Learning Challenges in Natural Language Processing

Swabha Swayamdipta
Jan 22, 2019

Carnegie Mellon University
Language Technologies Institute

Good News!

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.

Good News!


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.

Language Models $\sim p(\textbf{word} \mid \textbf{context})$

Good News!

Language Models ~ $p(\textbf{word} \mid \textbf{context})$

Contextualized Representations

Good News!

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.

Language Models $\sim p(\textbf{word} \mid \textbf{context})$

Contextualized Representations

Unsupervised

↓

Large Datasets

Good News!

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.

(Peters et. al., 2018)

(Devlin et. al., 2018)

Language Models $\sim p(\textbf{word} \mid \textbf{context})$

Contextualized Representations

Unsupervised

↓

Large Datasets

Downstream Supervised Tasks

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?

BERT (Devlin et. al., arXiv 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 old Acadian colonists declined after the year **1675**.

The number of new Huguenot colonists declined after what year?

Jia & Liang (EMNLP 17)

Percy Liang (AI Frontiers 18)

BERT (Devlin et. al., arXiv 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 old Acadian colonists declined after the year **1675**.

The number of new Huguenot colonists declined after what year?

Jia & Liang (EMNLP 17)

Percy Liang (AI Frontiers 18)

BERT (Devlin et. al., arXiv 2018)

Challenges

Part I

Can we incorporate some priors about language into deep learning 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.

Learning Challenge #1

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

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.

Language is Structured

- Semantics : Who did what to whom?

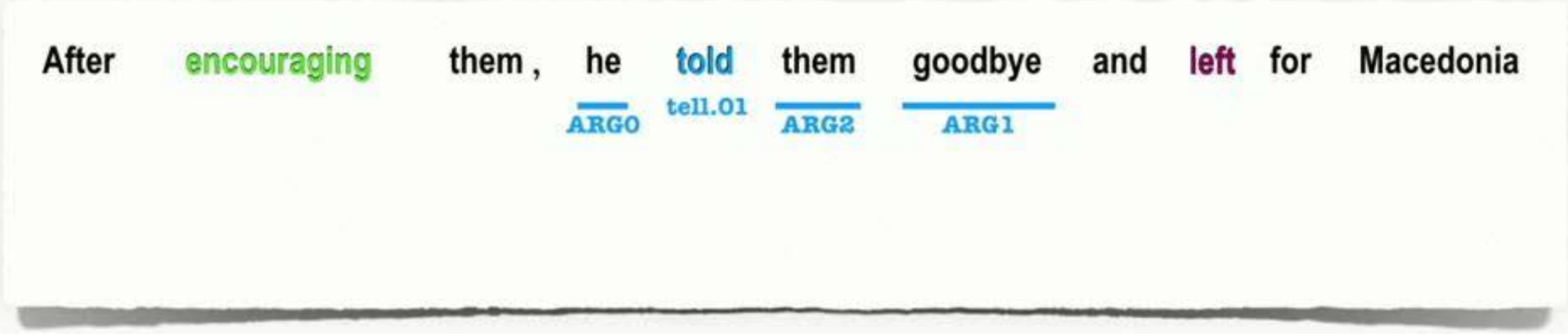
Language is Structured

- Semantics : Who did what to whom?

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

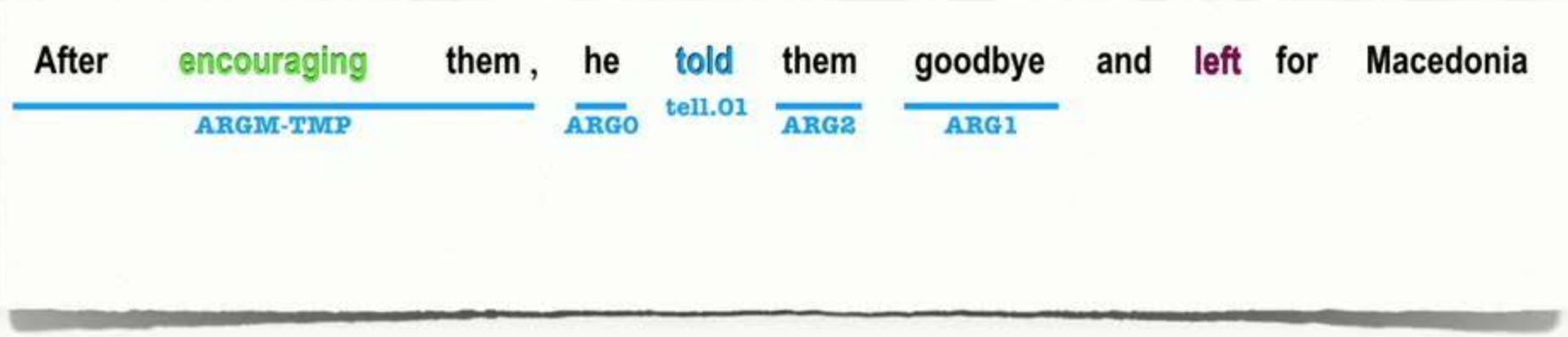
Language is Structured

- Semantics : Who did what to whom?



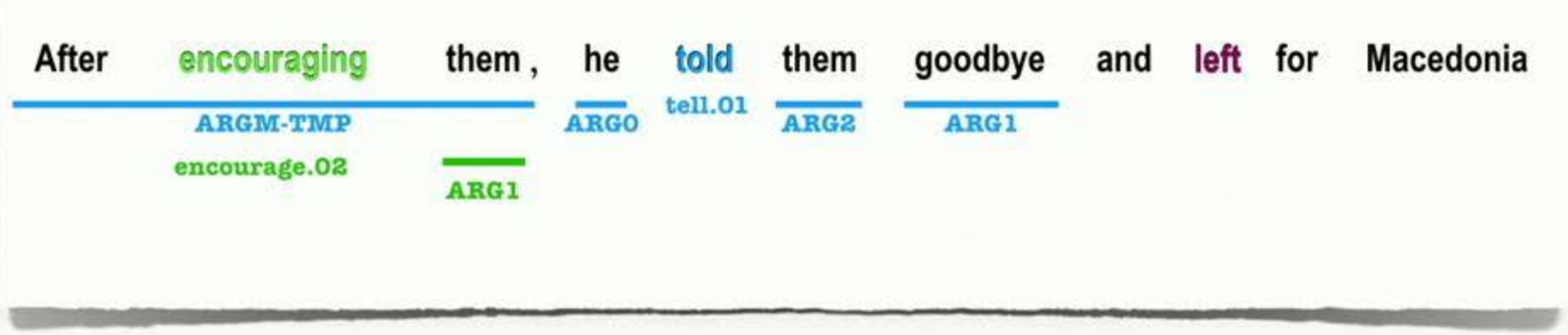
Language is Structured

- Semantics : Who did what to whom?



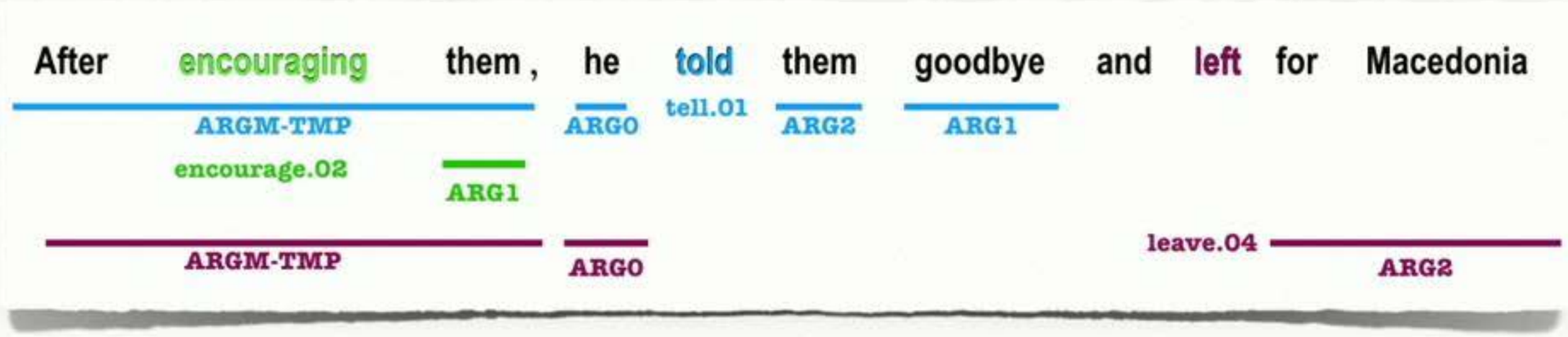
Language is Structured

- Semantics : Who did what to whom?



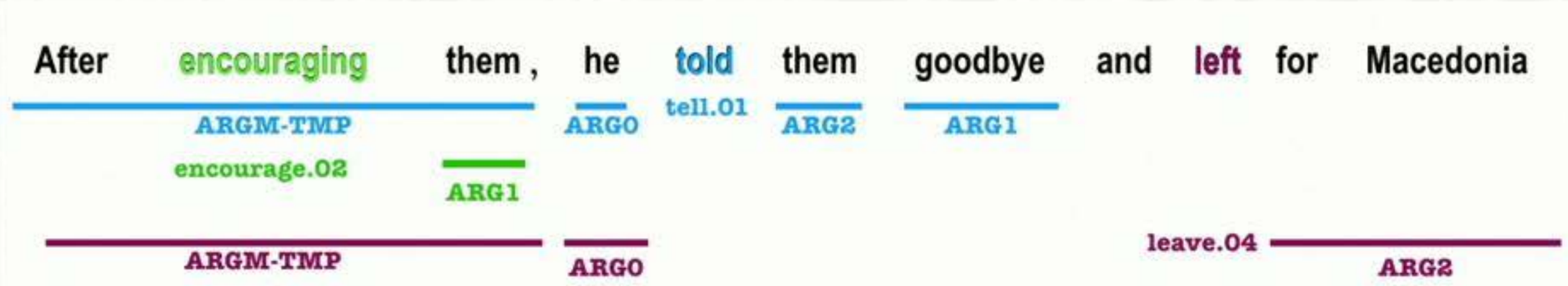
Language is Structured

- Semantics : Who did what to whom?



Language is Structured

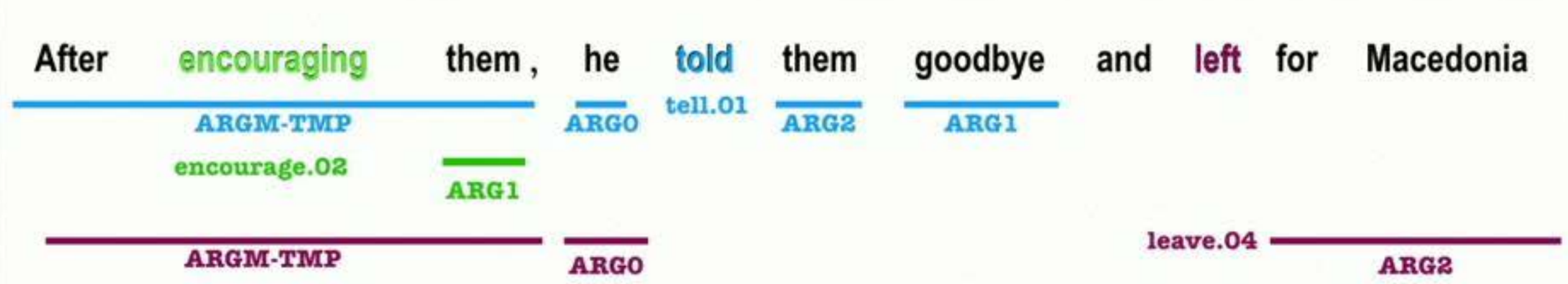
- Semantics : Who did what to whom?
- This talk: **Span**-based, broad-coverage semantic **structured** prediction.



PropBank (Palmer et. al., 05)

Language is Structured

- Semantics : Who did what to whom?
- This talk: **Span**-based, broad-coverage semantic **structured** prediction.
- Availability of data...



PropBank (Palmer et. al., 05)

Syntax & Semantics

- Syntax - a foundation for sentence meaning / semantics

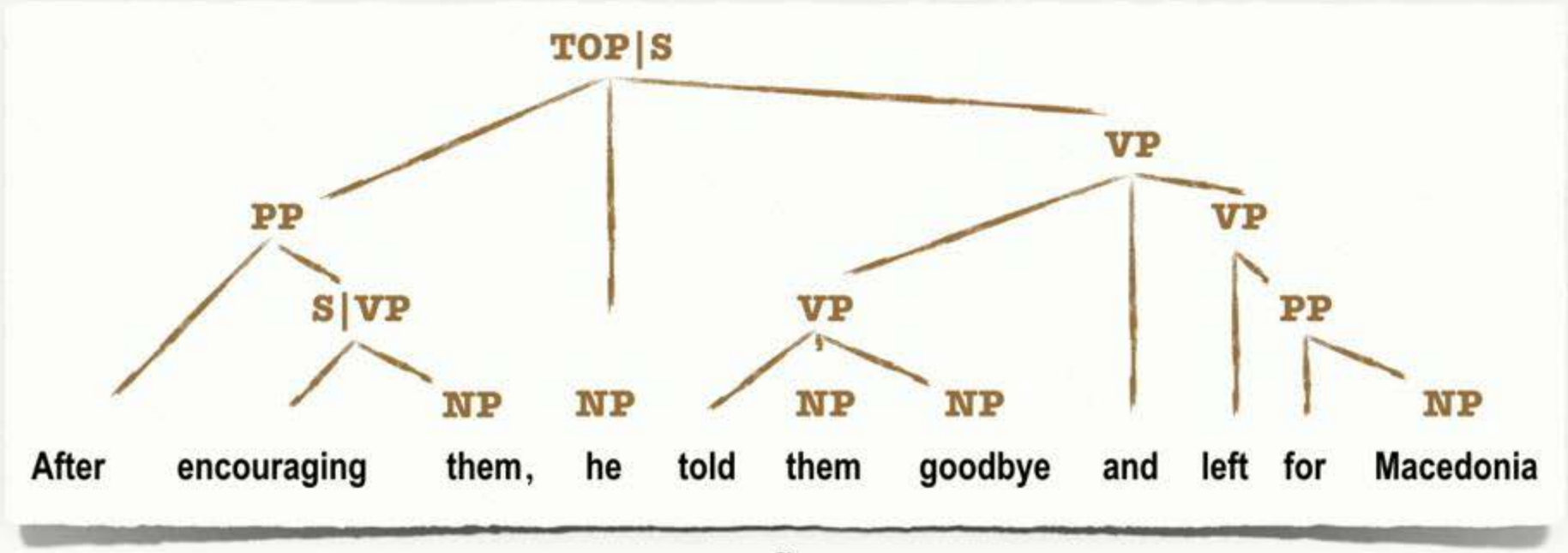
Syntax & Semantics

- Syntax - a foundation for sentence meaning / semantics

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

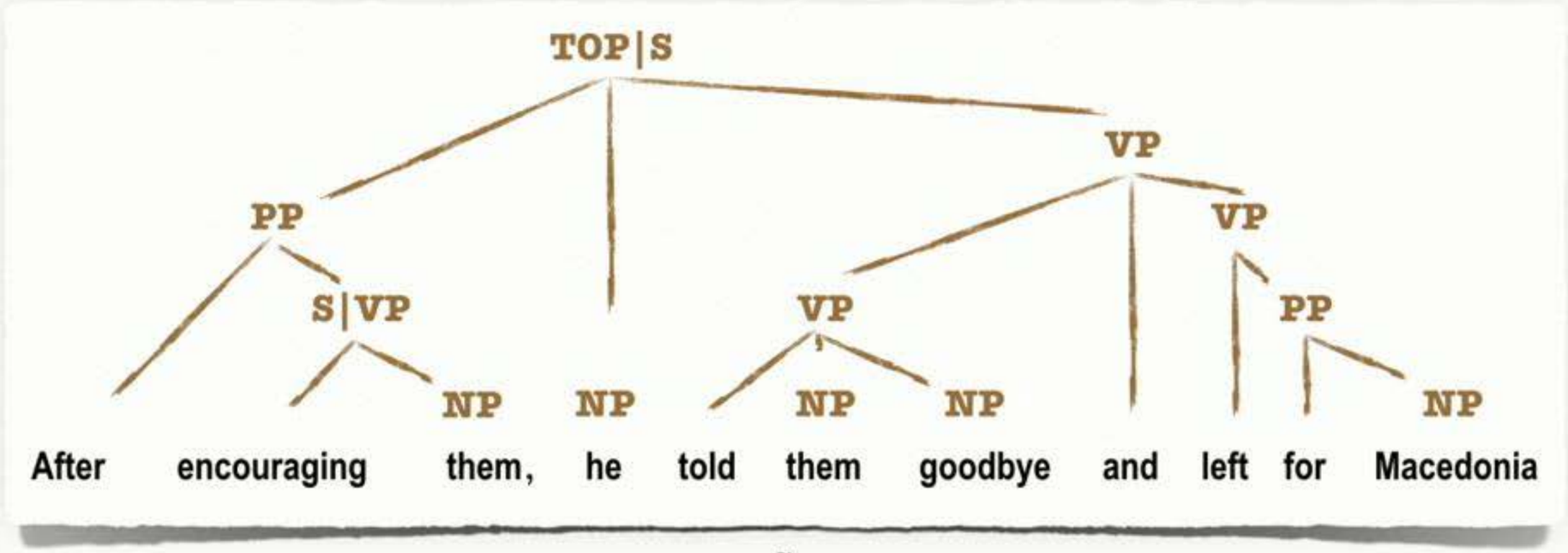
Syntax & Semantics

- Syntax - a foundation for sentence meaning / semantics



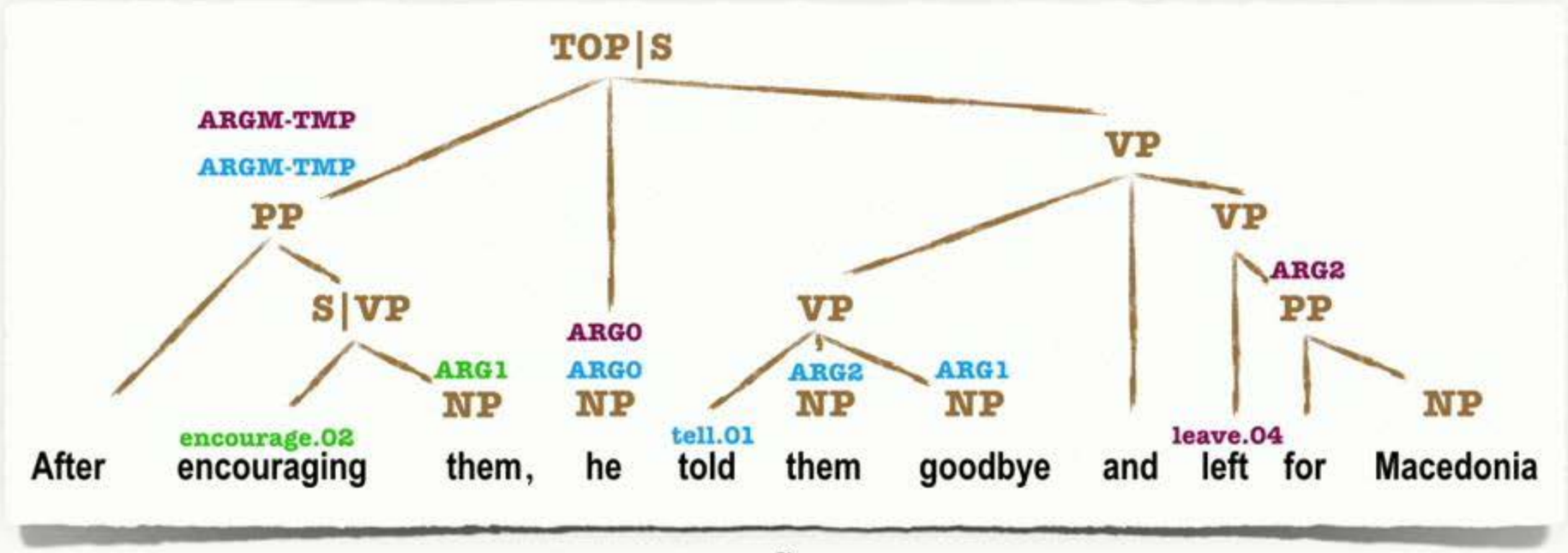
Syntax & Semantics

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



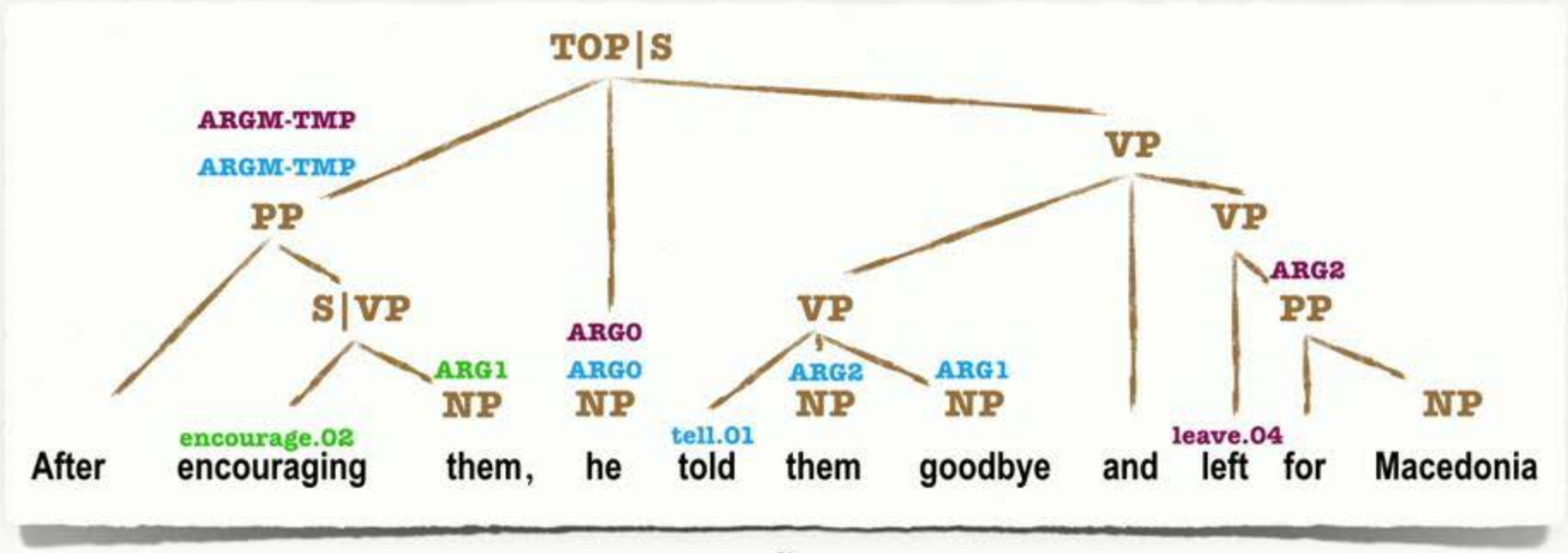
Syntax & Semantics

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



Syntax & Semantics

- Syntax - a foundation for sentence meaning / semantics
- Phrase-based syntax (node → span)
- Key Intuition: Learning from **multiple, complementary** structures results in stronger representations.



Syntactic Scaffolds for Semantic Structures

EMNLP 2018

S.

Sam Thomson

Kenton Lee

Luke Zettlemoyer

Chris Dyer

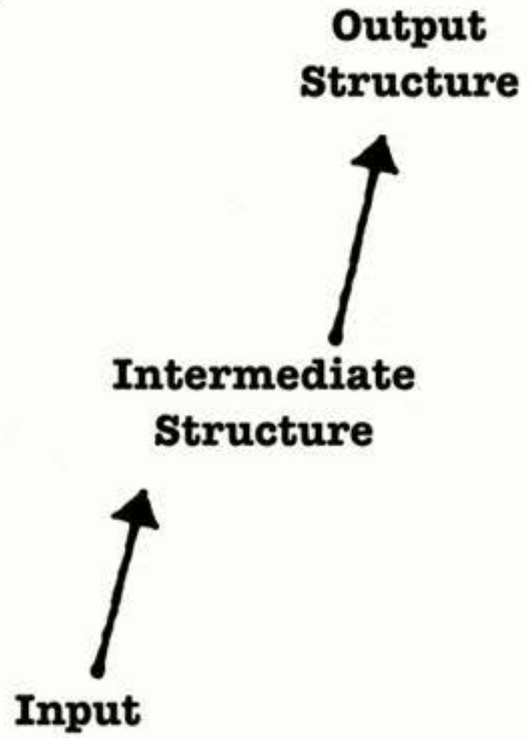
Noah A. Smith

Structured prediction with intermediate structures

- Intermediate syntactic structures.

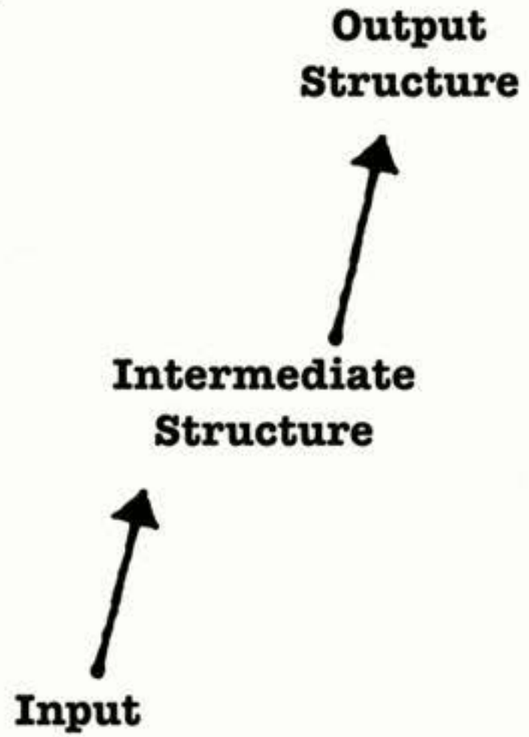
Structured prediction with intermediate structures

- Intermediate syntactic structures.
- Traditionally a pipeline, both at train and test time.



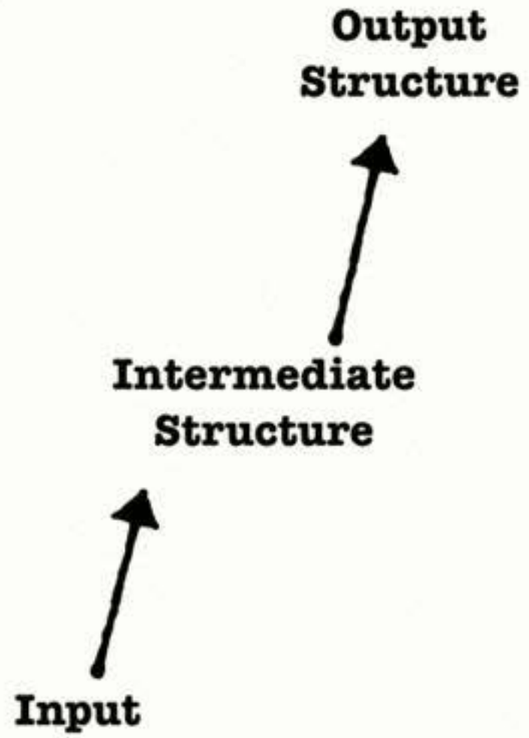
Structured prediction with intermediate structures

- Intermediate syntactic structures.
- Traditionally a pipeline, both at train and test time.
  - More structured data



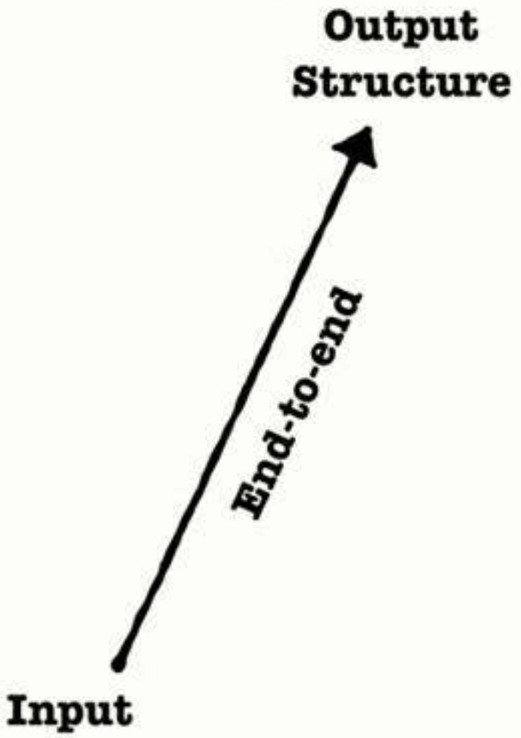
Structured prediction with intermediate structures

- Intermediate syntactic structures.
- Traditionally a pipeline, both at train and test time.
  - More structured data
  - Cascading errors



Structured prediction with intermediate structures

- Intermediate syntactic structures.
- Traditionally a pipeline, both at train and test time.
  - More structured data
  - Cascading errors
- Forsaken in most end-to-end models, but at a cost (He et. al, 2017).



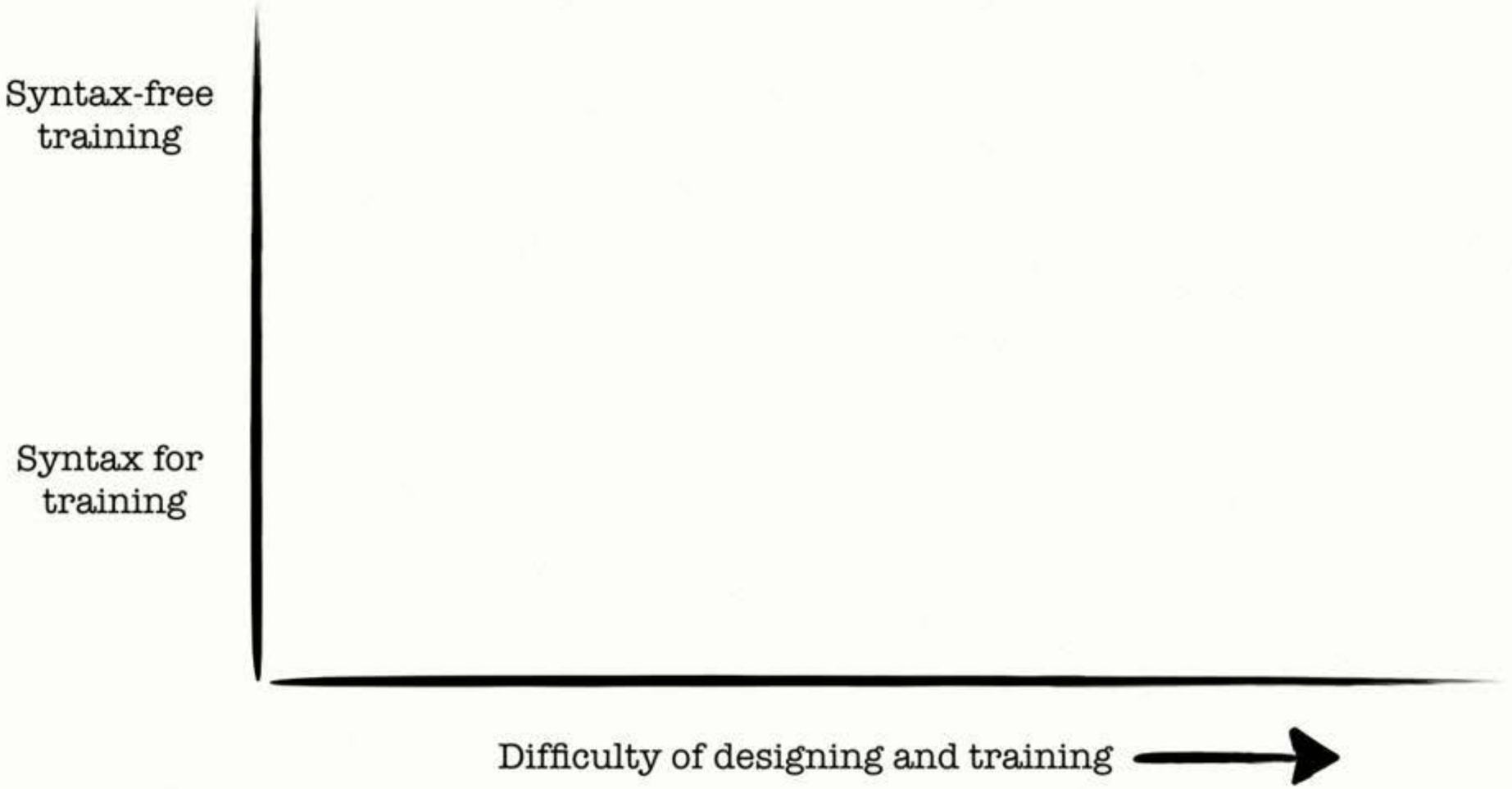
Questions

1. Can we leverage intermediate structure, but avoid **full** intermediate structured prediction?

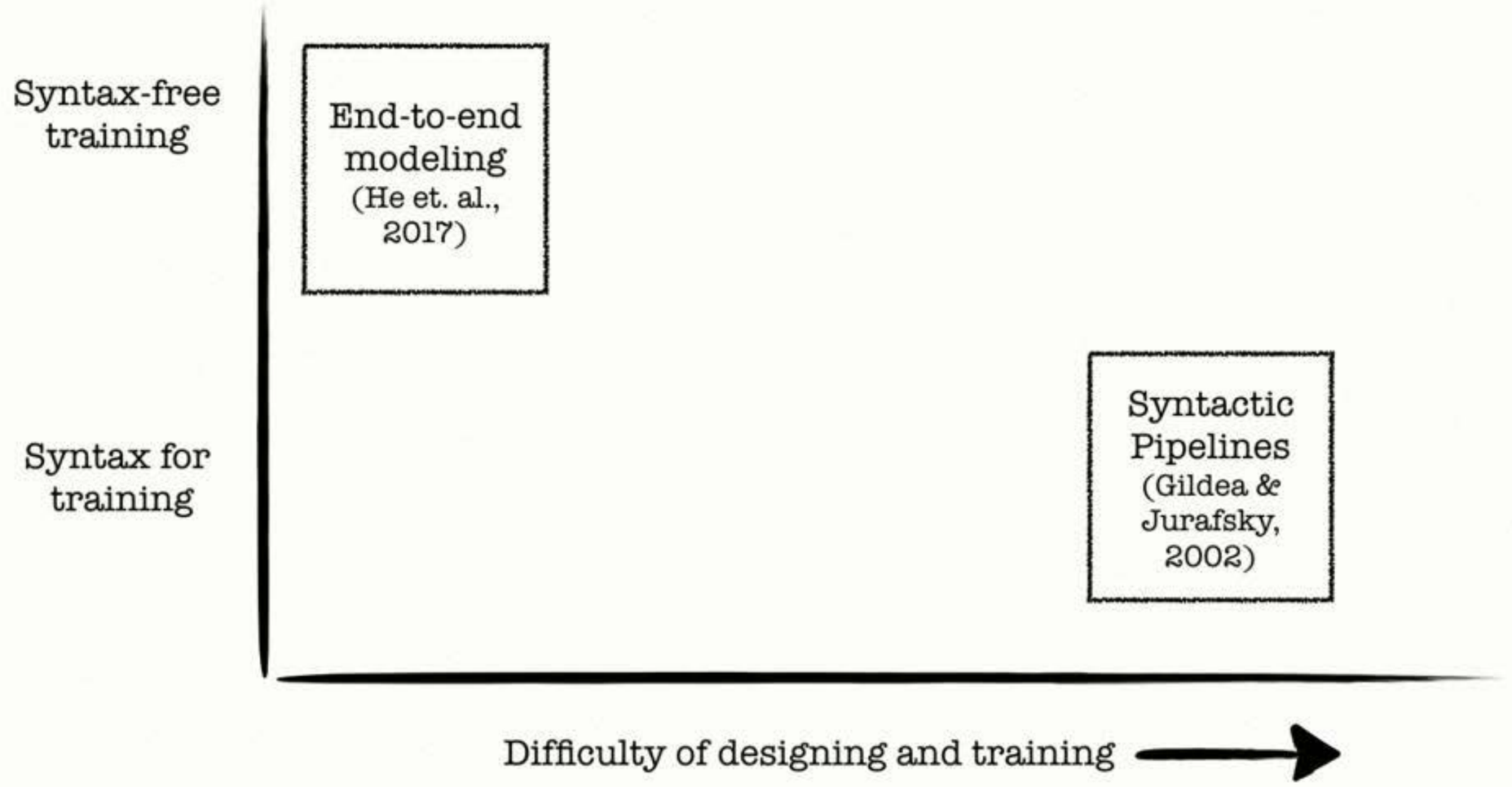
2. Can we avoid intermediate structured prediction altogether at test time?

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

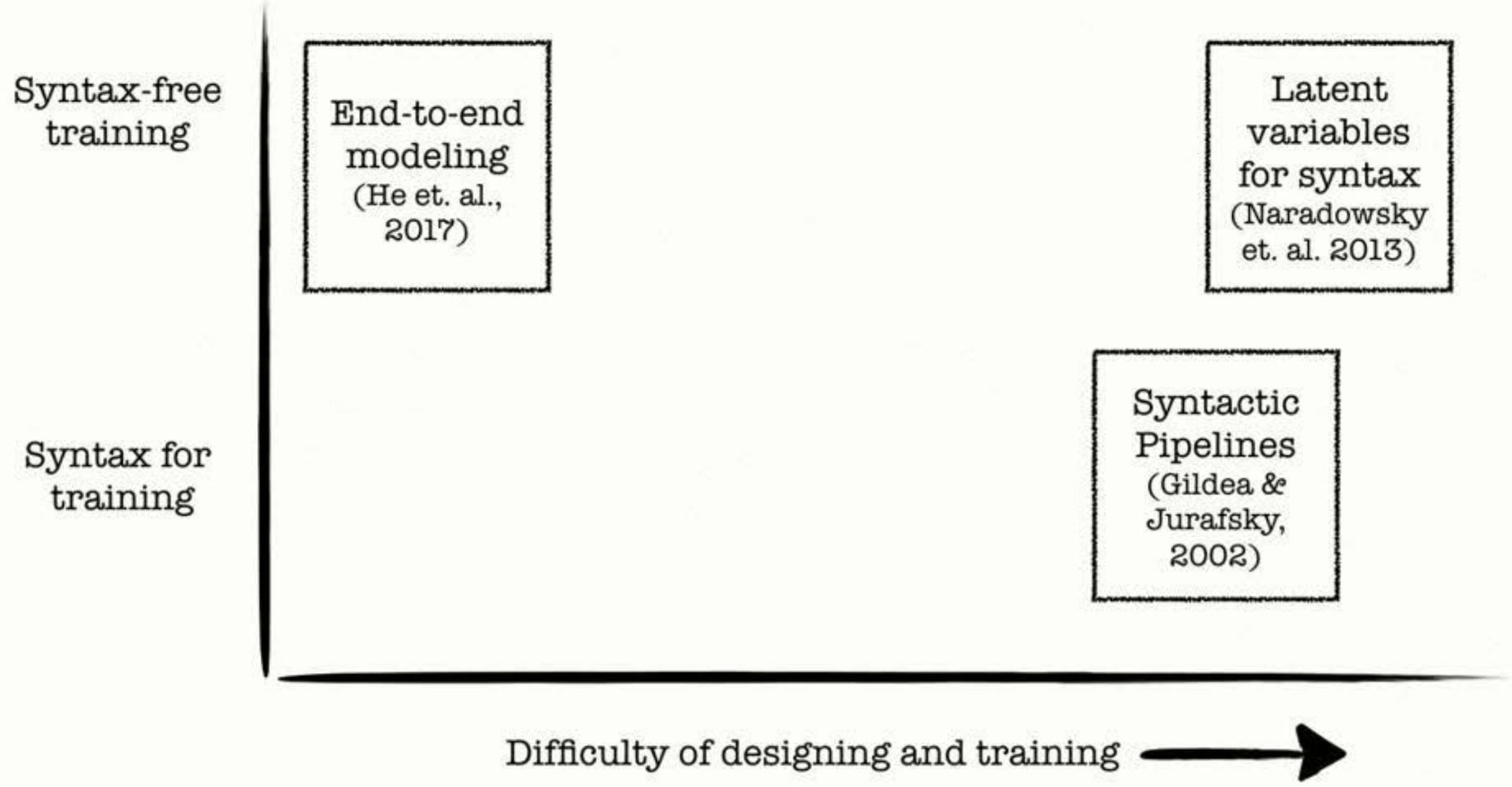
Training Paradigms



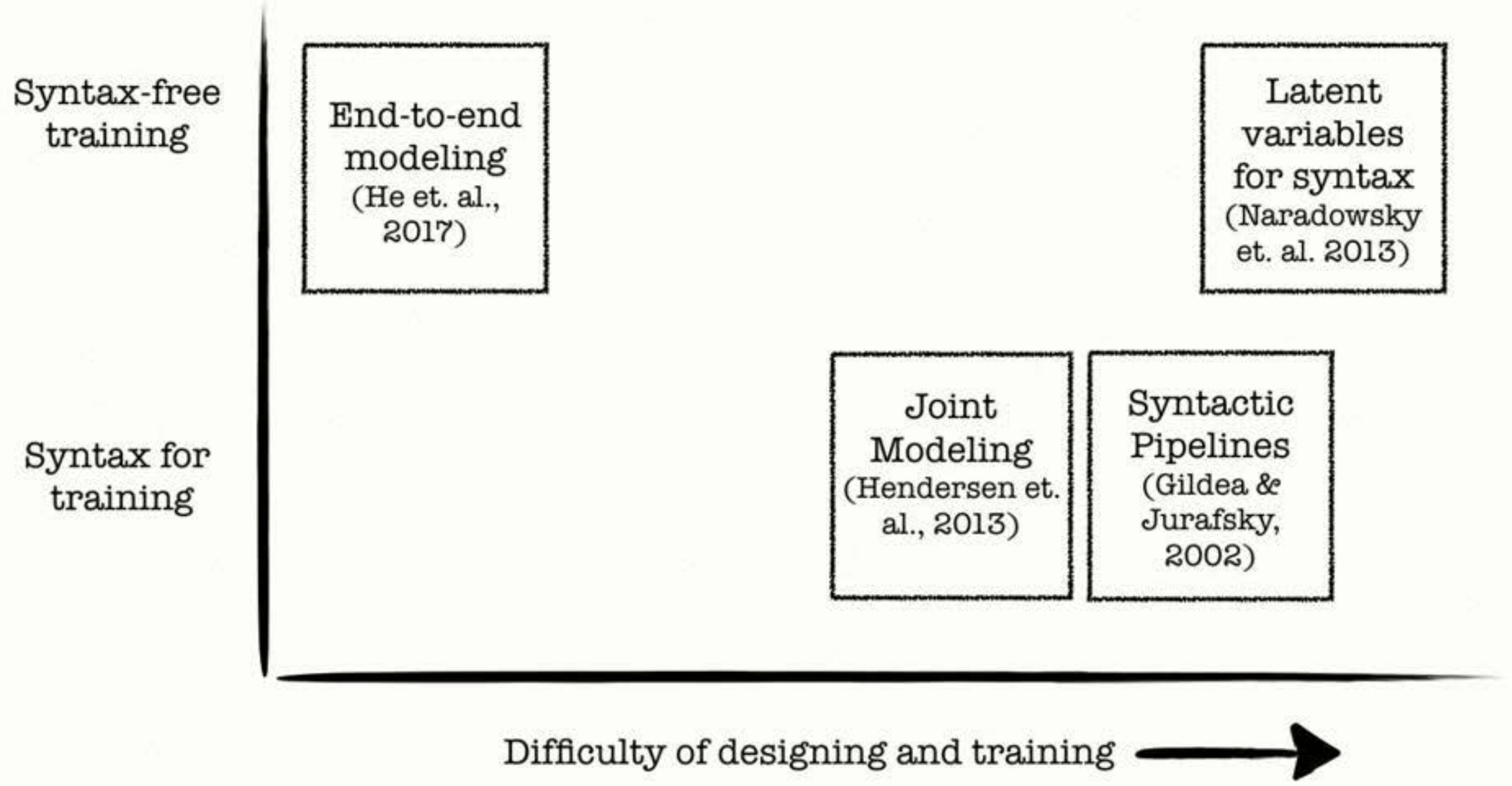
Training Paradigms



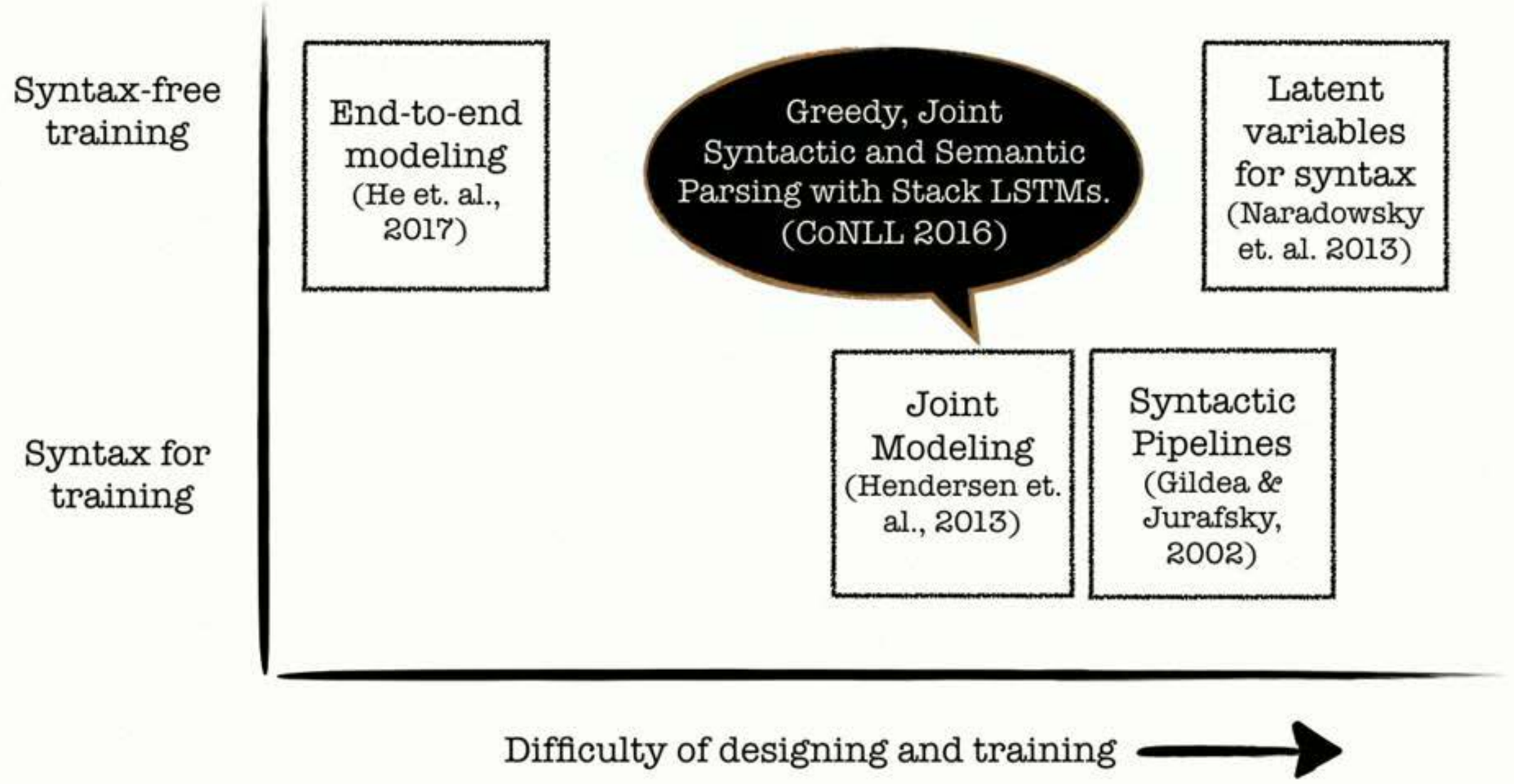
Training Paradigms



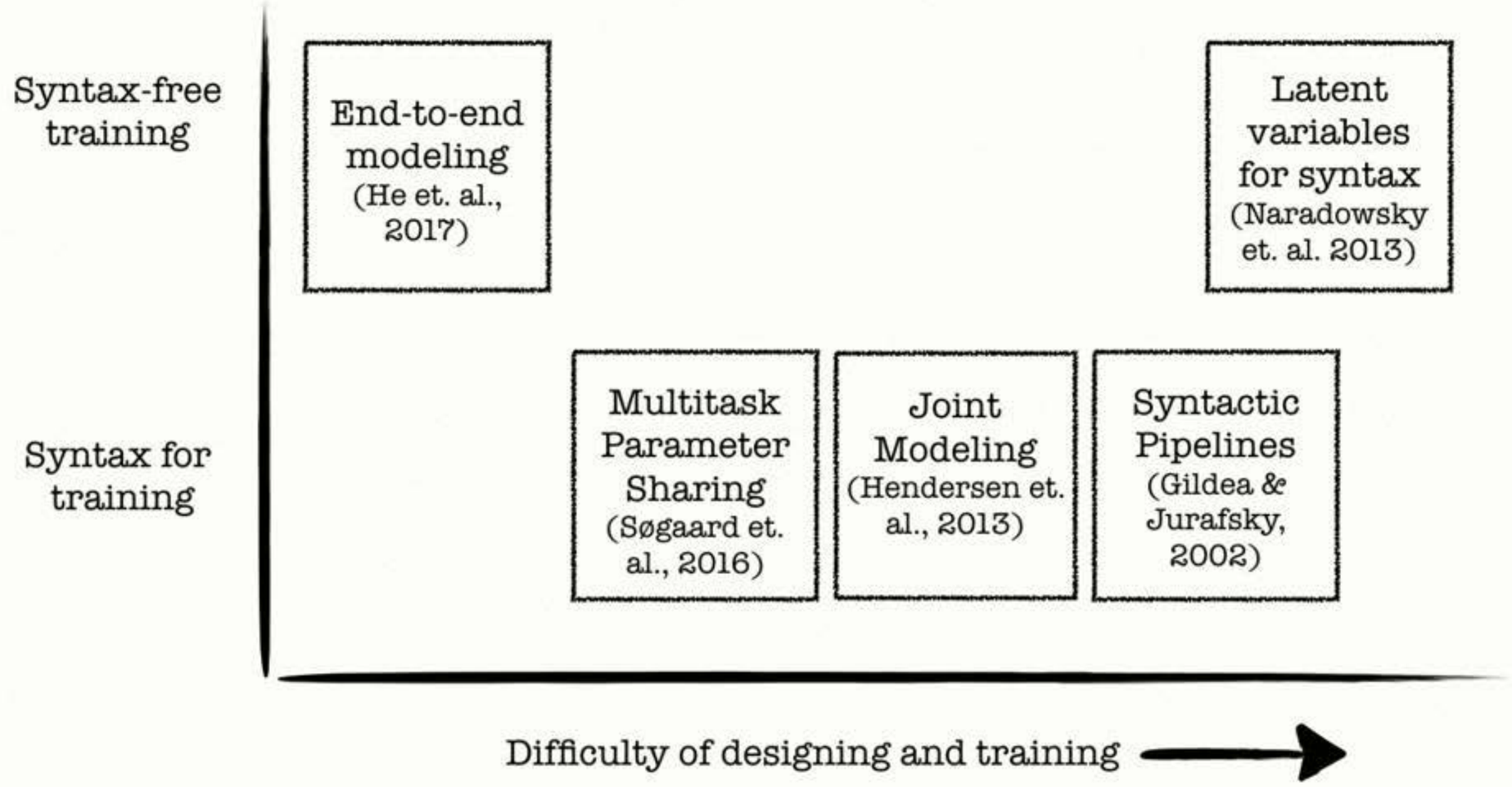
Training Paradigms



Training Paradigms

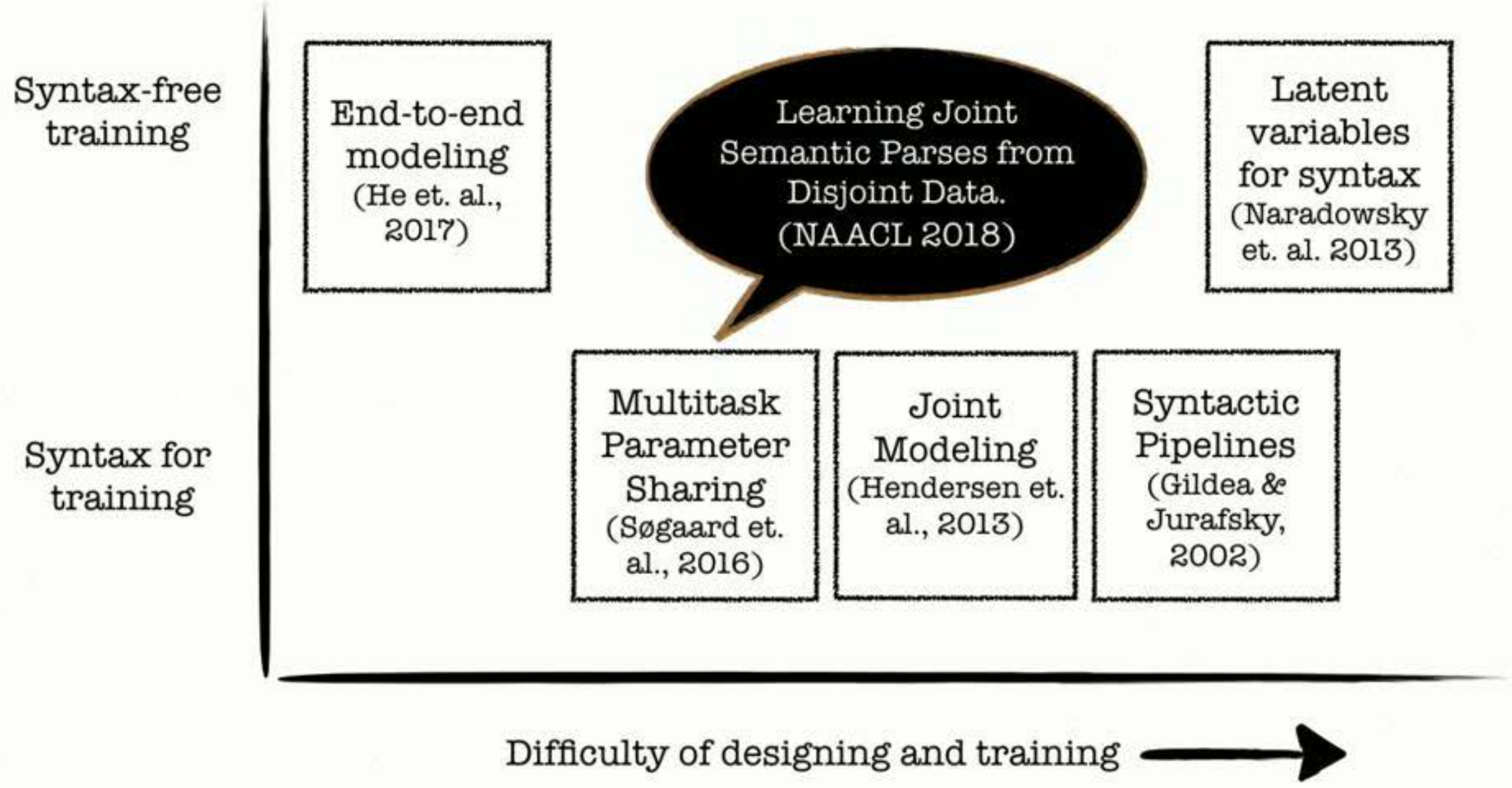


Training Paradigms



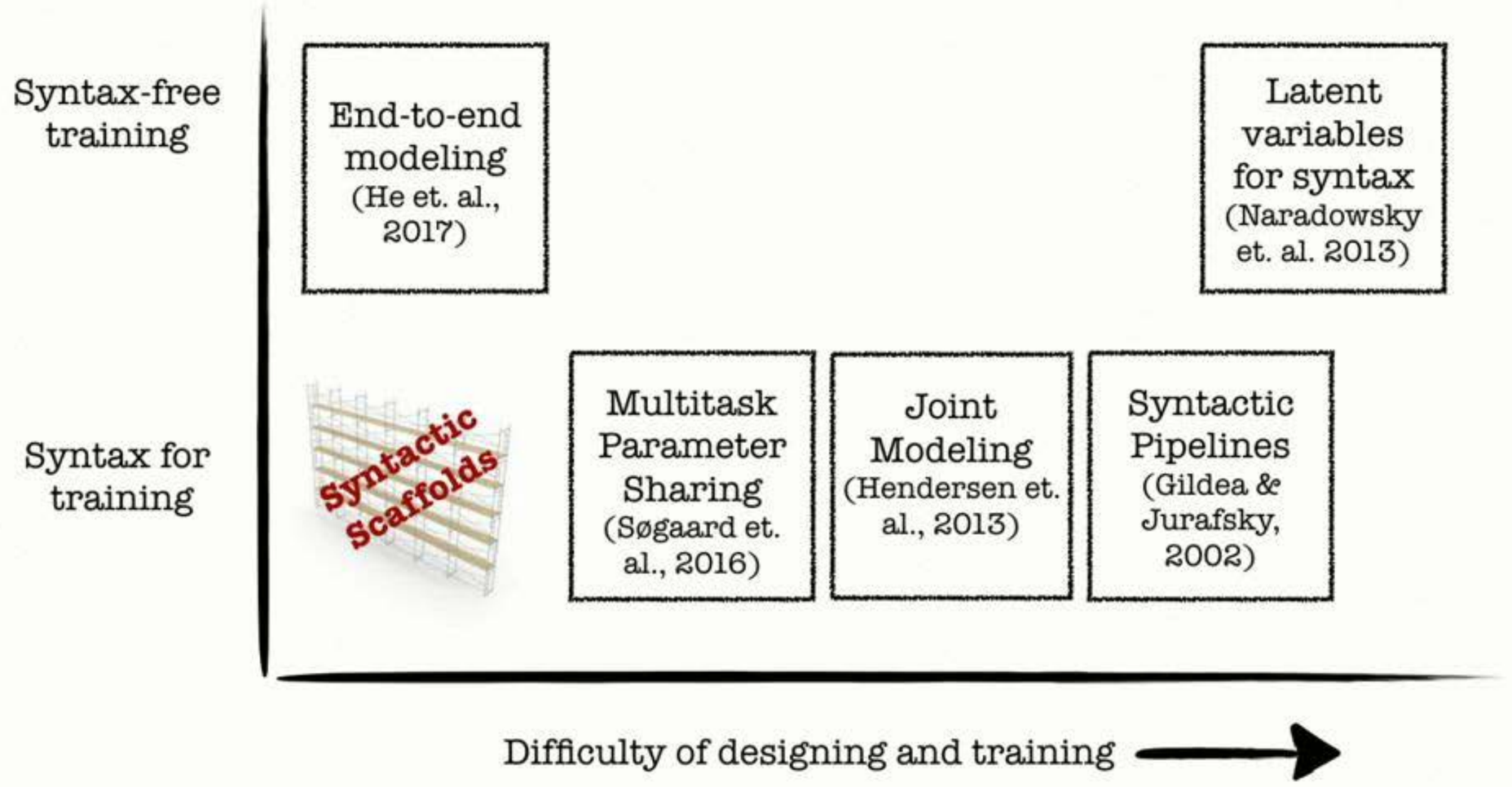
FREDA (Daumé, 09)

Training Paradigms



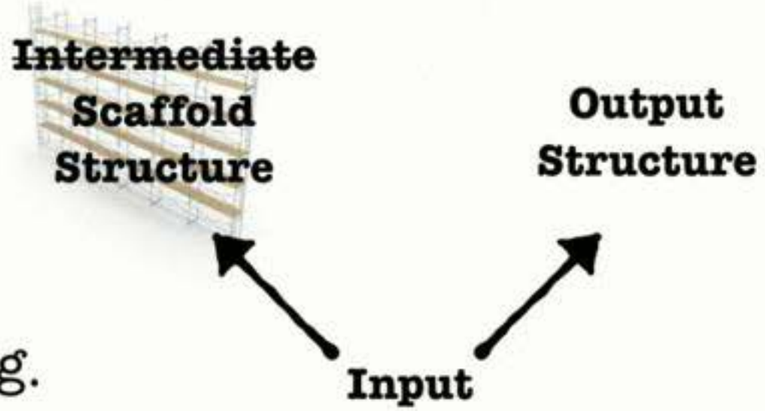
FREDA (Daumé, 09)

Training Paradigms



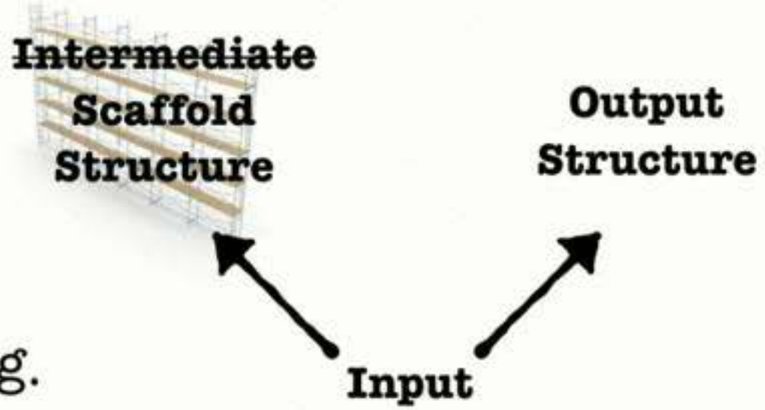
FREDA (Daumé, 09)

Syntactic Scaffolds



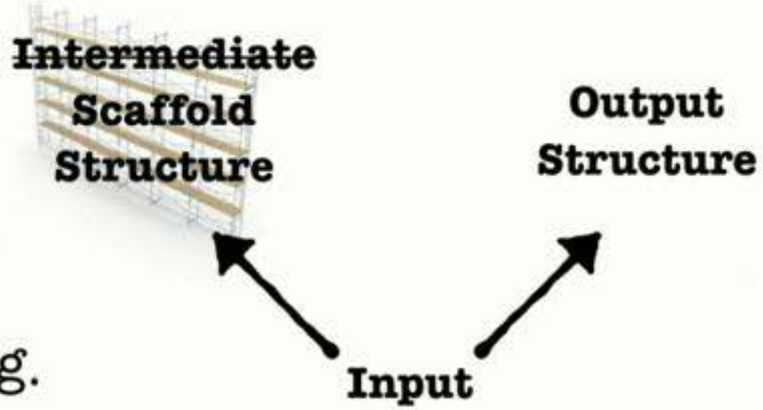
- Multitask setting.
- Shallow intermediate structure prediction.

Syntactic Scaffolds



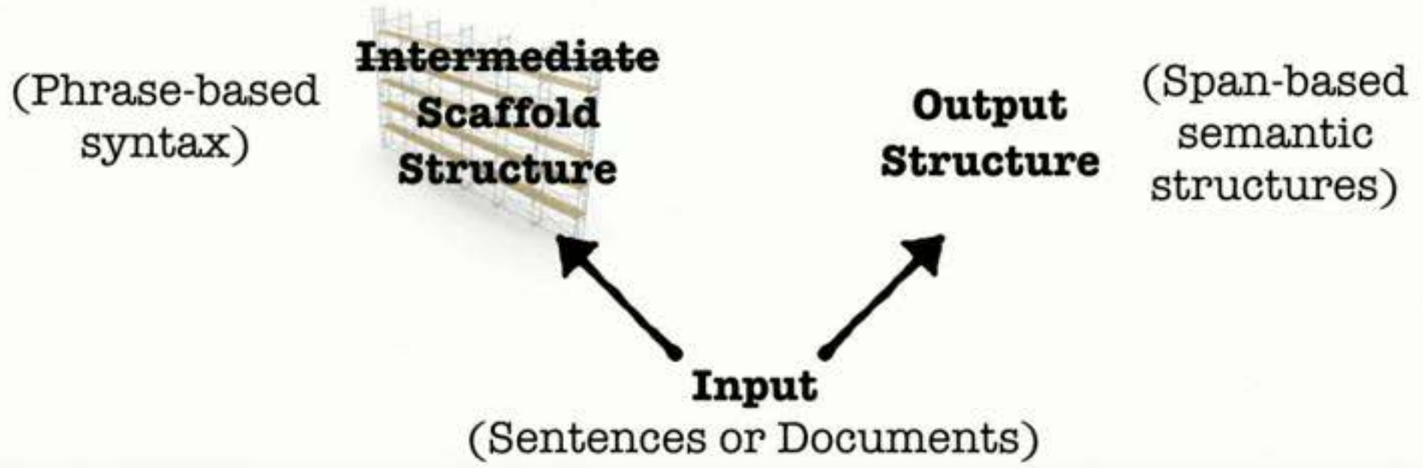
- Multitask setting.
- Shallow intermediate structure prediction.
- Learn (soft) syntax-aware representations, avoid cascaded errors.

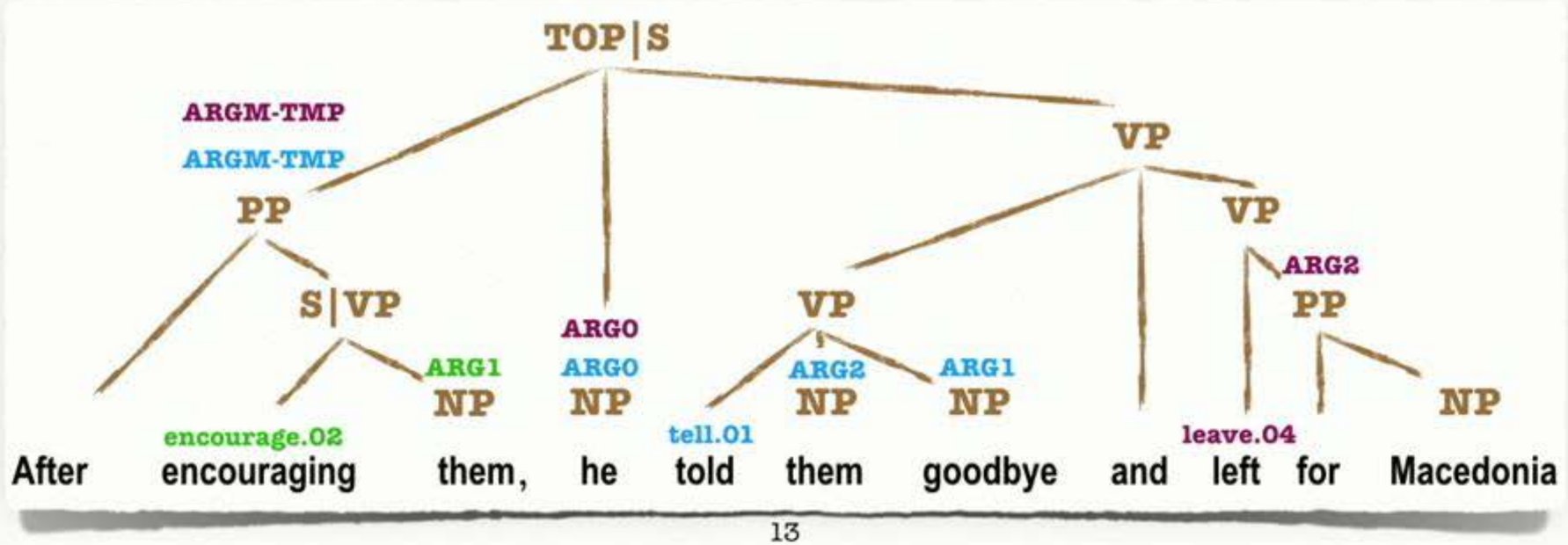
Syntactic Scaffolds



- Multitask setting.
- Shallow intermediate structure prediction.
- Learn (soft) syntax-aware representations, avoid cascaded errors.
- Not required during test.

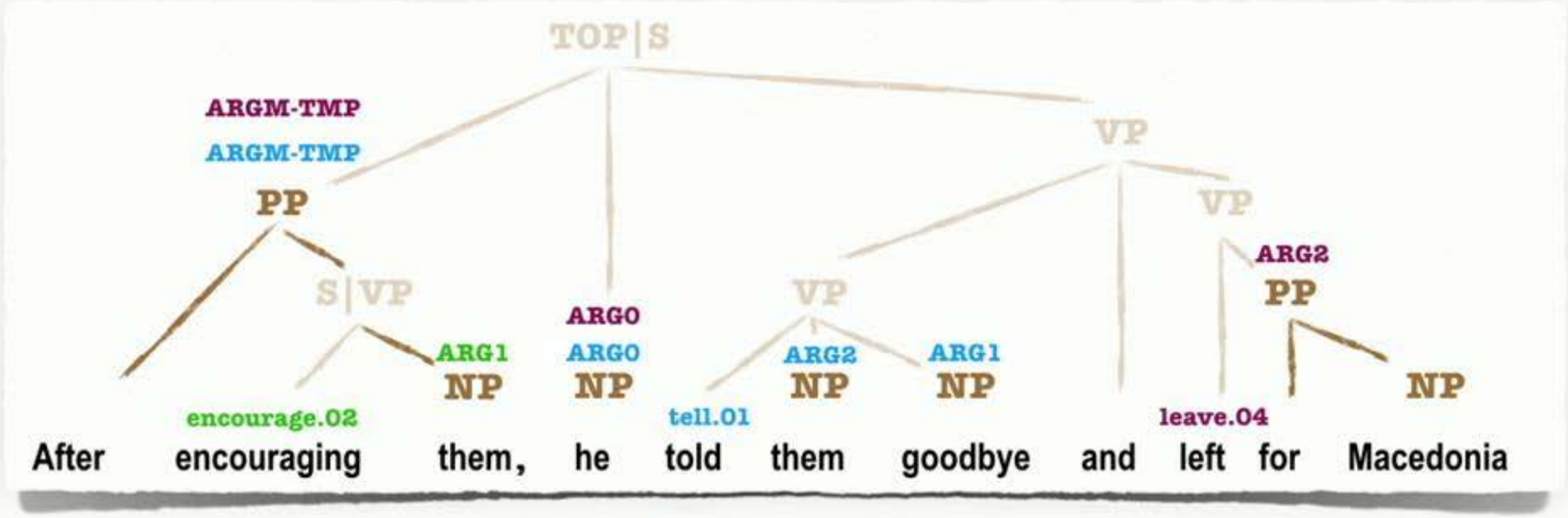
Task Overview





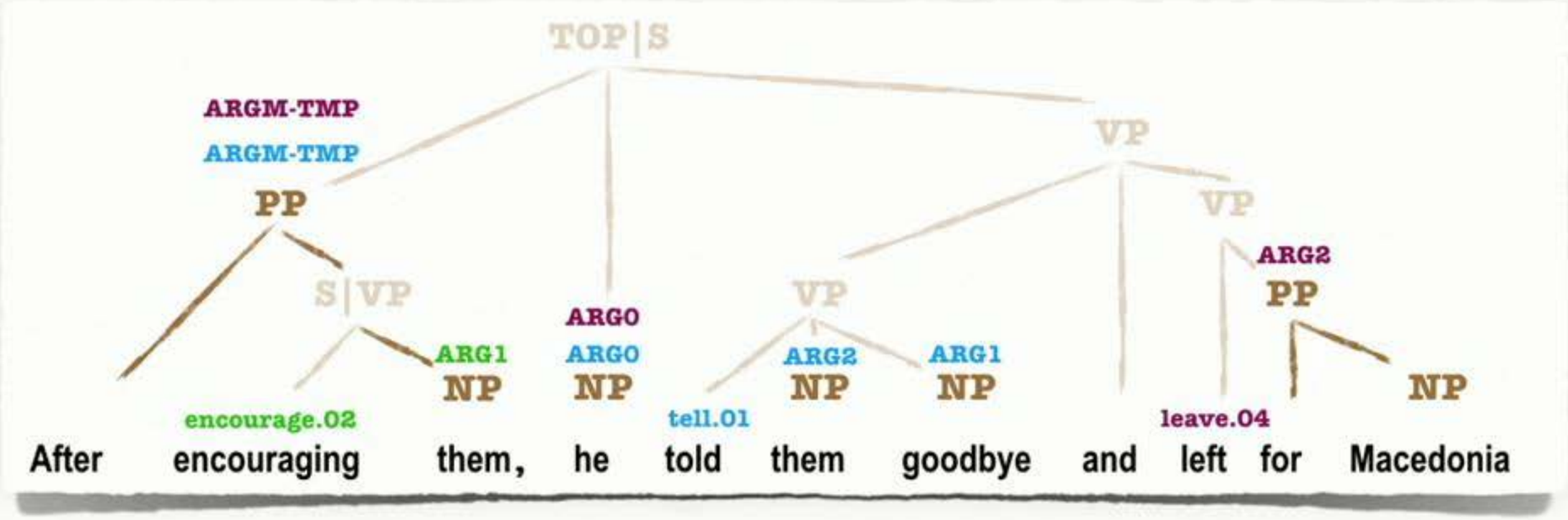
How to build a syntactic scaffold?

- **Desired** parts of syntactic tree:



How to build a syntactic scaffold?

- **Desired** parts of syntactic tree:



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

Training with syntactic scaffolds

x = Input
y = Output Structure
z = Scaffold Structure

$$\sum_{(\mathbf{x},\mathbf{y})\in\mathcal{D}_1} \mathcal{L}_1(\mathbf{x}, \mathbf{y}; \theta, \phi)$$

Primary Task Objective

Primary Dataset

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

Scaffold Task Objective

Scaffold Dataset

Training with syntactic scaffolds

x = Input
y = Output Structure
z = Scaffold Structure

Mixing Ratio

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

Primary Task Objective

Scaffold Task Objective

Primary Dataset

Scaffold Dataset

Training with syntactic scaffolds

x = Input
y = Output Structure
z = Scaffold Structure

Mixing Ratio

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

Primary Task Objective

Scaffold Task Objective

Primary Dataset

Scaffold Dataset

Shared parameters for input representations

Training with syntactic scaffolds

x = Input
y = Output Structure
z = Scaffold Structure

Mixing Ratio

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

Primary Task Objective

Scaffold Task Objective

Primary Dataset

Scaffold Dataset

Shared parameters for input representations

Alternate between primary and scaffold batches

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	

- 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 (**s**) of a sentence (**x**).

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

- Generalization of CRFs:
  - length of an input segment
  - in addition to its label.

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

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**).

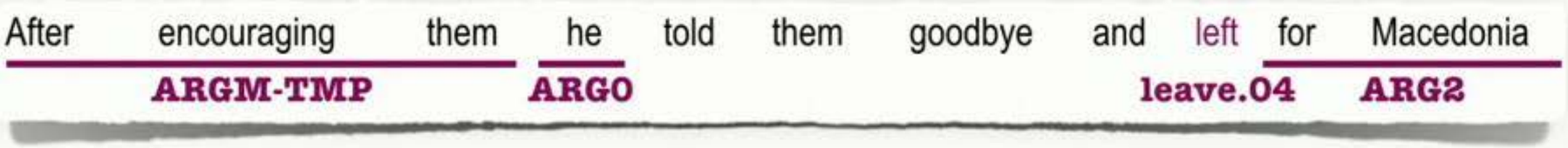
$$p(\mathbf{s} \mid \mathbf{x})$$

- Generalization of CRFs:
  - length of an input segment
  - in addition to its label.

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

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

Semi-Markov CRFs



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

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

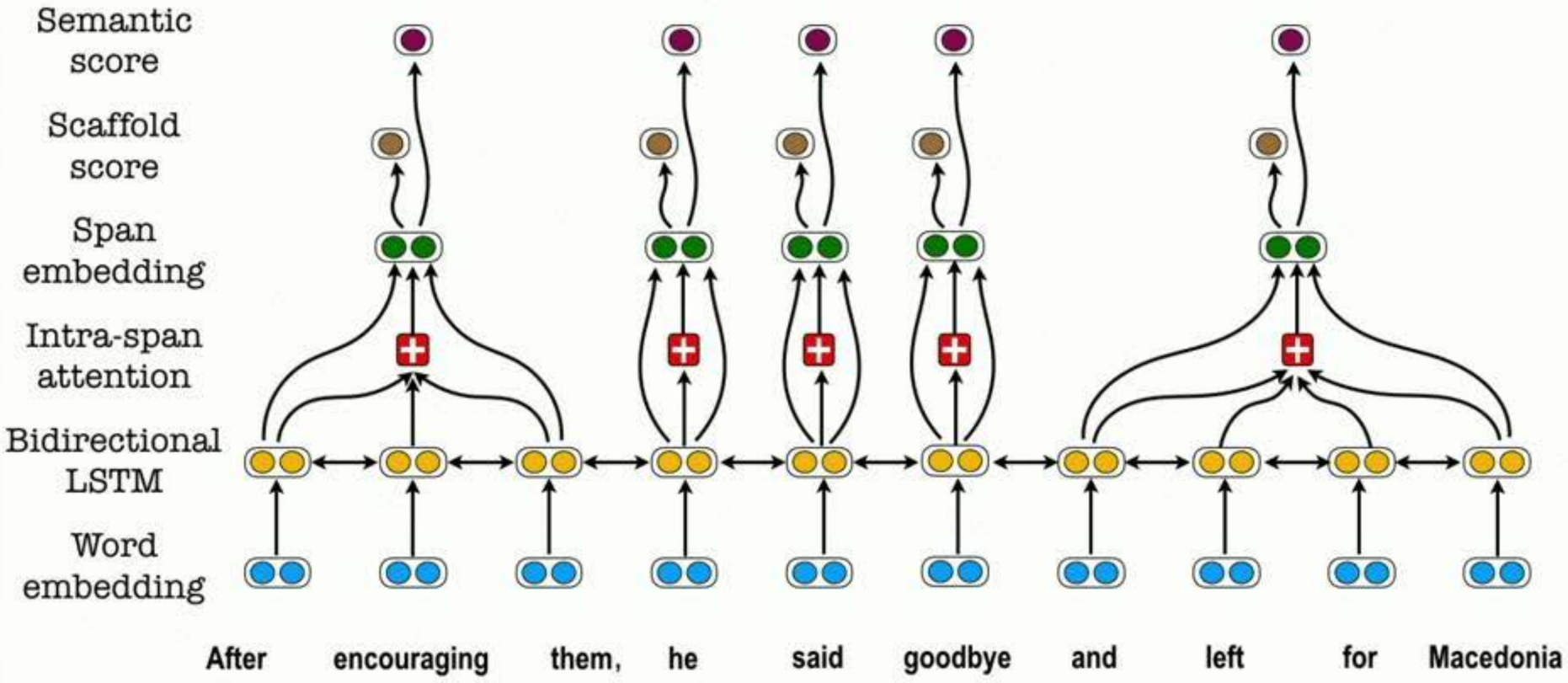
- Generalization of CRFs:
  - length of an input segment
  - in addition to its label.

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

- Training and inference given by O(ndl) dynamic programs, with a 0-order Markovian assumption.

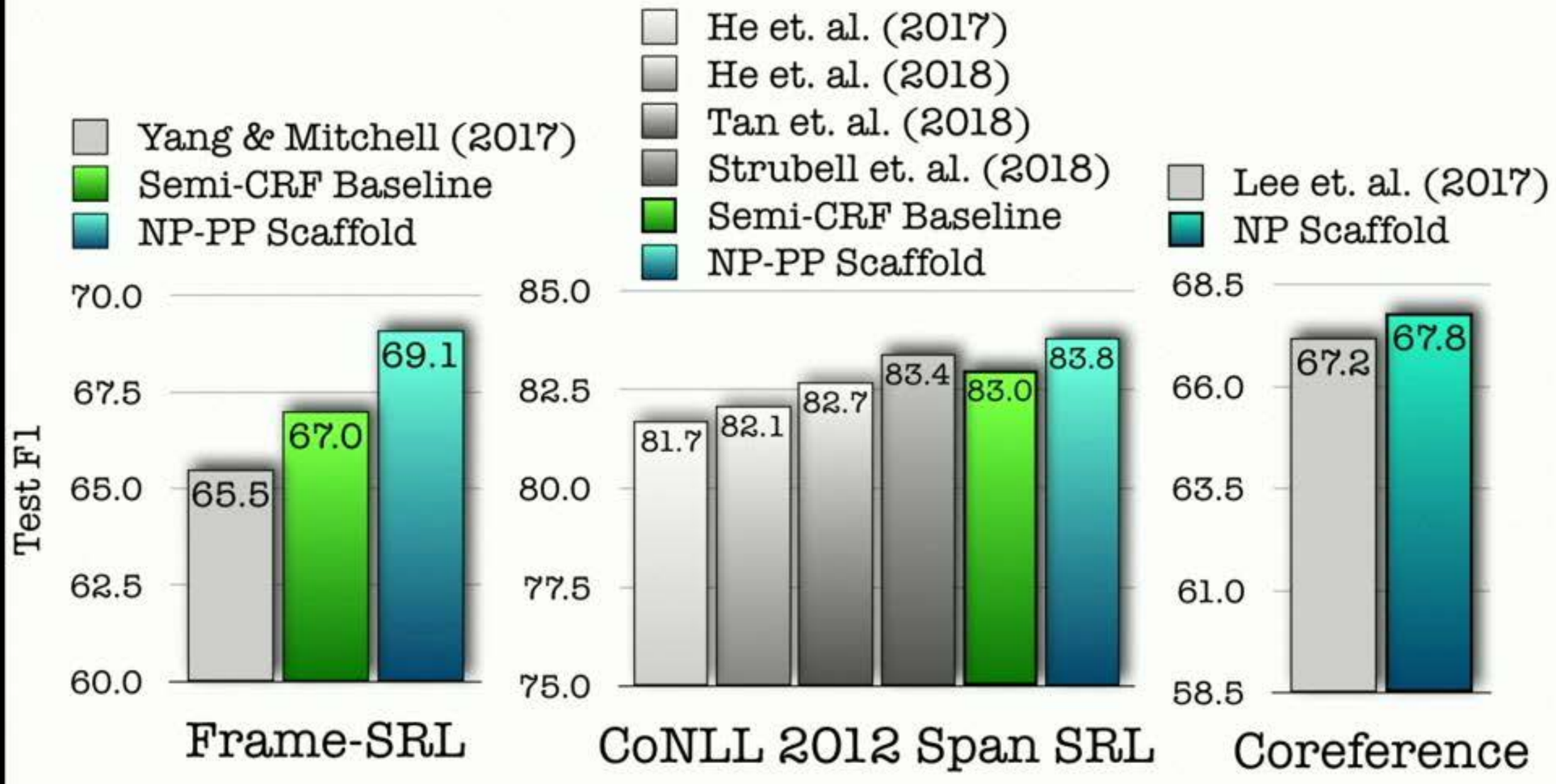
$$\Phi(\mathbf{x}, \mathbf{s}) = \sum_{k=1}^{m} \phi(s_k, x_{i_k:j_k})$$

Model architecture



- Learn scaffold score when syntactic annotations available.

Results


Yang & Mitchell (2017)
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
He et. al. (2017)
He et. al. (2018)
Tan et. al. (2018)
Strubell et. al. (2018)
Semi-CRF Baseline
NP-PP Scaffold
85.0
82.5
80.0
77.5
75.0
81.7
82.1
82.7
83.4
83.0
83.8
CoNLL 2012 Span SRL
Lee et. al. (2017)
NP Scaffold
68.5
66.0
63.5
61.0
58.5
67.2
67.8
Coreference

Recap: Challenge #1

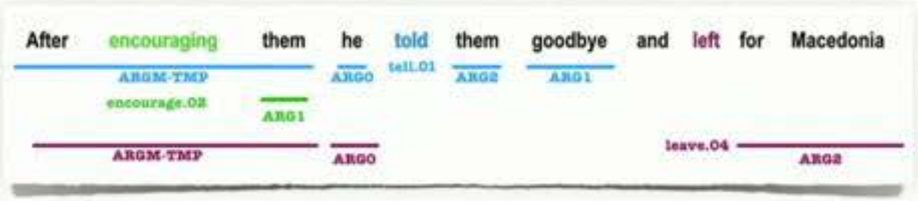
Can linguistic structure act as an informative prior for deep learning?

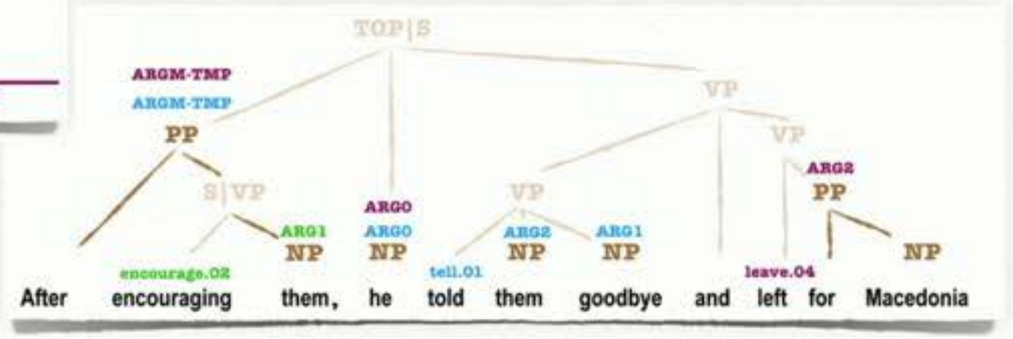
Recap: Challenge #1

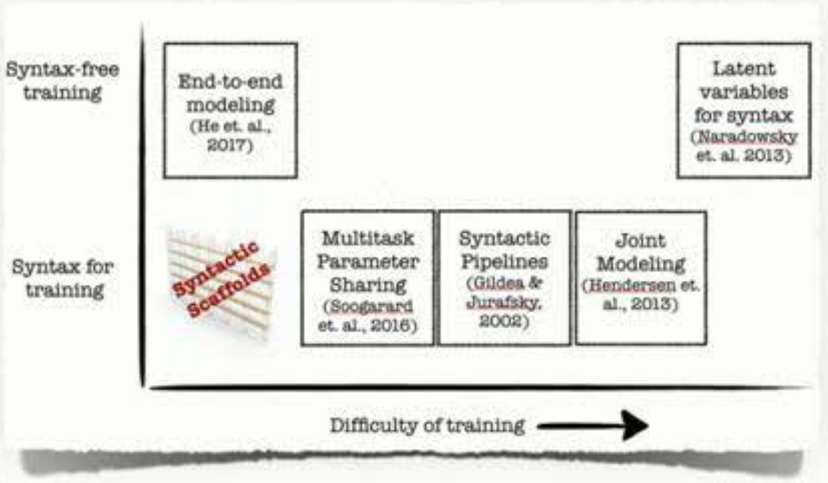
Can linguistic structure act as an informative prior for deep learning?

Recap: Challenge #1

Can linguistic structure act as an informative prior for deep learning?

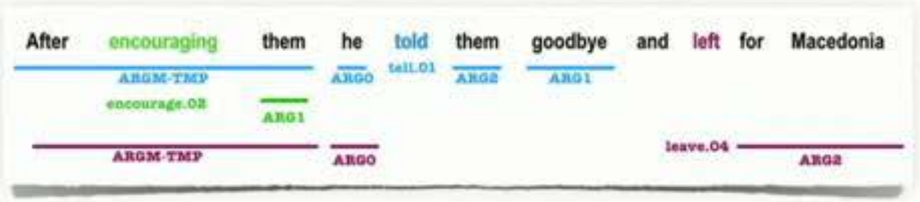


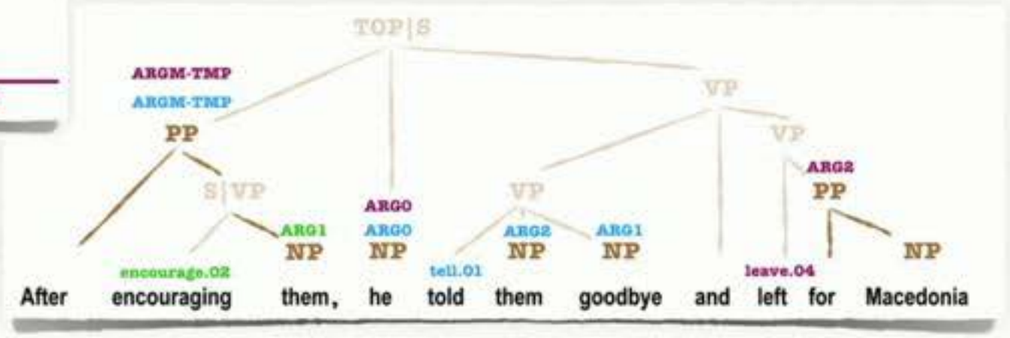


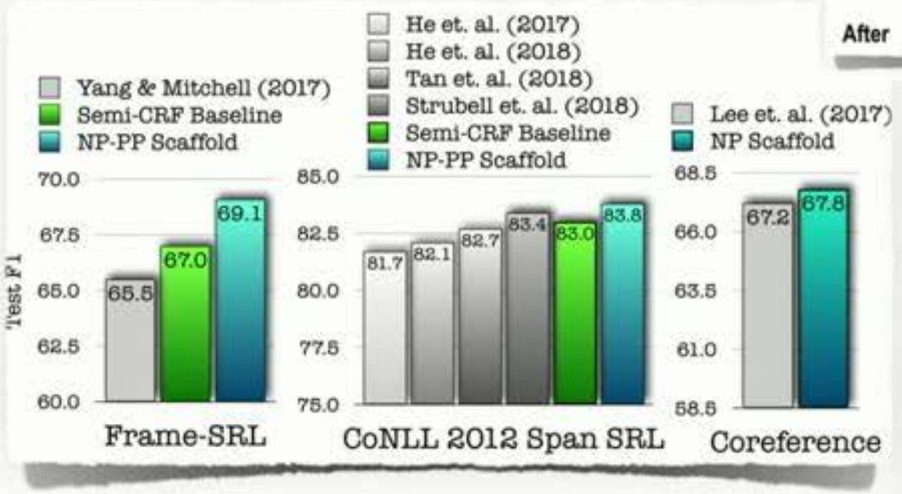


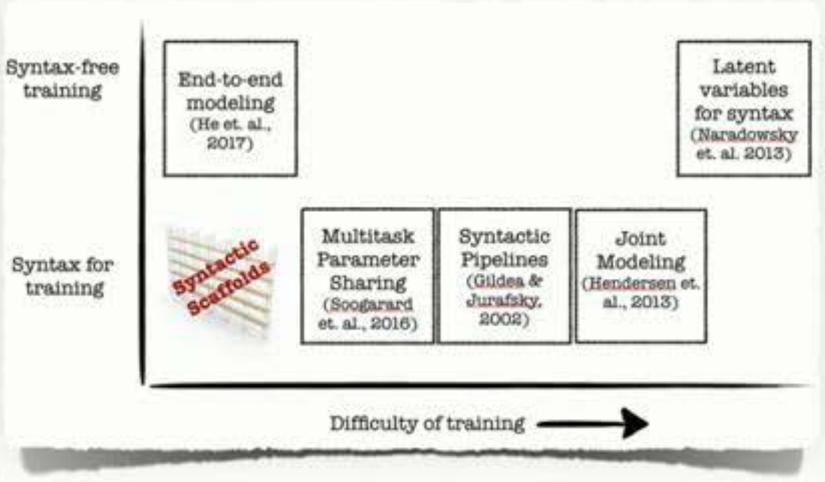
Recap: Challenge #1

Can linguistic structure act as an informative prior for deep learning?

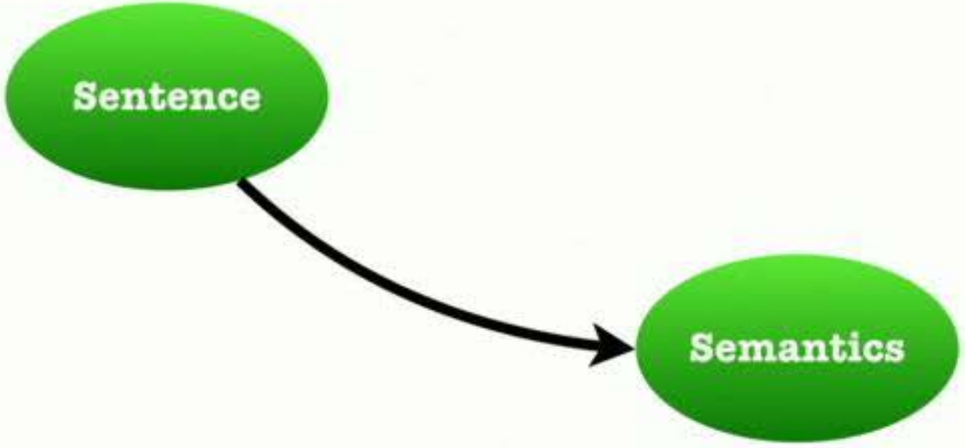




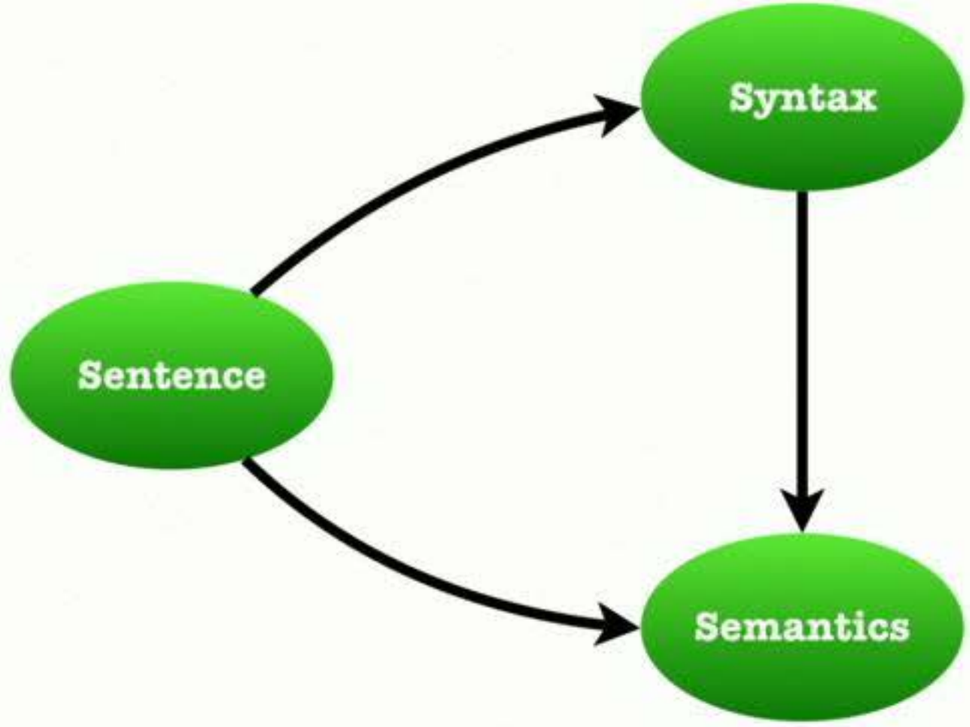




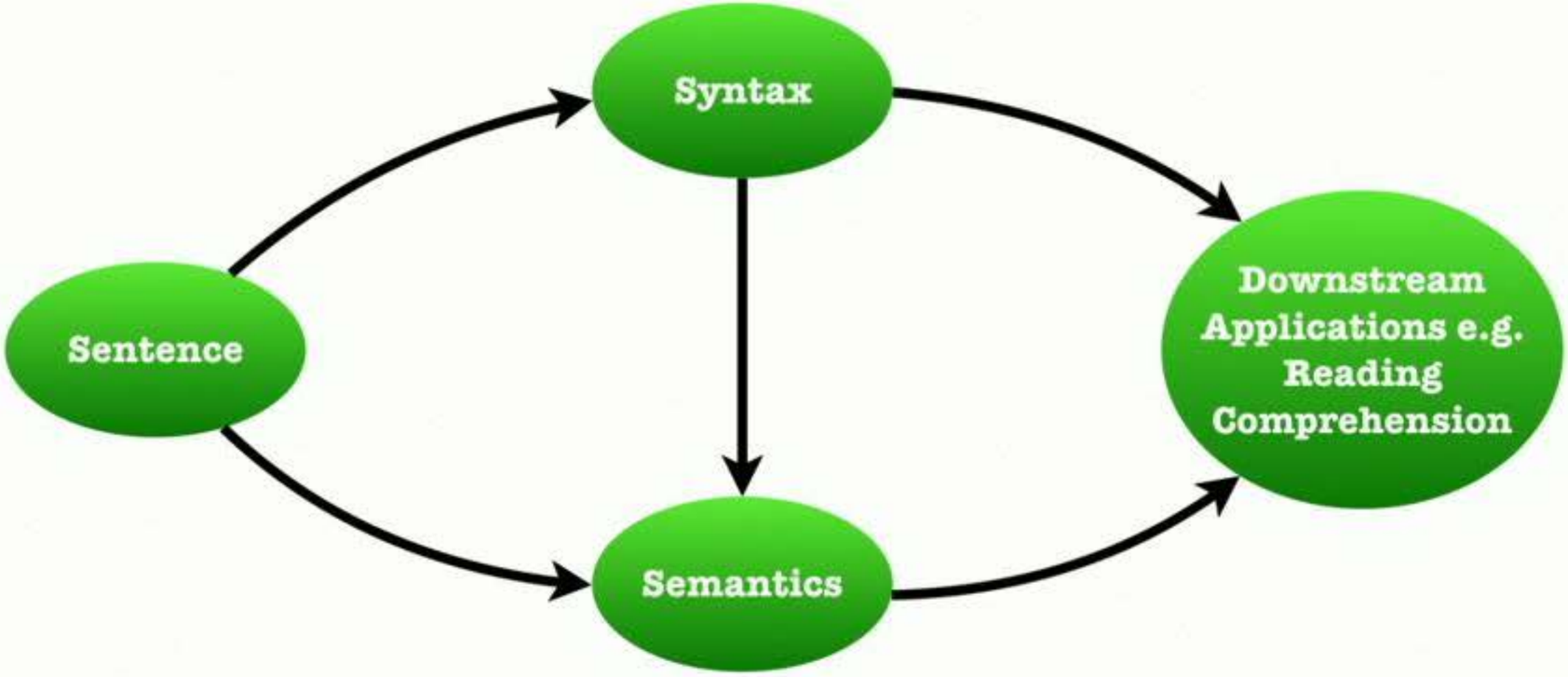
Looking ahead: Predicted Structure



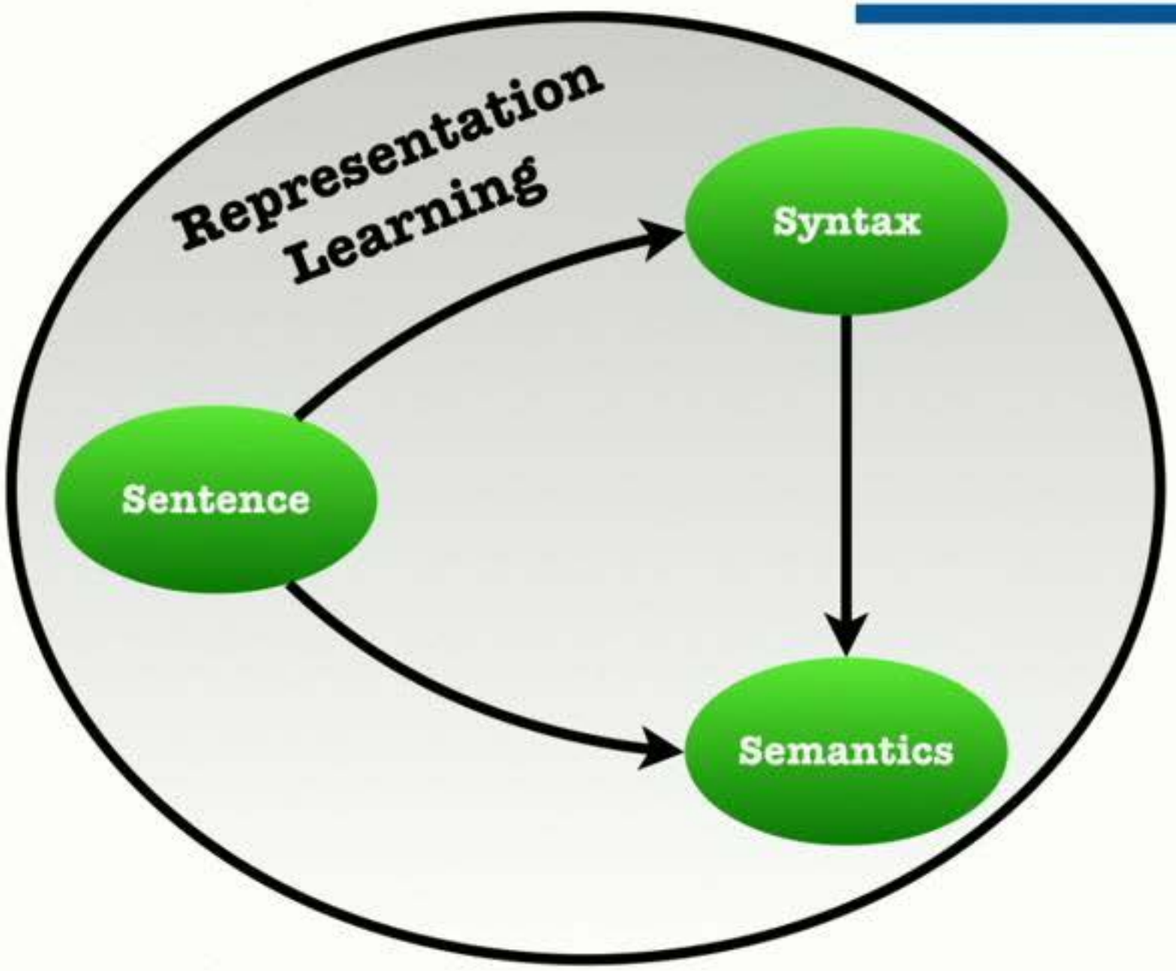
Looking ahead: Predicted Structure



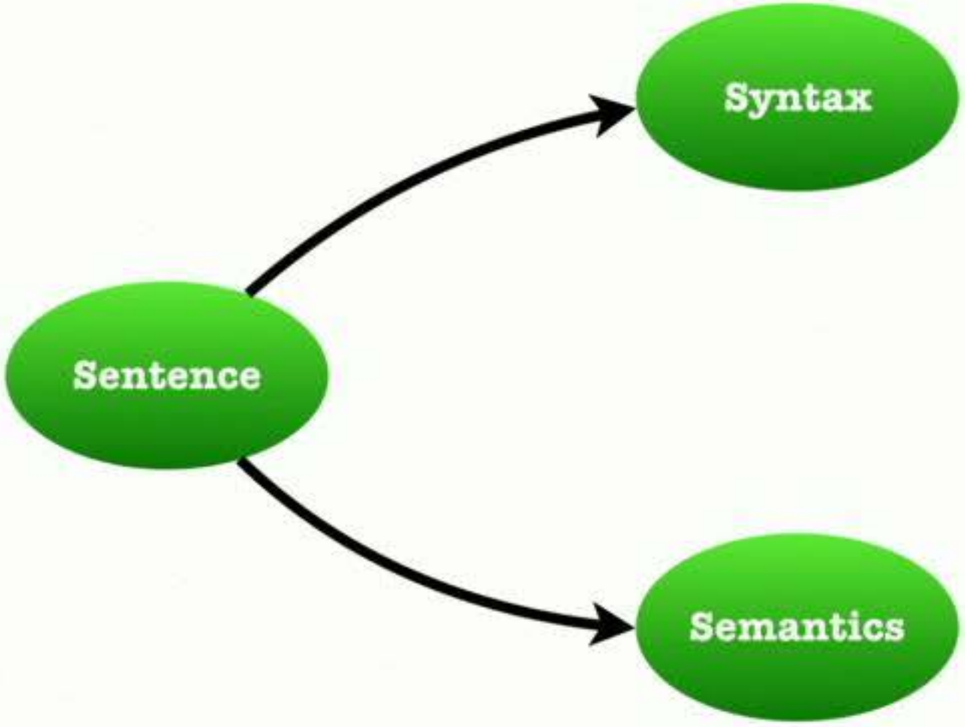
Looking ahead: Predicted Structure



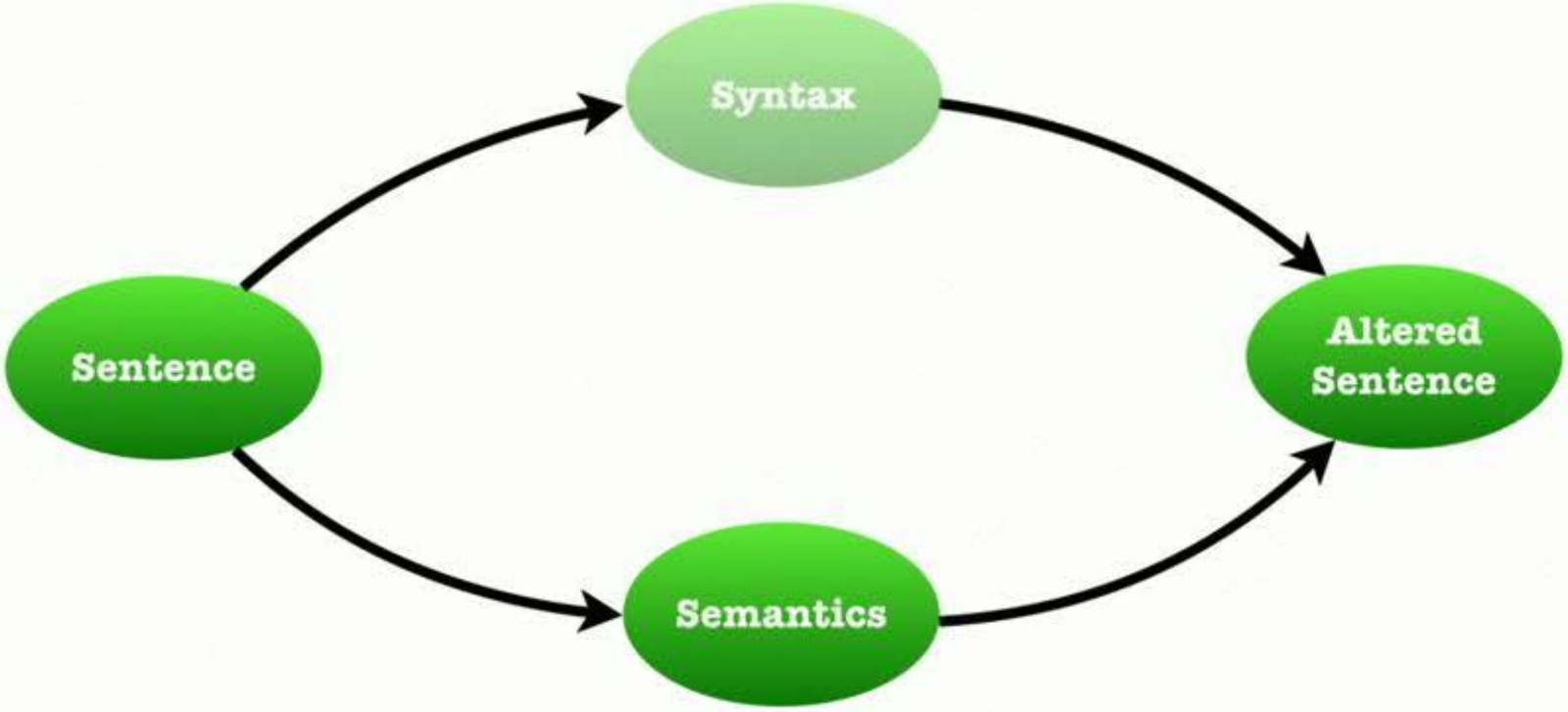
Looking ahead: Predicted Structure



Looking ahead: Controlled Generation

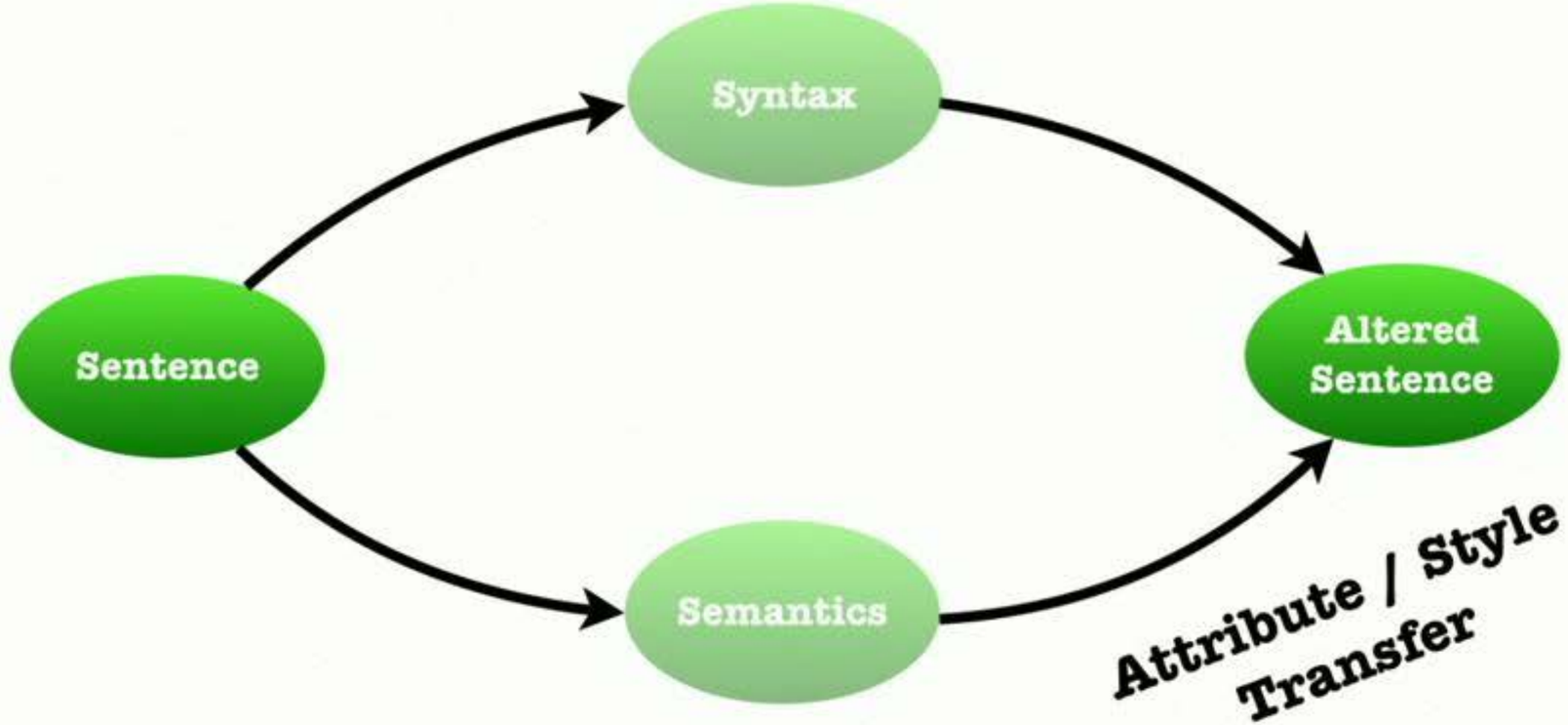


Looking ahead: Controlled Generation



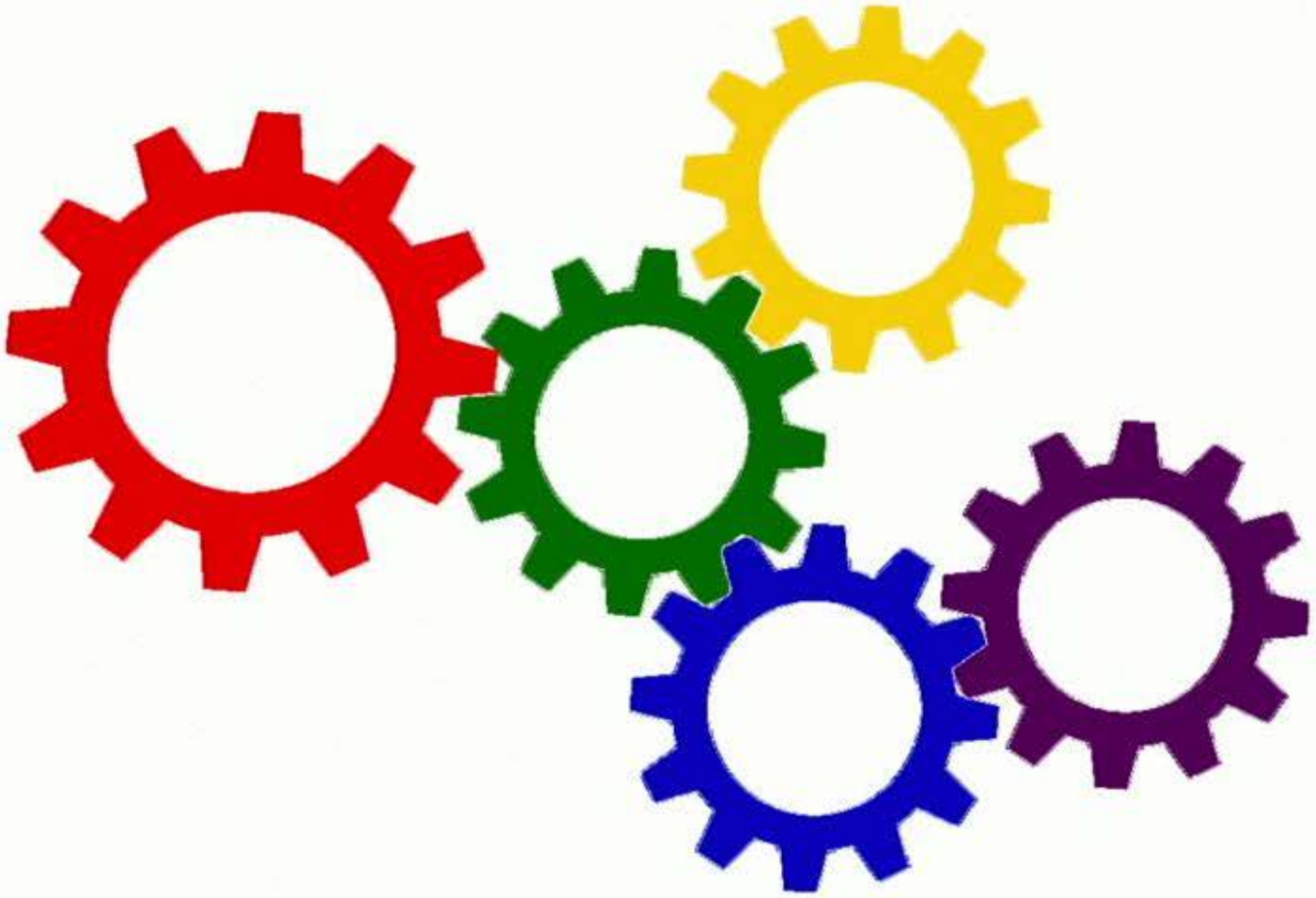
Iyyer et. al. (NAACL 2018)

Looking ahead: Controlled Generation



Iyyer et. al. (NAACL 2018)

Part II



Recap: BERT's confusion

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 (EMNLP 17)

Percy Liang (AI Frontiers 18)

BERT (Devlin et. al., arXiv 2018)

Challenges

Part I

Can linguistic structure act as an informative prior for deep learning?

- ☑ 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.

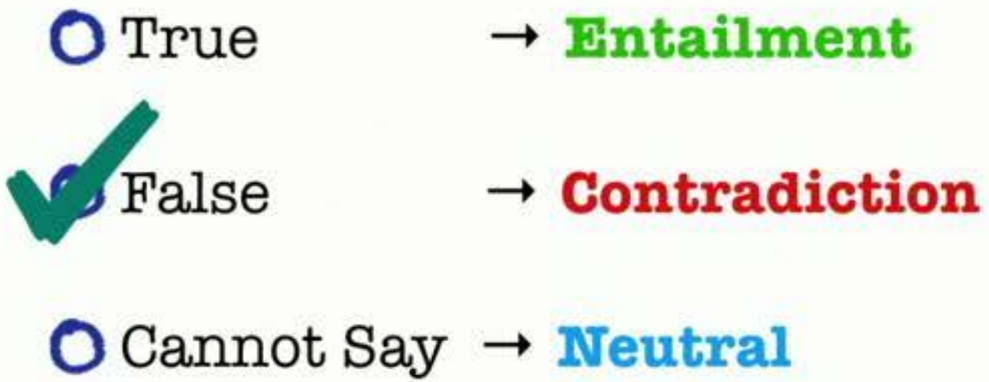
- True → **Entailment**
- False → **Contradiction**
- Cannot Say → **Neutral**

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

Two dogs are running through a field.

Premise

NLI Datasets

Two dogs are running through a field.

Premise

Entailment

There are animals outdoors.

NLI Datasets

Two dogs are running through a field.

Premise

Entailment → There are animals outdoors.

Neutral → Some puppies are running to catch a stick.

NLI Datasets

Two dogs are running through a field.

Premise

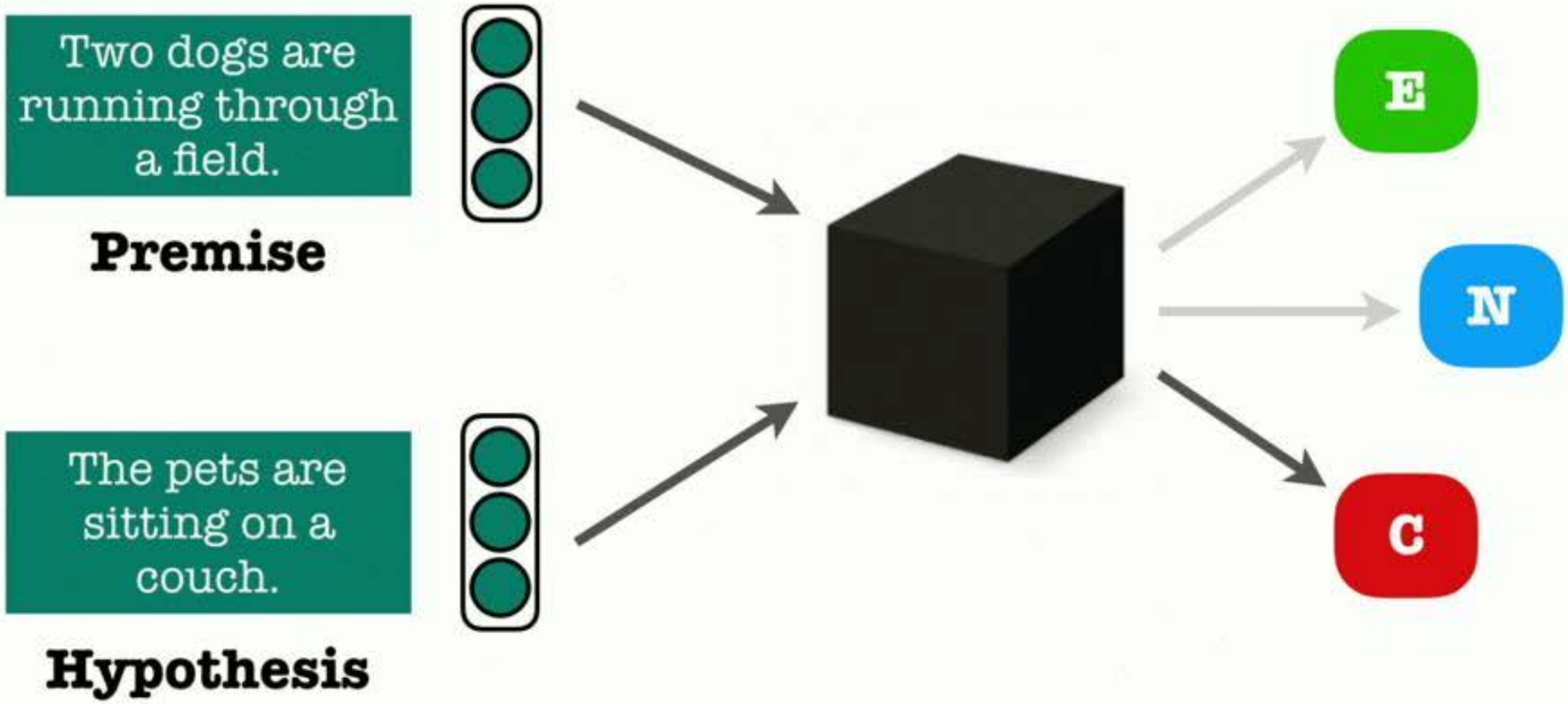
Entailment → There are animals outdoors.

Neutral → Some puppies are running to catch a stick.

Contradiction → The pets are sitting on a couch.

SNLI (Bowman et. al, 2015) 570 K
MultiNLI (Williams et. al., 2017) 433 K

NLI as Text Classification

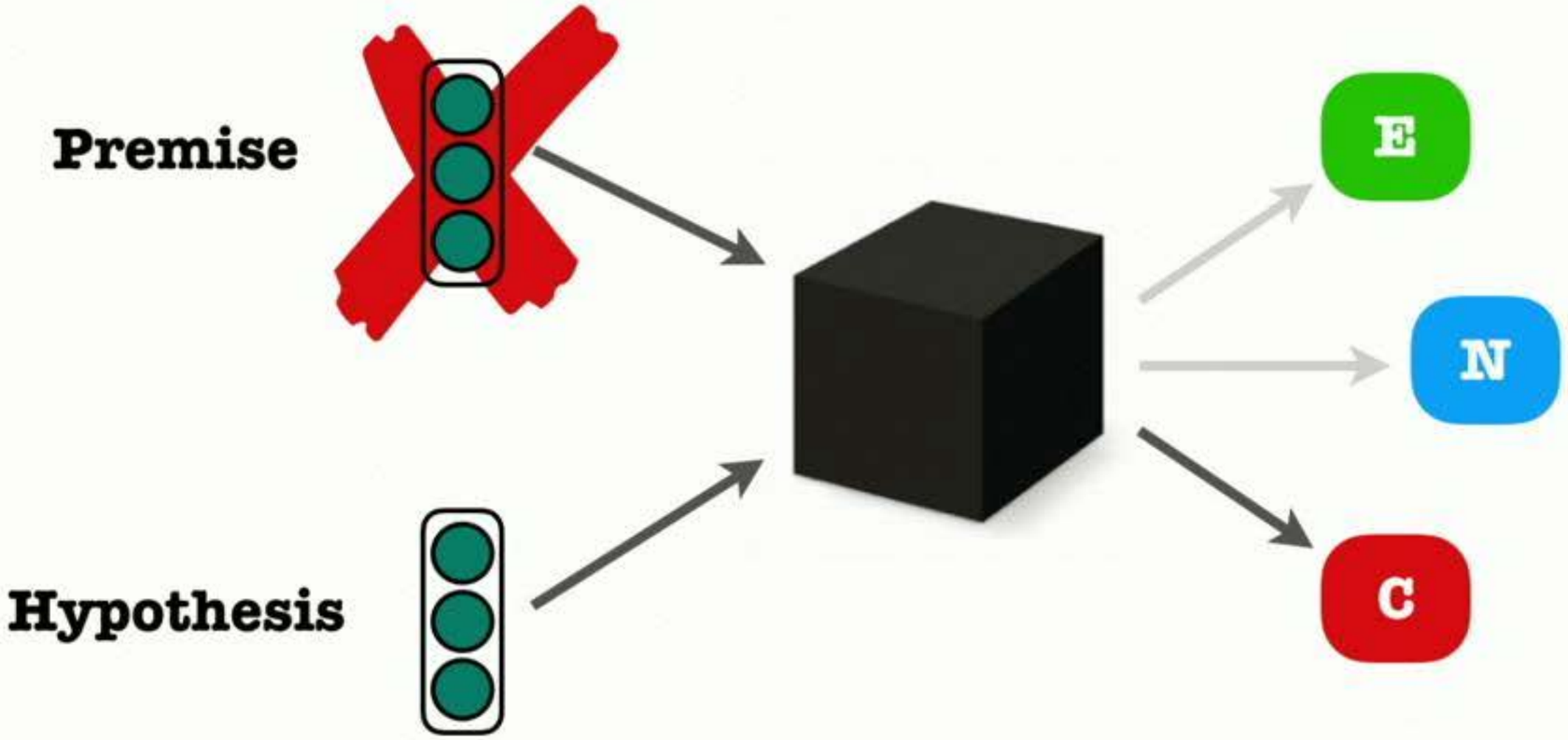


Lots of progress

Publication	Model	Parameters	Train (% acc)	Test (% acc)
Feature-based models				
Bowman et al. '15	Unlexicalized features		49.4	50.4
Bowman et al. '15	+ Unigram and bigram features		99.7	*78.2*
	⋮			
Peters et al. '18	ESIM + ELMo	8.0m	91.6	88.7
Boyuan Pan et al. '18	300D DMAN	9.2m	95.4	88.8
Zhiguo Wang et al. '17	BiMPM **Ensemble**	6.4m	93.2	88.8
Yichen Gong et al. '17	448D Densely Interactive Inference Network (DIIN, code) **Ensemble**	17m	92.3	88.9
Seonhoon Kim et al. '18	Densely-Connected Recurrent and Co-Attentive Network	6.7m	93.1	88.9
Zhuosheng Zhang et al. '18	SLRC	6.1m	89.1	89.1
Qian Chen et al. '17	KIM **Ensemble**	43m	93.6	89.1
Ghaeini et al. '18	450D DR-BiLSTM **Ensemble**	45m	94.8	89.3
Peters et al. '18	ESIM + ELMo **Ensemble**	40m	92.1	89.3
Yi Tay et al. '18	300D CAFE **Ensemble**	17.5m	92.5	89.3
Chuanqi Tan et al. '18	150D Multiway Attention Network **Ensemble**	58m	95.5	89.4
Boyuan Pan et al. '18	300D DMAN **Ensemble**	79m	96.1	89.6
Radford et al. '18	Fine-Tuned LM-Pretrained Transformer	85m	96.6	**89.9**
Seonhoon Kim et al. '18	Densely-Connected Recurrent and Co-Attentive Network **Ensemble**	53.3m	95.0	**90.1**

A simple experiment

A simple experiment



fastText (Joulin et. al. 2017)

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.

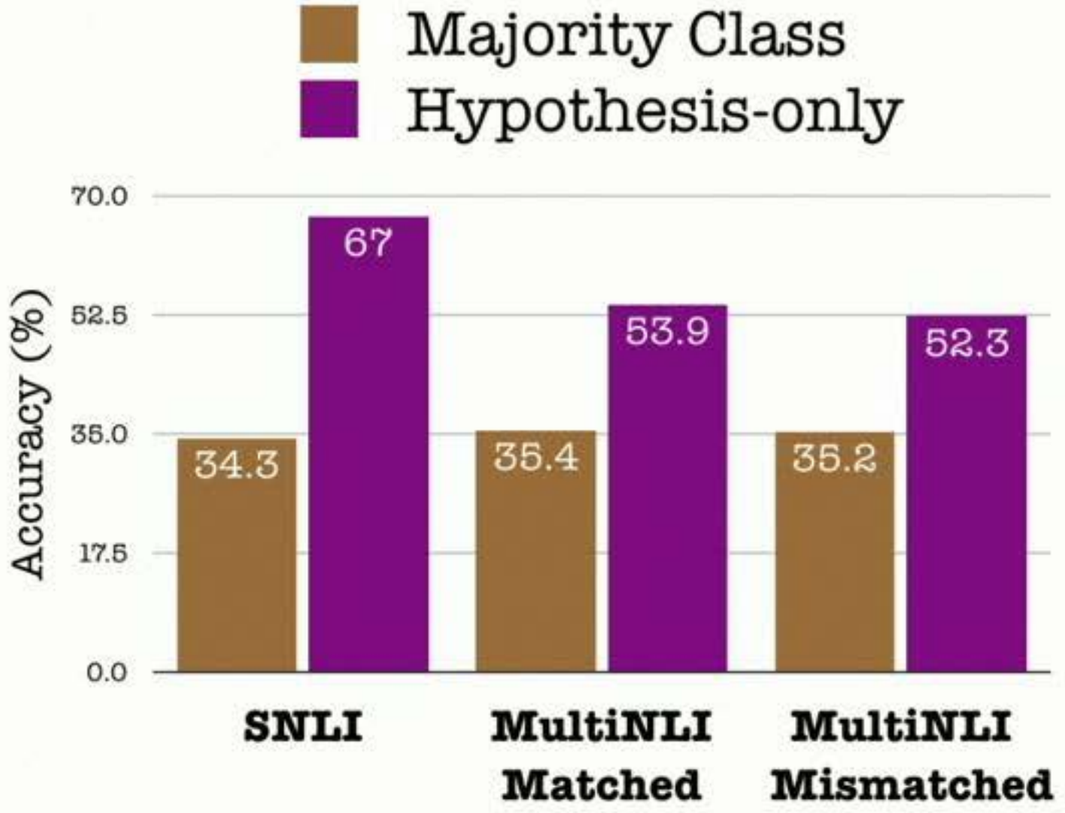
A simple experiment

- 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!

Entailment Artifacts

Entailment Artifacts

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

Premise

Generalization

People play frisbee **outdoors**.

Entailment

Entailment Artifacts

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

Premise

Generalization

People play frisbee **outdoors**.

Entailment

A person in a red ~~**shirt**~~ is mowing the grass with a ~~**green**~~ riding mower.

Premise

Shortening

A person in red is cutting the grass on a riding mower.

Entailment

Neutral Artifacts

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

Neutral Artifacts

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

A group of female athletes are huddled together and excited.

Premise

Purpose Clauses

They are huddled together **because** they are working together.

Neutral

Contradiction Artifacts

Older man with white hair and a red cap painting the golden gate bridge on the shore with the golden gate bridge in the distance.

Premise

Negation

Nobody wears a cap.

Contradiction

Contradiction Artifacts

Older man with white hair and a red cap painting the golden gate bridge on the shore with the golden gate bridge in the distance.

Premise

Negation

Nobody wears a cap.

Contradiction

Three dogs racing on racetrack.

Premise

Cats!

Three **cats** race on a track.

Contradiction

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

Two dogs are running through a field.

Premise

Entailment

There are **animals** outdoors.

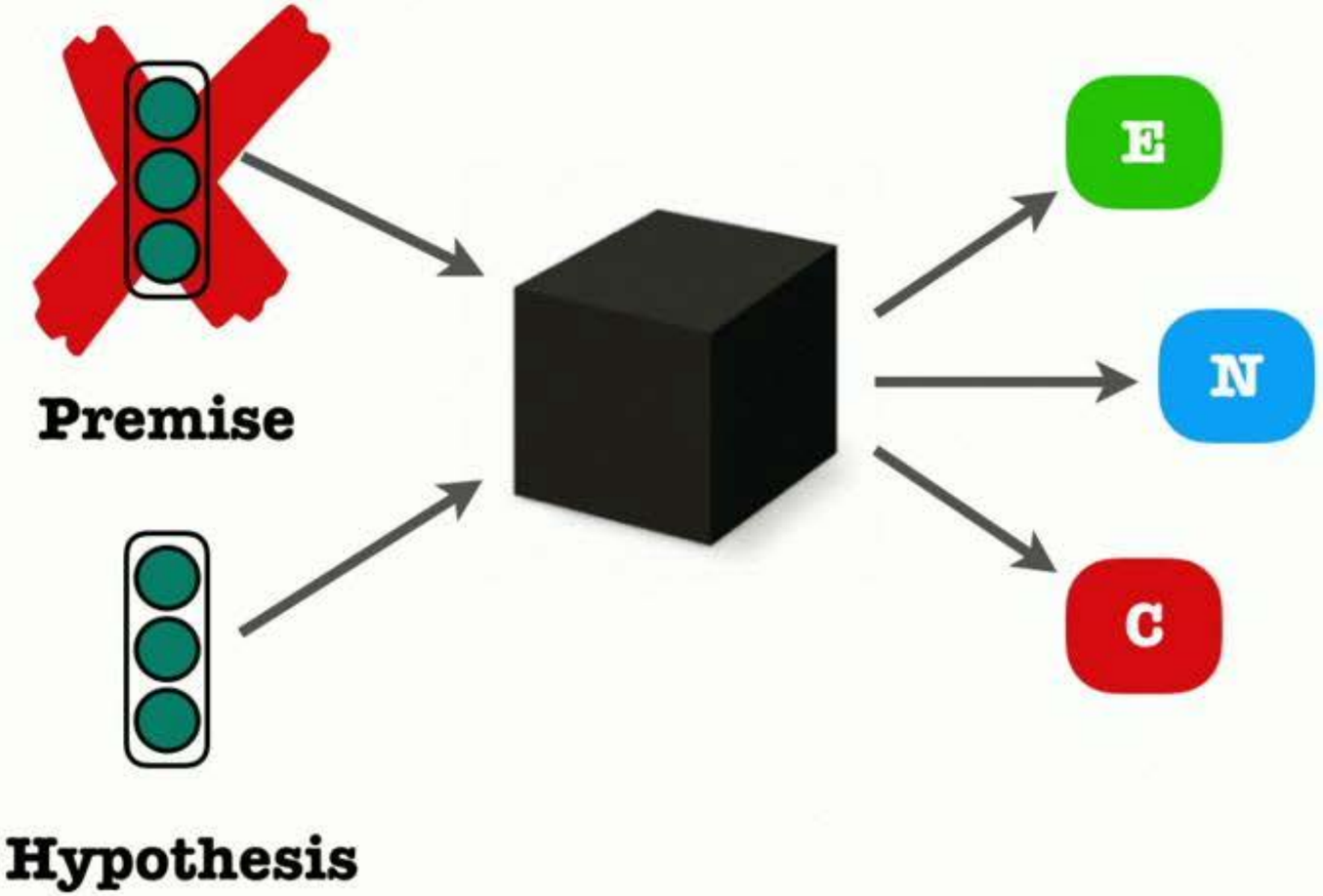
Neutral

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

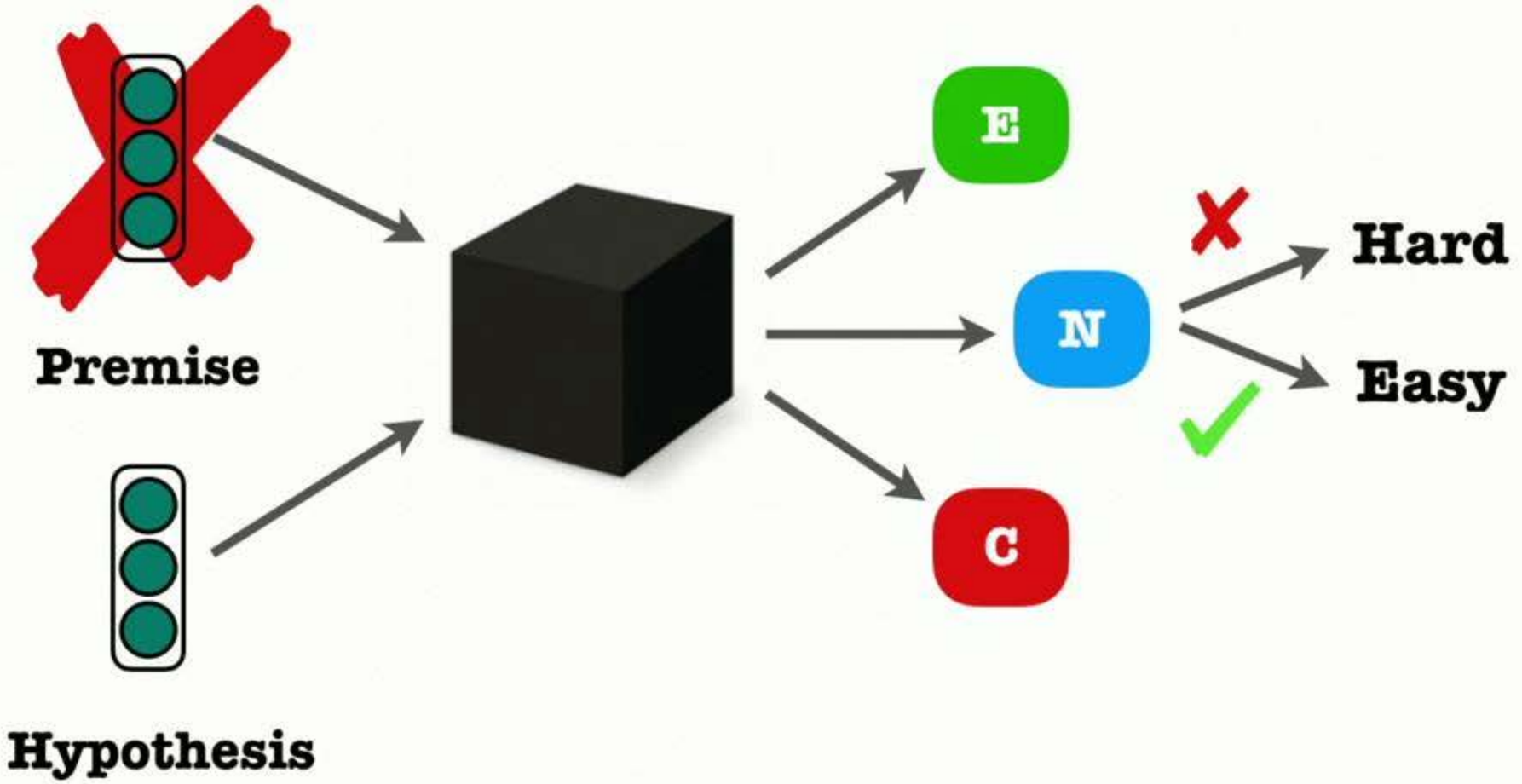
Contradiction

The pets are sitting on a couch.

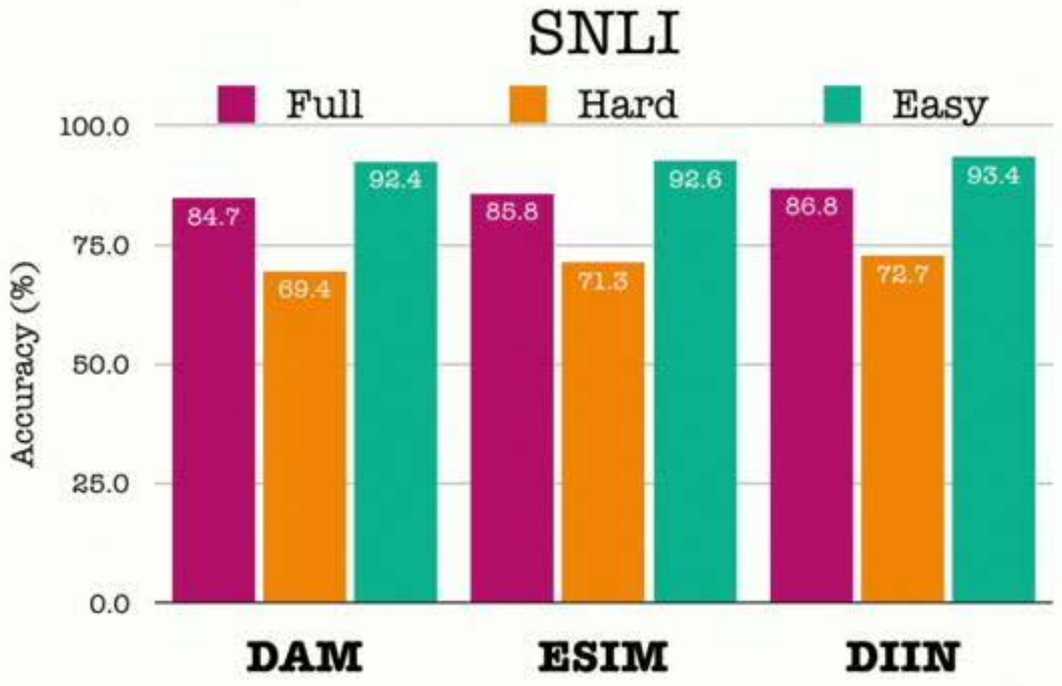
Can we filter out examples with artifacts?



Can we filter out examples with artifacts?

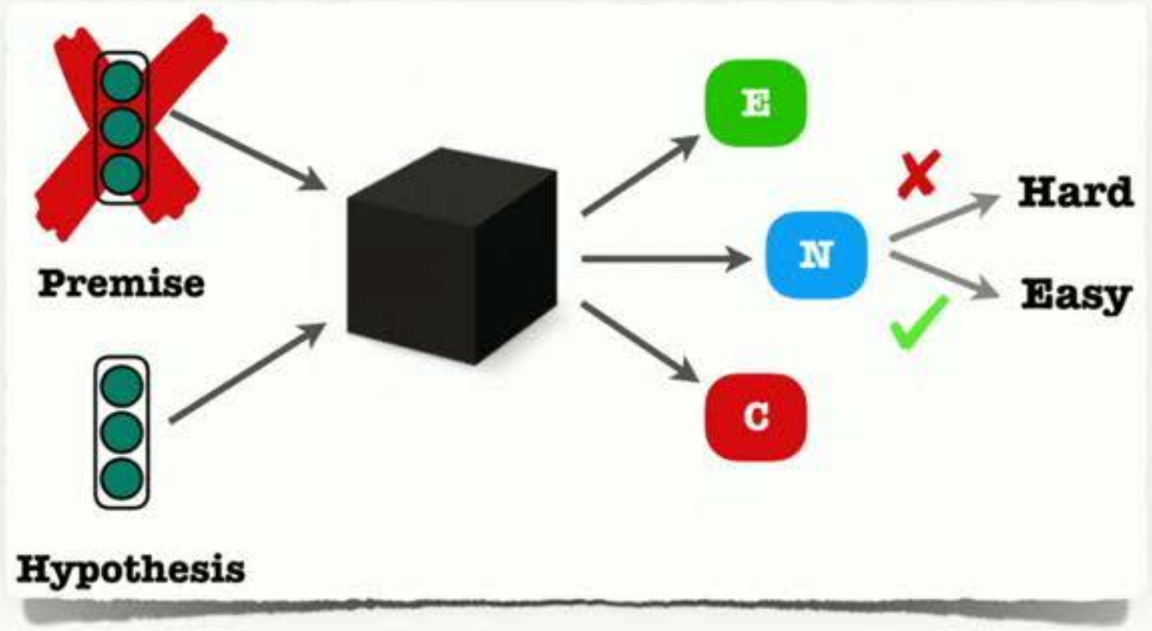


What makes NLI models succeed?

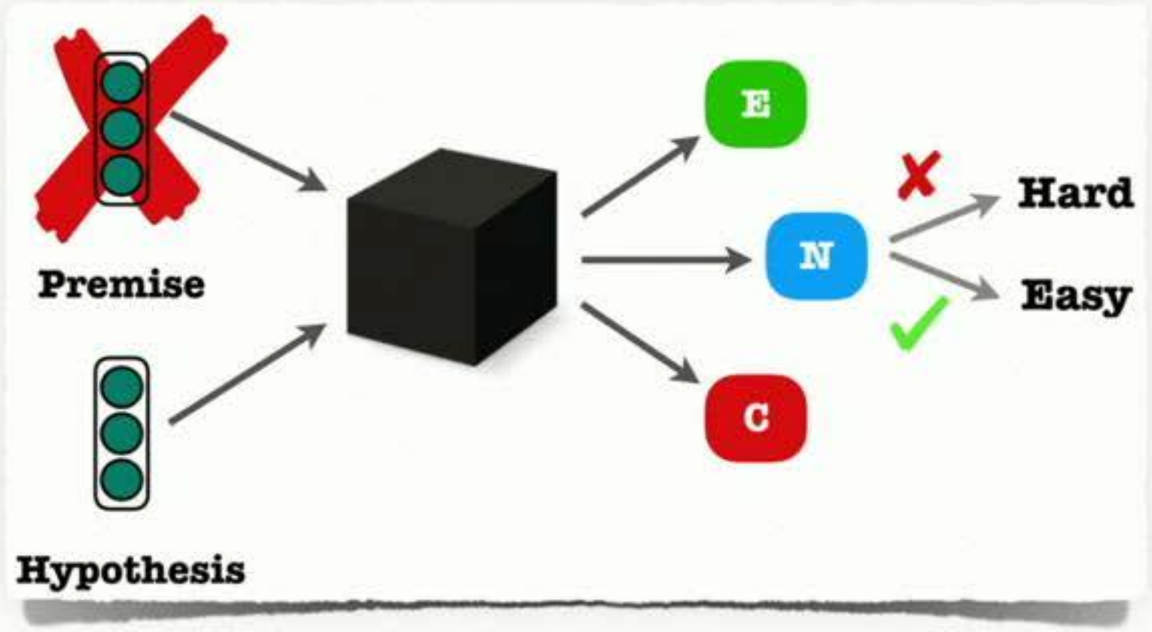


DAM - Decomposable Attention Model (Parikh et. al. 2016)
ESIM - Enhanced Sequential Inference Model (Chen et. al., 2017)
DIIN - Densely Interactive Inference Network (Gong et. al. 2018)

Can we filter out examples with artifacts?

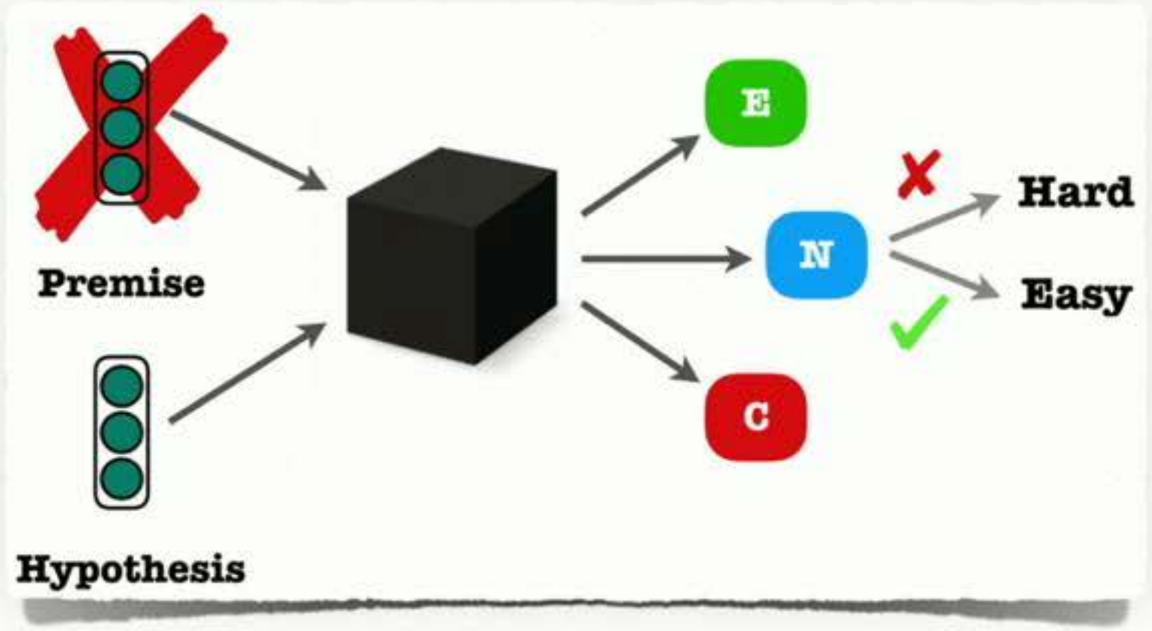


Can we filter out examples with artifacts?



- Hard examples exhibit their own artifacts!

Can we filter out examples with artifacts?



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

Not unique to NLI...

- CNN/DailyMail [Chen et al., 2017]
- Visual QA [Jabri et al., 2016]
- SQuAD [Jia and Liang et al., 2017]
- RoC Story [Schwartz et al., 2017, Cai et al., 2017]
- Recognizing Textual Entailment [Levy et al., 2015]

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 hard it could be (Beygelzimer et. al., 2015, Coleman et. al., 2018).

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 hard it could be (Beygelzimer et. al., 2015, Coleman et. al., 2018).

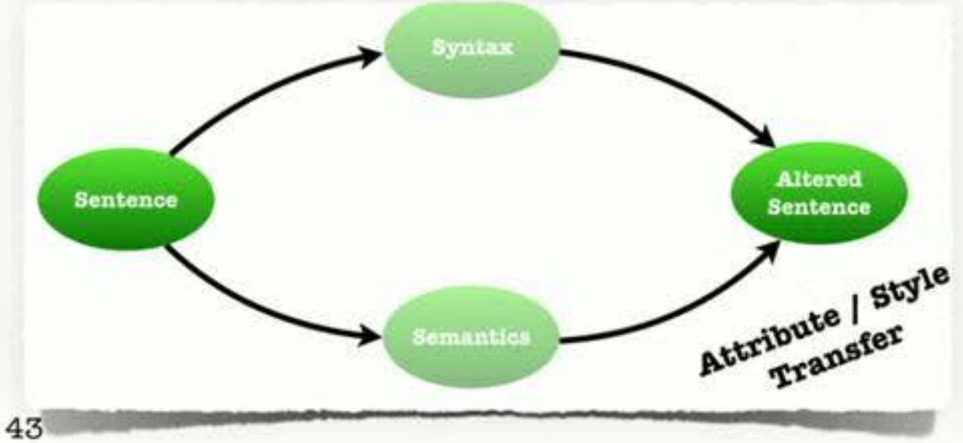
Easy  **Hard**

Looking Ahead: Improved Data Collection

- Annotation Instruction: Avoid simple heuristics!
- Real-time rewards and batch reviews of annotated examples

Looking Ahead: Improved Data Collection

- Annotation Instruction: Avoid simple heuristics!
- Real-time rewards and batch reviews of annotated examples
- Alternatives to human elicitation for building datasets?



Challenges

Part I

Can linguistic structure act as an informative prior for deep learning?

- ☑ 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)

Take home message

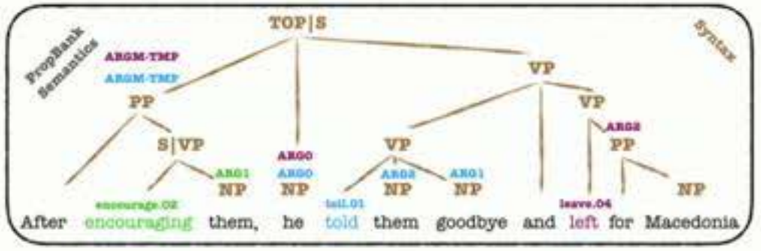
The New York Times

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.

Take home message

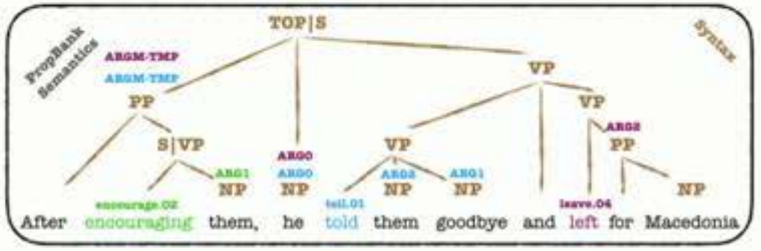
Structural inductive biases for better language understanding.



Predicted structure can help representation learning, generation.

Take home message

Structural inductive biases for better language understanding.



Predicted structure can help representation learning, generation.

Imperative to look closely at what models learn.

Three dogs racing on racetrack.

Premise

Cats!

Three cats race on a track.

Contradiction

Robust annotation, and models robust to artifacts

Thanks!

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

swabhs

swabhz