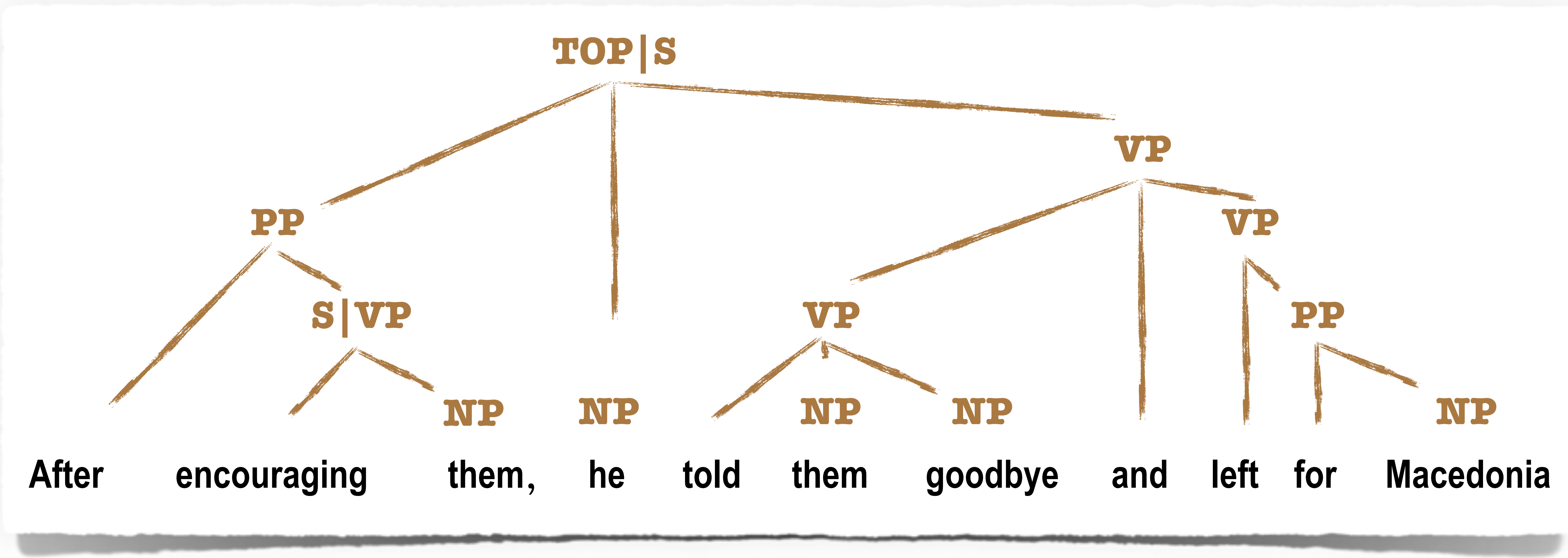
A Prior for Semantics

- **Syntax** - a foundation for sentence meaning / semantics

After encouraging them, he told them goodbye and left for Macedonia

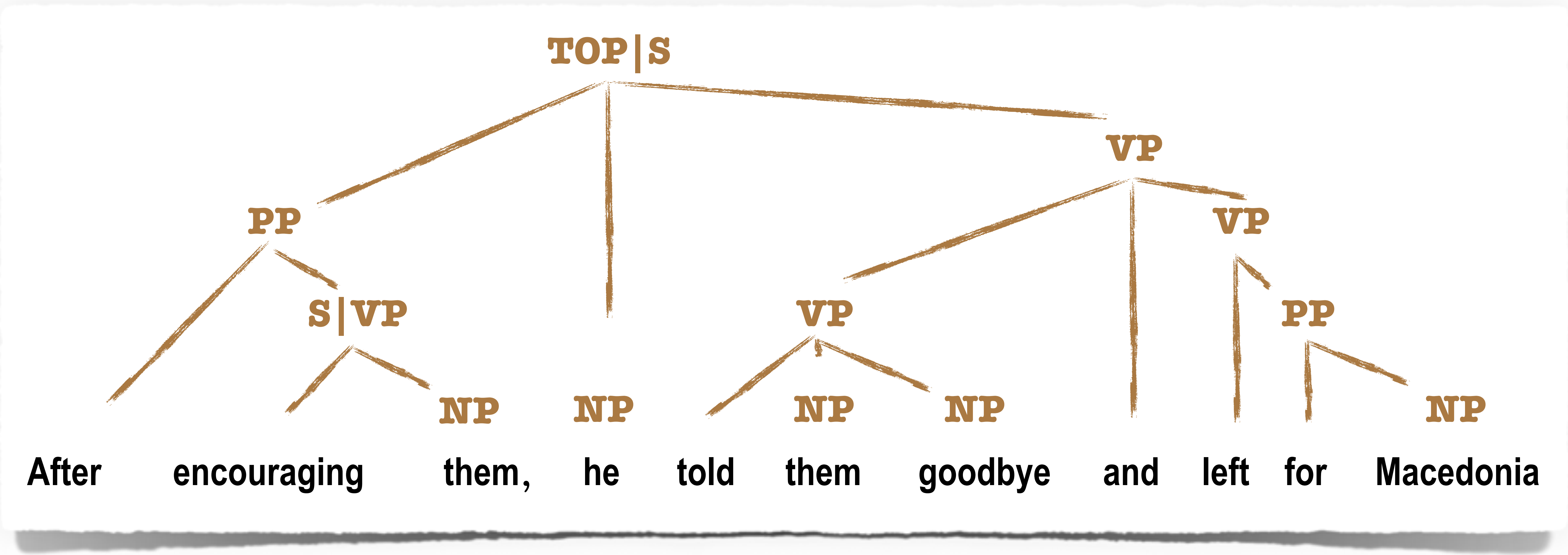
A Prior for Semantics

- **Syntax** - a foundation for sentence meaning / semantics



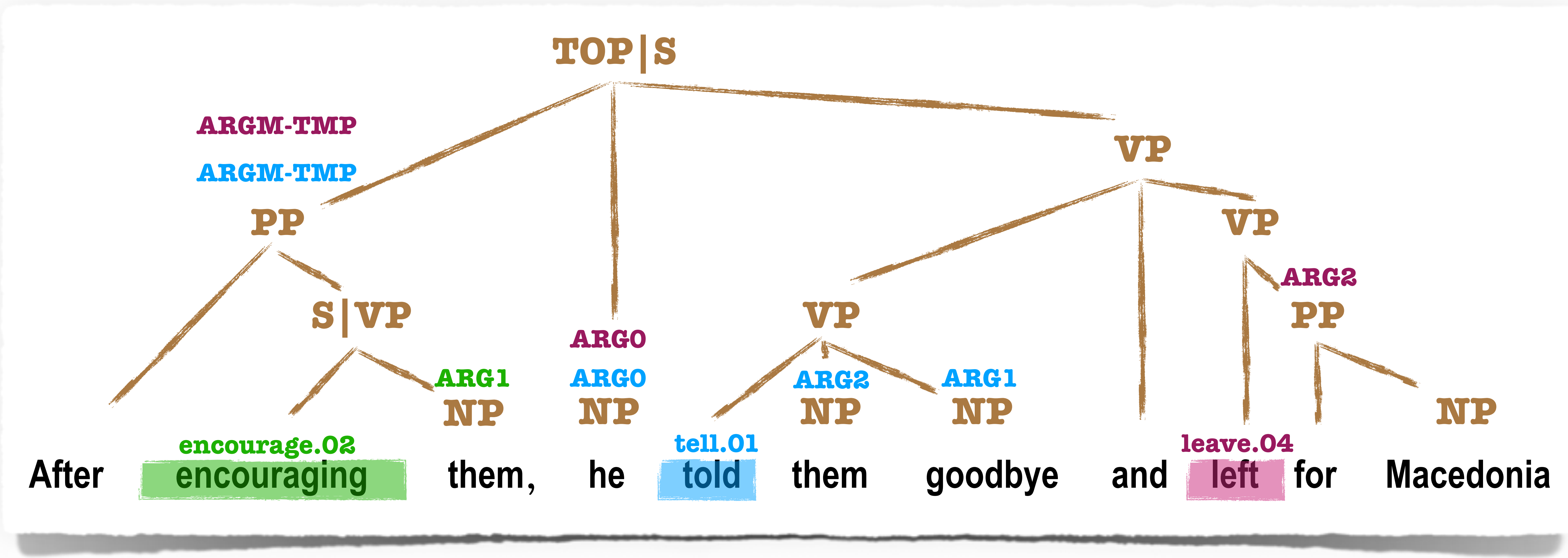
A Prior for Semantics

- **Syntax** - a foundation for sentence meaning / semantics
- Phrase-based syntax (node → span)



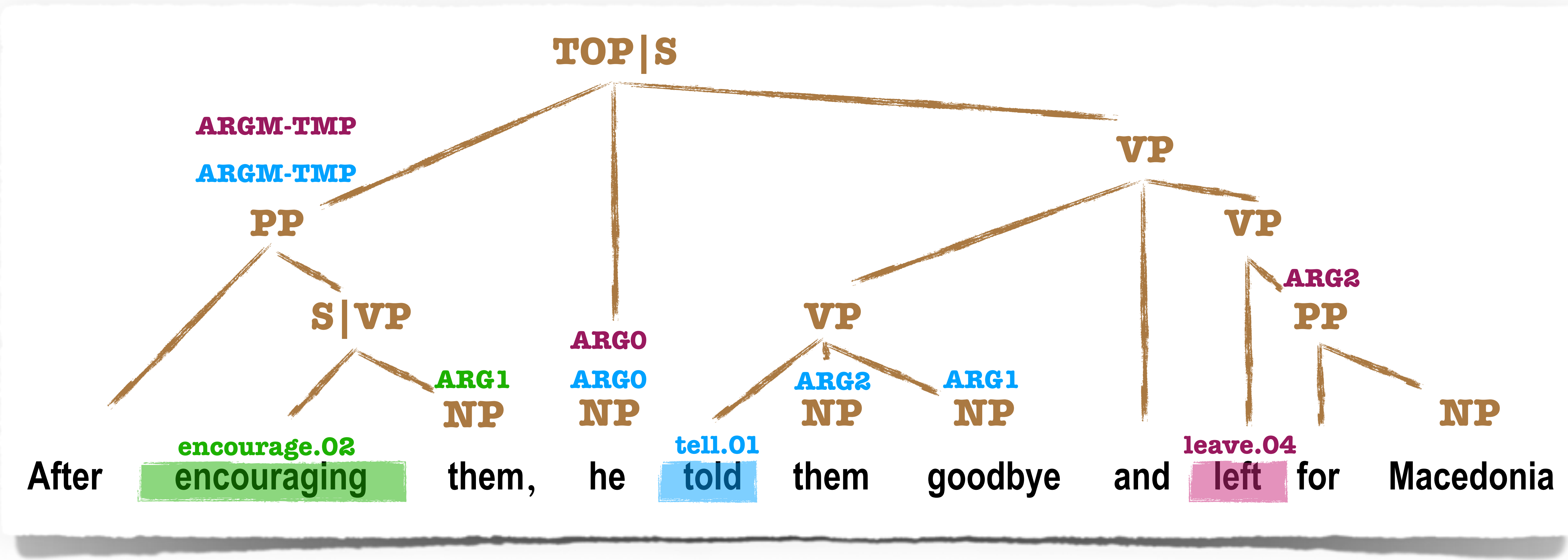
A Prior for Semantics

- **Syntax** - a foundation for sentence meaning / semantics
- Phrase-based syntax (node → span)

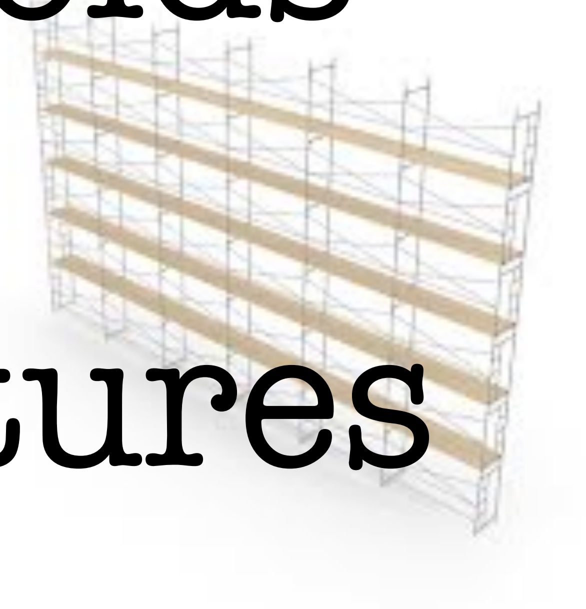


A Prior for Semantics

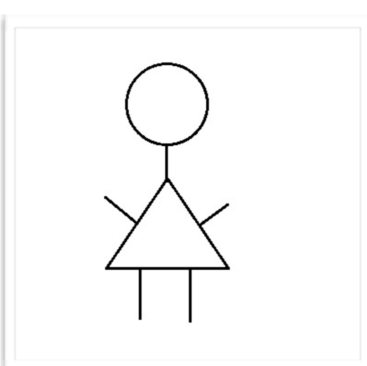
- **Syntax** - a foundation for sentence meaning / semantics
- Phrase-based syntax (node → span)
- Key Intuition: Learn from a **complementary** structure



Syntactic Scaffolds for Semantic Structures



EMNLP 2018

S.

Sam Thomson

Kenton Lee

Luke Zettlemoyer

Chris Dyer

Noah A. Smith

Structured prediction with an auxiliary structure

Structured prediction with an auxiliary structure

- Auxiliary structure: **syntax**

Structured prediction with an auxiliary structure

- Auxiliary structure: **syntax**

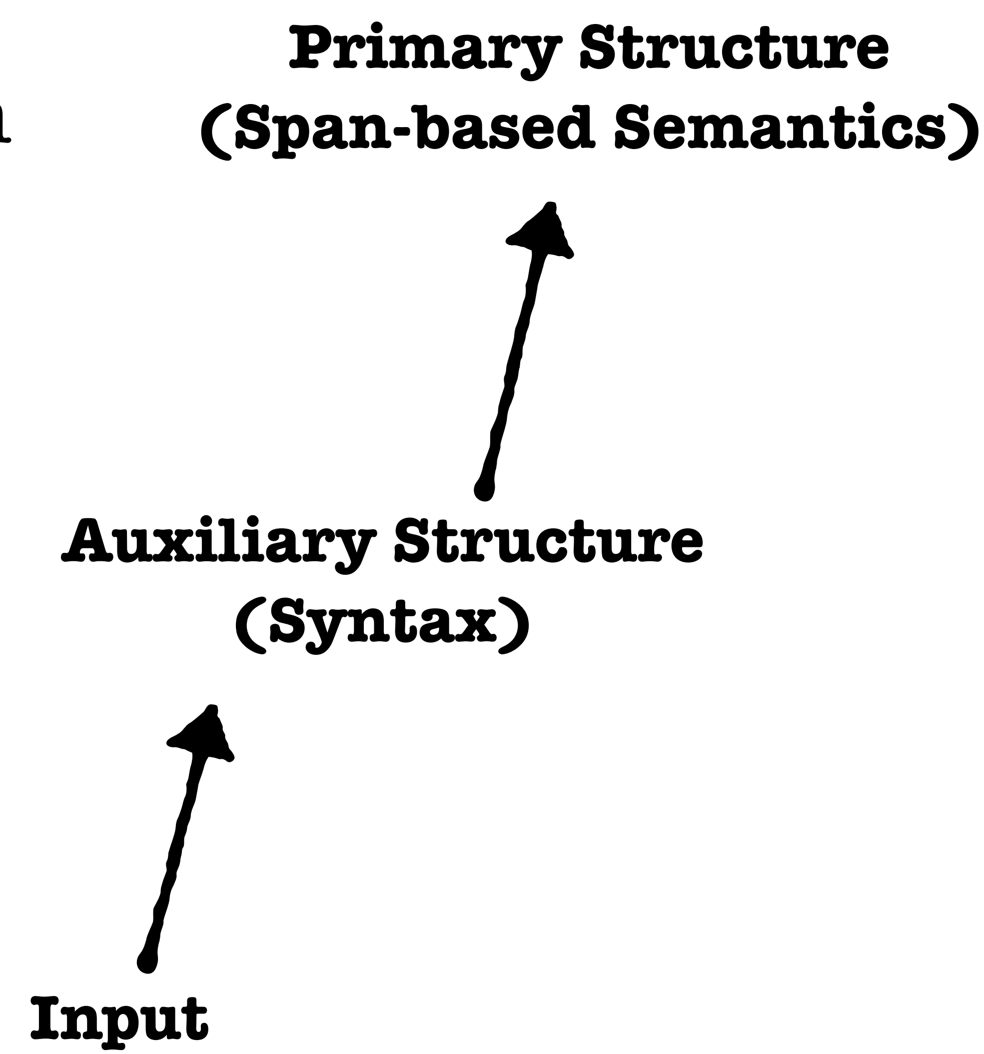
Primary Structure
(Span-based Semantics)

Auxiliary Structure
(Syntax)

Input

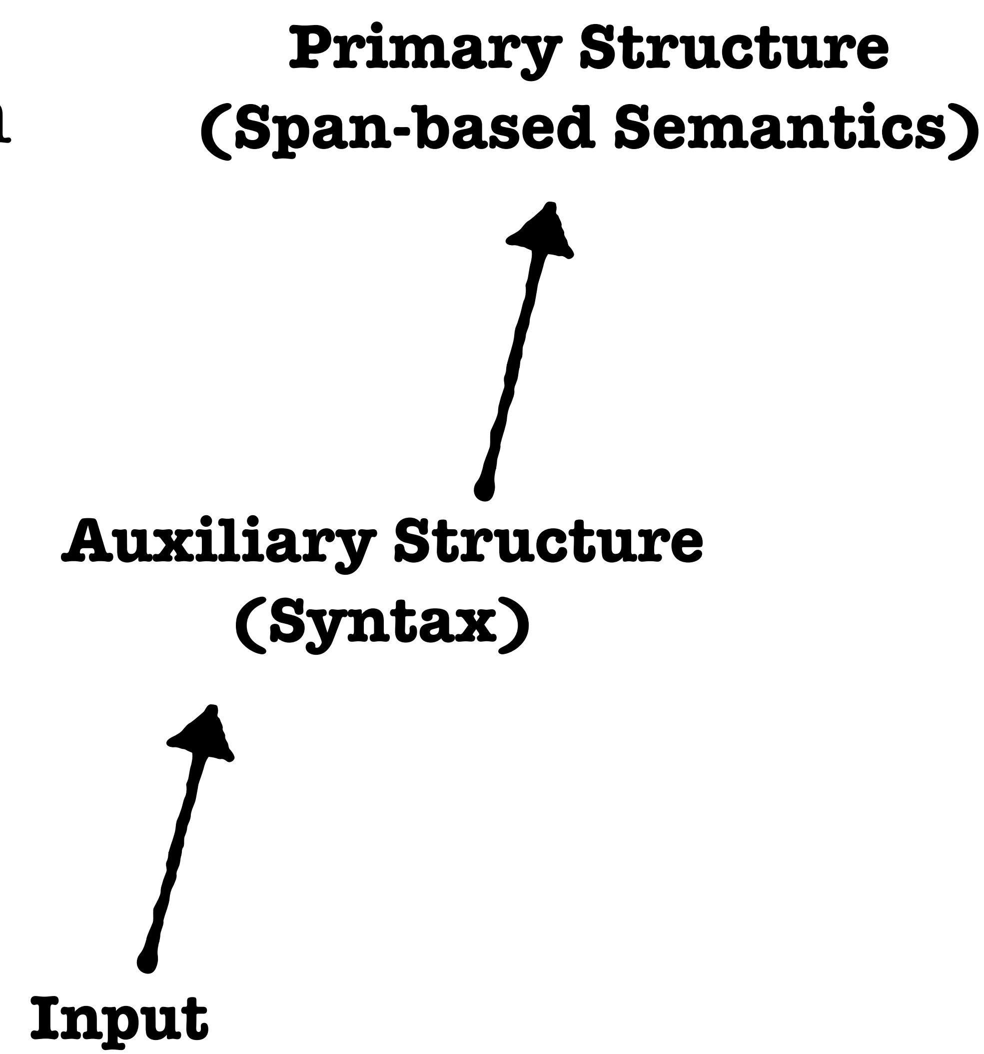
Structured prediction with an auxiliary structure

- Auxiliary structure: **syntax**
- Traditionally a pipeline, both at train and test time [Gildea & Jurafsky, 2002]



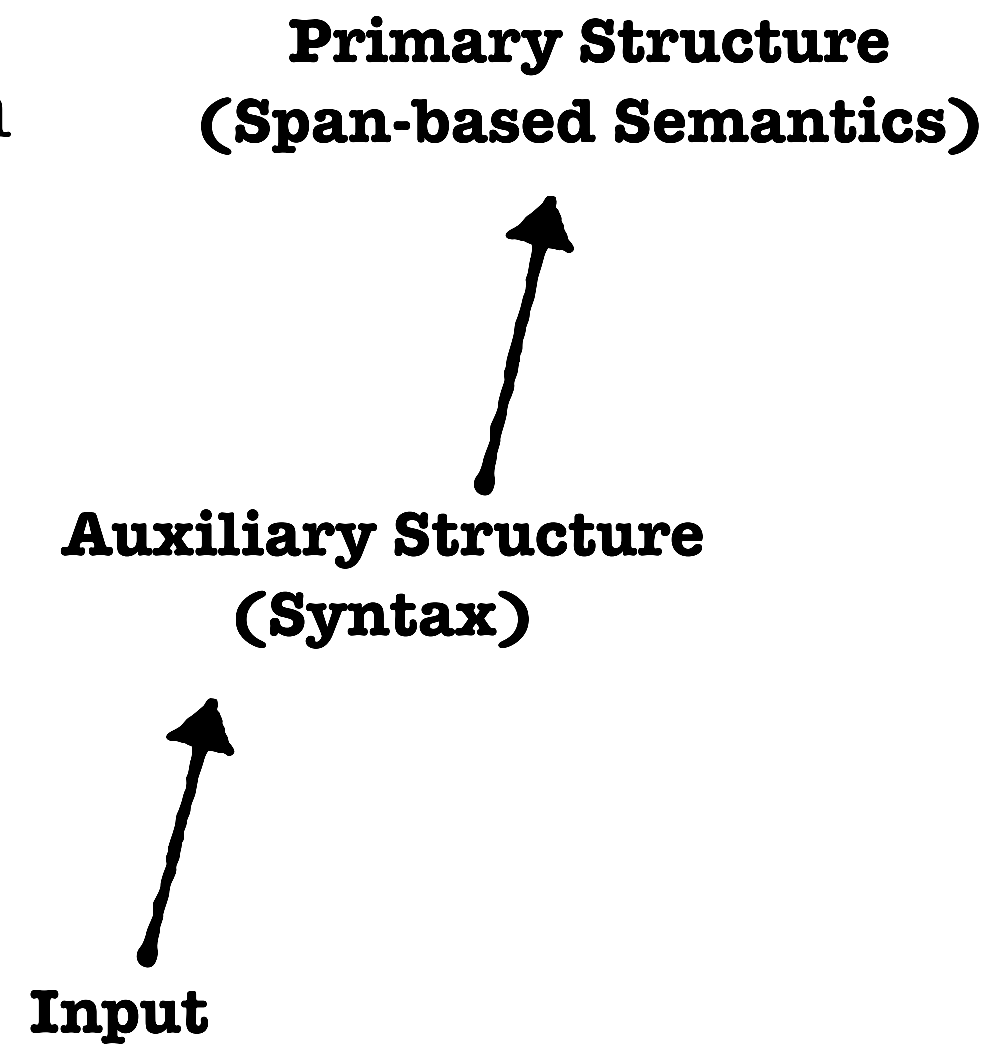
Structured prediction with an auxiliary structure

- Auxiliary structure: **syntax**
- Traditionally a pipeline, both at train and test time [Gildea & Jurafsky, 2002]
    - More structured data



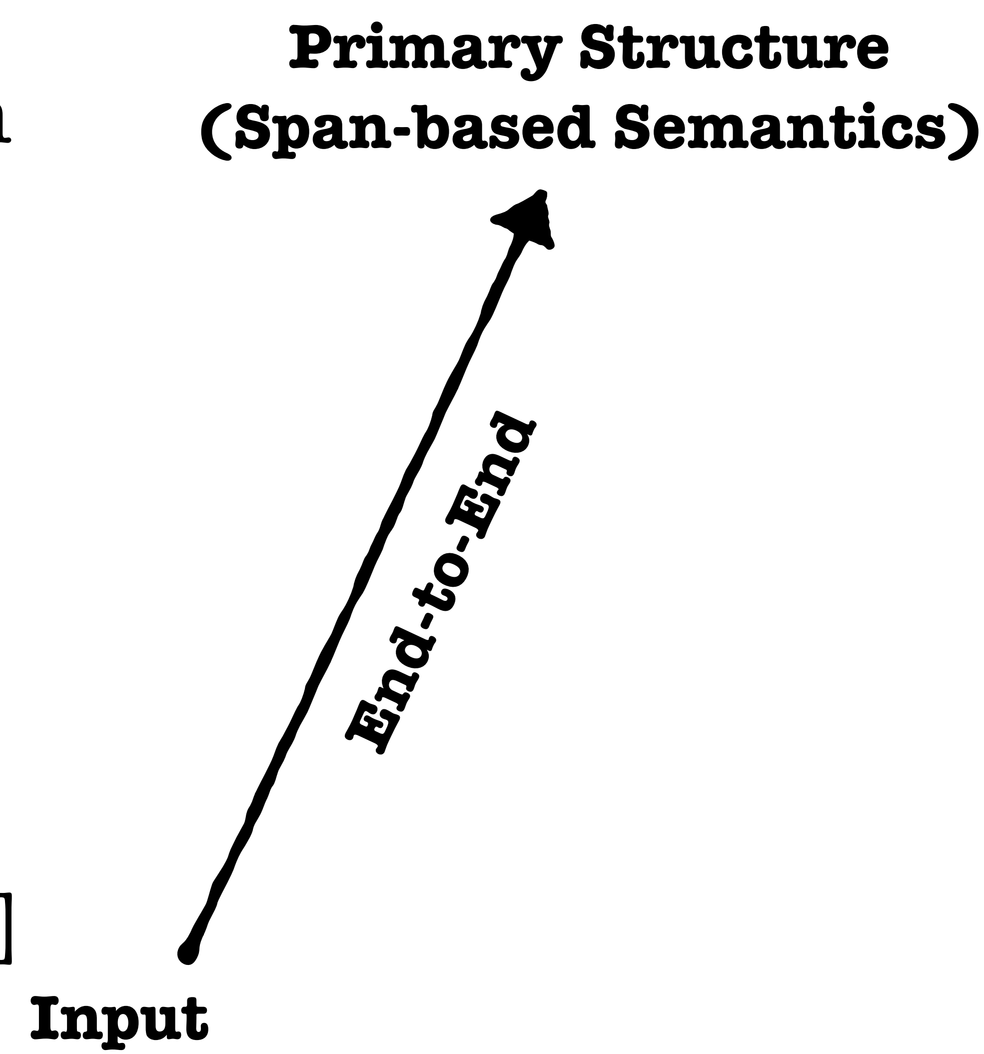
Structured prediction with an auxiliary structure

- Auxiliary structure: **syntax**
- Traditionally a pipeline, both at train and test time [Gildea & Jurafsky, 2002]
  - More structured data
  - Cascading errors

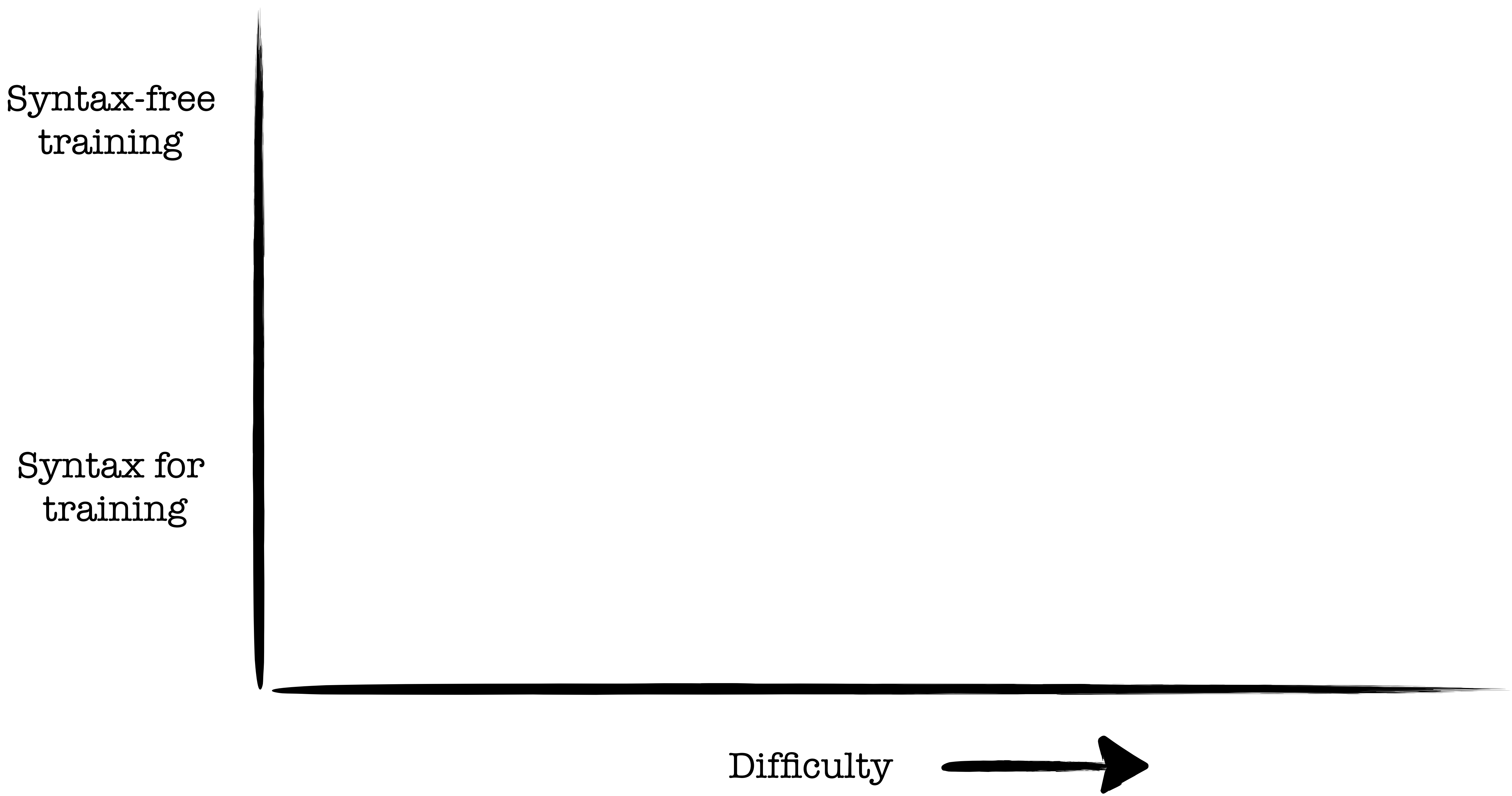


Structured prediction with an auxiliary structure

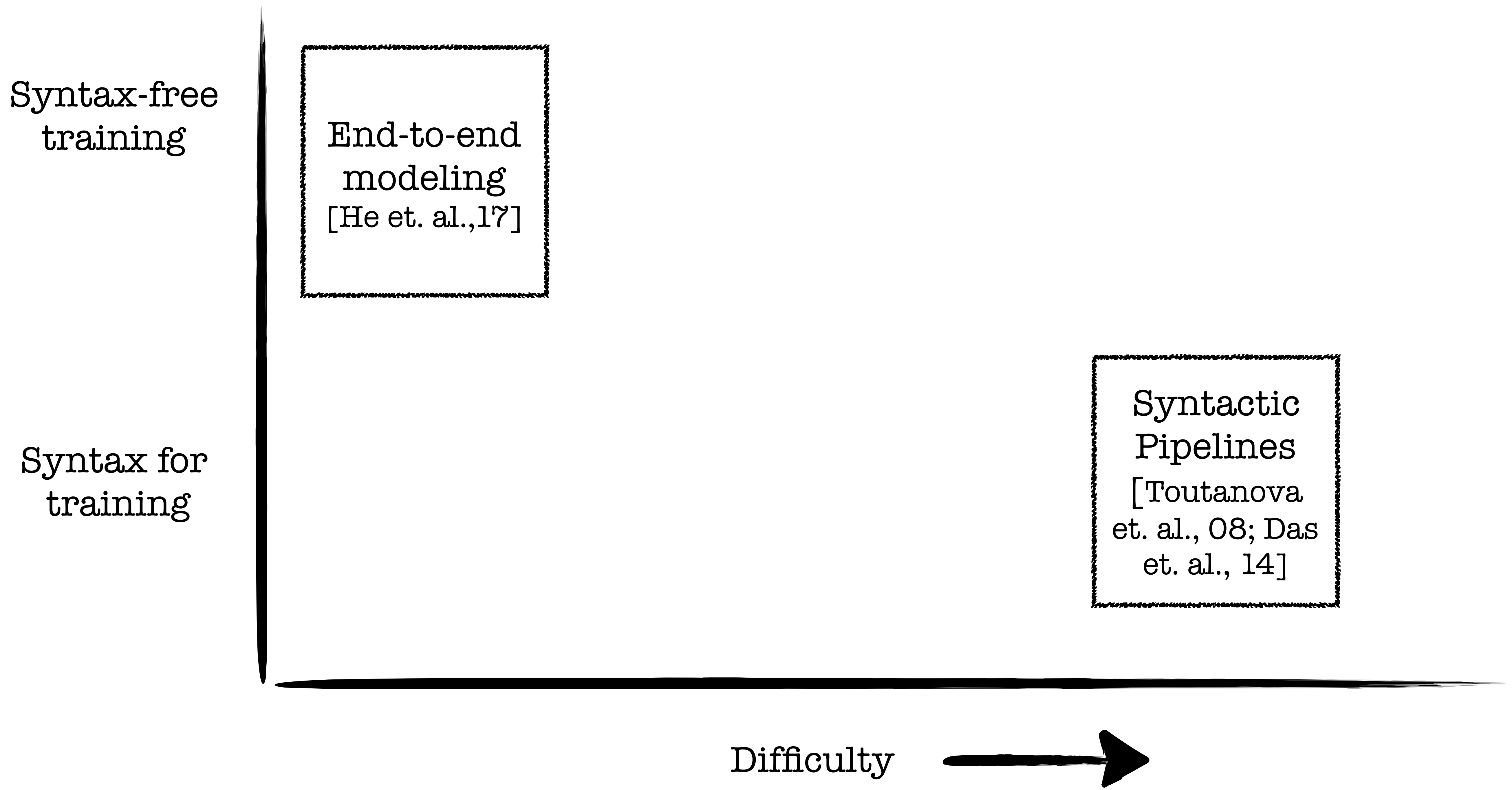
- Auxiliary structure: **syntax**
- Traditionally a pipeline, both at train and test time [Gildea & Jurafsky, 2002]
  - More structured data
  - Cascading errors
- Forsaken in most end-to-end models, but at a cost [He et. al, 17; Strubell et. al., 18]



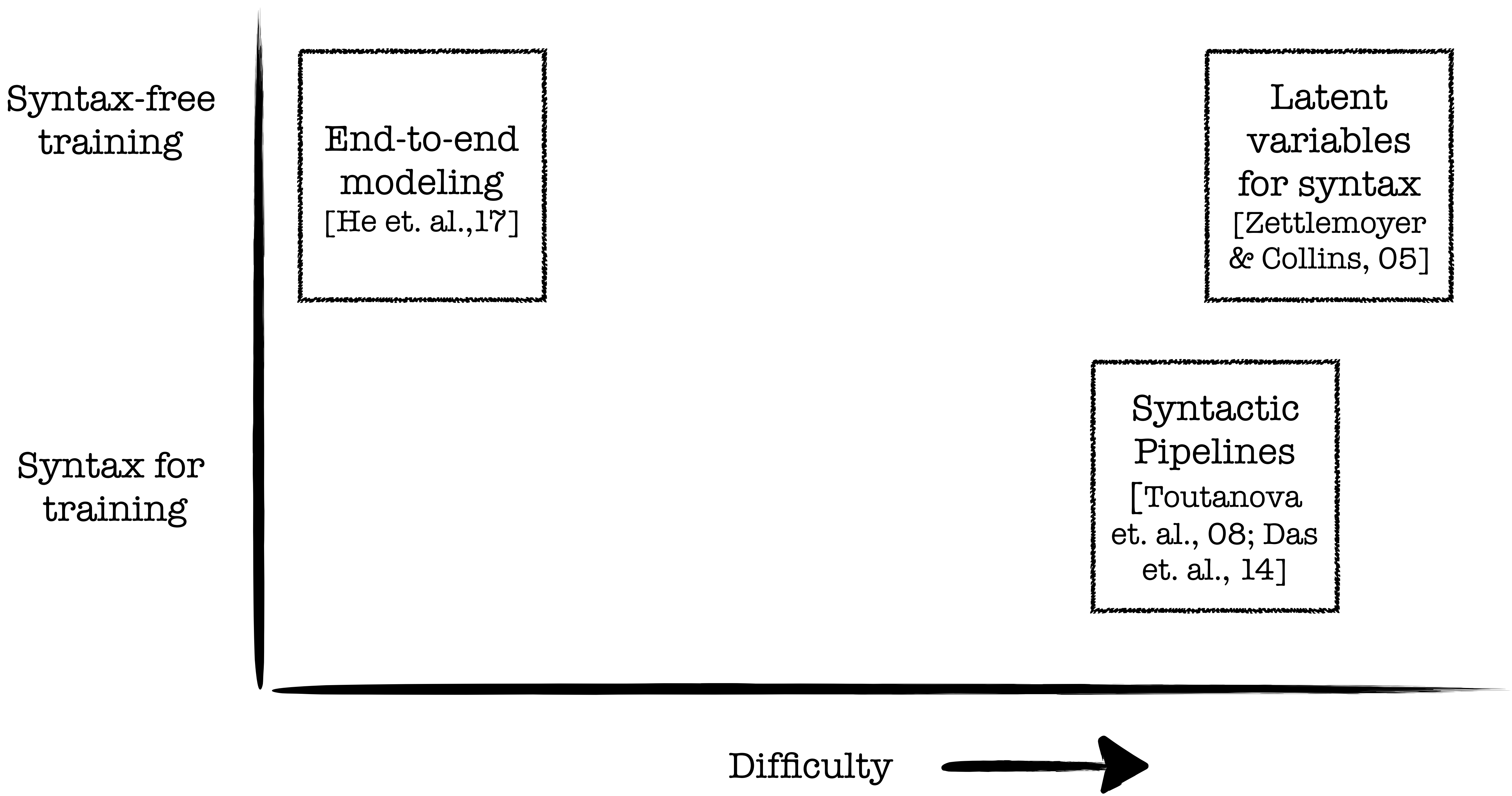
Training Paradigms



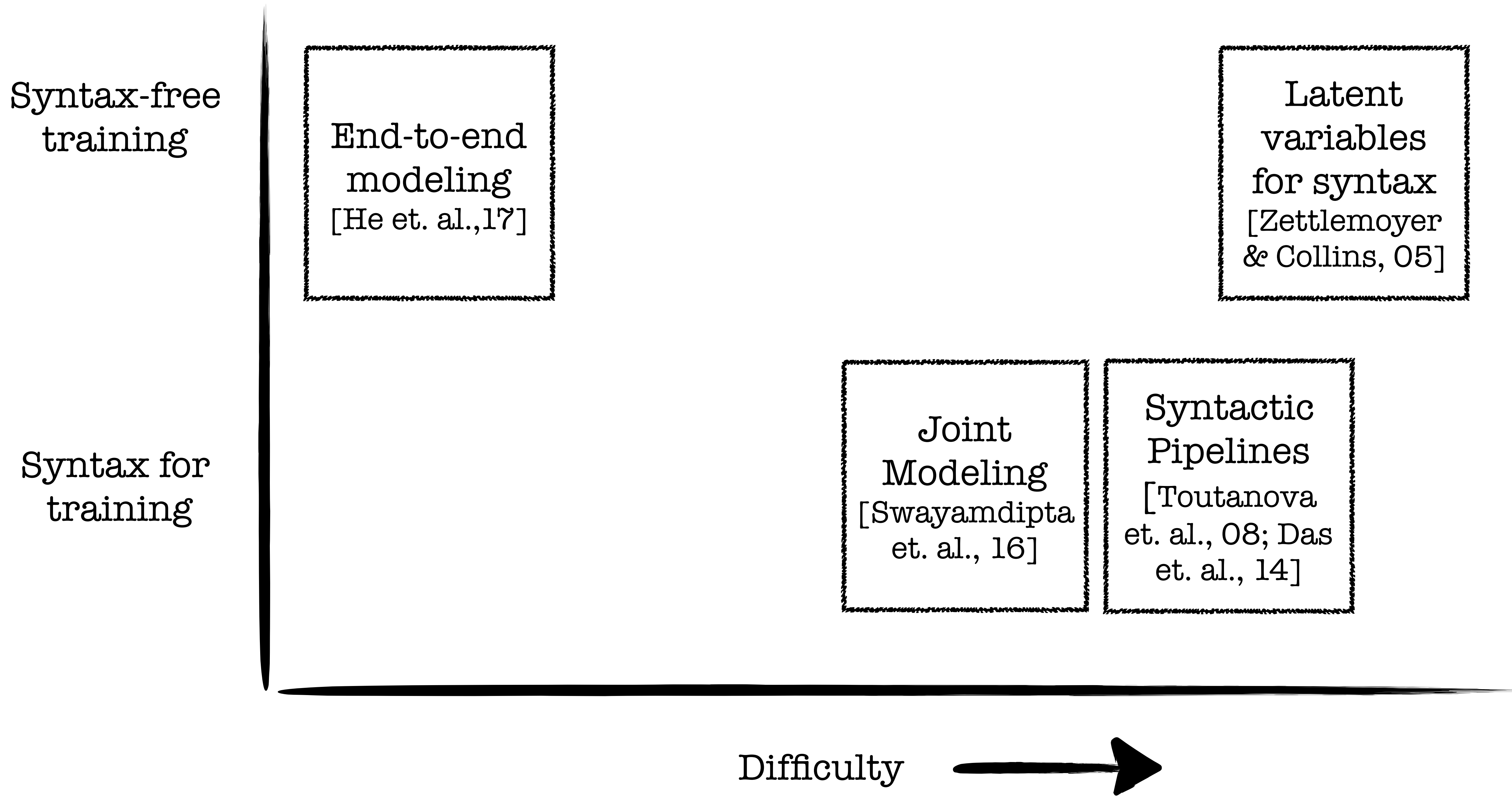
Training Paradigms



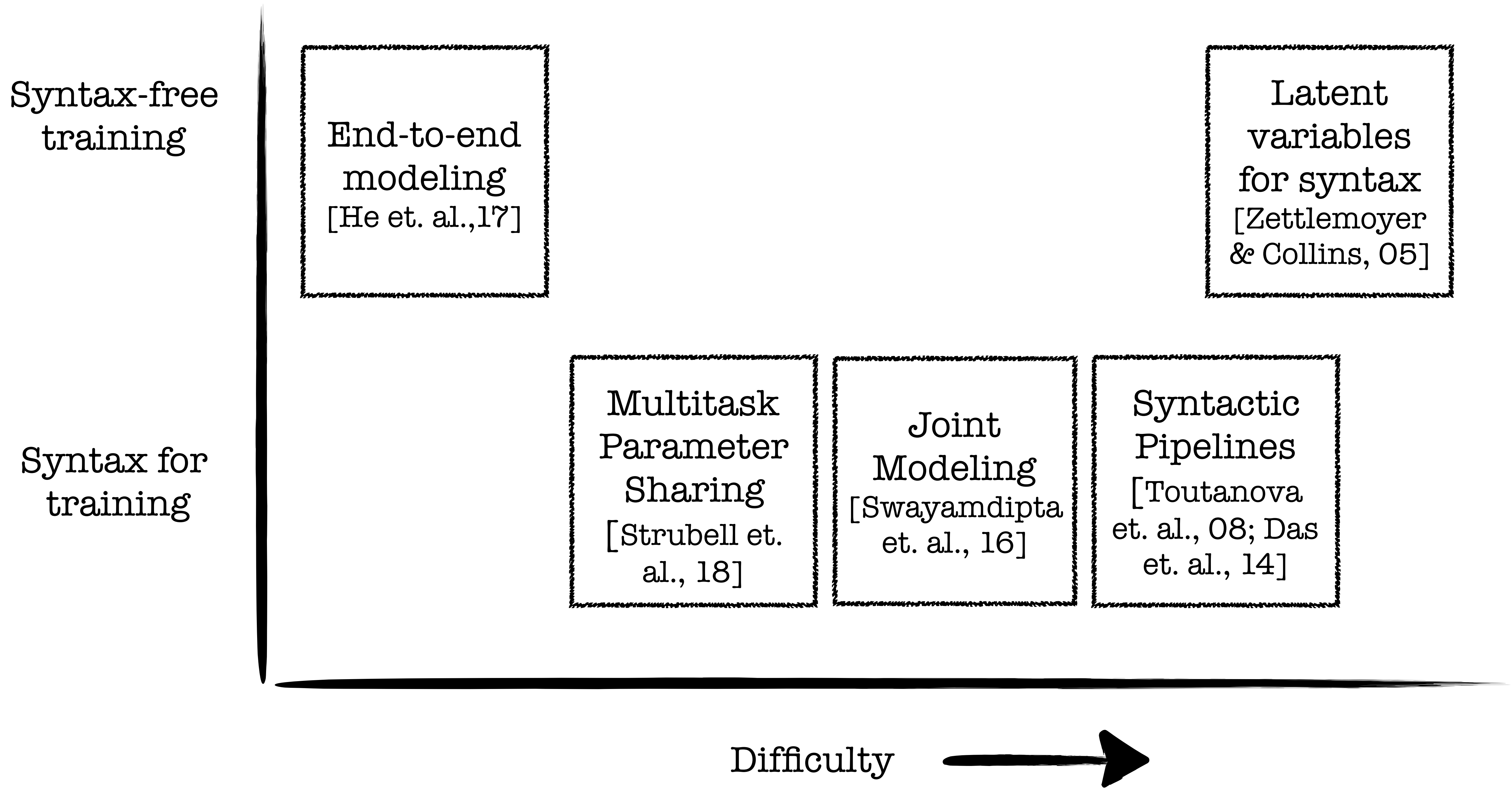
Training Paradigms



Training Paradigms



Training Paradigms



FREDA [Daumé III, 2009]

Training Paradigms

Syntax-free training

End-to-end modeling [He et. al.,17]

Latent variables for syntax [Zettlemoyer & Collins, 05]

Syntax for training

Multitask Parameter Sharing [Strubell et. al., 18]

Joint Modeling [Swayamdipta et. al., 16]

Syntactic Pipelines [Toutanova et. al., 08; Das et. al., 14]

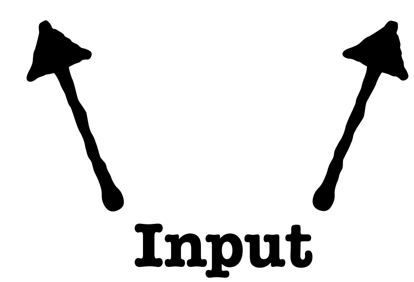
Difficulty →

FREDA [Daumé III, 2009]

Syntactic Scaffolds

Syntactic Scaffolds

- Multitask setting



Syntactic Scaffolds

- Multitask setting
  - Primary Task → Span-based Semantics

- ☑ PropBank Semantic Role Labeling
- ☑ Frame-Semantic Role Labeling
- ☑ Coreference Resolution

Span-based Semantics

Input

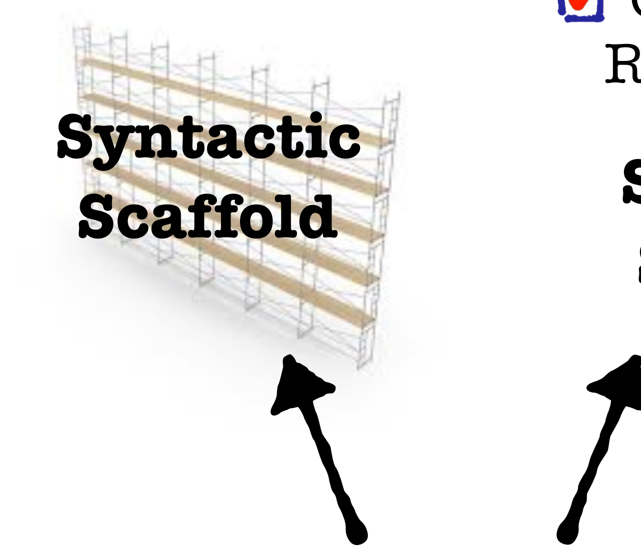
Syntactic Scaffolds

- Multitask setting
  - Primary Task → Span-based Semantics
  - Scaffold "Task" → Syntax

☑ PropBank Semantic Role Labeling

☑ Frame-Semantic Role Labeling

☑ Coreference Resolution



Span-based Semantics

Input

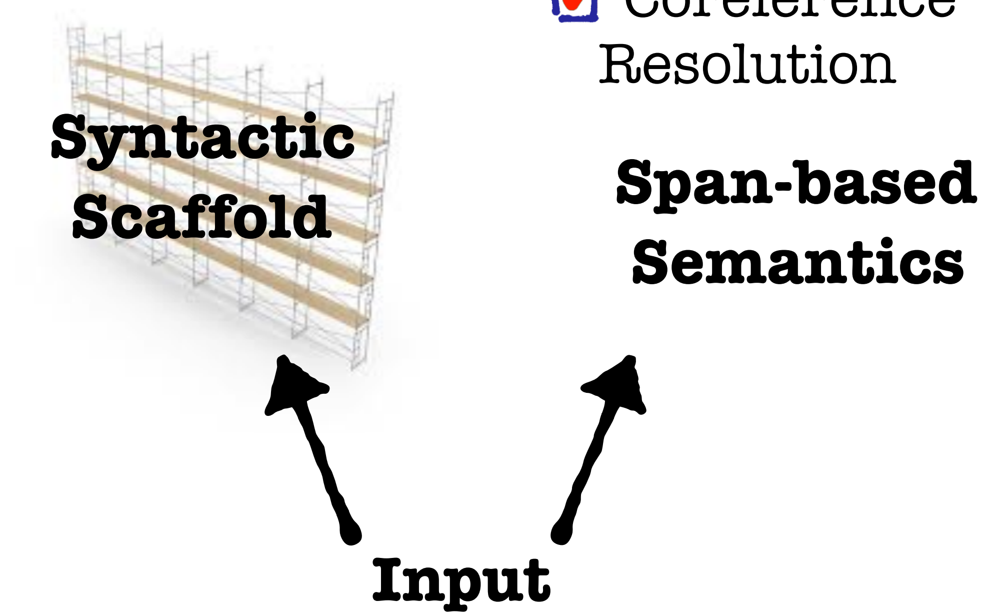
Syntactic Scaffolds

- Multitask setting
  - Primary Task → Span-based Semantics
  - Scaffold "Task" → Syntax
    - ~~Full Trees~~ Shallow syntax

☑ PropBank Semantic Role Labeling

☑ Frame-Semantic Role Labeling

☑ Coreference Resolution



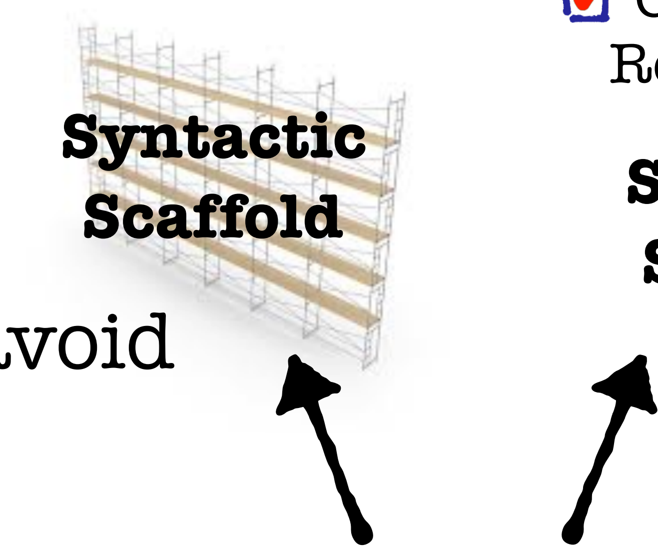
Syntactic Scaffolds

- Multitask setting
  - Primary Task→ Span-based Semantics
  - Scaffold “Task”→Syntax
    - ~~Full Trees~~ Shallow syntax
- Soft syntax-aware representations avoid cascaded errors

☑ PropBank Semantic Role Labeling

☑ Frame-Semantic Role Labeling

☑ Coreference Resolution



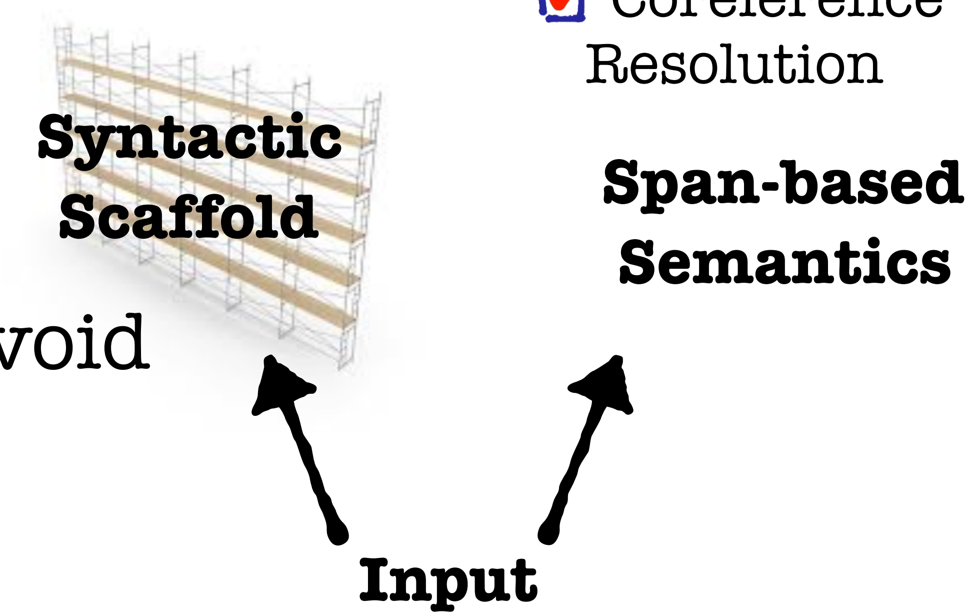
Syntactic Scaffolds

- Multitask setting
  - Primary Task→ Span-based Semantics
  - Scaffold "Task"→Syntax
    - ~~Full Trees~~ Shallow syntax
- Soft syntax-aware representations avoid cascaded errors
- Not required during test

☑ PropBank Semantic Role Labeling

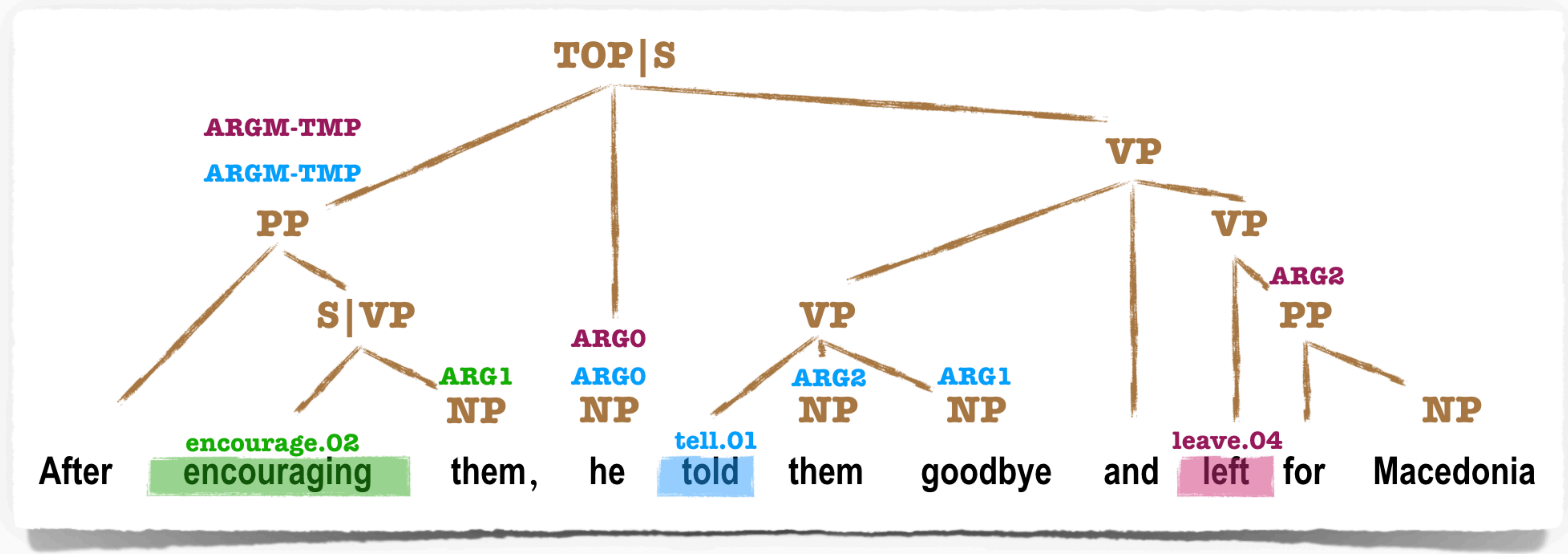
☑ Frame-Semantic Role Labeling

☑ Coreference Resolution



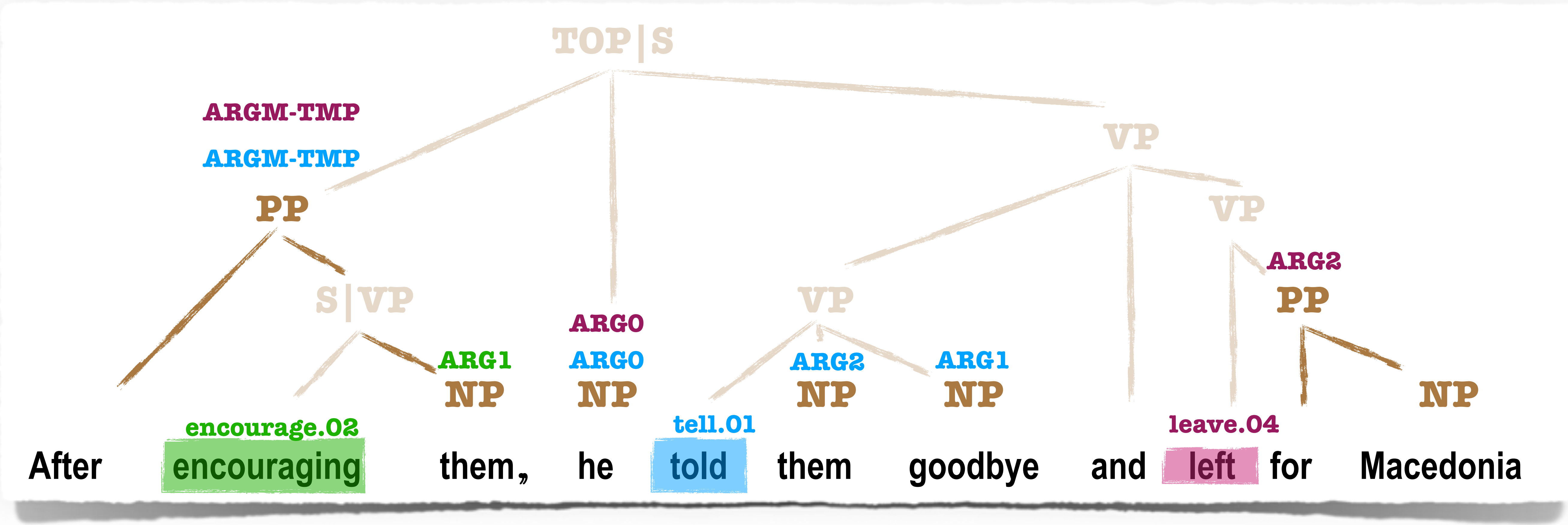
Shallow Syntactic Prediction

- **Desired** parts of syntactic tree:



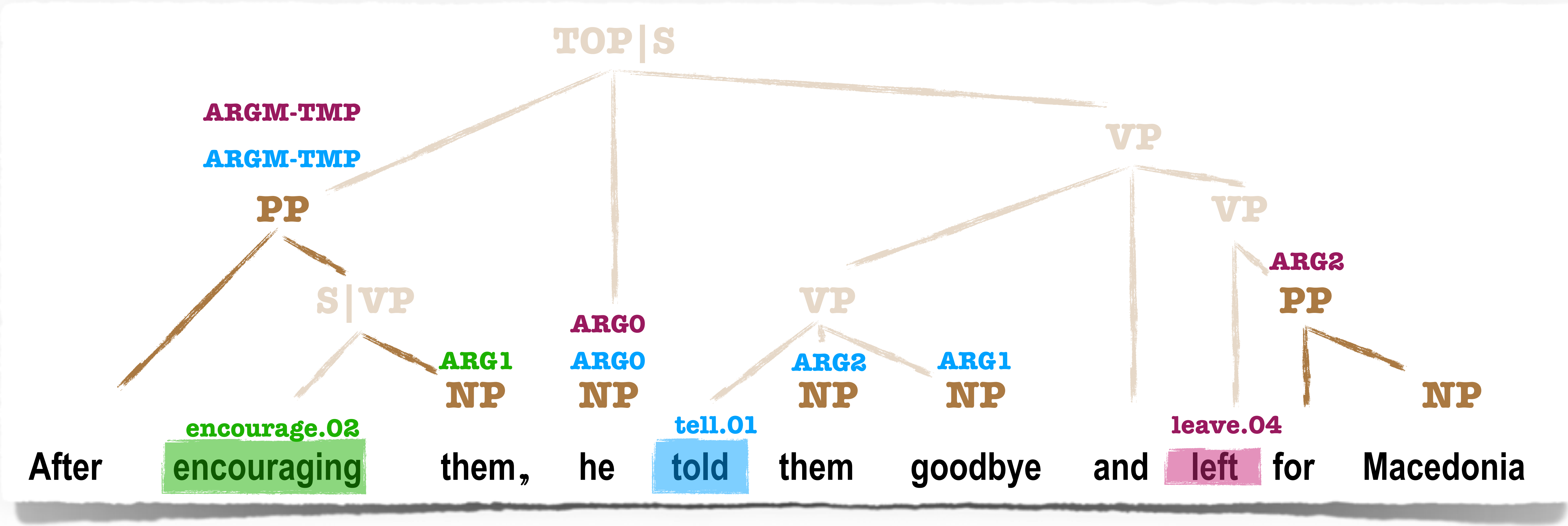
Shallow Syntactic Prediction

- **Desired** parts of syntactic tree:



Shallow Syntactic Prediction

- **Desired** parts of syntactic tree:



- Span-level classification: For every span, predict phrase category

$$\mathscr{L}_2(\mathbf{x}, \mathbf{z}) = - \sum_{1 \leqslant i \leqslant j \leqslant n} \log p(z_{i:j} \mid \mathbf{x}_{i:j})$$

Training with syntactic scaffolds

x = Input
y = Output Structure
z = Scaffold Structure

Training with syntactic scaffolds

x = Input
y = Output Structure
z = Scaffold Structure

$$\sum_{(\mathbf{x},\mathbf{z}) \in \mathscr{D}_2} \mathscr{L}_2(\mathbf{x}, \mathbf{z}; \theta, \psi)$$

Scaffold Dataset

Scaffold Task Objective

Training with syntactic scaffolds

x = Input
y = Output Structure
z = Scaffold Structure

$$\sum_{(\mathbf{x},\mathbf{y})\in\mathscr{D}_1} \mathscr{L}_1(\mathbf{x},\mathbf{y};\theta,\phi) \qquad \sum_{(\mathbf{x},\mathbf{z})\in\mathscr{D}_2} \mathscr{L}_2(\mathbf{x},\mathbf{z};\theta,\psi)$$

Primary Dataset

Primary Task Objective

Scaffold Dataset

Scaffold Task Objective

Training with syntactic scaffolds

x = Input
y = Output Structure
z = Scaffold Structure

$$\sum_{(\mathbf{x},\mathbf{y})\in\mathcal{D}_1} \mathcal{L}_1(\mathbf{x},\mathbf{y};\theta,\phi) + \delta \sum_{(\mathbf{x},\mathbf{z})\in\mathcal{D}_2} \mathcal{L}_2(\mathbf{x},\mathbf{z};\theta,\psi)$$

Primary Dataset

Primary Task Objective

Mixing Ratio

Scaffold Dataset

Scaffold Task Objective

Training with syntactic scaffolds

x = Input
y = Output Structure
z = Scaffold Structure

$$\sum_{(\mathbf{x},\mathbf{y})\in\mathscr{D}_1} \mathscr{L}_1(\mathbf{x},\mathbf{y};\theta,\phi) + \delta \sum_{(\mathbf{x},\mathbf{z})\in\mathscr{D}_2} \mathscr{L}_2(\mathbf{x},\mathbf{z};\theta,\psi)$$

Primary Dataset

Primary Task Objective

Mixing Ratio

Scaffold Dataset

Scaffold Task Objective

Shared input parameters

The primary objective

The primary objective

Same structures must be scored in both the primary and the scaffold task.

The primary objective

Same structures must be scored in both the primary and the scaffold task.

- Span-based classification, with aggressive pruning [Lee et. al., 2017]

The primary objective

Same structures must be scored in both the primary and the scaffold task.

- Span-based classification, with aggressive pruning [Lee et. al., 2017]
- Semi-Markov Conditional Random Fields [Sarawagi et. al. 2004]

Semi-Markov CRFs

After encouraging them he told them goodbye and left for Macedonia

ARGM-TMP ARG0 leave.04 ARG2

Semi-Markov CRFs

After encouraging them he told them goodbye and left for Macedonia

ARGM-TMP ARG0 leave.04 ARG2

- Globally normalized model for segmentations (**s**) of a sentence (**x**)

Semi-Markov CRFs

After encouraging them he told them goodbye and left for Macedonia

ARGM-TMP ARG0 leave.04 ARG2

- Globally normalized model for segmentations ($\mathbf{s}$) of a sentence ($\mathbf{x}$)

$$p(\mathbf{s} \mid \mathbf{x})$$

Semi-Markov CRFs

After encouraging them he told them goodbye and left for Macedonia

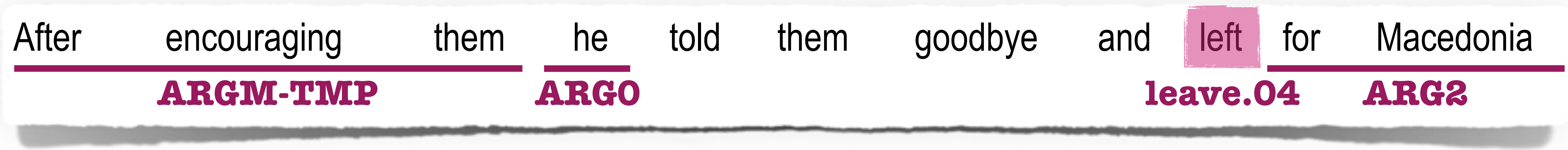
ARGM-TMP ARG0 leave.04 ARG2

- Globally normalized model for segmentations ($\mathbf{s}$) of a sentence ($\mathbf{x}$)

$$p(\mathbf{s} \mid \mathbf{x})$$

- Generalization of CRFs [Lafferty et. al., 01]:

Semi-Markov CRFs



- Globally normalized model for segmentations ($\mathbf{s}$) of a sentence ($\mathbf{x}$)

$$p(\mathbf{s} \mid \mathbf{x})$$

- Generalization of CRFs [Lafferty et. al., 01]:
  - label and length of an input segment

Semi-Markov CRFs

After encouraging them he told them goodbye and left for Macedonia

ARGM-TMP ARG0 leave.04 ARG2

- Globally normalized model for segmentations (**s**) of a sentence (**x**)

$$p(\mathbf{s} \mid \mathbf{x})$$

- Generalization of CRFs [Lafferty et. al., 01]:
  - label and length of an input segment

$$s = \langle i, j, y_{i:j} \rangle$$

Semi-Markov CRFs

After encouraging them he told them goodbye and left for Macedonia

ARGM-TMP ARG0 leave.04 ARG2

- Globally normalized model for segmentations (**s**) of a sentence (**x**)

$$p(\mathbf{s} \mid \mathbf{x})$$

- Generalization of CRFs [Lafferty et. al., 01]:
  - label and length of an input segment

$$s = \langle i, j, y_{i:j} \rangle$$

$$\Phi(\mathbf{x}, \mathbf{s}) = \sum_{k=1}^{m} \phi(s_k, x_{i_k:j_k})$$

Semi-Markov CRFs

After encouraging them he told them goodbye and left for Macedonia

ARGM-TMP ARG0 leave.04 ARG2

- Globally normalized model for segmentations (**s**) of a sentence (**x**)

$$p(\mathbf{s} \mid \mathbf{x})$$

- Generalization of CRFs [Lafferty et. al., 01]:
  - label and length of an input segment

$$s = \langle i, j, y_{i:j} \rangle$$

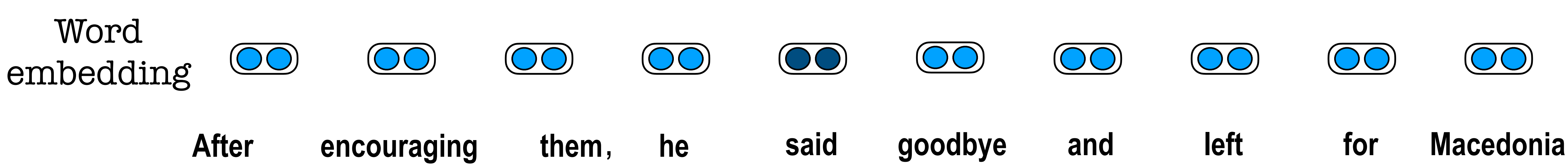
- Training and inference → O(ndl) dynamic programs, with a 0th-order Markovian assumption

$$\Phi(\mathbf{x}, \mathbf{s}) = \sum_{k=1}^{m} \phi(s_k, x_{i_k:j_k})$$

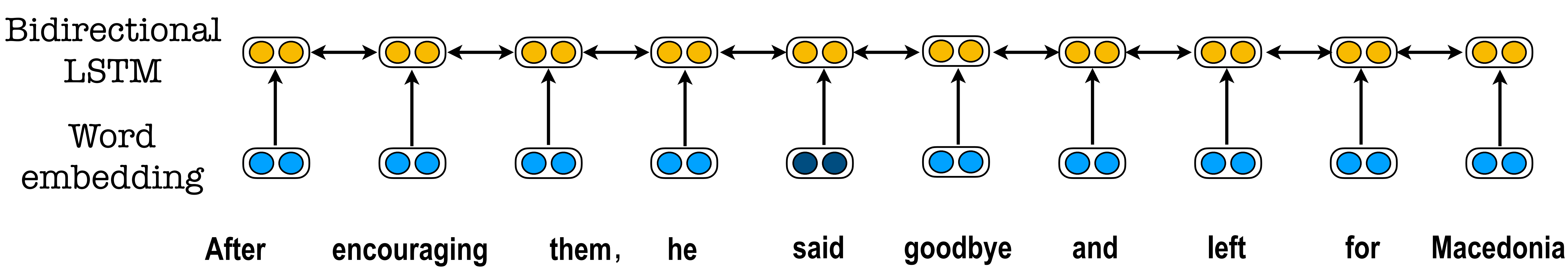
Model architecture

After encouraging them, he said goodbye and left for Macedonia

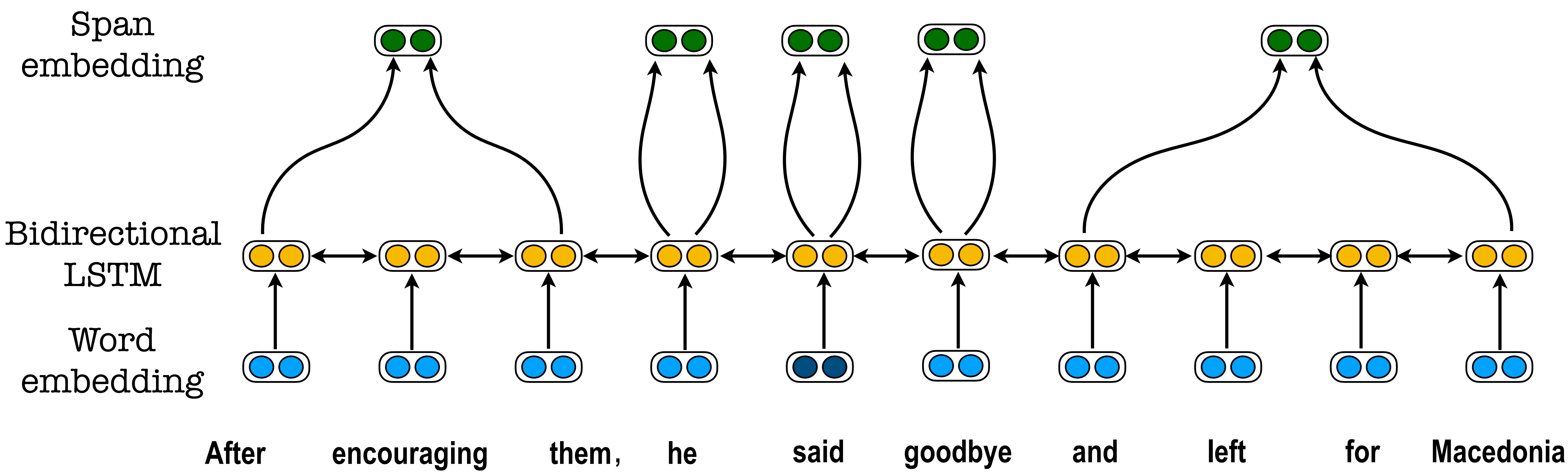
Model architecture



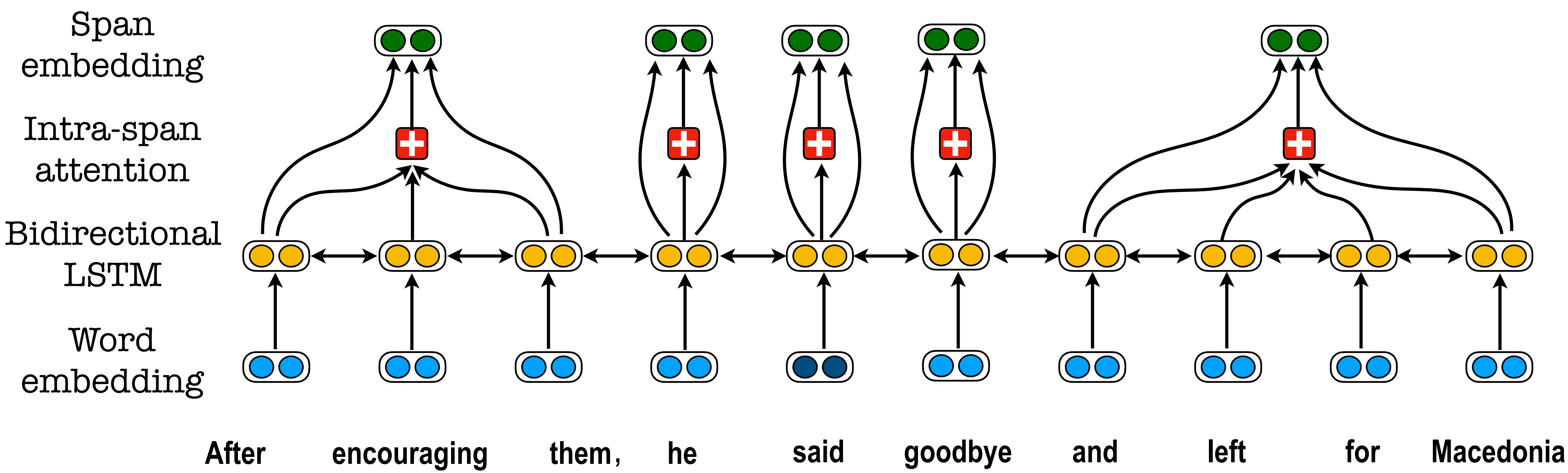
Model architecture



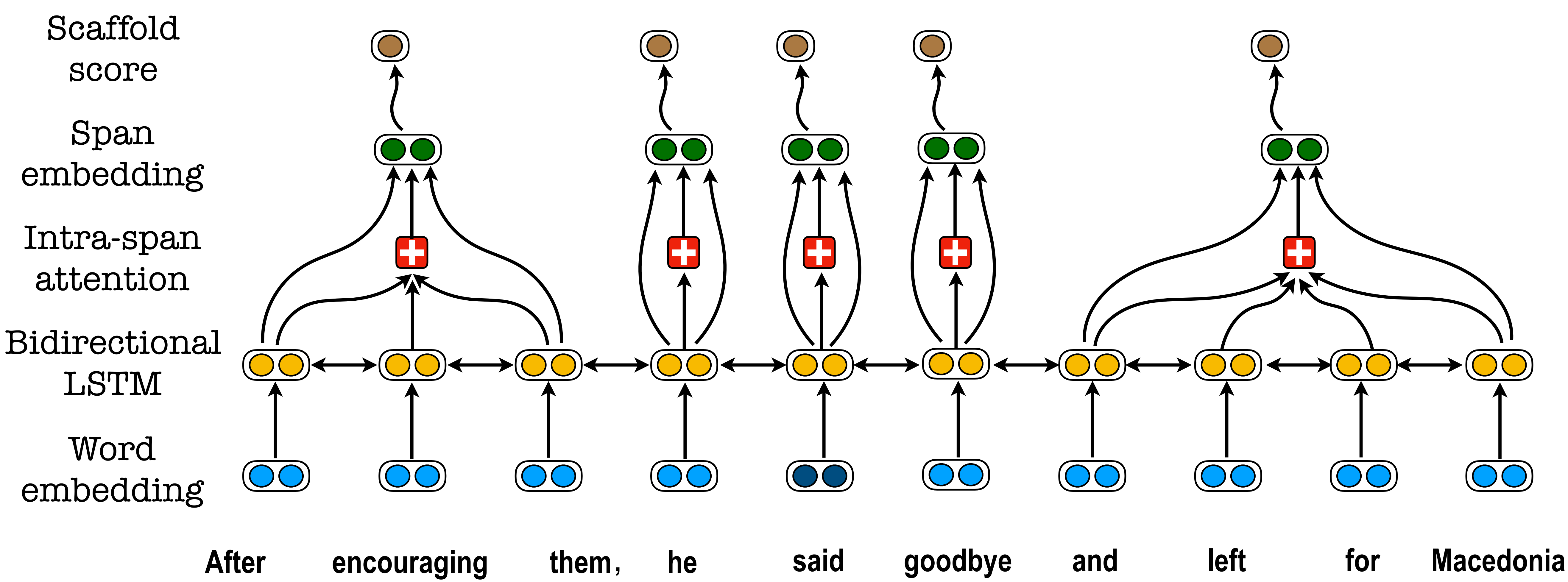
Model architecture



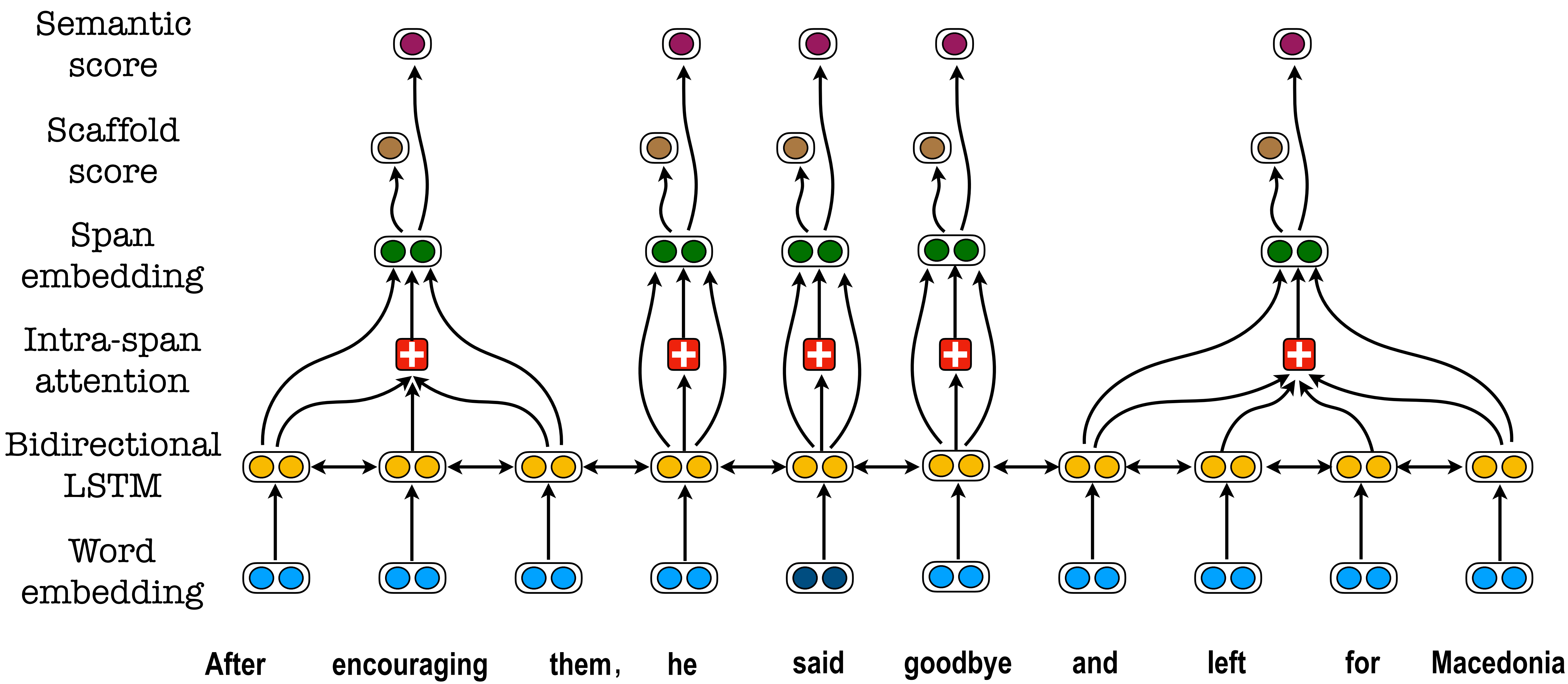
Model architecture



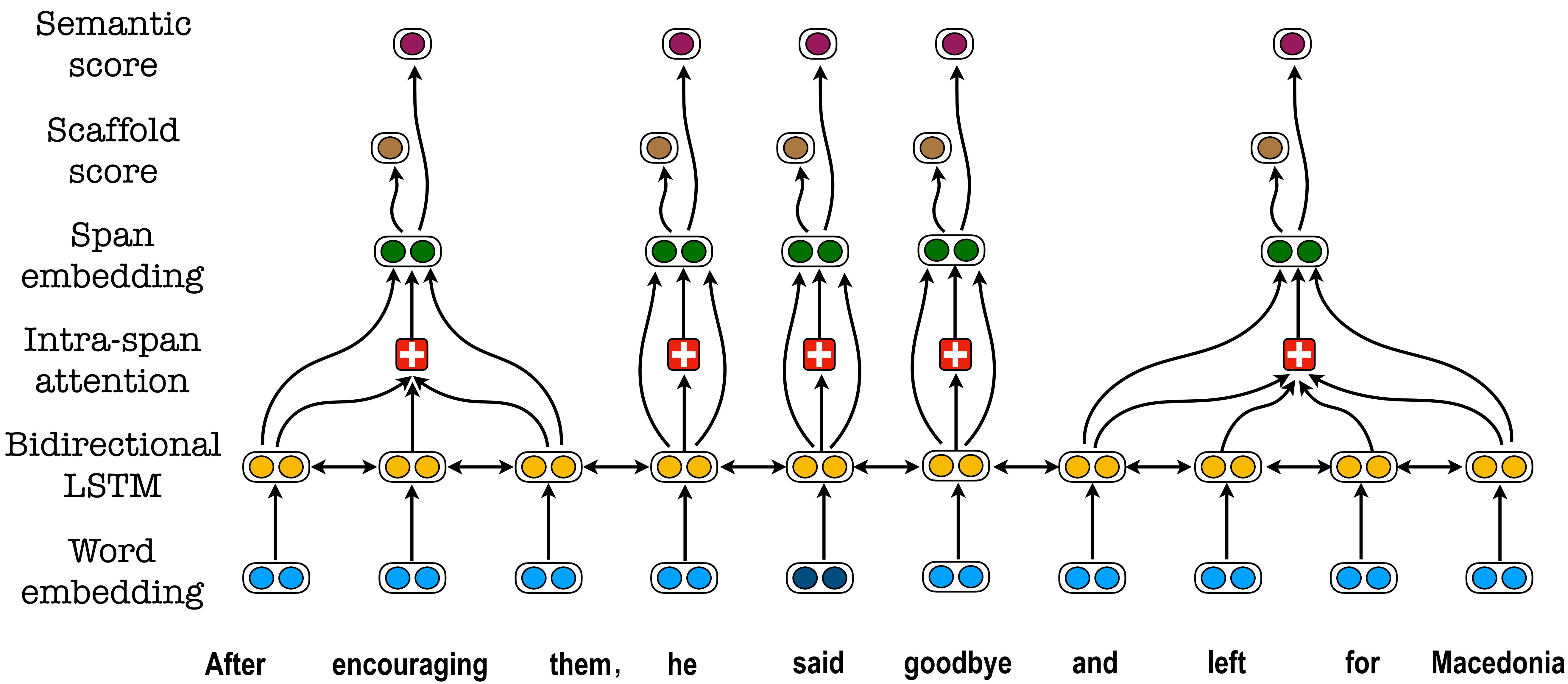
Model architecture



Model architecture



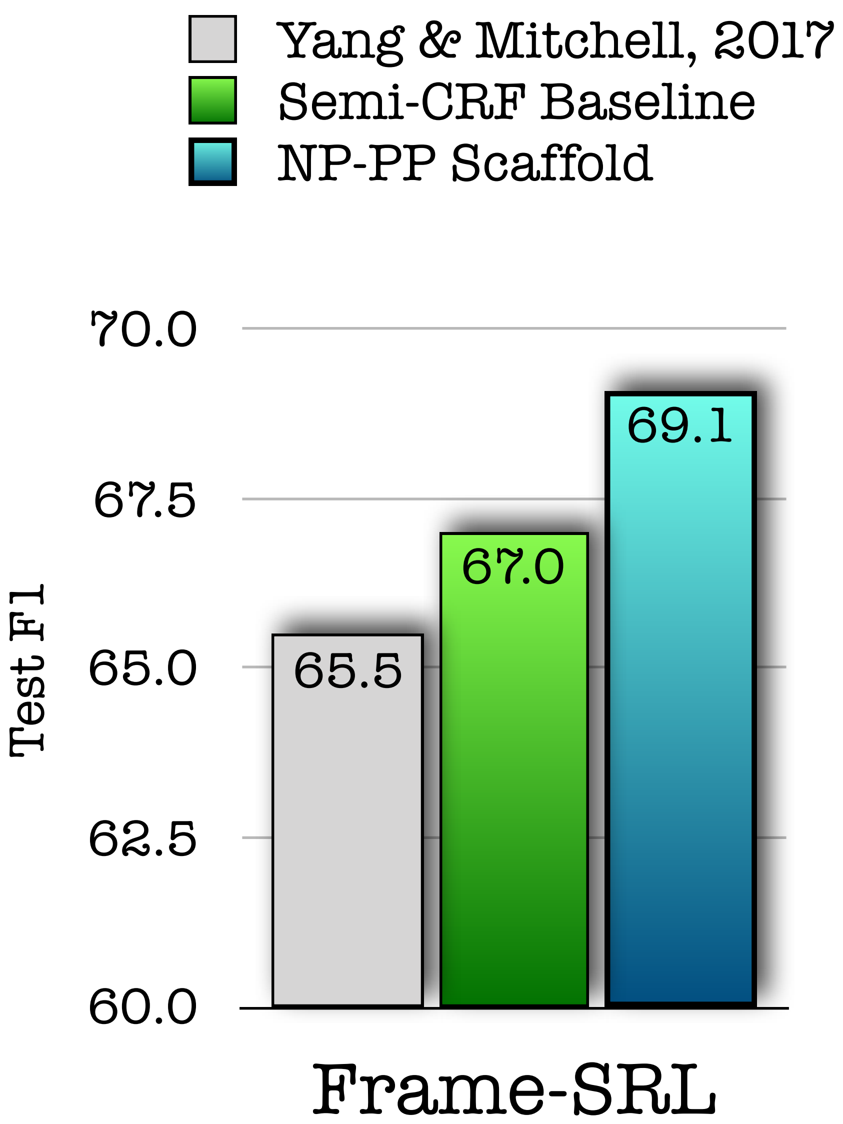
Model architecture



- Learn scaffold score when syntactic annotations available.

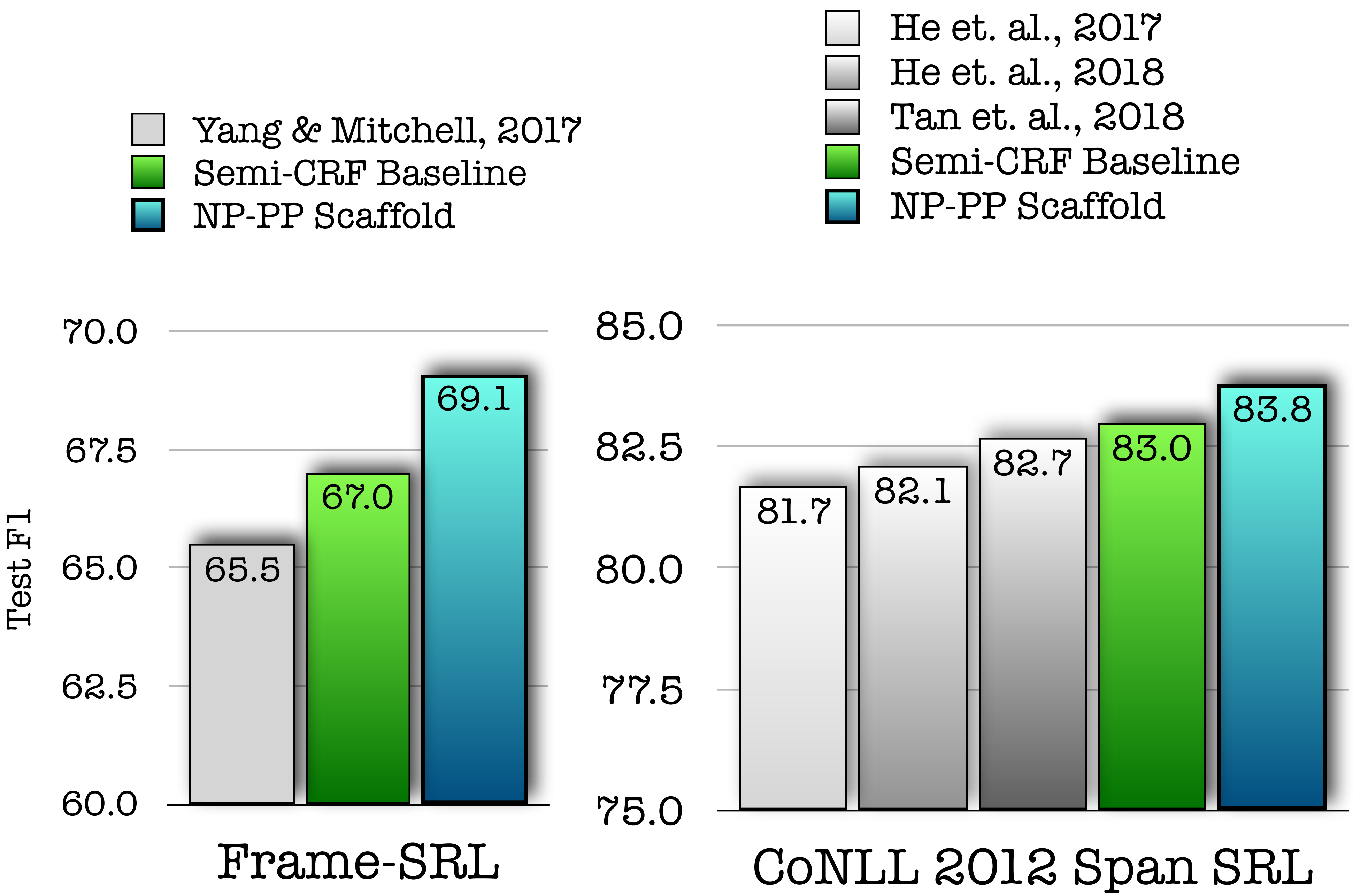
Results

Results



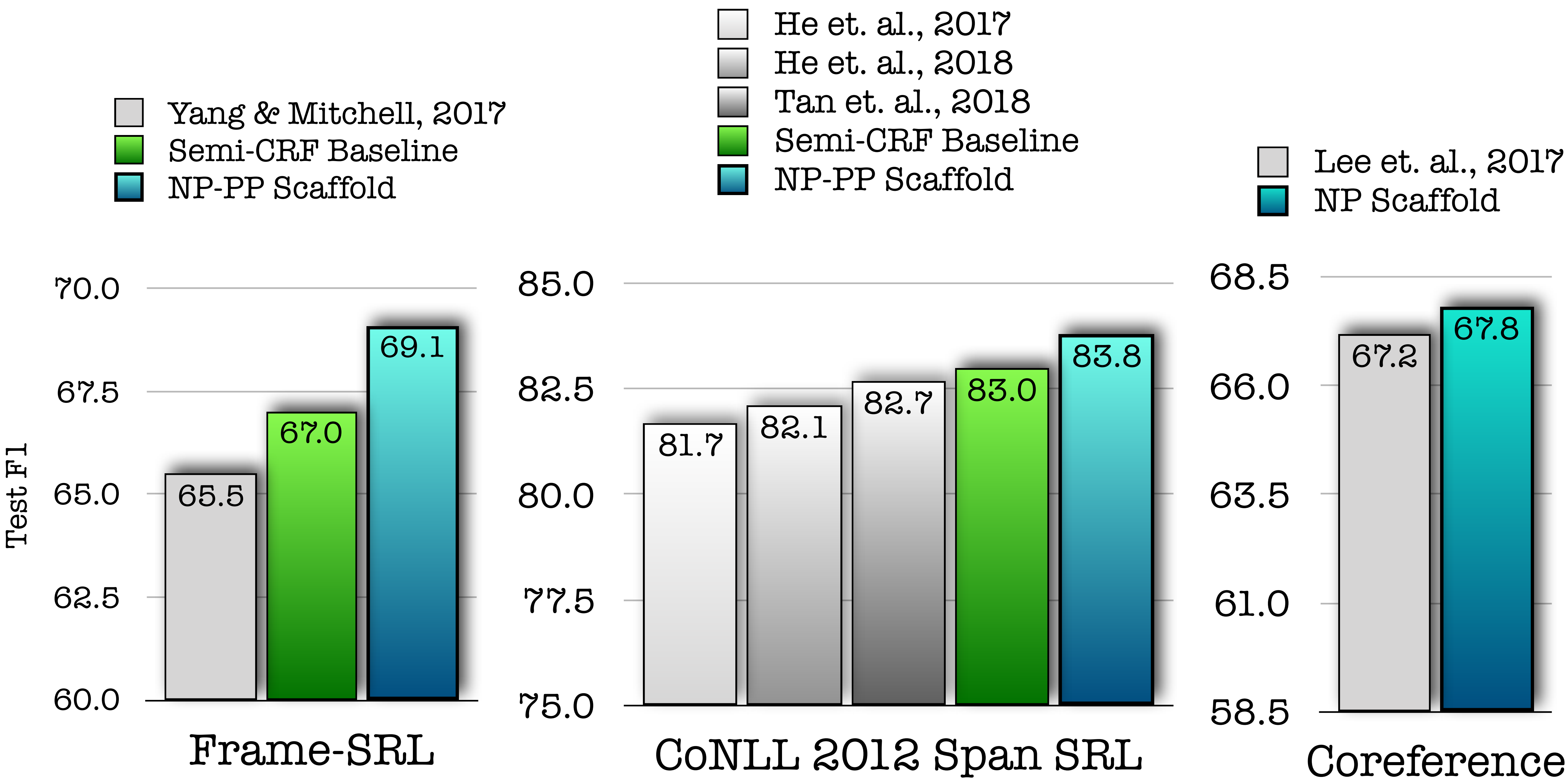
Yang & Mitchell, 2017
Semi-CRF Baseline
NP-PP Scaffold
70.0
67.5
65.0
62.5
60.0
Test F1
65.5
67.0
69.1
Frame-SRL

Results

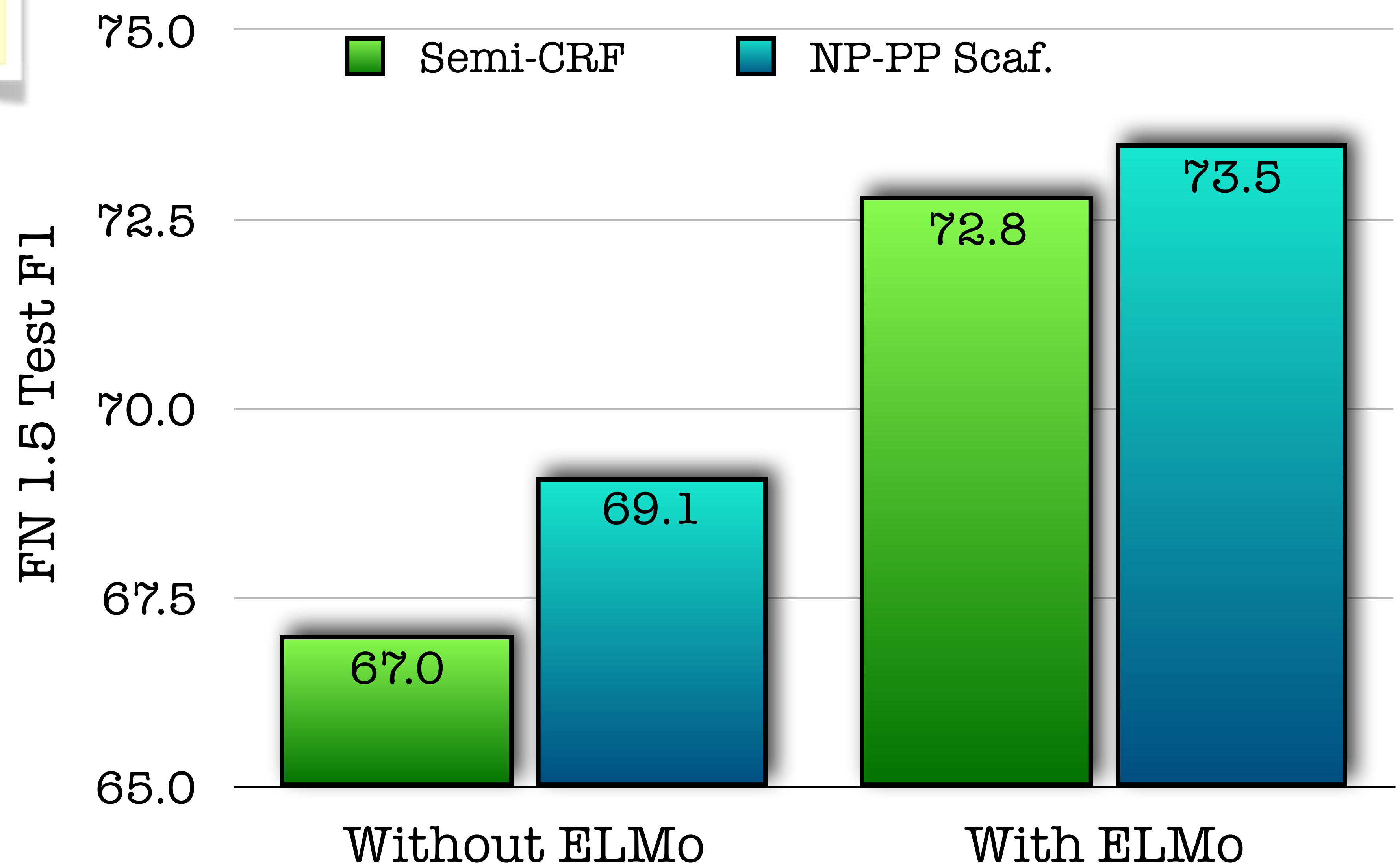


Yang & Mitchell, 2017
Semi-CRF Baseline
NP-PP Scaffold
He et. al., 2017
He et. al., 2018
Tan et. al., 2018
Semi-CRF Baseline
NP-PP Scaffold
Test F1
70.0
67.5
65.0
62.5
60.0
65.5
67.0
69.1
Frame-SRL
85.0
82.5
80.0
77.5
75.0
81.7
82.1
82.7
83.0
83.8
CoNLL 2012 Span SRL

Results


Yang & Mitchell, 2017
Semi-CRF Baseline
NP-PP Scaffold
He et. al., 2017
He et. al., 2018
Tan et. al., 2018
Semi-CRF Baseline
NP-PP Scaffold
Lee et. al., 2017
NP Scaffold
Test F1
70.0
67.5
65.0
62.5
60.0
65.5
67.0
69.1
Frame-SRL
85.0
82.5
80.0
77.5
75.0
81.7
82.1
82.7
83.0
83.8
CoNLL 2012 Span SRL
68.5
66.0
63.5
61.0
58.5
67.2
67.8
Coreference

Effect of Contextualized Representations



- Note: These results are not included in the paper.

Recap: Learning Challenge #1

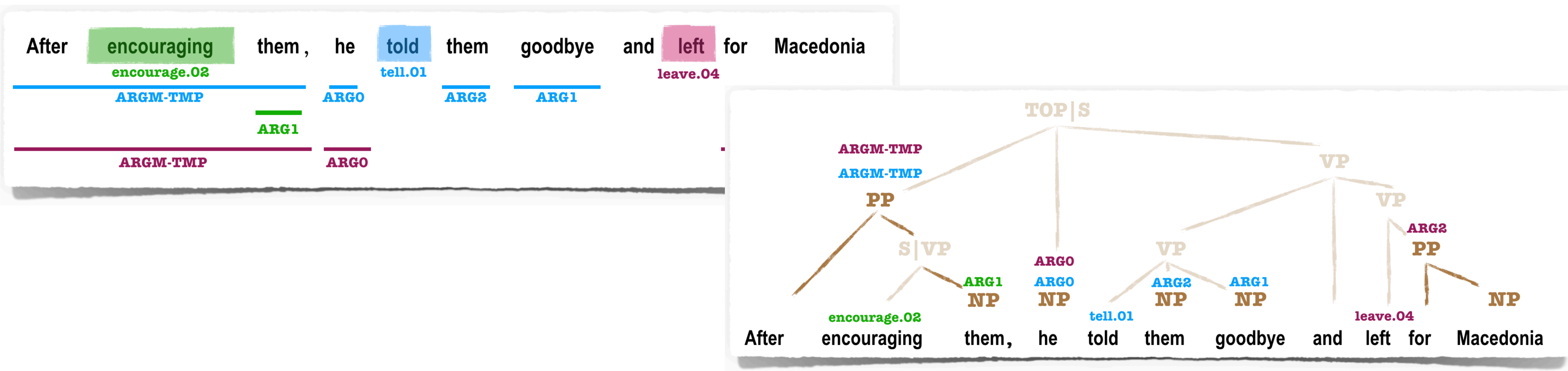
Can linguistic structure act as an informative prior for improving our models?

Recap: Learning Challenge #1

Can linguistic structure act as an informative prior for improving our models?

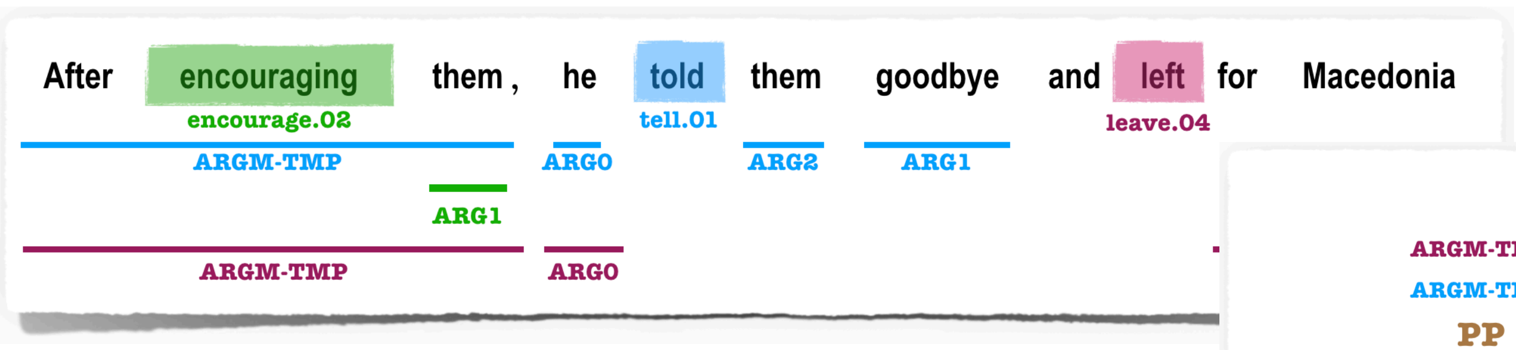
Recap: Learning Challenge #1

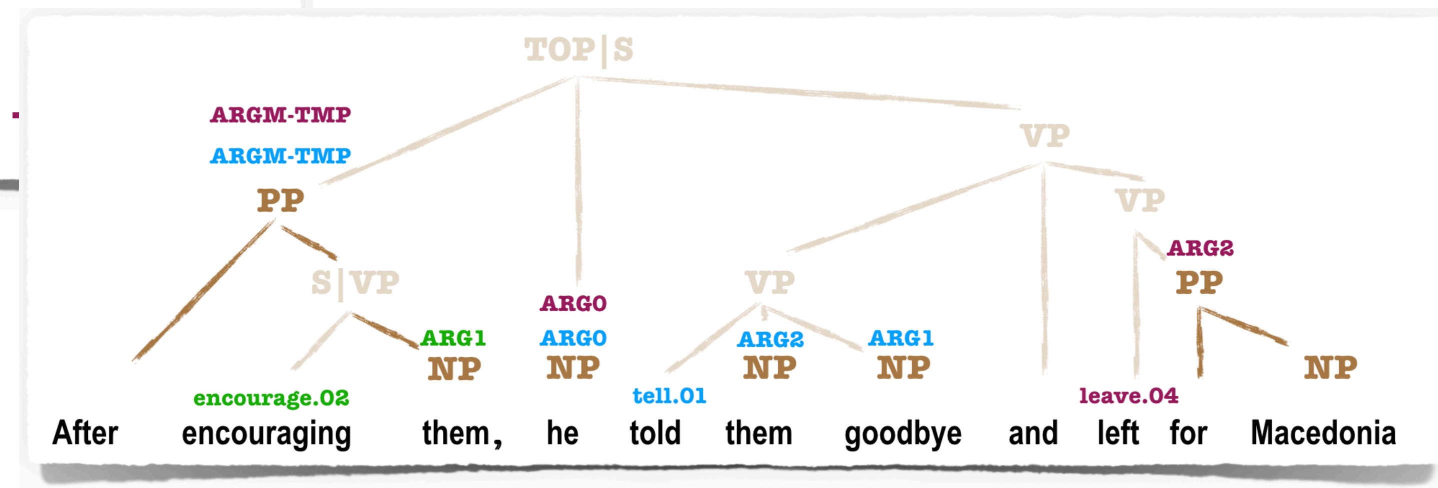
Can linguistic structure act as an informative prior for improving our models?

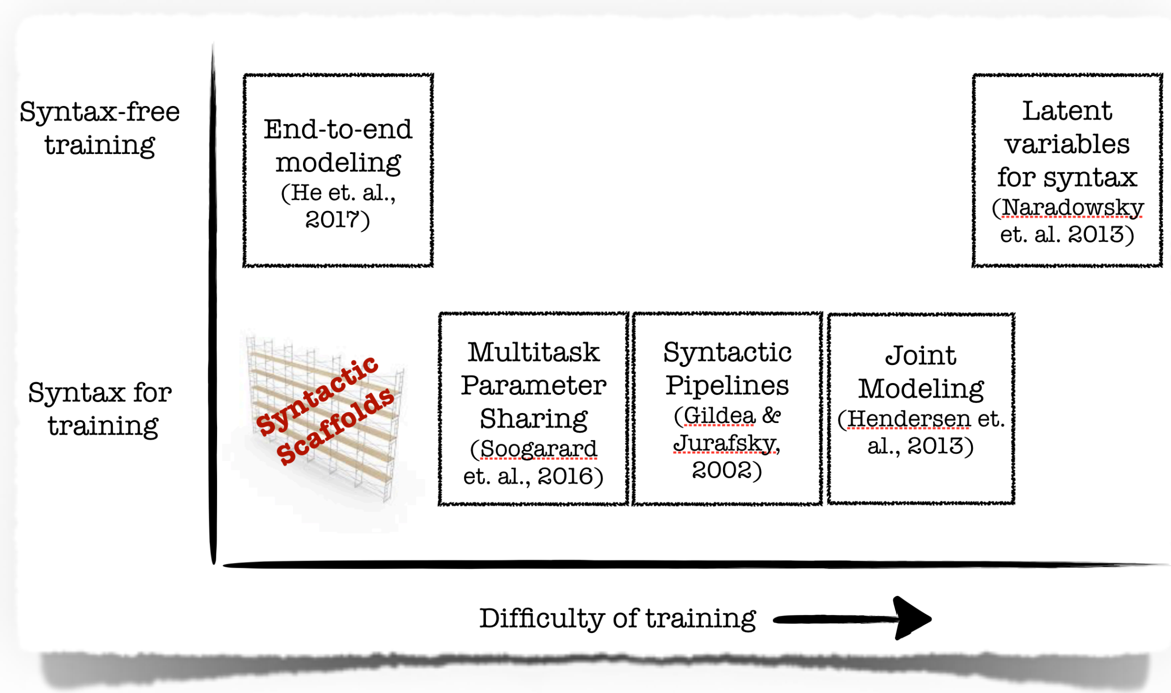


Recap: Learning Challenge #1

Can linguistic structure act as an informative prior for improving our models?

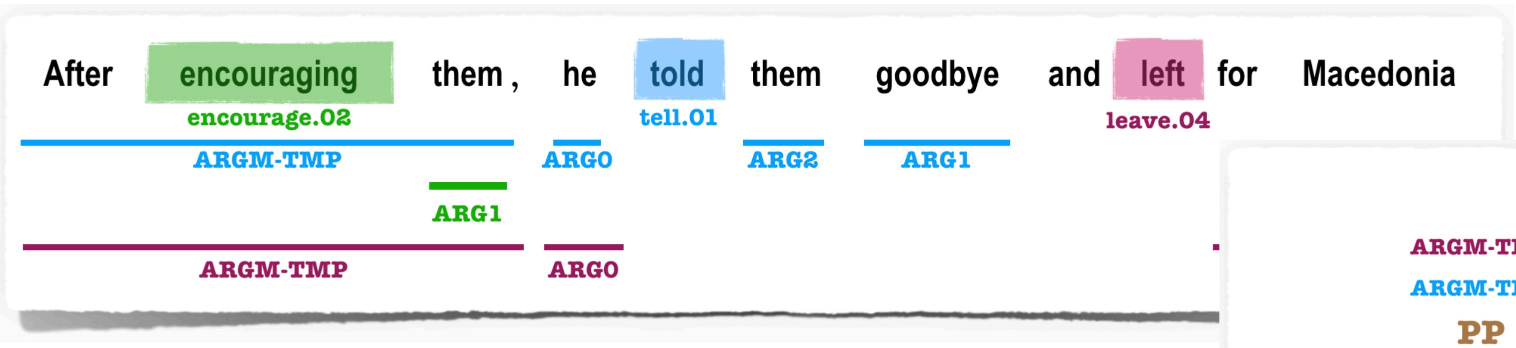


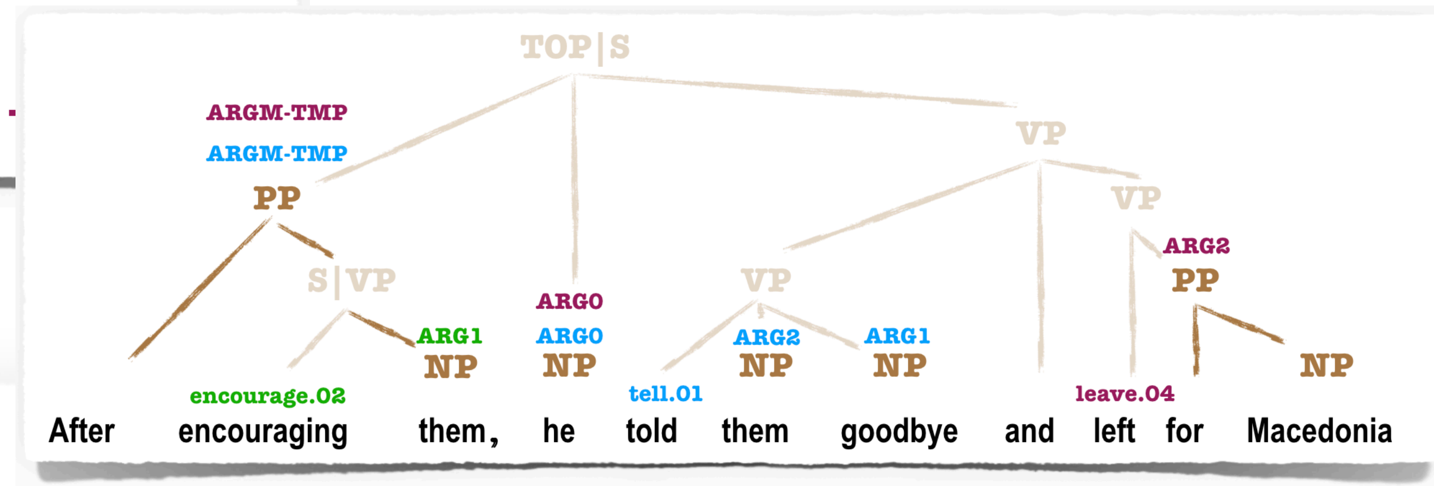


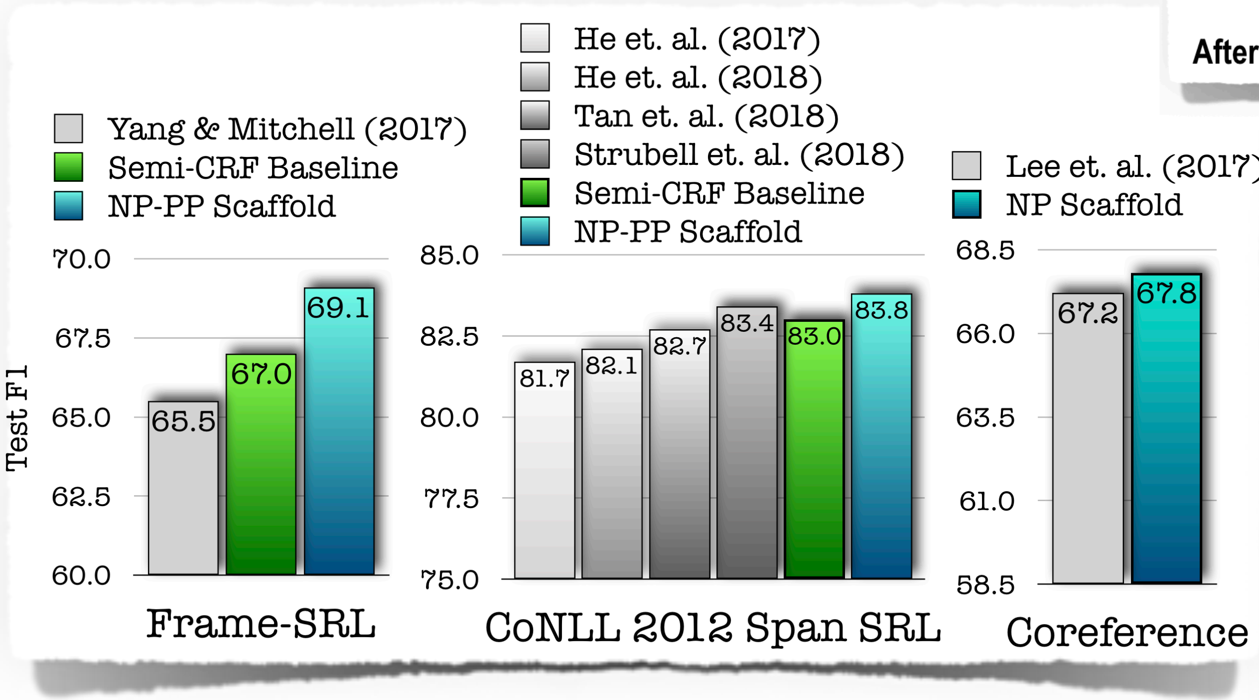


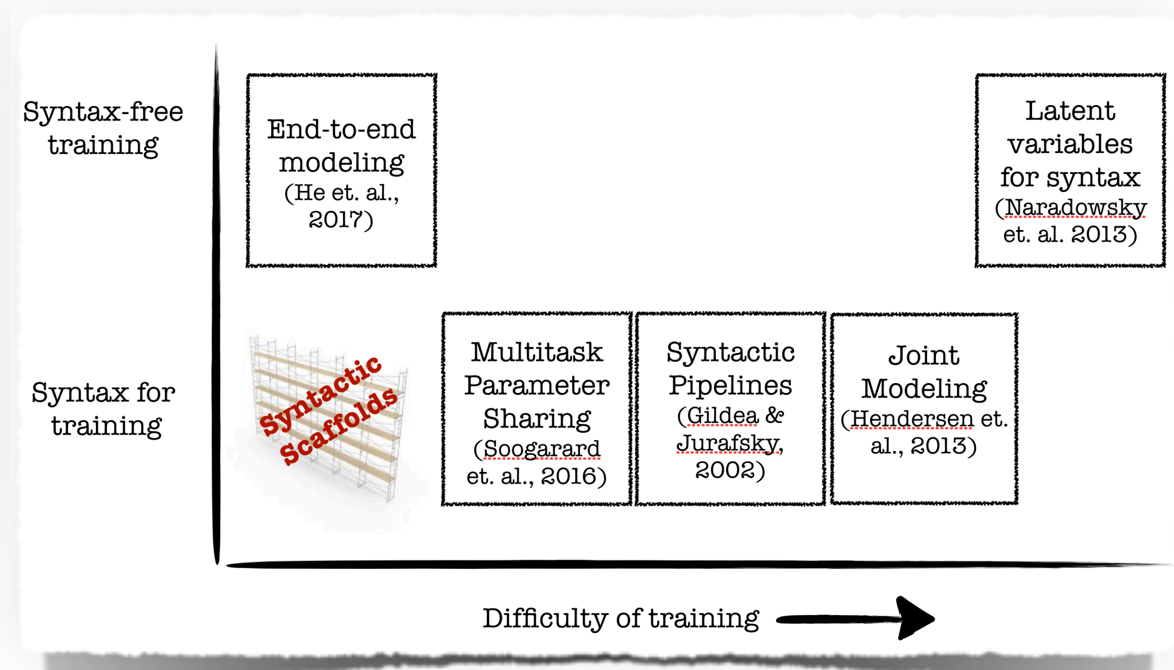
Recap: Learning Challenge #1

Can linguistic structure act as an informative prior for improving our models?

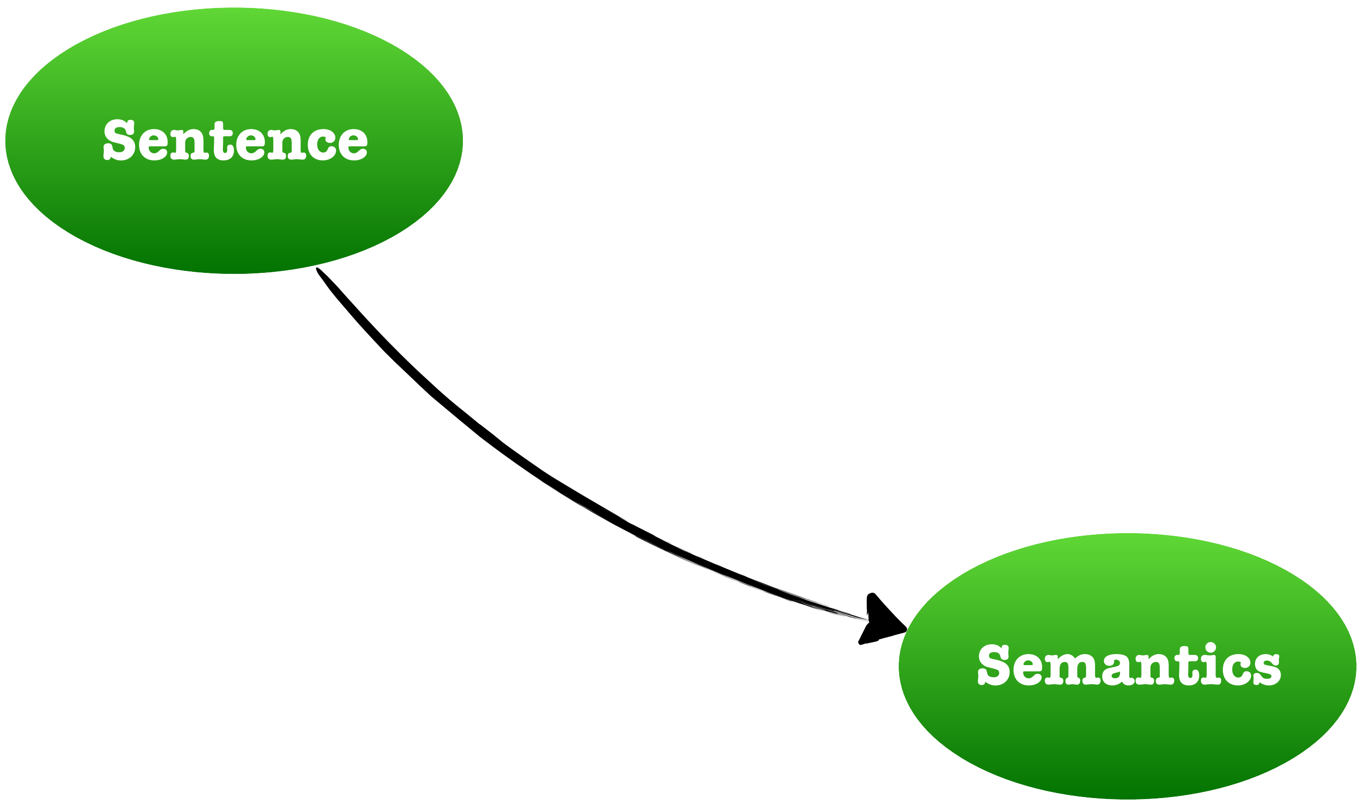




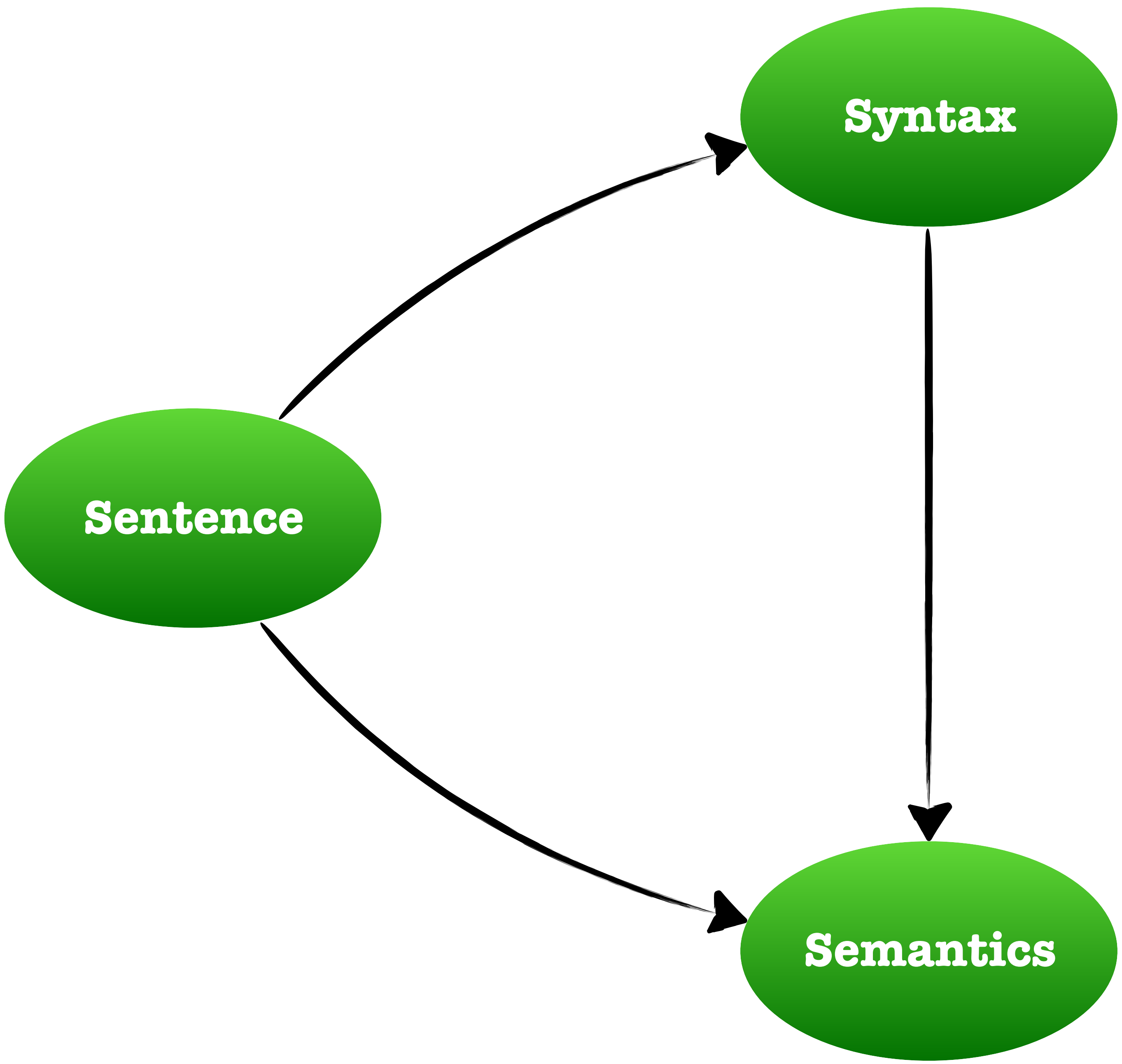




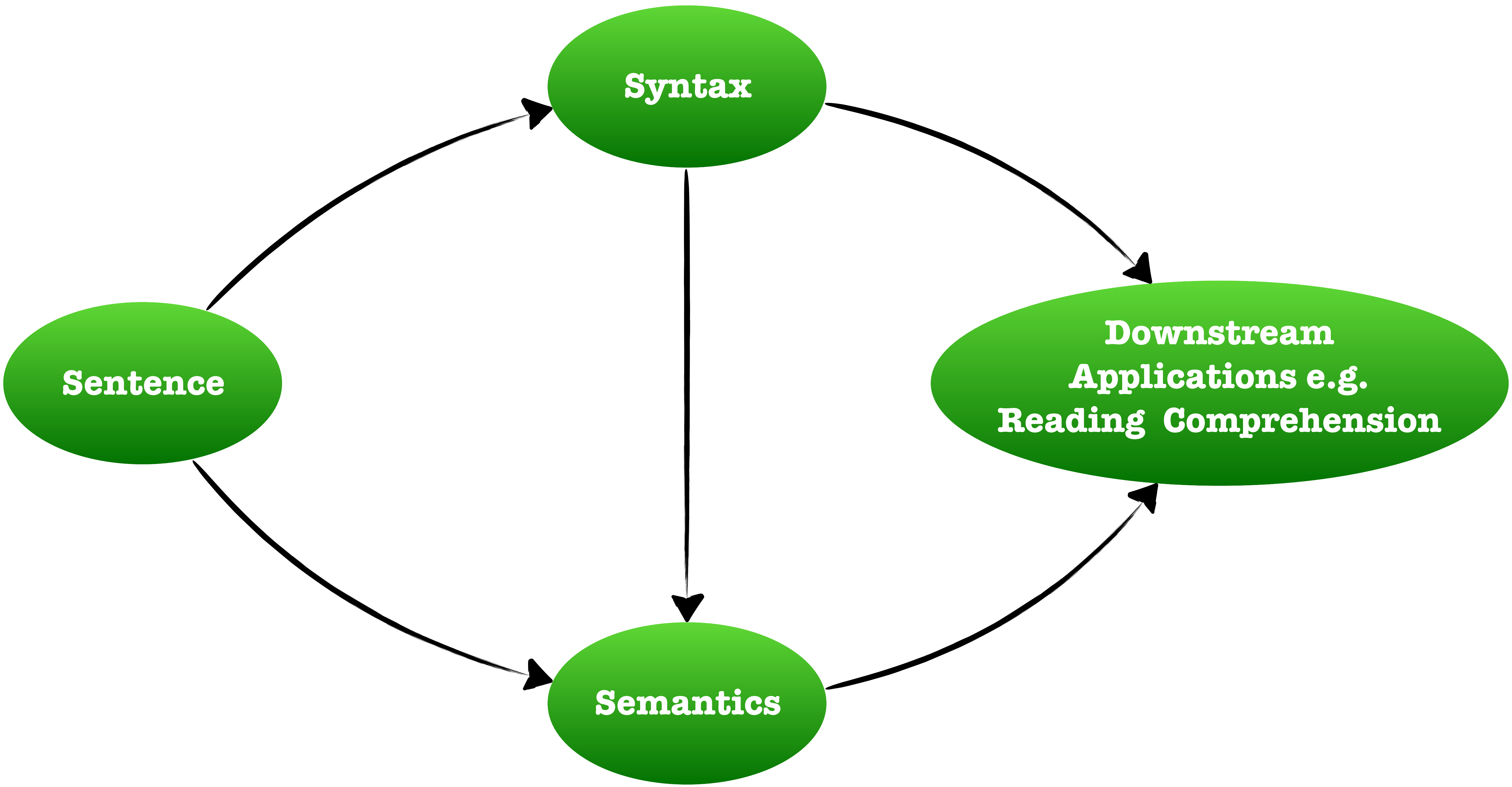
Looking ahead: Predicted Structure



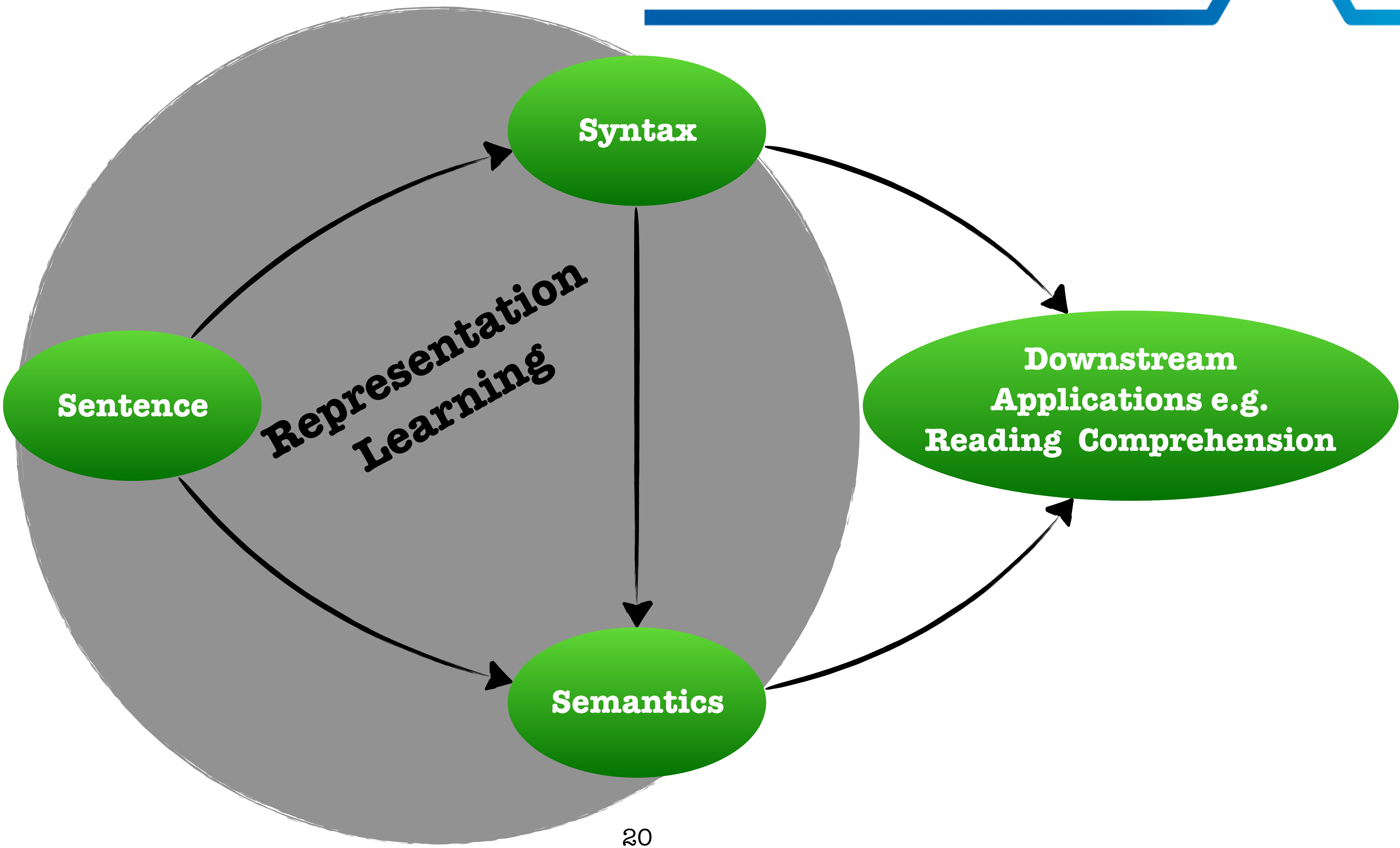
Looking ahead: Predicted Structure



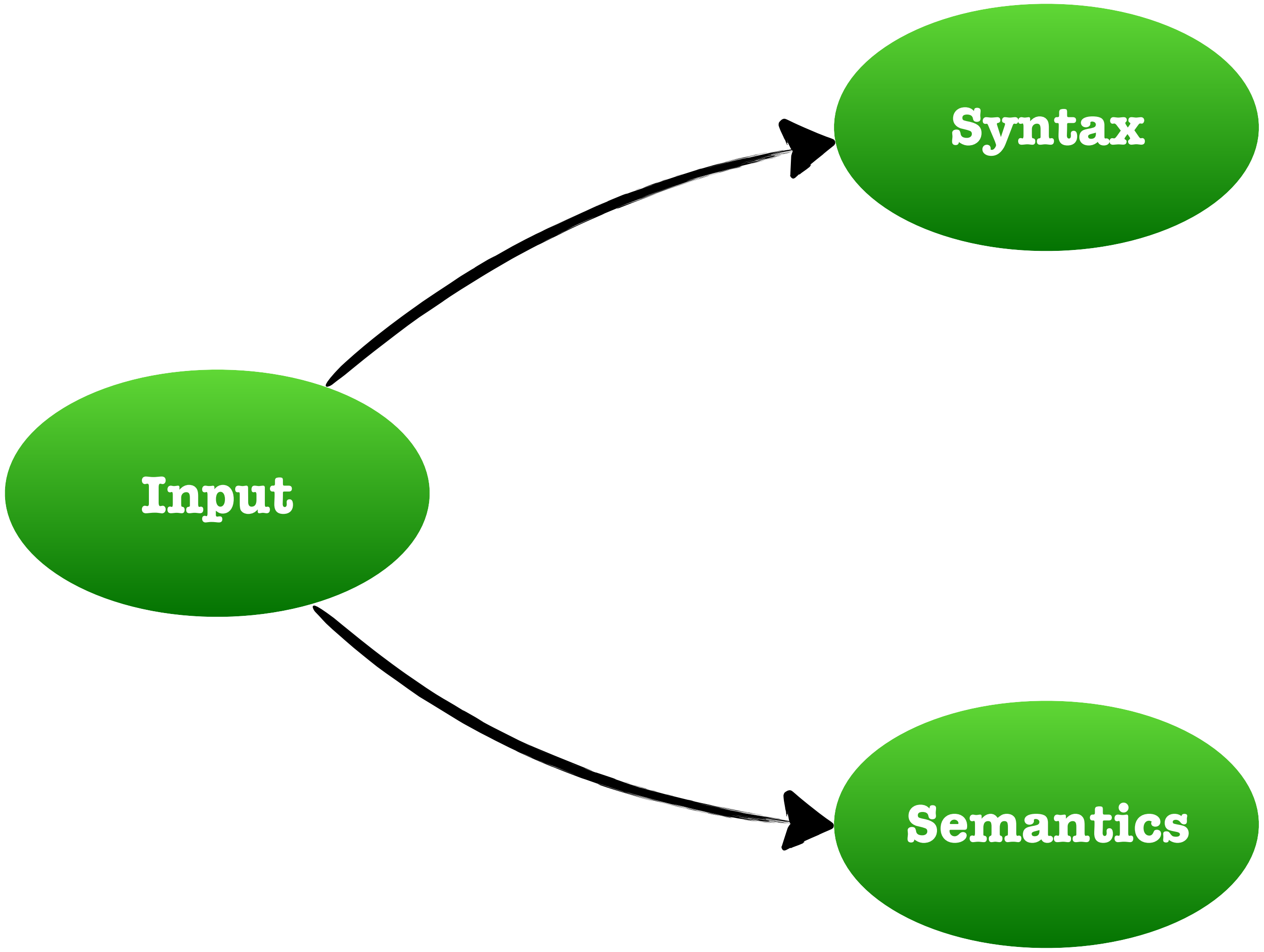
Looking ahead: Predicted Structure



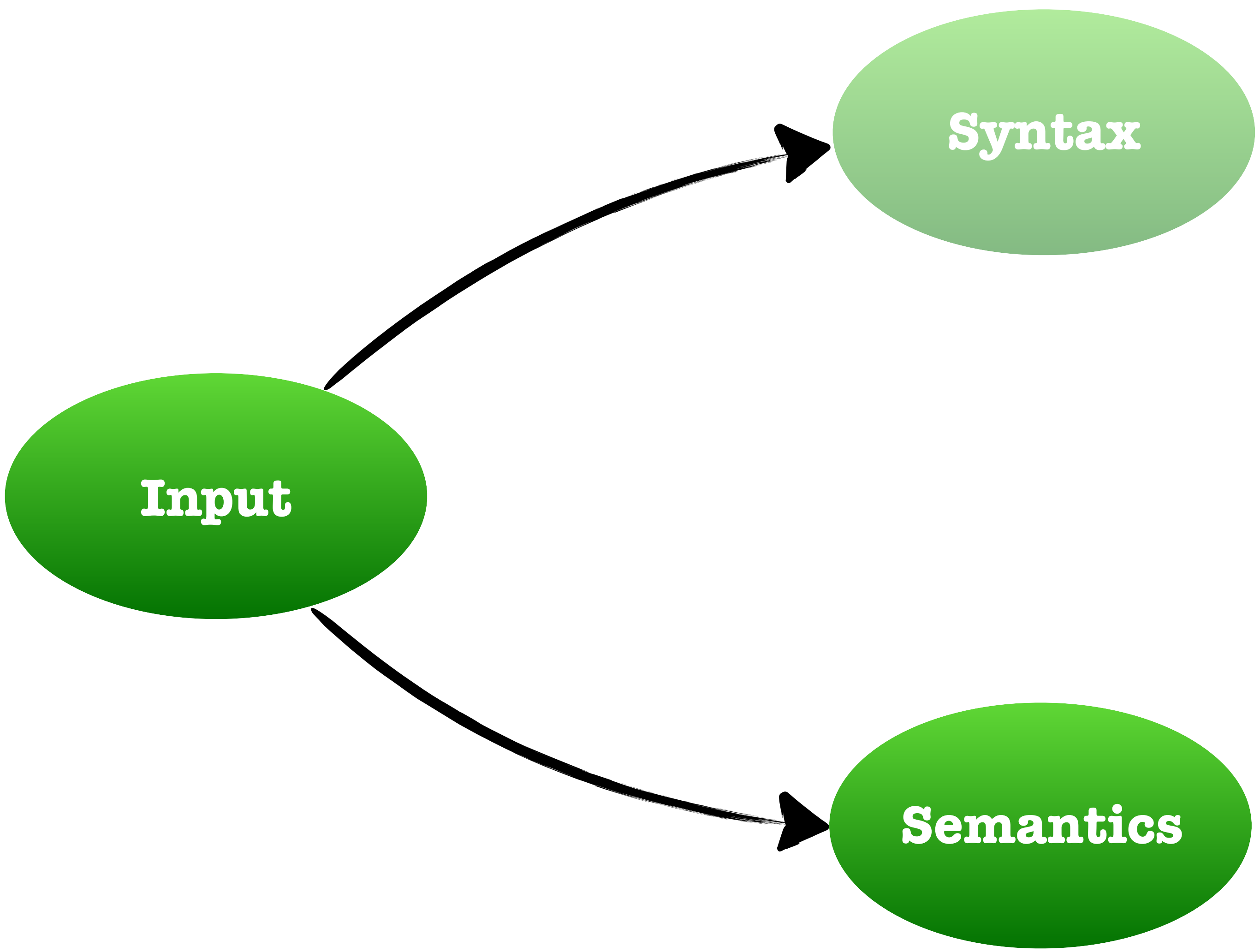
Looking ahead: Predicted Structure



Looking ahead: Structured Transformation

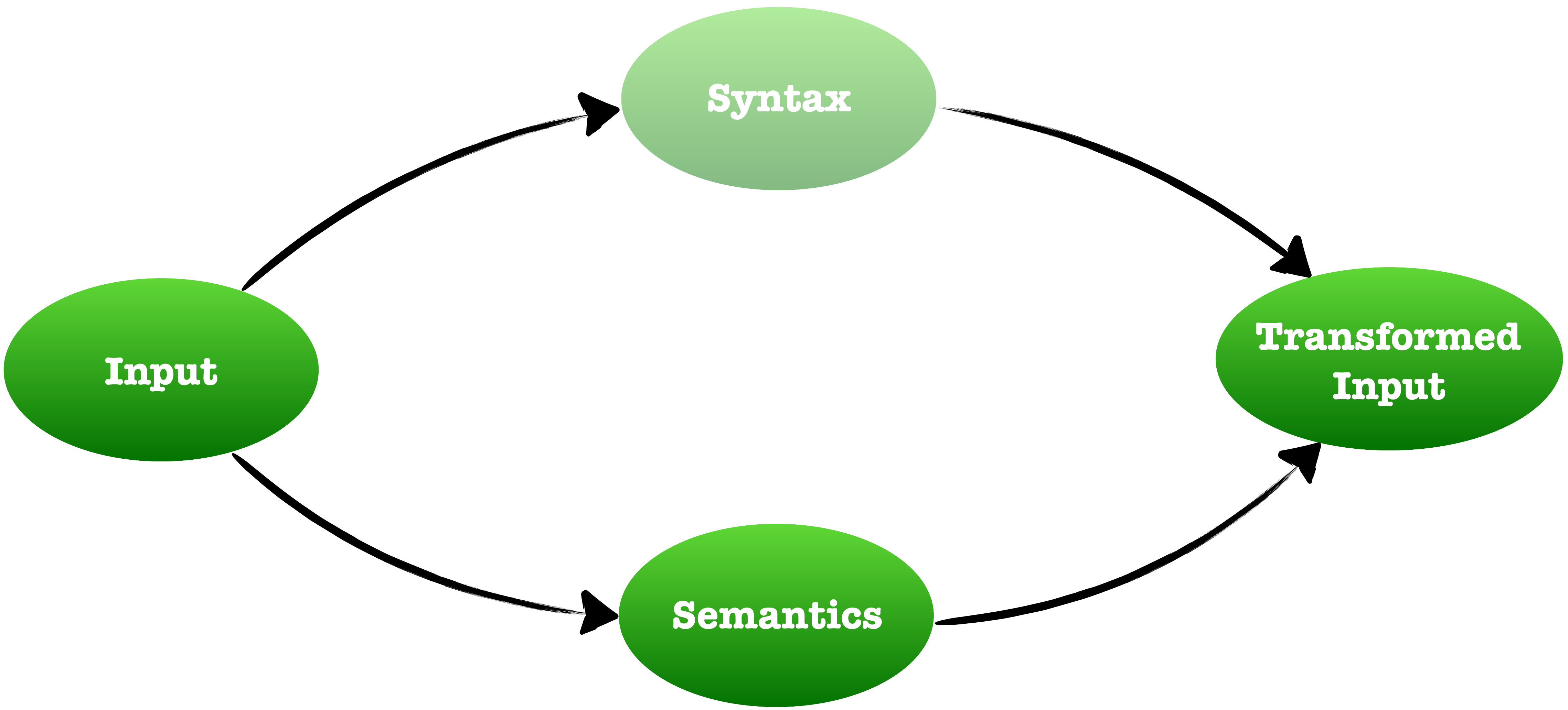


Looking ahead: Structured Transformation



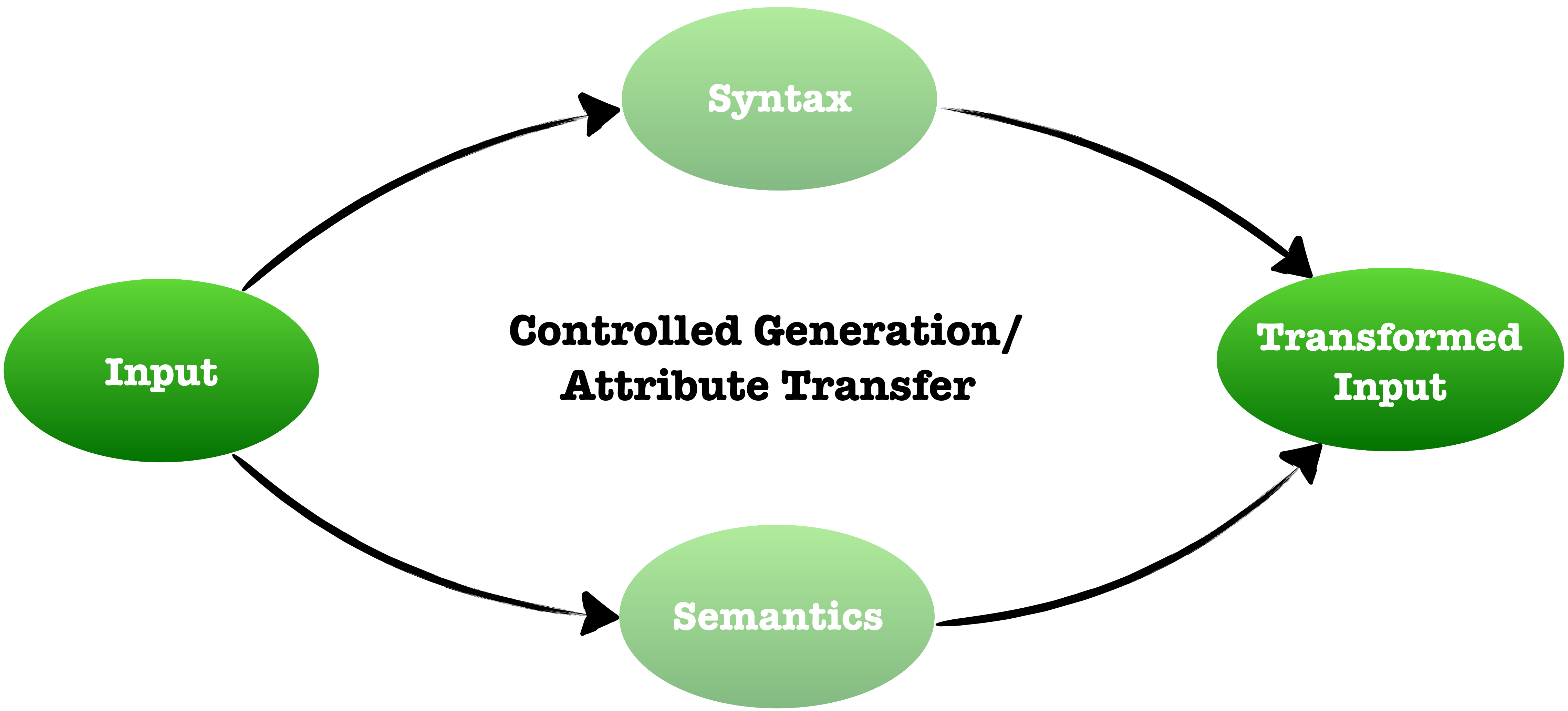
Iyyer et. al. [NAACL 2018]

Looking ahead: Structured Transformation



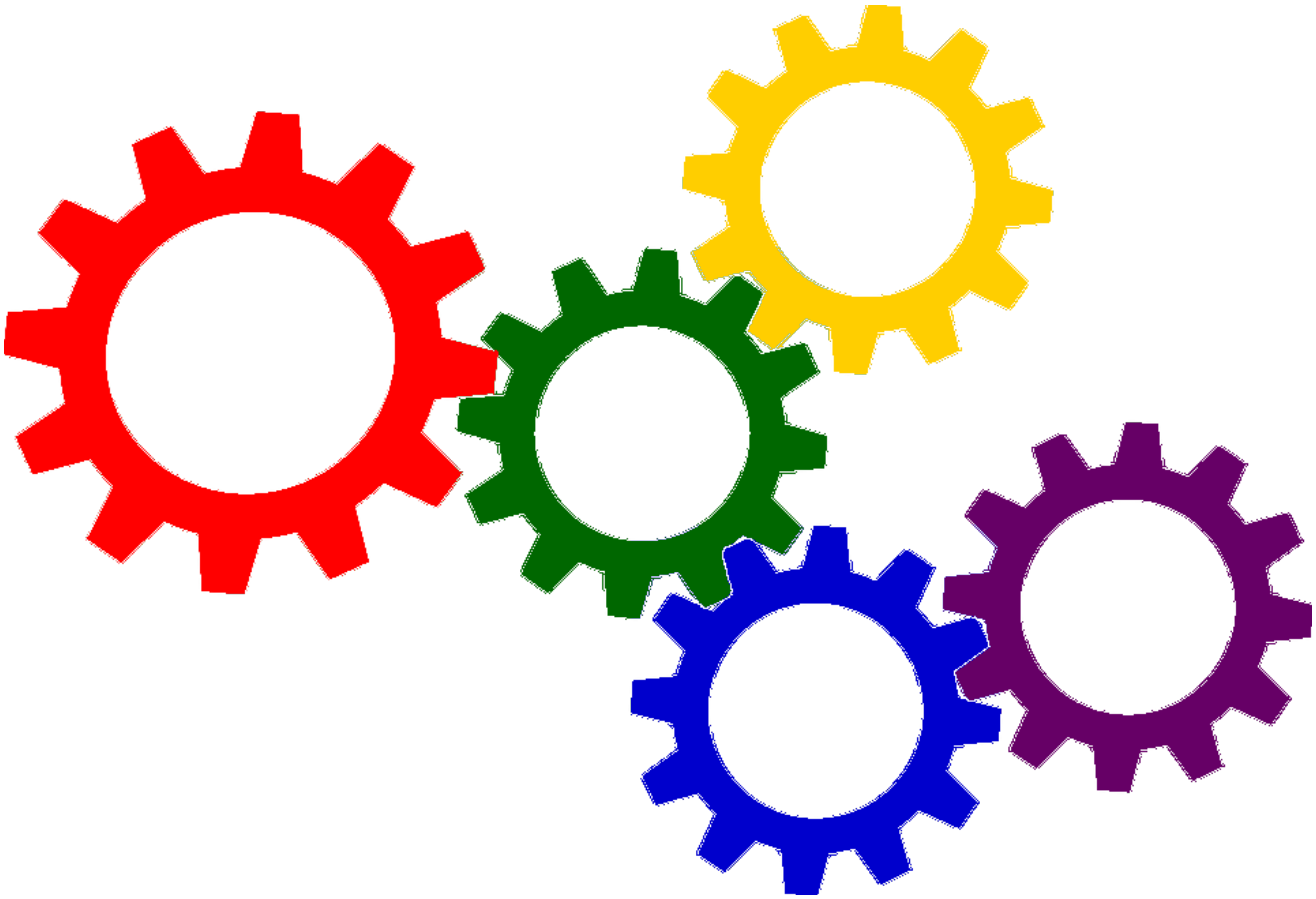
Iyyer et. al. [NAACL 2018]

Looking ahead: Structured Transformation



Iyyer et. al. [NAACL 2018]

Part II



Recap: Confusion of the Muppets

On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of old Acadian colonists declined after the year 16~~7~~5.

The number of new Huguenot colonists declined after what year?

[Jia & Liang, 2017]

Learning Challenges

Part I

Can linguistic structure act as an informative prior for improving our models?

- [x] Syntactic Scaffolds for Semantic Structures (EMNLP 2018)

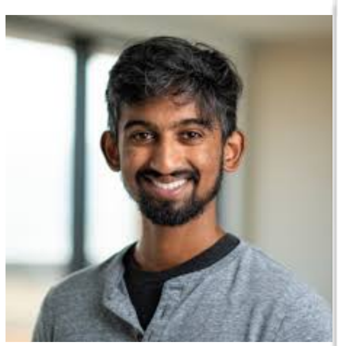
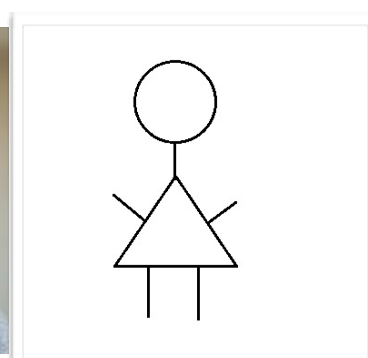
Part II

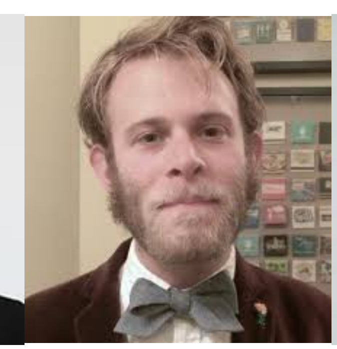
What in our data is causing models to achieve high performance?

- [ ] Annotation Artifacts in Natural Language Inference Data (NAACL 2018)

Annotation Artifacts in Natural Language Inference Data

NAACL 2018


Suchin Gururangan*

S.*

Omer Levy

Roy Schwartz

Sam Bowman

Noah A. Smith

*equal contribution

Natural Language Inference (NLI)

Given a premise, is a hypothesis true, false or neither?

Natural Language Inference (NLI)

Given a premise, is a hypothesis true, false or neither?

Premise Two dogs are running through a field.

Hypothesis The pets are sitting on a couch.

- True → **Entailment**
- False → **Contradiction**
- Cannot Say → **Neutral**

Natural Language Inference (NLI)

Given a premise, is a hypothesis true, false or neither?

Premise Two dogs are running through a field.

Hypothesis The pets are sitting on a couch.

- True → **Entailment**
- False → **Contradiction**
- Cannot Say → **Neutral**

NLI Datasets

Stanford NLI [Bowman et. al, 2015]   570 K
Multi-genre NLI [Williams et. al., 2017] 433 K

NLI Datasets

Two dogs are running through a field.

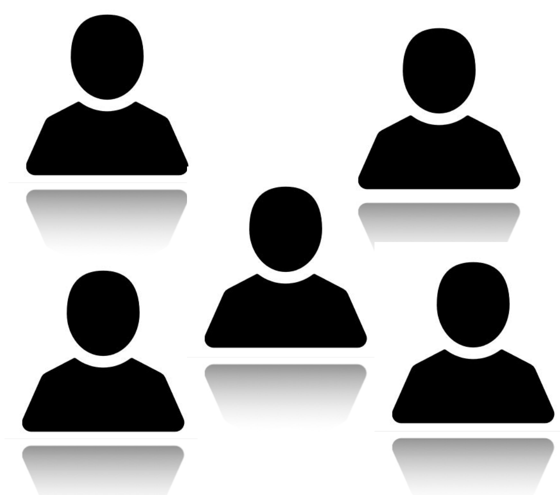
Premise

Stanford NLI [Bowman et. al, 2015] 570 K

Multi-genre NLI [Williams et. al., 2017] 433 K

NLI Datasets

Two dogs are running through a field.



Premise

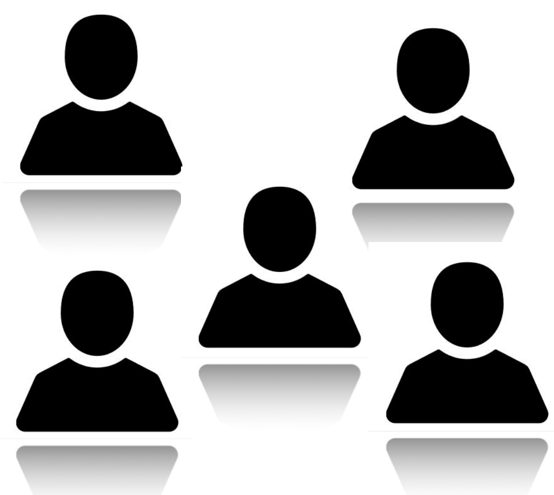
Stanford NLI [Bowman et. al, 2015] 570 K

Multi-genre NLI [Williams et. al., 2017] 433 K

NLI Datasets

Two dogs are running through a field.

Premise



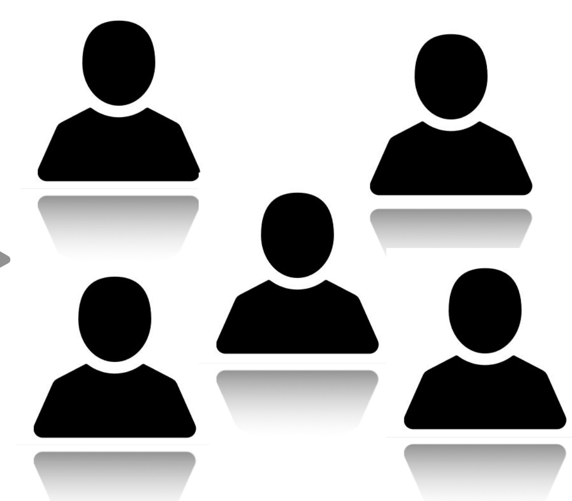
Stanford NLI [Bowman et. al, 2015] 570 K

Multi-genre NLI [Williams et. al., 2017] 433 K

NLI Datasets

Two dogs are running through a field.

Premise



Entailment

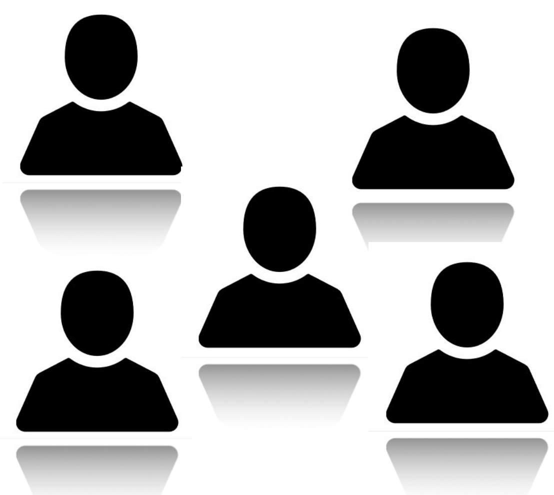
There are animals outdoors.

Stanford NLI [Bowman et. al, 2015] 570 K
Multi-genre NLI [Williams et. al., 2017] 433 K

NLI Datasets

Two dogs are running through a field.

Premise



Entailment

There are animals outdoors.

Neutral

Some puppies are running to catch a stick.

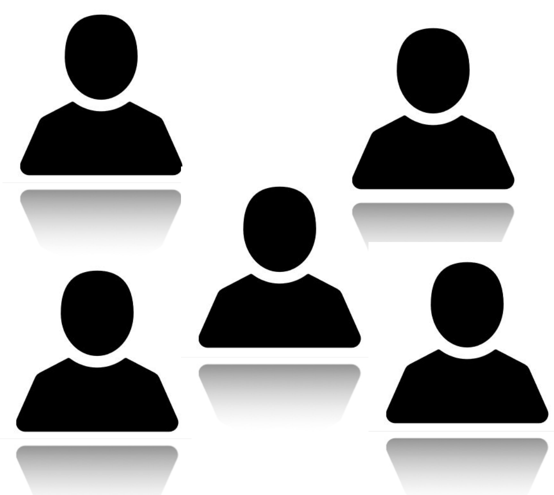
Stanford NLI [Bowman et. al, 2015] 570 K

Multi-genre NLI [Williams et. al., 2017] 433 K

NLI Datasets

Two dogs are running through a field.

Premise



Entailment

There are animals outdoors.

Neutral

Some puppies are running to catch a stick.

Contradiction

The pets are sitting on a couch.

Stanford NLI [Bowman et. al, 2015] 570 K

Multi-genre NLI [Williams et. al., 2017] 433 K

Lots of progress

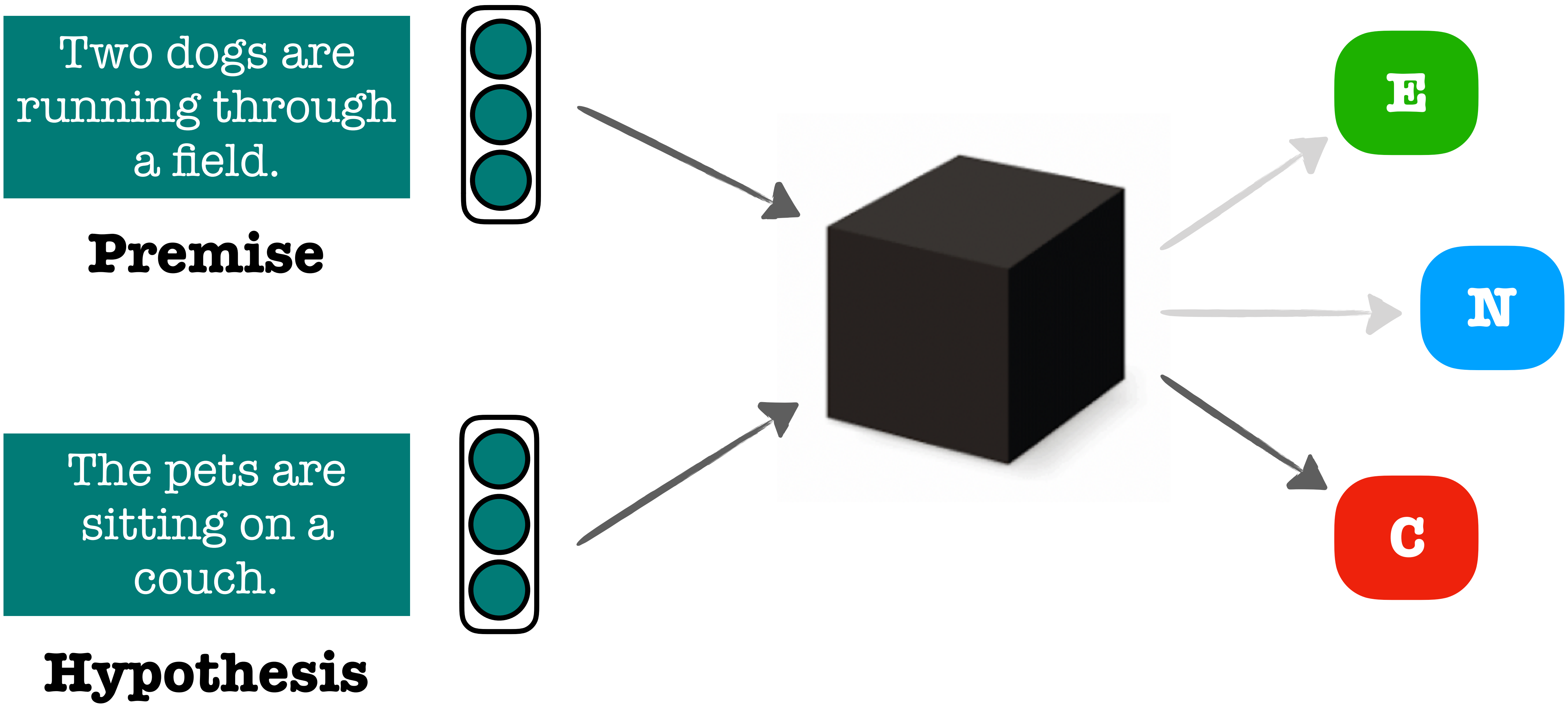
#	Team Name	Kernel	Team Members	Score	Entries	Last
1	**Allen Lao**			0.86443	4	3mo
2	**Anonymous**			0.86351	2	4mo
3	**sherry77**			0.85034	2	12d
4	**Ariel**			0.84953	10	13d
5	**ysffirst**			0.84718	6	13d
6	**ArielY**			0.84687	4	12d
7	**mattpeters**			0.84595	7	3mo
⋮						
	Bidirectional LSTM			0.67507		
104	**gabrielalmeida**			0.67313	5	8mo
105	**Zippy**			0.67160	2	1y
106	**kudkudak**			0.66435	2	1y
107	**Shawn Tan**			0.65271	1	6d
	CBOW			0.65200		

Lots of progress

#	Team Name	Kernel	Team Members	Score	Entries	Last
1	**Allen Lao**			0.86443	4	3mo
2	**Anonymous**			0.86351	2	4mo
3	**sherry77**			0.85034	2	12d
4	**Ariel**			0.84953	10	13d
5	**ysffirst**			0.84718	6	13d
6	**ArielY**			0.84687	4	12d
7	**mattpeters**			0.84595	7	3mo
⋮						
	Bidirectional LSTM			0.67507		
104	**gabrielalmeida**			0.67313	5	8mo
105	**Zippy**			0.67160	2	1y
106	**kudkudak**			0.66435	2	1y
107	**Shawn Tan**			0.65271	1	6d
	CBOW			0.65200		

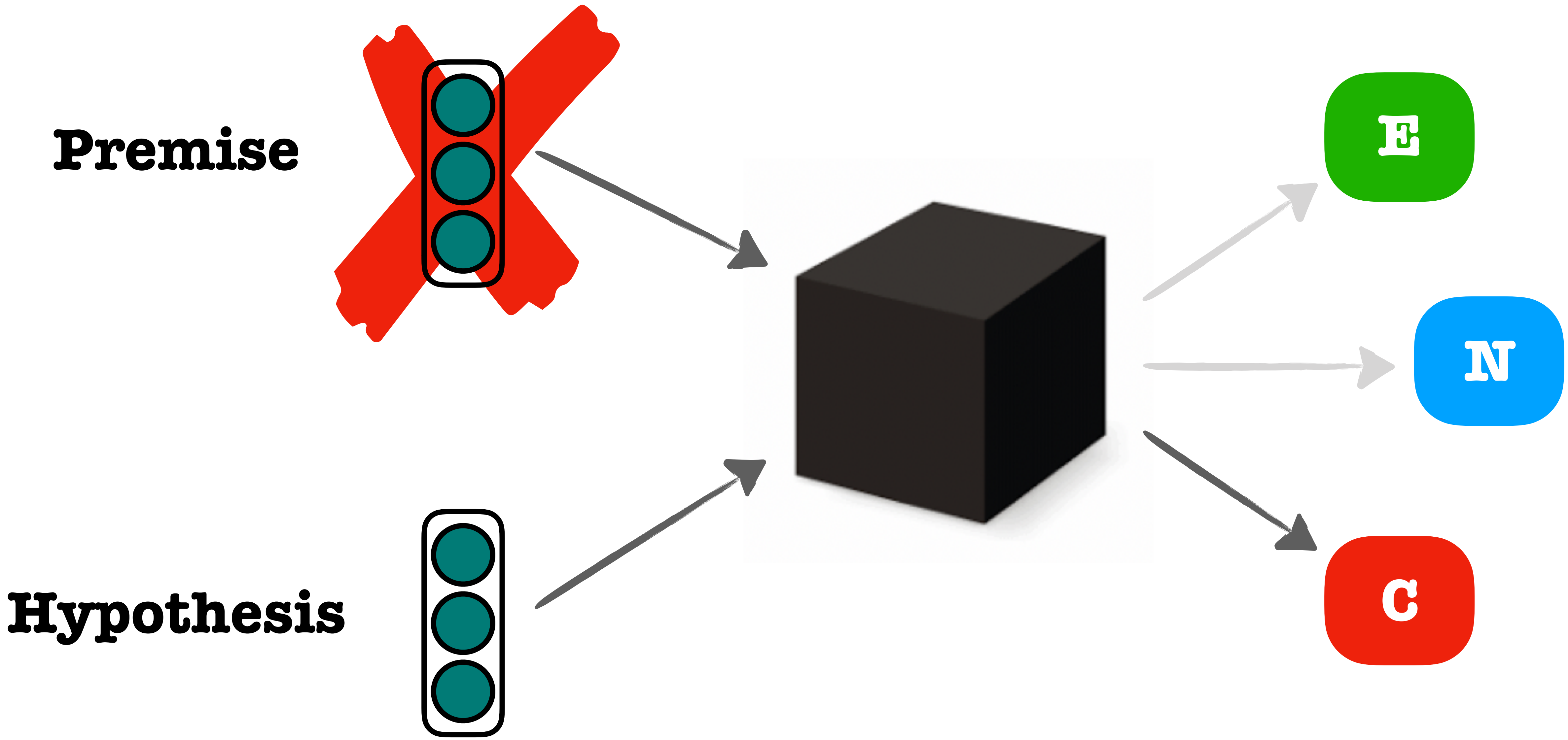
MNLI Leaderboard

NLI as Text Classification



A simple experiment

A simple experiment



fastText [Joulin et. al. 2017]

A simple experiment

Given **no** premise, is a hypothesis true, false or neither?

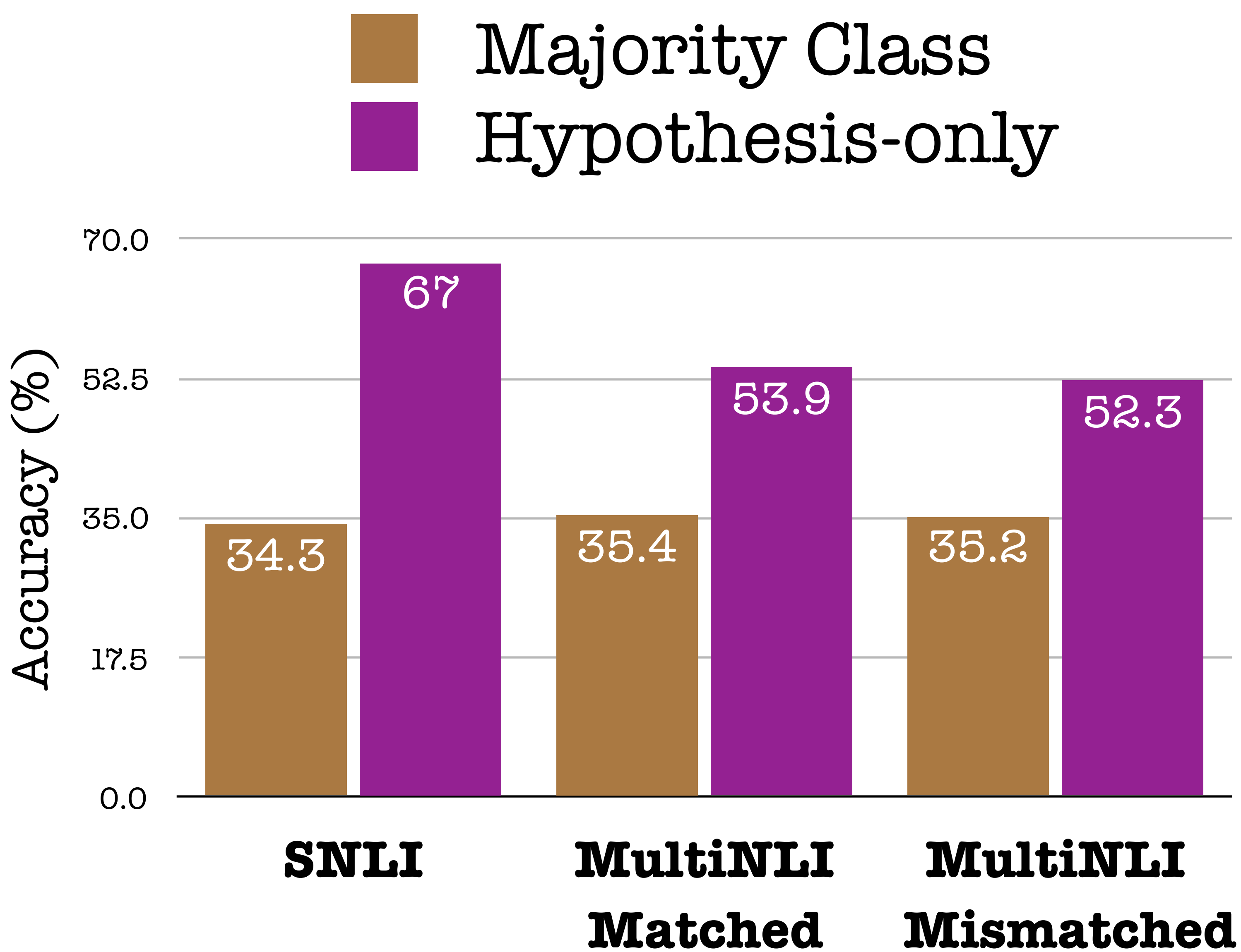
A simple experiment

Given **no** premise, is a hypothesis true, false or neither?

Hypothesis	The little boy is diving off the diving board because he is an excellent swimmer.

- True → **Entailment**
- False → **Contradiction**
- Cannot Say → **Neutral**

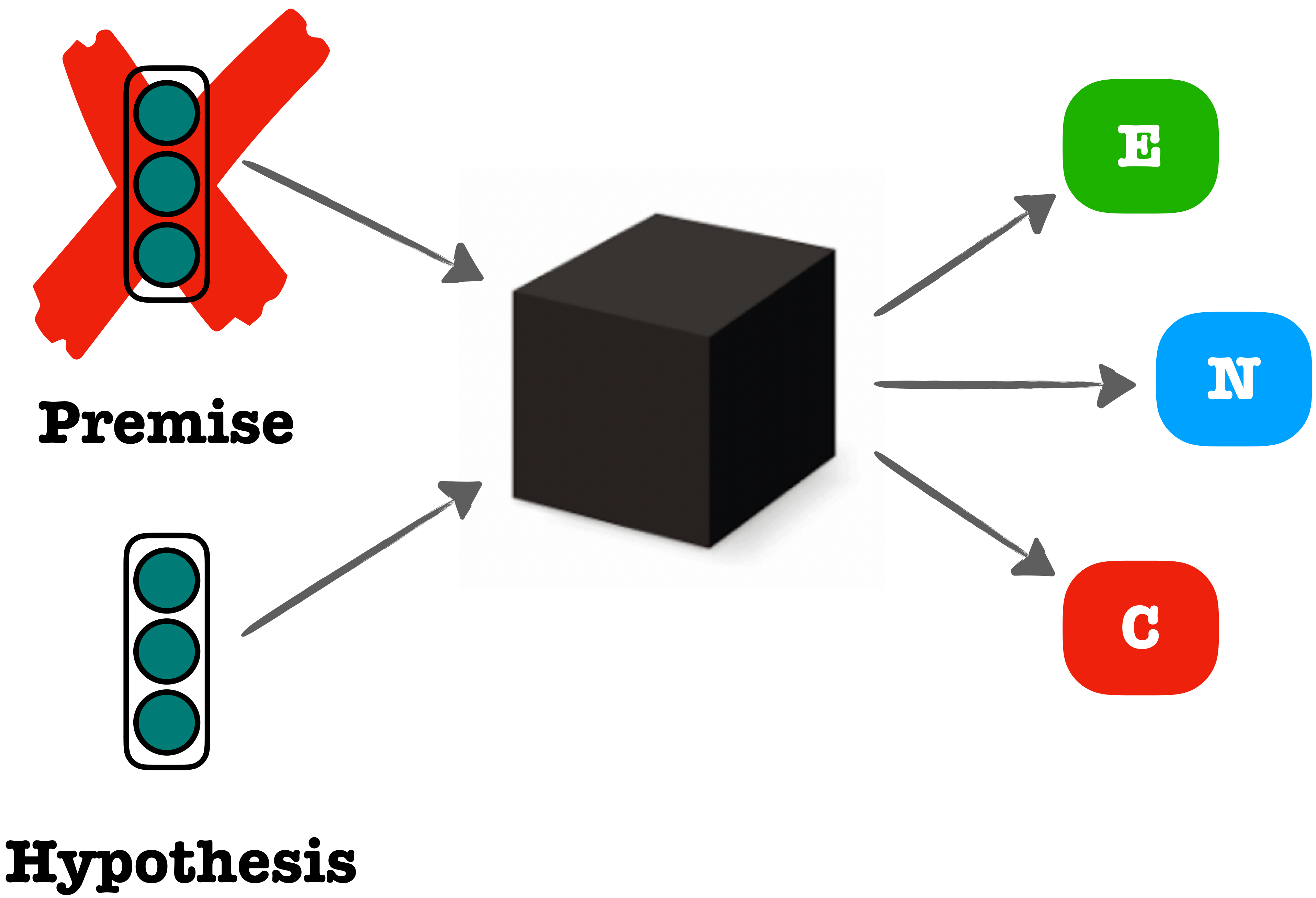
Surprising Results!



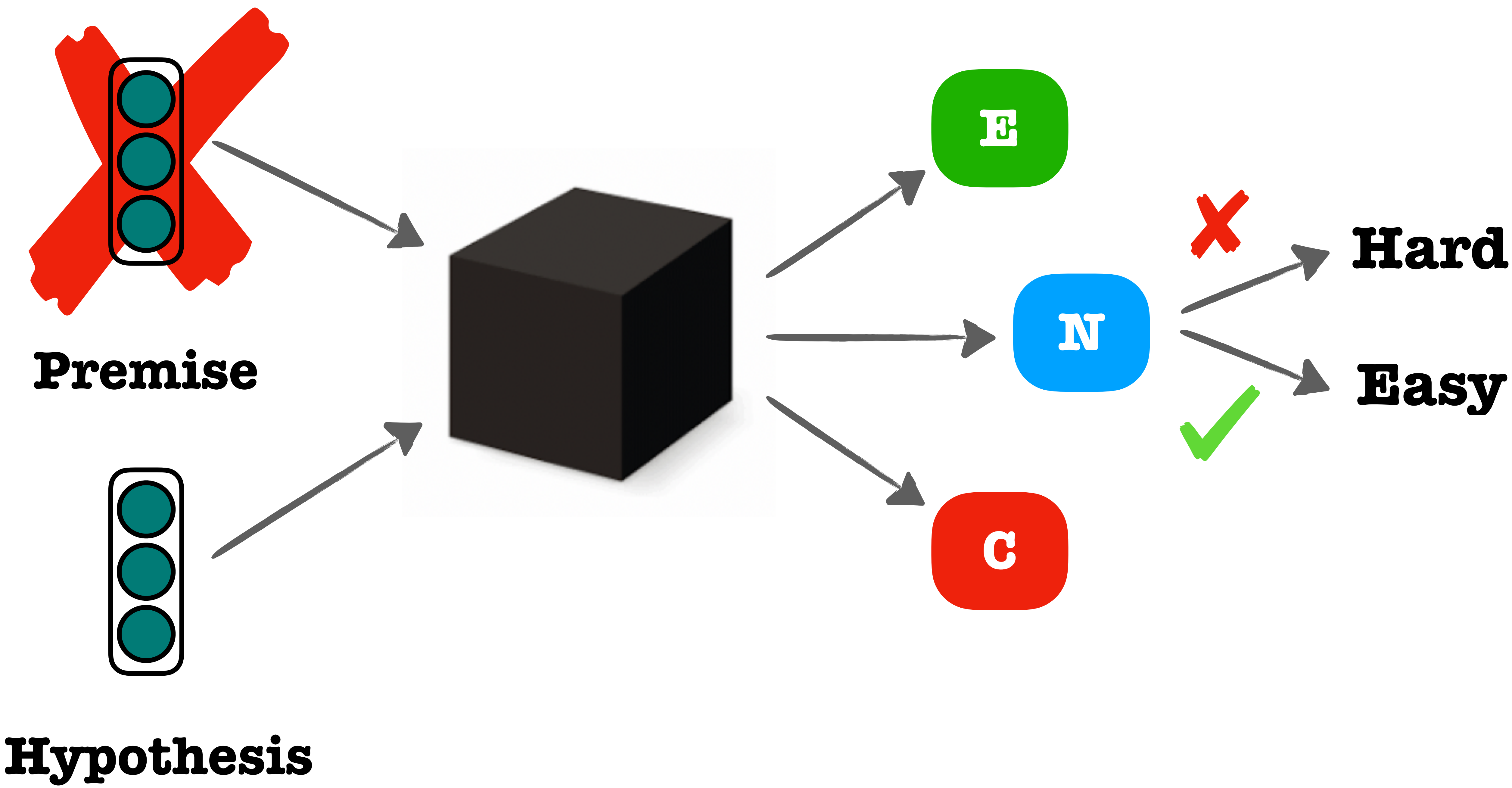
Over 50% of NLI examples can be correctly classified **without** ever observing the premise!

[Poliak et. al., 2018, Glockner et. al., 2018]

Can we filter out examples with artifacts?



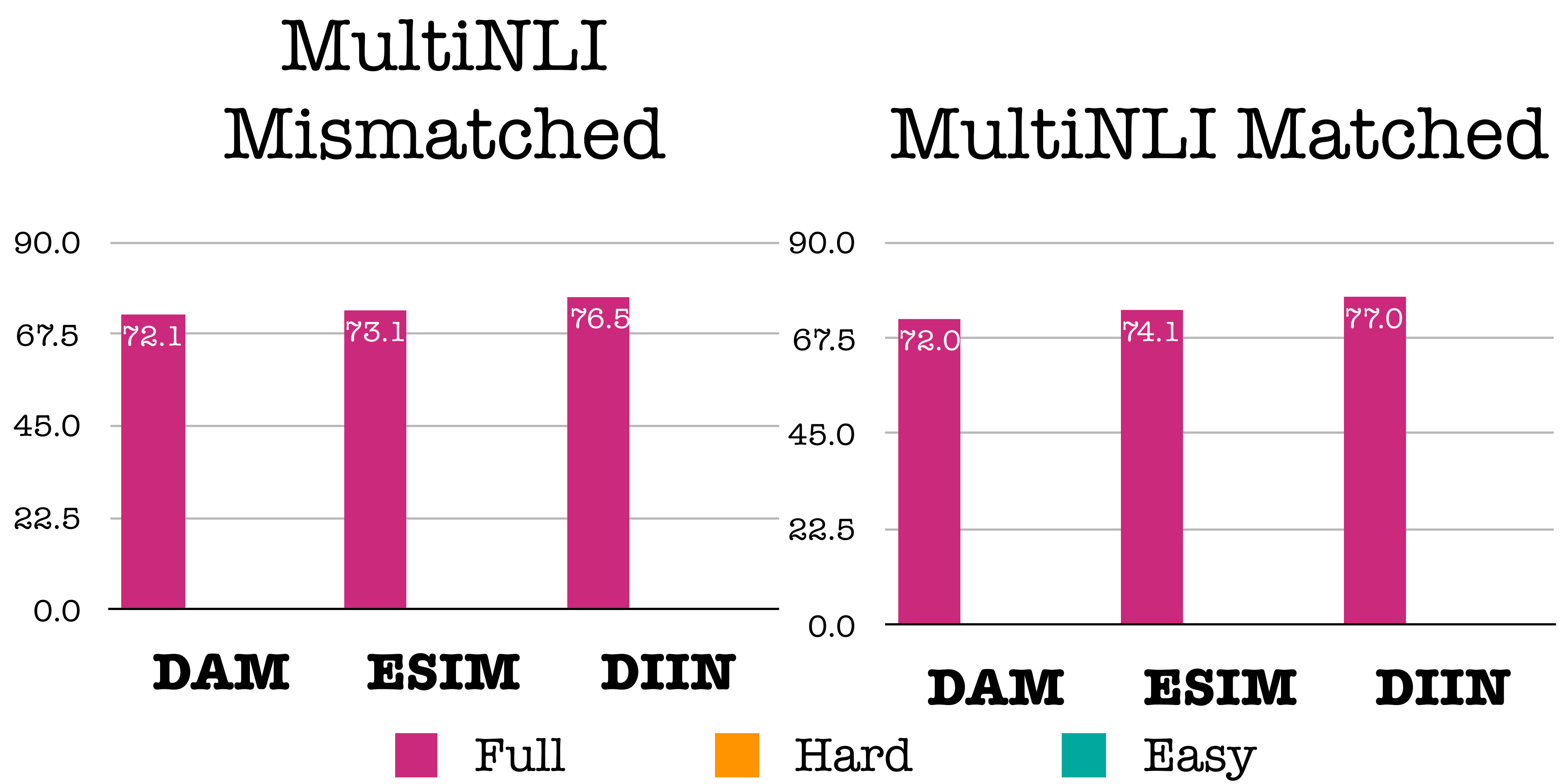
Can we filter out examples with artifacts?



Revisiting NLI models

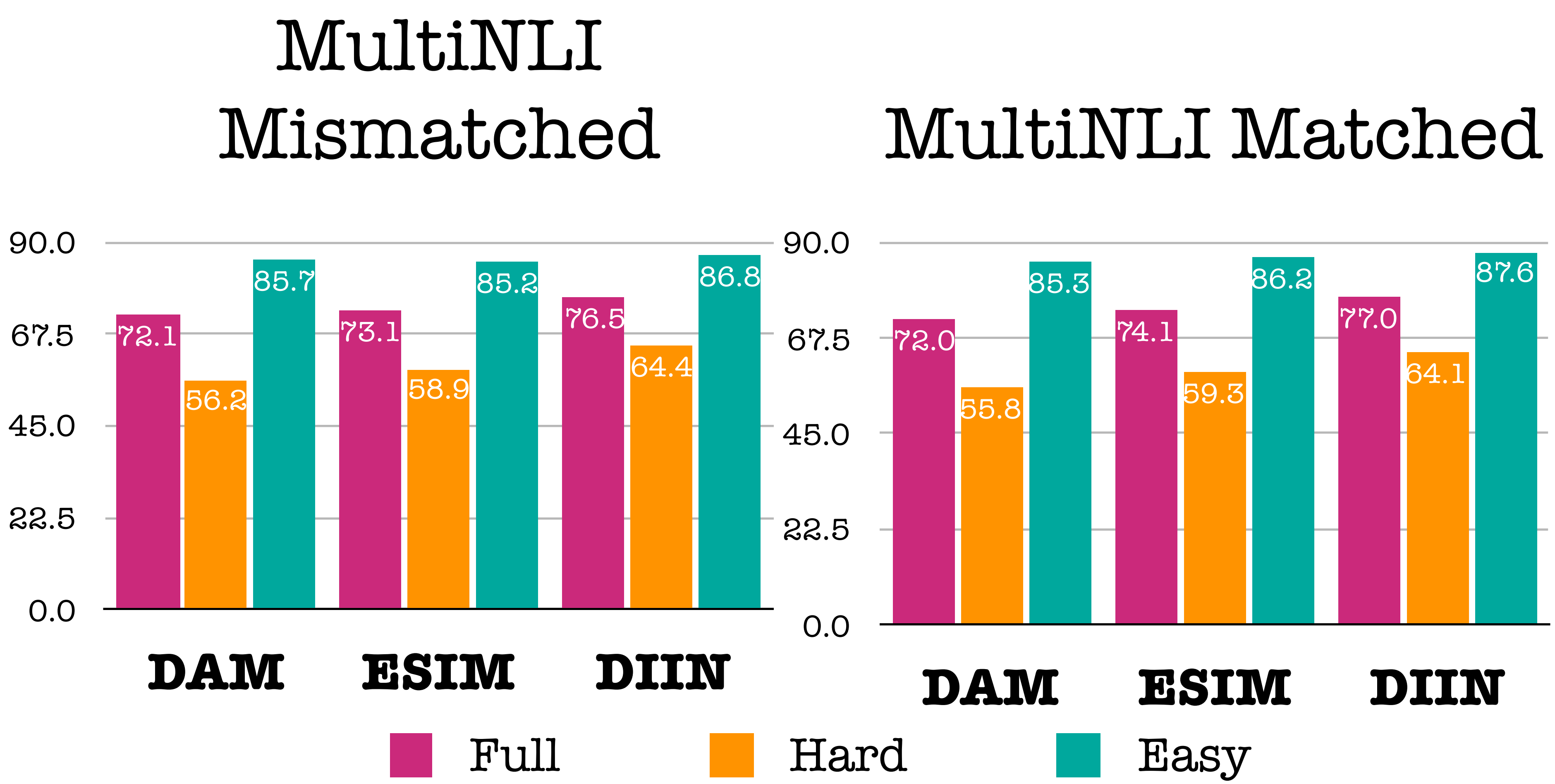
DAM - Decomposable Attention Model [Parikh et. al. 2016]
ESIM - Enhanced Sequential Inference Model [Chen et. al., 2017]
DIIN - Densely Interactive Inference Network [Gong et. al. 2018]

Revisiting NLI models



DAM - Decomposable Attention Model [Parikh et. al. 2016]
ESIM - Enhanced Sequential Inference Model [Chen et. al., 2017]
DIIN - Densely Interactive Inference Network [Gong et. al. 2018]

Revisiting NLI models



DAM - Decomposable Attention Model [Parikh et. al. 2016]
ESIM - Enhanced Sequential Inference Model [Chen et. al., 2017]
DIIN - Densely Interactive Inference Network [Gong et. al. 2018]

Artifacts by NLI Class

Artifacts by NLI Class

Some men and boys are playing frisbee in a grassy area.

Premise

Generalization

People play frisbee **outdoors**.

Entailment Hypothesis

Artifacts by NLI Class

Some men and boys are playing frisbee in a grassy area.

Premise

Generalization →

People play frisbee **outdoors**.

Entailment Hypothesis

A middle-aged man works under the engine of a train on rail tracks.

Premise

Modifiers →

A man is doing work on a **black** Amtrak train.

Neutral Hypothesis

Artifacts by NLI Class

Some men and boys are playing frisbee in a grassy area.

Premise

Generalization

People play frisbee **outdoors**.

Entailment Hypothesis

A middle-aged man works under the engine of a train on rail tracks.

Premise

Modifiers

A man is doing work on a **black** Amtrak train.

Neutral Hypothesis

Three dogs racing on racetrack.

Premise

Cats!

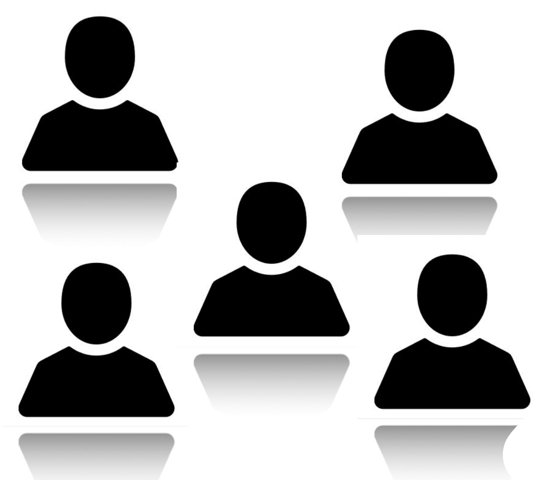
Three **cats** race on a track.

Contradiction Hypothesis

Annotation Artifacts

Two dogs are running through a field.

Premise



Entailment

There are **animals** outdoors.

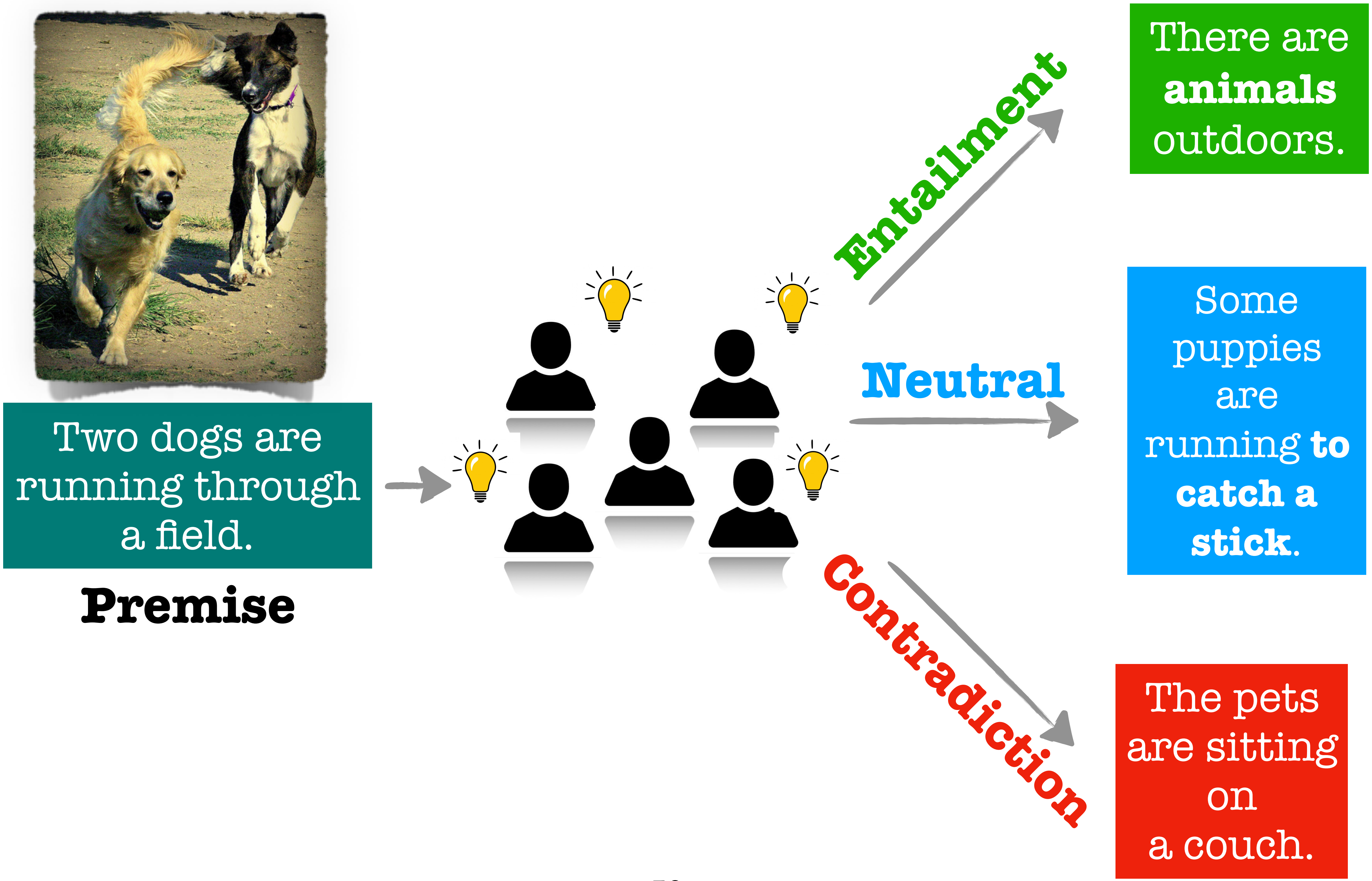
Neutral

Some puppies are running **to catch a stick**.

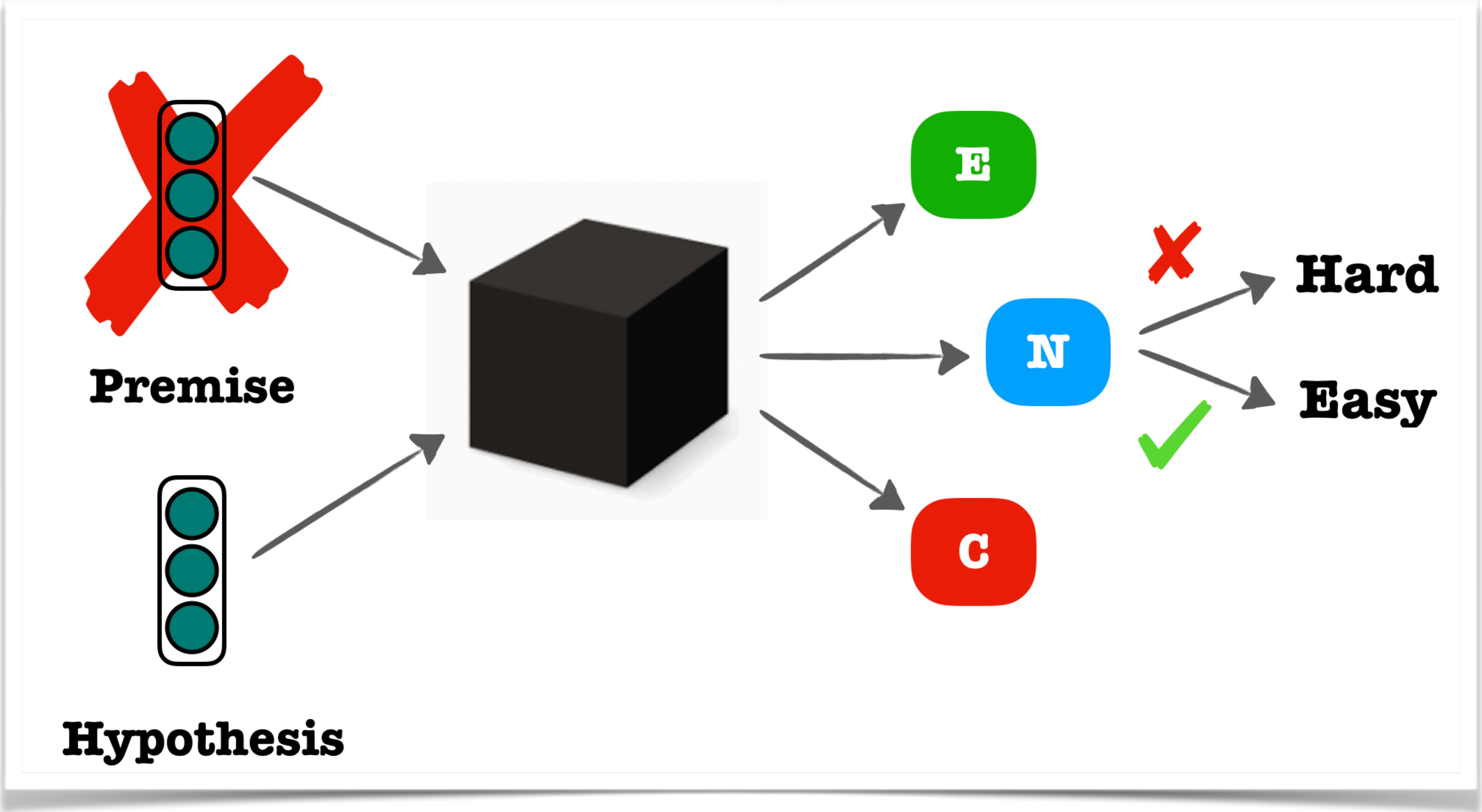
Contradiction

The pets are sitting on a couch.

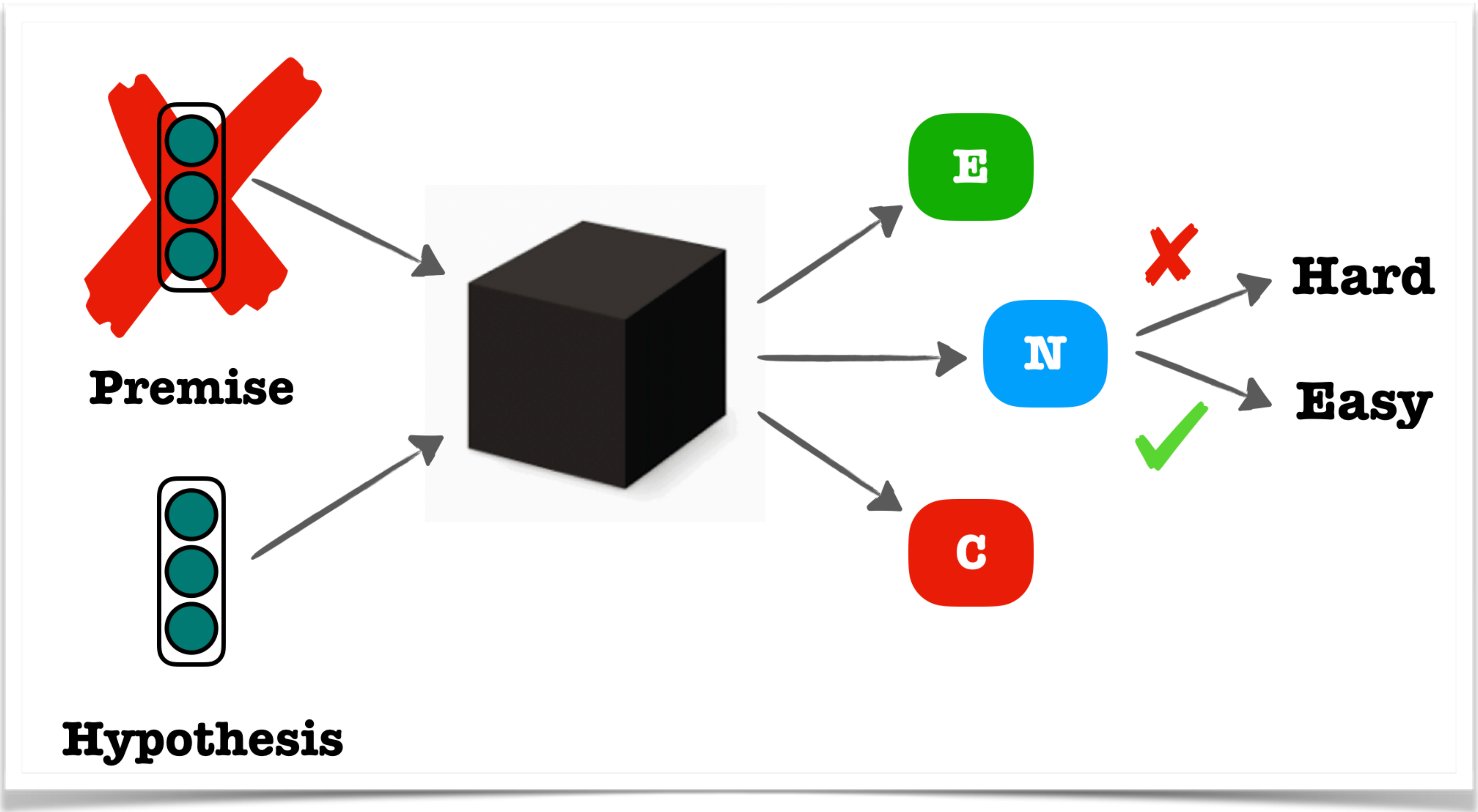
Annotation Artifacts



Can we filter out examples with artifacts?

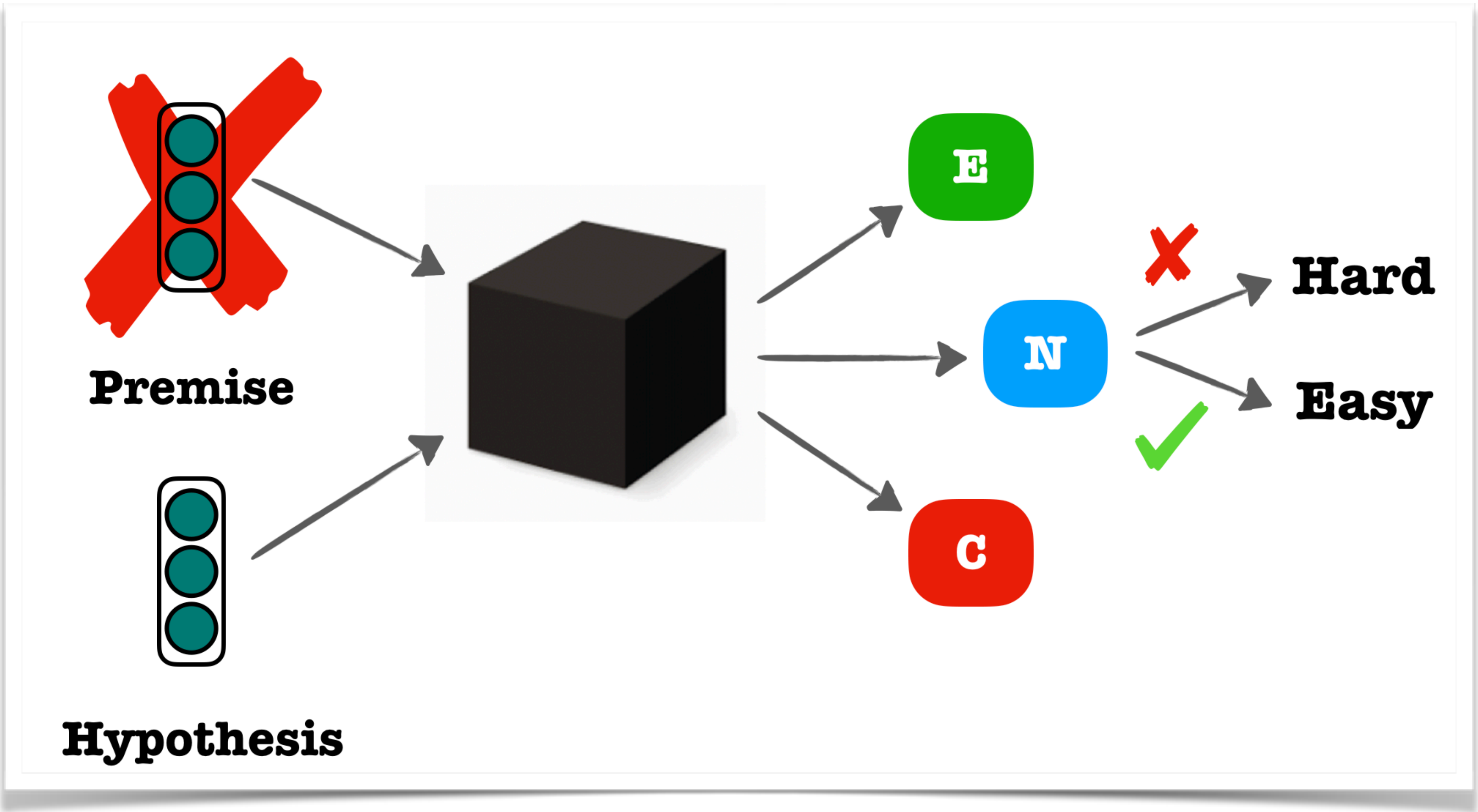


Can we filter out examples with artifacts?



- Hard examples exhibit their own artifacts!

Can we filter out examples with artifacts?



- Hard examples exhibit their own artifacts!
- Artifacts are still valid examples...

Looking ahead:
Learning from Datasets with Artifacts

Looking ahead: Learning from Datasets with Artifacts

- Intuition: Models which exploit artifacts == models which can detect artifacts

Looking ahead: Learning from Datasets with Artifacts

- Intuition: Models which exploit artifacts == models which can detect artifacts
    - Stylistic global features

Looking ahead: Learning from Datasets with Artifacts

- Intuition: Models which exploit artifacts == models which can detect artifacts
  - Stylistic global features
- Subsampling large datasets → weight each example based on how representative it could be [Coleman et. al., 2018]

Easy  **Hard**

Looking Ahead: Improved Data Collection

Looking Ahead: Improved Data Collection

- Partial input baselines. E.g. SWAG [Zellers et. al., 2018], DROP [Dua et. al., 2019], Diverse NLI [Poliak et. al., 2018]

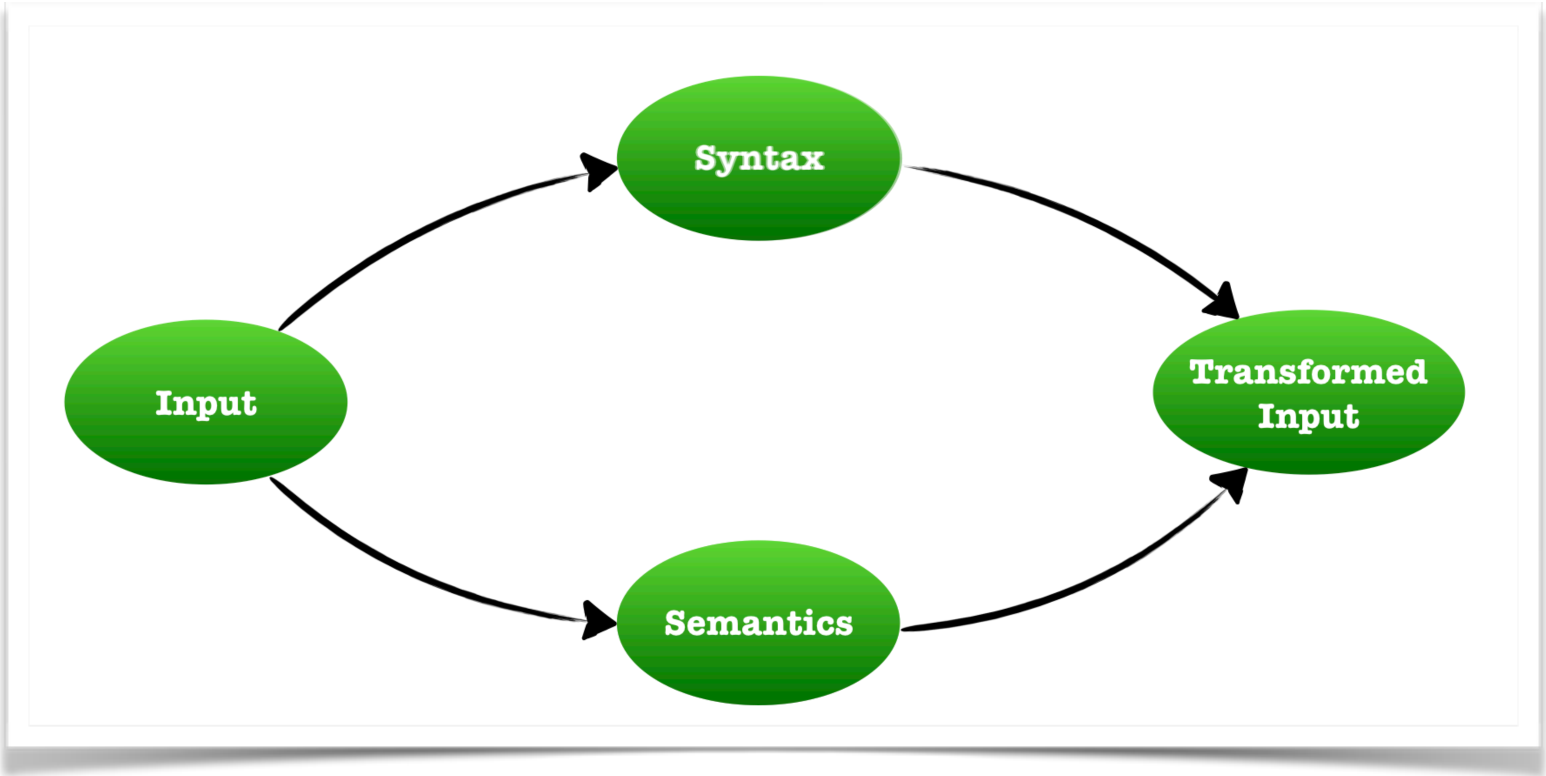
Looking Ahead: Improved Data Collection

- Partial input baselines. E.g. SWAG [Zellers et. al., 2018], DROP [Dua et. al., 2019], Diverse NLI [Poliak et. al., 2018]
- Alternatives to human elicitation for building datasets?

Looking Ahead: Improved Data Collection

- Partial input baselines. E.g. SWAG [Zellers et. al., 2018], DROP [Dua et. al., 2019], Diverse NLI [Poliak et. al., 2018]

- Alternatives to human elicitation for building datasets?



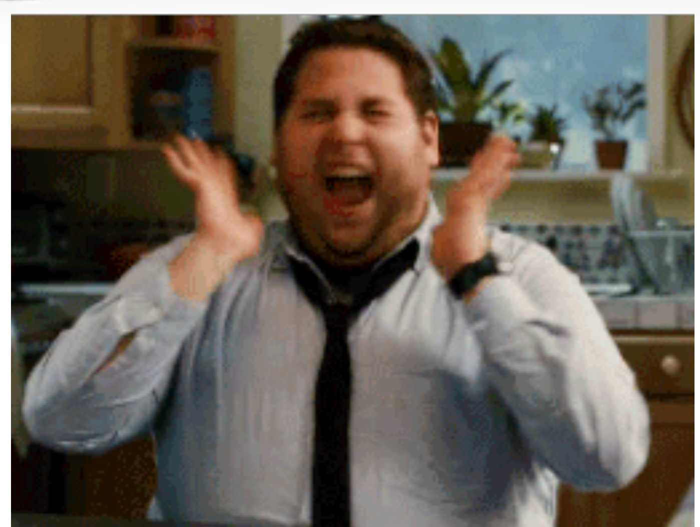
In conclusion : It’s an exciting time for NLP!

In conclusion :
It’s an exciting time for NLP!

The New York Times

Finally, a Machine That Can Finish Your Sentence

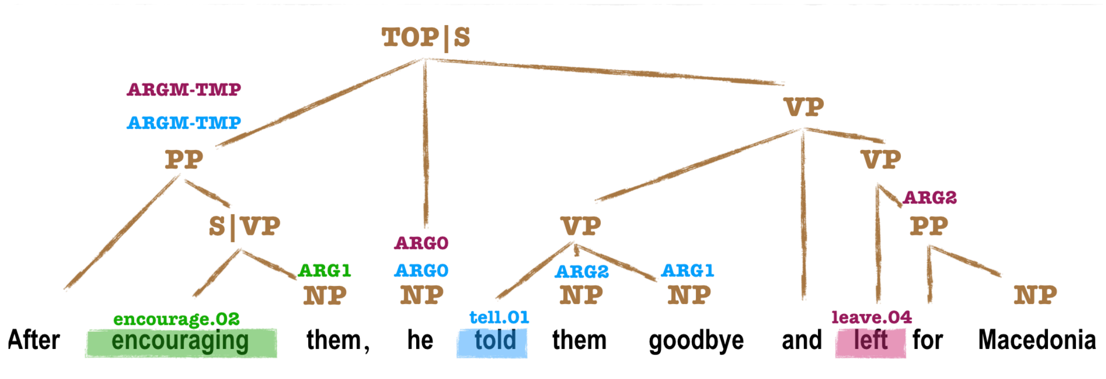
Completing someone else’s thought is not an easy trick for A.I. But new systems are starting to crack the code of natural language.



In conclusion - Learning Challenges

Part I

Can linguistic structure act as an informative prior to improve our models?



Predicted structure can help representation learning.

Part II

What in our data is causing models to achieve high performance?

Need models robust to artifacts.

Thanks!

http://www.cs.cmu.edu/~sswayamd

swabhs

swabhz