### Large Scale QA-SRL Parsing

Nicholas FitzGerald, Julian Michael, Luheng He, and Luke Zettlemoyer

#### **ACL 2018**









http://qasrl.org/





**Subject Manner Verb Object Time** 

John surreptitiously ate the burrito at 2am.

Subject Manner Verb Object Time

John surreptitiously ate the burrito at 2am.

Applied to improve state-of-the-art in NLP tasks such as Question Answering
[Shen 2007] and Machine Translation [Liu and Gildea, 2010]

Subject Manner Verb Object Time

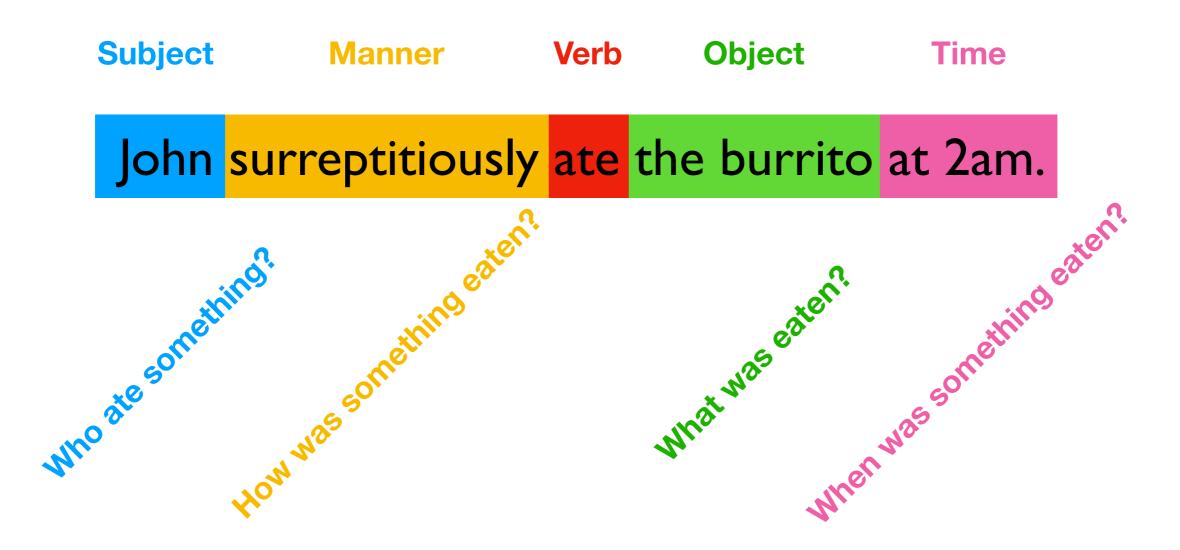
John surreptitiously ate the burrito at 2am.

- Applied to improve state-of-the-art in NLP tasks such as Question Answering
  [Shen 2007] and Machine Translation [Liu and Gildea, 2010]
- Commonly used interface to facilitate Data Exploration and Information Extraction [Stanovsky et al 2018] [Chiticariu et al. 2018]

Subject Manner Verb Object Time

John surreptitiously ate the burrito at 2am.

- Applied to improve state-of-the-art in NLP tasks such as Question Answering
  [Shen 2007] and Machine Translation [Liu and Gildea, 2010]
- Commonly used interface to facilitate Data Exploration and Information Extraction [Stanovsky et al 2018] [Chiticariu et al. 2018]
- Considerable interest in general-purpose SRL parsers



**Time Subject Manner Object Verb** John surreptitiously ate the burrito at 2am. **QA-SRL 1.0 Small dataset** 

Only explored sub-problems

[He et al. 2015]

**Trained annotators** 

# Goal

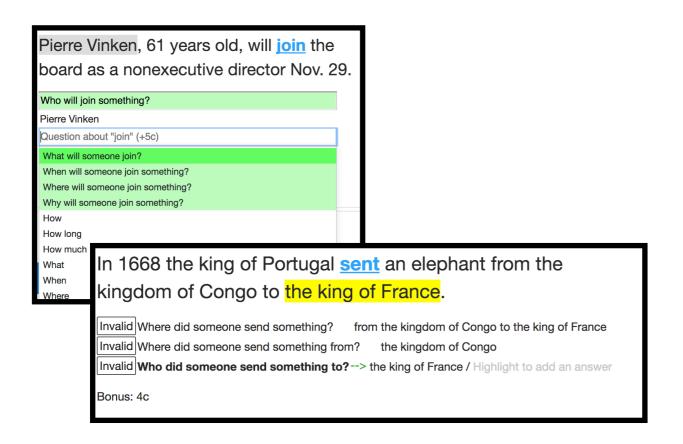
# A high-quality, large-scale parser for QA-SRL

# In 1950 Alan M. Turing *published* "Computing machinery and intelligence" in Mind, in which he *proposed* that machines could be *tested* for intelligence *using* questions and answers.

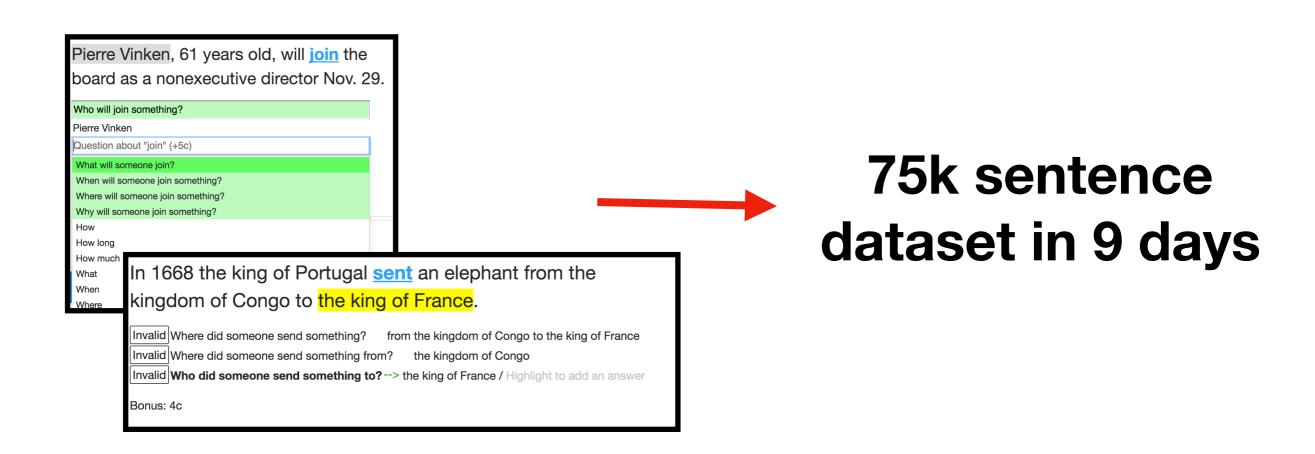
	When was something published?	In 1950
	Who published something?	Alan M. Turing
published	What was published?	"Computing Machinery and Intelligence"
	Where was something published?	in Mind
	When did someone propose something?	In 1950
	Who proposed something?	Alan M. Turing
proposed	What did someone propose?	that machines could be tested for intelligent using questions and answers
	What did someone propose something in?	"Computing Machinery and Intelligence"
	What can be tested?	machines
tested	What can something be tested for?	intelligence
	How can something be tested?	using questions and answers
	What was being used?	questions and answers
using	Why was something being used?	tested for intelligence

1. Scale up QA-SRL data annotation

1. Scale up QA-SRL data annotation

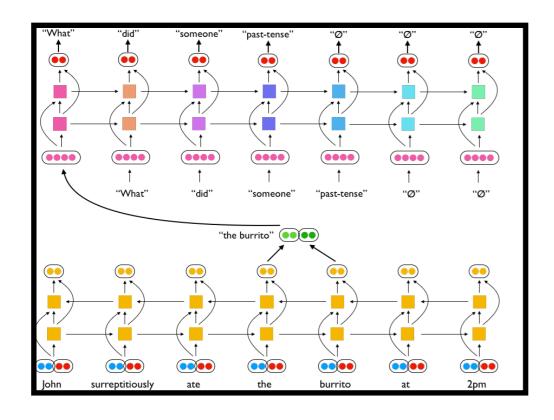


1. Scale up QA-SRL data annotation



#### 1. Scale up QA-SRL data annotation

#### 2. Train a QA-SRL Parser



In the video, the perpetrators never appeared to look at the camera.

Some of the vegetarians he met were members of the Theosophical Society, which had been founded in 1875 to further universal brotherhood, and which was devoted to the study of Buddhist and Hindu literature. met

founded

	What produced something?	A much larger super eruption	
Produced	Where did something produce something?	in Colorado	
	What did something produce?	over 5,000 cubic kilometers of material	
	Where didn't someone appear to do something?	In the video	
	Who didn't appear to do something?	the perpetrators	
appeared	When did someone appear?	never	
	What didn't company appear to do?	look at the camera	
	What didn't someone appear to do?	to look at the camera	
	Where didn't someone look at something?	In the video	
look	Who didn't look?	the perpetrators	
	What didn't someone look at?	the camera	

Who met someone?

Who met?

What did someone meet?

What had been founded?

When was something founded?

Why has something been founded?

Some of the vegetarians

vegetarians

members of the Theosophical Society

members of the Theosophical Society

the Theosophical Society

in 1875

1875

to further universal brotherhood

- 1. Scale up QA-SRL data annotation
- 2. Train a QA-SRL Parser
- 3. Improve Recall



# In 1950 Alan M. Turing *published* "Computing machinery and intelligence" in Mind, in which he *proposed* that machines could be *tested* for intelligence *using* questions and answers.

	When was something published?	In 1950
	Who published something?	Alan M. Turing
published	What was published?	"Computing Machinery and Intelligence"
	Where was something published?	in Mind
	When did someone propose something?	In 1950
	Who proposed something?	Alan M. Turing
proposed	What did someone propose?	that machines could be tested for intelligent using questions and answers
	What did someone propose something in?	"Computing Machinery and Intelligence"
	What can be tested?	machines
tested	What can something be tested for?	intelligence
	How can something be tested?	using questions and answers
	What was being used?	questions and answers
using	Why was something being used?	tested for intelligence

# Large-scale QA-SRL Parsing

- 1. Scale up QA-SRL data annotation
- 2. Train a QA-SRL Parser
- 3. Improve Recall

### **Easier Annotation**

UCCA [Abend and Rapaport 2013]	~6k sentences	4 Trained Annotators
Semantic Proto-roles [Reisinger et al. 2015]	~7k sentences	MTurk
Groningen Meaning Bank [Basile et al. 2012]	~40k sentences	Trained annotators/ GWAP
QASRL 1.0 [He et al. 2015]	~3k sentences	Trained annotators

#### **Questions:**

Wh	Aux	Subj	Verb	Obj	Prep	Obj2
Who	Ø	Ø	Ø	Ø	Ø	Ø
What	did	someone	stem	someone	on	someone
Where	didn't	something	past	something	to	something
When	might		past participle		by	
Why	will		present		from	
How	•••					

John surreptitiously ate the burrito at 2am.

#### **Questions:**

Wh	Aux	Subj	Verb	Obj	Prep	Obj2
Who	Ø	Ø	Ø	Ø	Ø	Ø
What	did	someone	stem	someone	on	someone
Where	didn't	something	past	something	to	something
When	might		past participle		by	
Why	will		present		from	
How	•••					

John surreptitiously ate the burrito at 2am.

#### **Questions:**

Who ate something?

Wh	Aux	Subj	Verb	Obj	Prep	Obj2
Who	Ø	Ø	Ø	Ø	Ø	Ø
What	did	someone	stem	someone	on	someone
Where	didn't	something	past	something	to	something
When	might		past participle		by	1 1 1 1 1 1
Why	will		present		from	
How						
						1 1 1 1 1 1

John surreptitiously ate the burrito at 2am.

#### **Questions:**

Who ate something?

What did someone eat?

	Wh	Aux	Subj	Verb	Obj	Prep	Obj2
	Who	Ø	Ø	Ø	Ø	Ø	Ø
	What	did	someone	stem	someone	on	someone
١	Where	didn't	something	past	something	to	something
1	When	might		past participle		by	
	Why	will		present		from	
	How						

John surreptitiously ate the burrito at 2am.

#### **Questions:**

Who ate something?

What did someone eat?

. . .

John surreptitiously ate the burrito at 2am.

Questions:

Answers:

Who ate something?

John

What did someone eat?

. . .

John surreptitiously ate the burrito at 2am.

**Questions:** 

Who ate something?

What did someone eat?

**Answers:** 

John

the burrito

. . .

# Annotation Pipeline

 $\mathcal{X}$ 

John surreptitiously ate the burrito at 2am



**Predicate** detection

Identify verbs with POS + heuristics

# Annotation Pipeline

 ${\mathcal X}$ 

John surreptitiously ate the burrito at 2am

**Predicate** detection

Identify verbs with POS + heuristics

Question annotation

One worker writes as many QA-SRL questions as possible, and provides the answer

# Annotation Pipeline

 $\mathcal{X}$ 

John surreptitiously ate the burrito at 2am

**Predicate** detection

Identify verbs with POS + heuristics

Question annotation

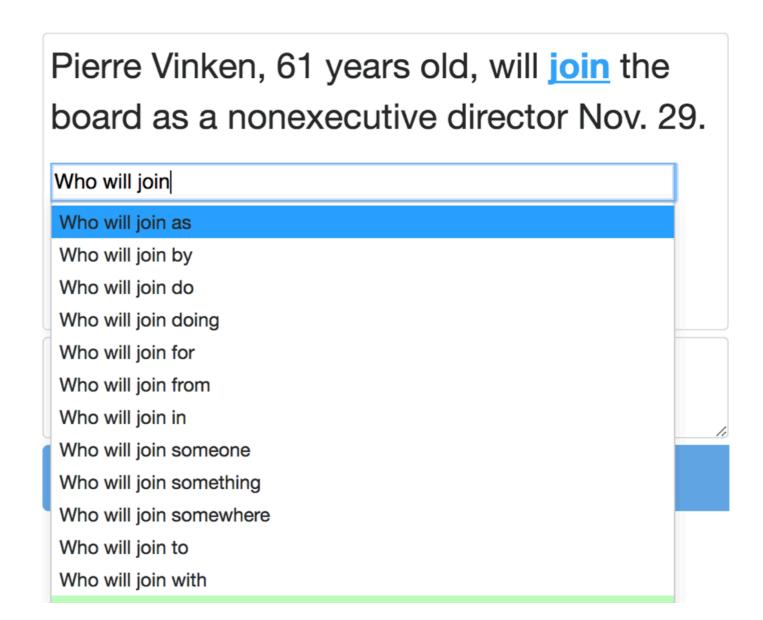
One worker writes as many QA-SRL questions as possible, and provides the answer

Validation

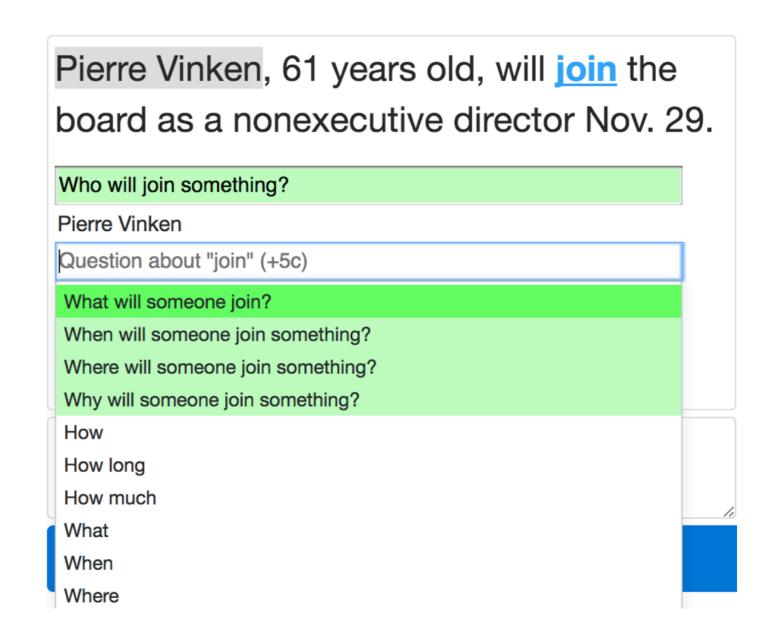
2 workers are shows questions, provide answers or mark as invalid

- Efficiency
- Recall

- Efficiency
  - Autocomplete



- Efficiency
  - Autocomplete
- Recall
  - Autosuggest



- Efficiency
  - Autocomplete
- Recall
  - Autosuggest
  - Financial Incentives

### Validation Interface

In 1668 the king of Portugal sent an elephant from the kingdom of Congo to the king of France.

Invalid Where did someone send something? from the kingdom of Congo to the king of France

Invalid Where did someone send something from? the kingdom of Congo

Invalid Who did someone send something to?--> the king of France / Highlight to add an answer

Bonus: 4c

### Dataset

- I annotator provides questions
- 2 annotators validate -> 3 spans / question
  - Question invalid if any annotator marks invalid

Additional 3 validators for small dense dev and test set

### Dataset

	Wikipedia	Wikinews	Science
Sentences	15,000	14,682	46,715
Verbs	32,758	34,026	66,653
Questions	75,867	80,081	143,388
Valid Qs	67,146	70,555	127,455

[He et al 2015]	This work
3000 sentences	75k sentences
Several weeks	9 days
~50c / verb	33c / verb
2.43 questions / verb	2.05 questions / verb

# In 1950 Alan M. Turing *published* "Computing machinery and intelligence" in Mind, in which he *proposed* that machines could be *tested* for intelligence *using* questions and answers.

	Who published something?	Alan M. Turing
published	What was published?	"Computing Machinery and Intelligence"
pasiiorioa	When was something published?	In 1950
	Where was something published?	in Mind
	Who proposed something?	Alan M. Turing
proposed	What did someone propose?	that machines could be tested for intelligent using questions and answers
	When did someone propose something?	In 1950
	What did someone propose something in?	"Computing Machinery and Intelligence"
	What can be tested?	machines
tested	What can something be tested for?	intelligence
	How can something be tested?	using questions and answers
	What was being used?	questions and answers
using	Why was something being used?	tested for intelligence

# In 1950 Alan M. Turing *published* "Computing machinery and intelligence" in Mind, in which he *proposed* that machines could be *tested* for intelligence *using* questions and answers.

	Who published something?	Alan M. Turing
published	What was published?	"Computing Machinery and Intelligence"
pasiiorioa	When was something published?	In 1950
	Where was something published?	in Mind
	Who proposed something?	Alan M. Turing
proposed	What did someone propose?	that machines could be tested for intelligent using questions and answers
' '	When did someone propose something?	In 1950
	What did someone propose something in?	"Computing Machinery and Intelligence"
	What can be tested?	machines
tested	What can something be tested for?	intelligence
	How can something be tested?	using questions and answers
	What was being used?	questions and answers
using	Why was something being used?	tested for intelligence

# In 1950 Alan M. Turing *published* "Computing machinery and intelligence" in Mind, in which he *proposed* that machines could be *tested* for intelligence *using* questions and answers.

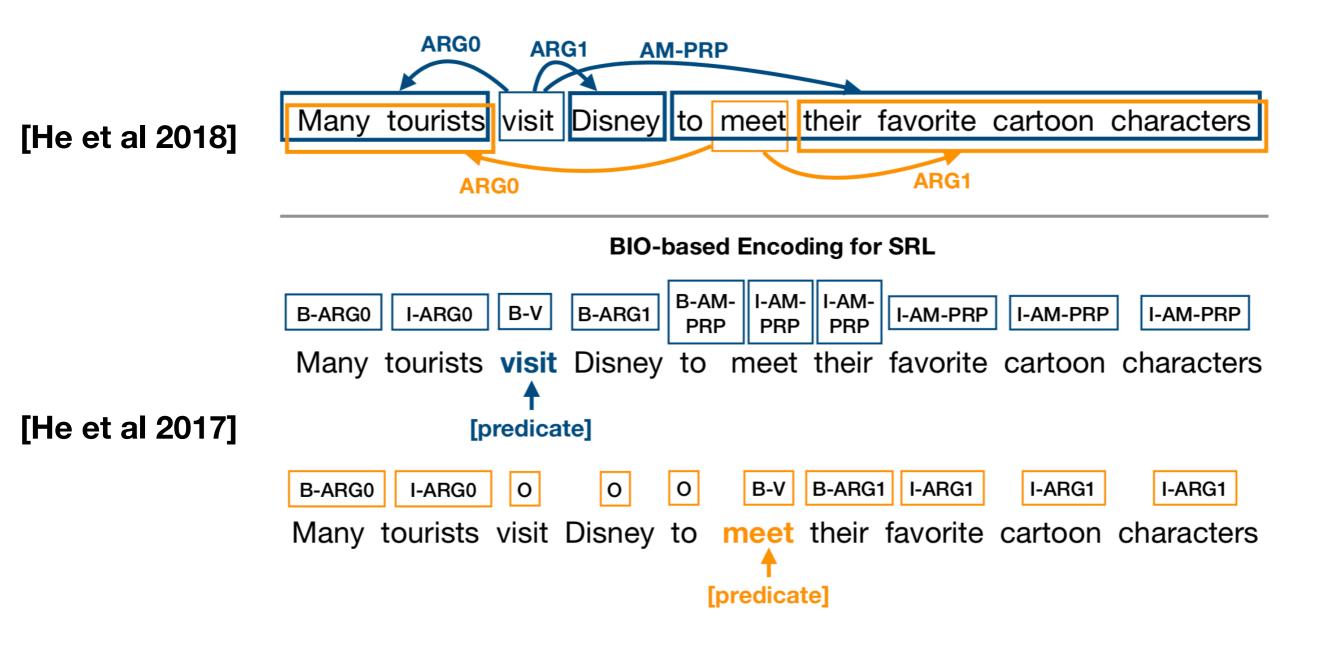
	Who published something?	Alan M. Turing
published	What was published?	"Computing Machinery and Intelligence"
pasiiorioa	When was something published?	In 1950
	Where was something published?	in Mind
	Who proposed something?	Alan M. Turing
proposed	What did someone propose?	that machines could be tested for intelligent using questions and answers
' '	When did someone propose something?	In 1950
	What did someone propose something in?	"Computing Machinery and Intelligence"
	What can be tested?	machines
tested	What can something be tested for?	intelligence
	How can something be tested?	using questions and answers
	What was being used?	questions and answers
using	Why was something being used?	tested for intelligence

# Large-scale QA-SRL Parsing

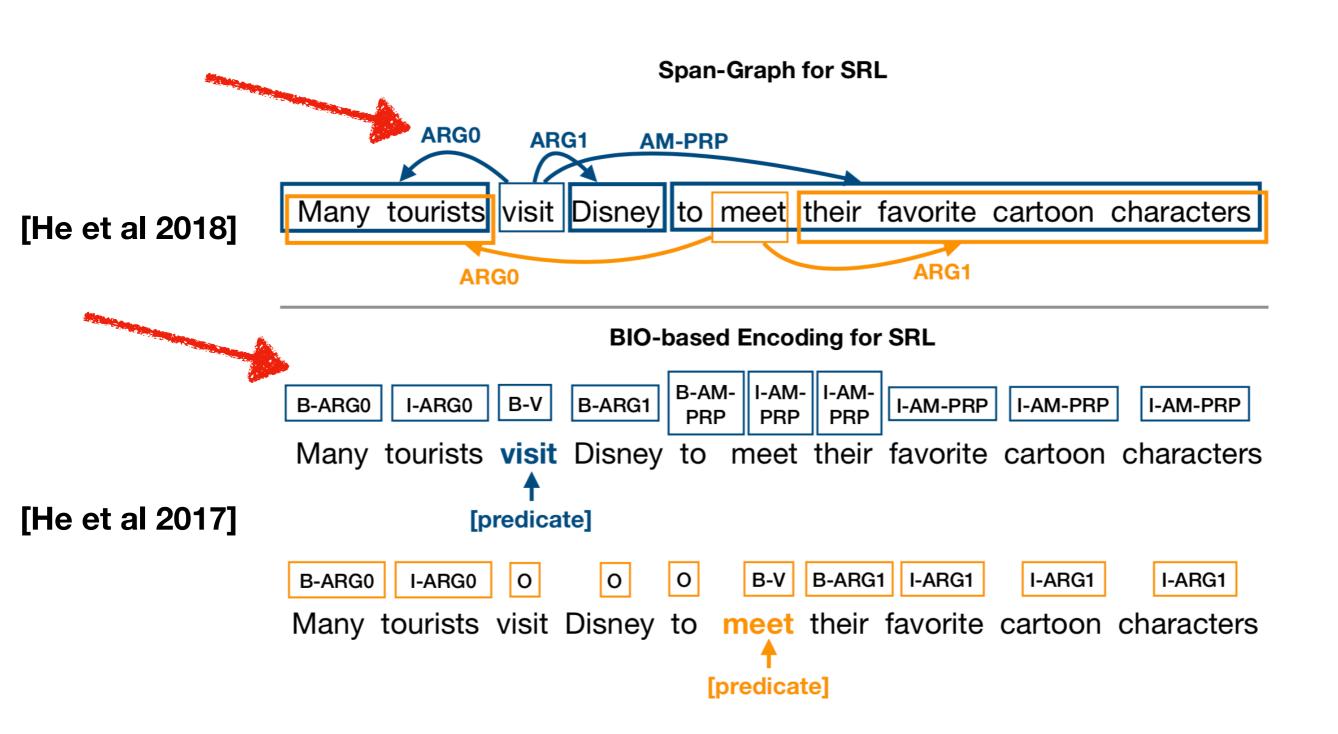
- 1. Scale up QA-SRL data annotation
- 2. Train a QA-SRL Parser
- 3. Improve Recall

### QA-SRL Parser

#### Span-Graph for SRL



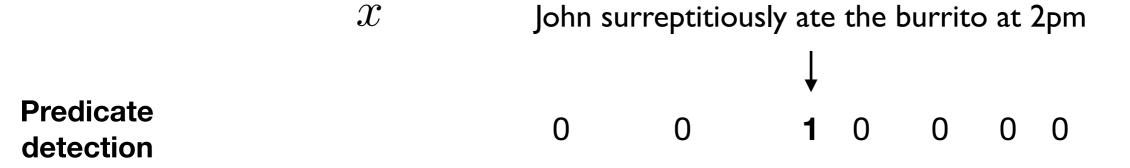
### QA-SRL Parser



### QA-SRL Parser

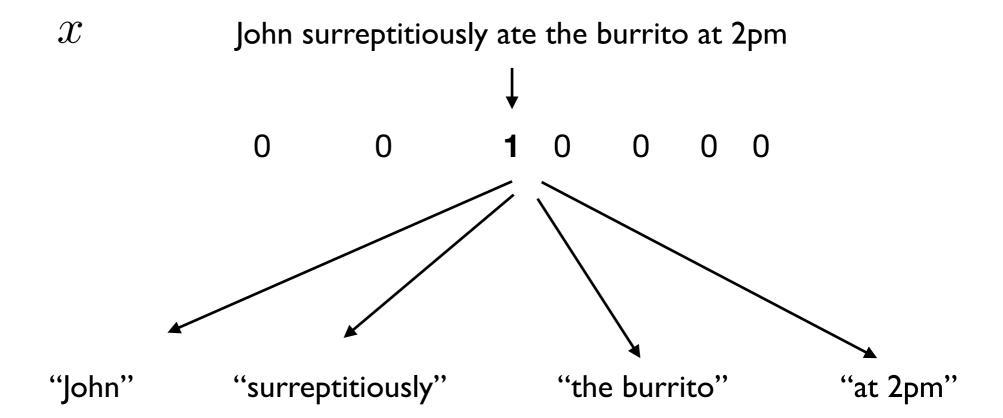
- Unlabeled Argument Detection
- Question generation

 ${\mathcal X}$  John surreptitiously ate the burrito at 2pm



**Predicate** detection

**Argument** detection



**Predicate** 

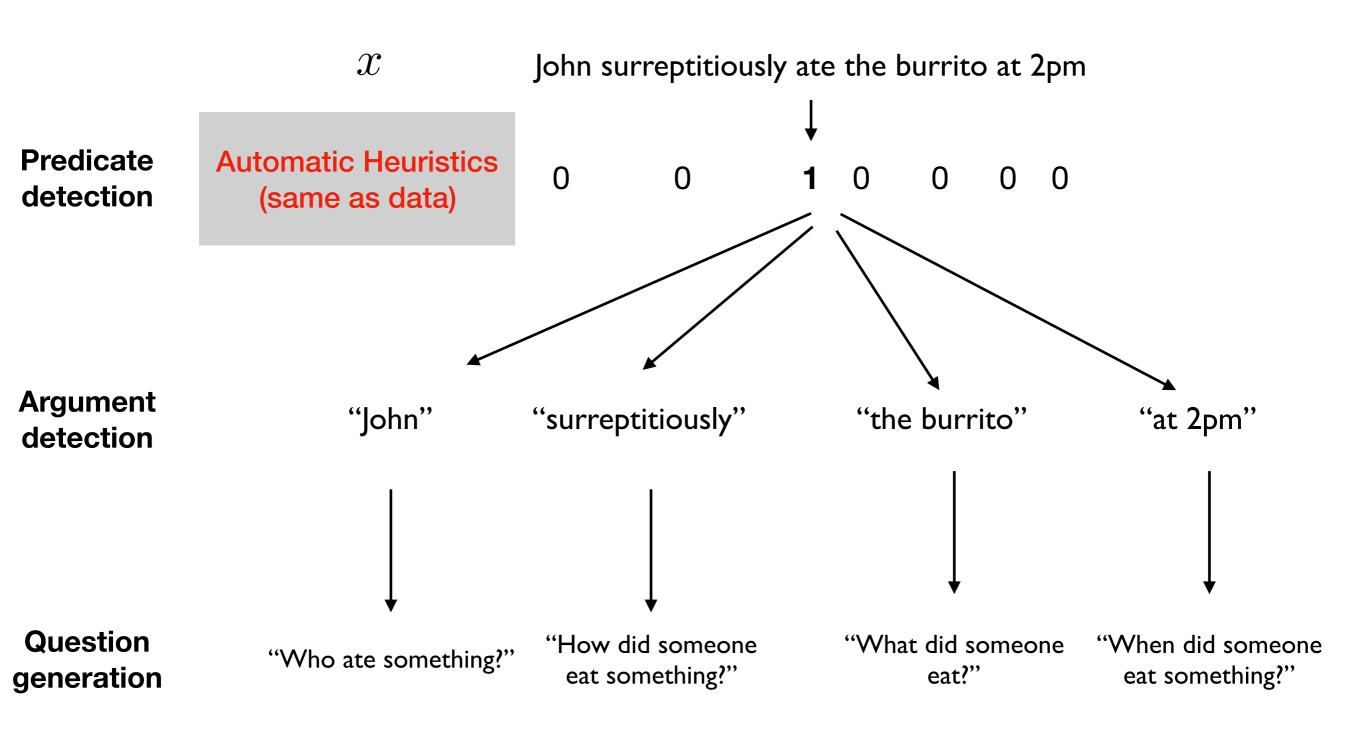
detection

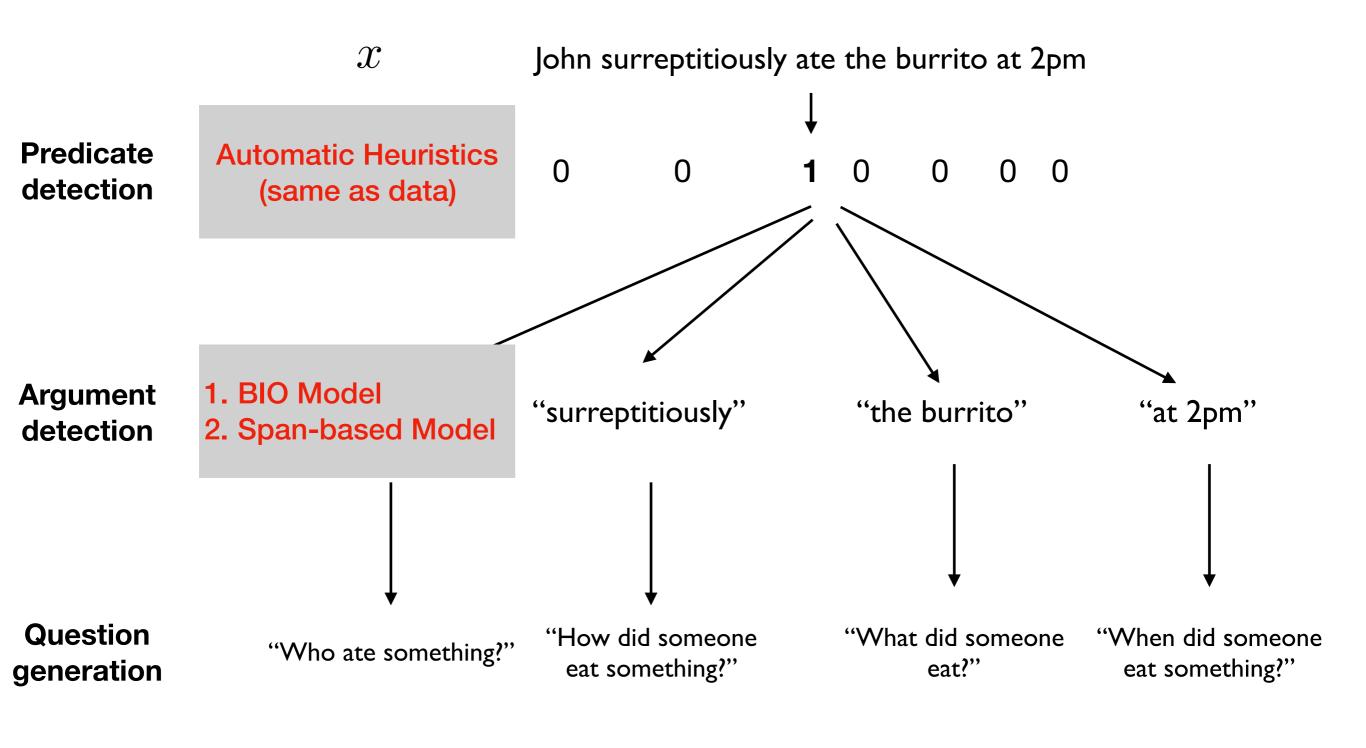
Argument

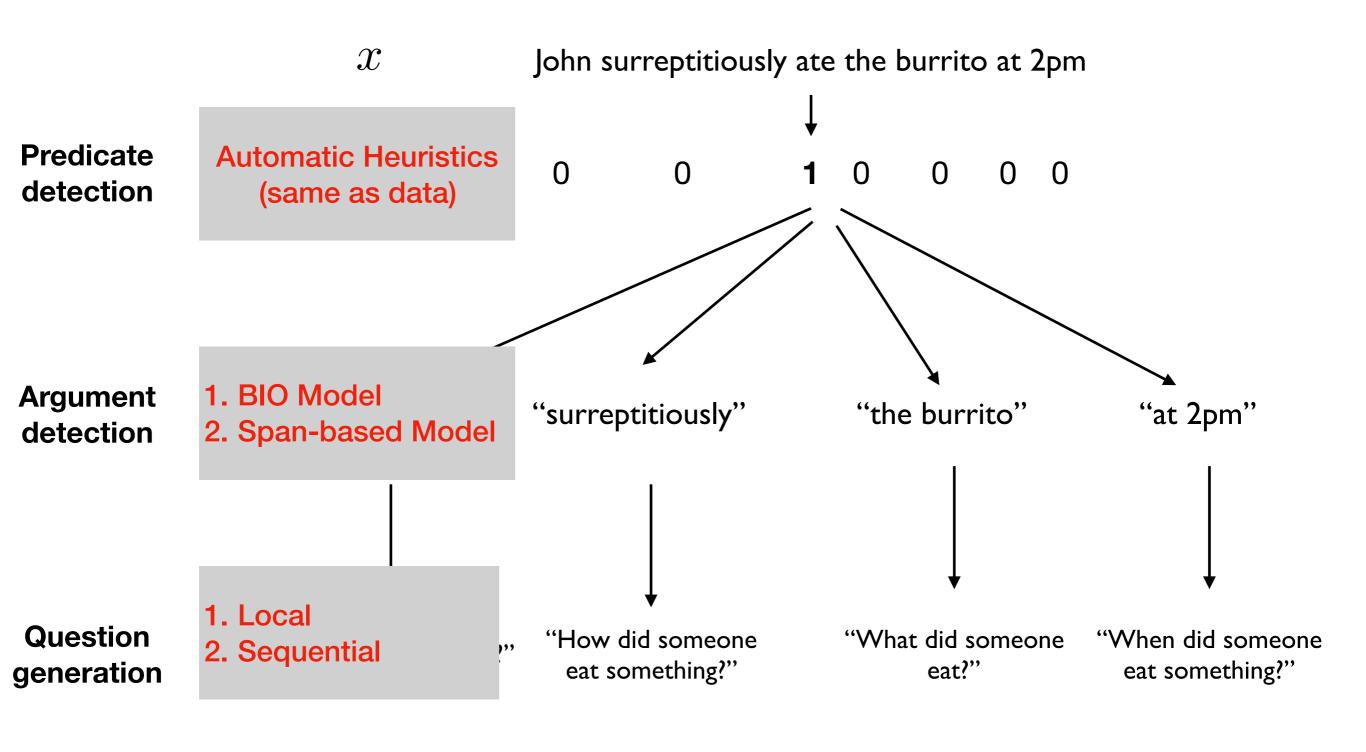
detection

Question

 $\mathcal{X}$ John surreptitiously ate the burrito at 2pm 0 "the burrito" "John" "surreptitiously" "at 2pm" "What did someone "How did someone "When did someone "Who ate something?" eat something?" eat?" eat something?" generation

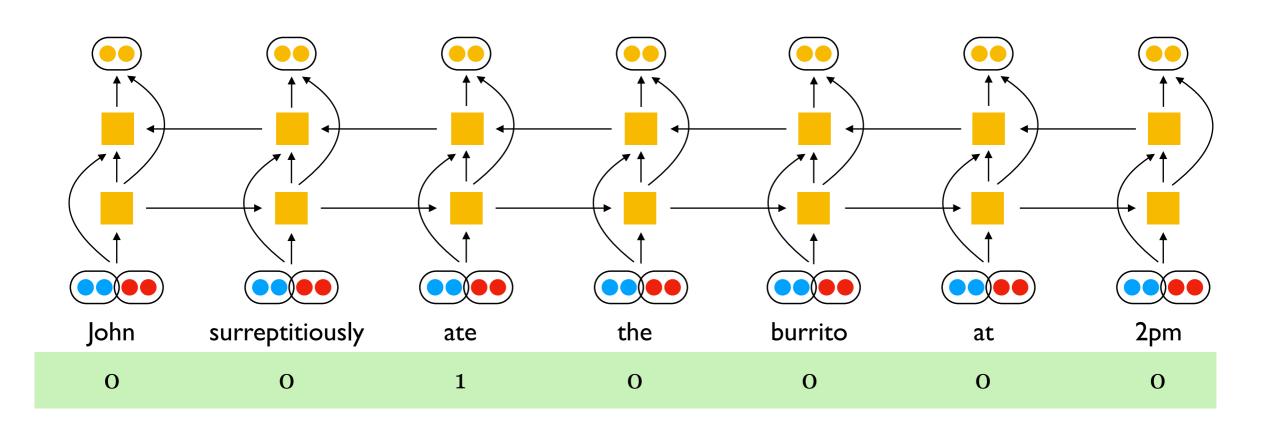




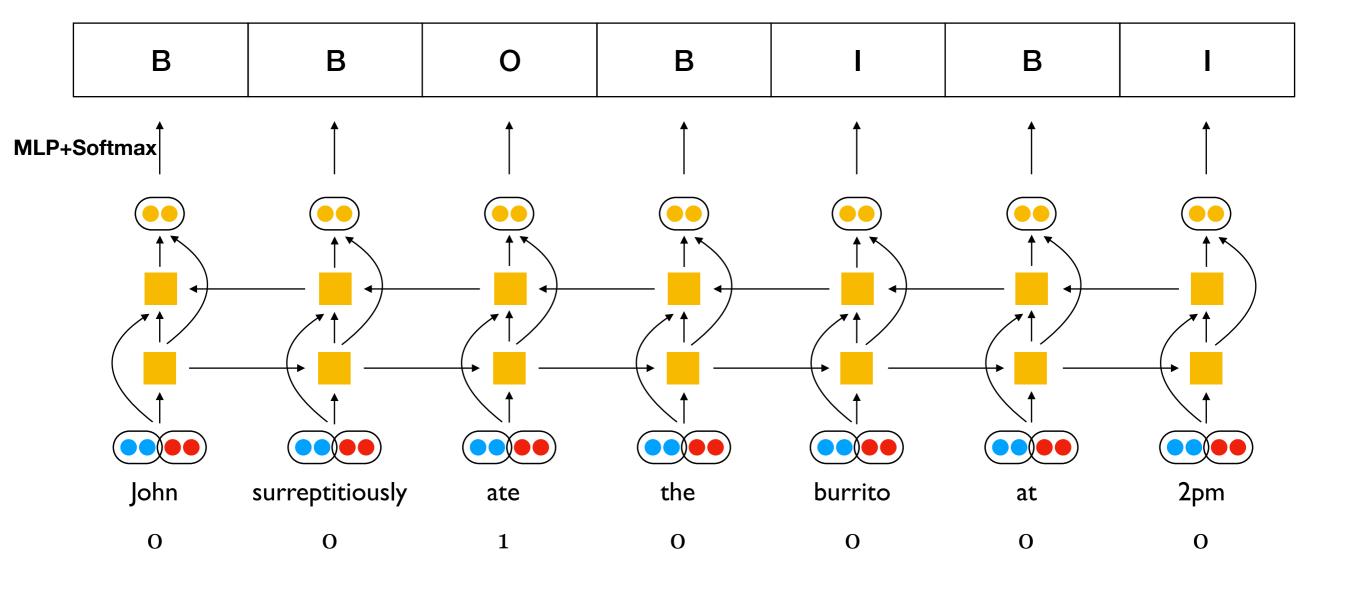


#### **Argument Detection - BIO Model**

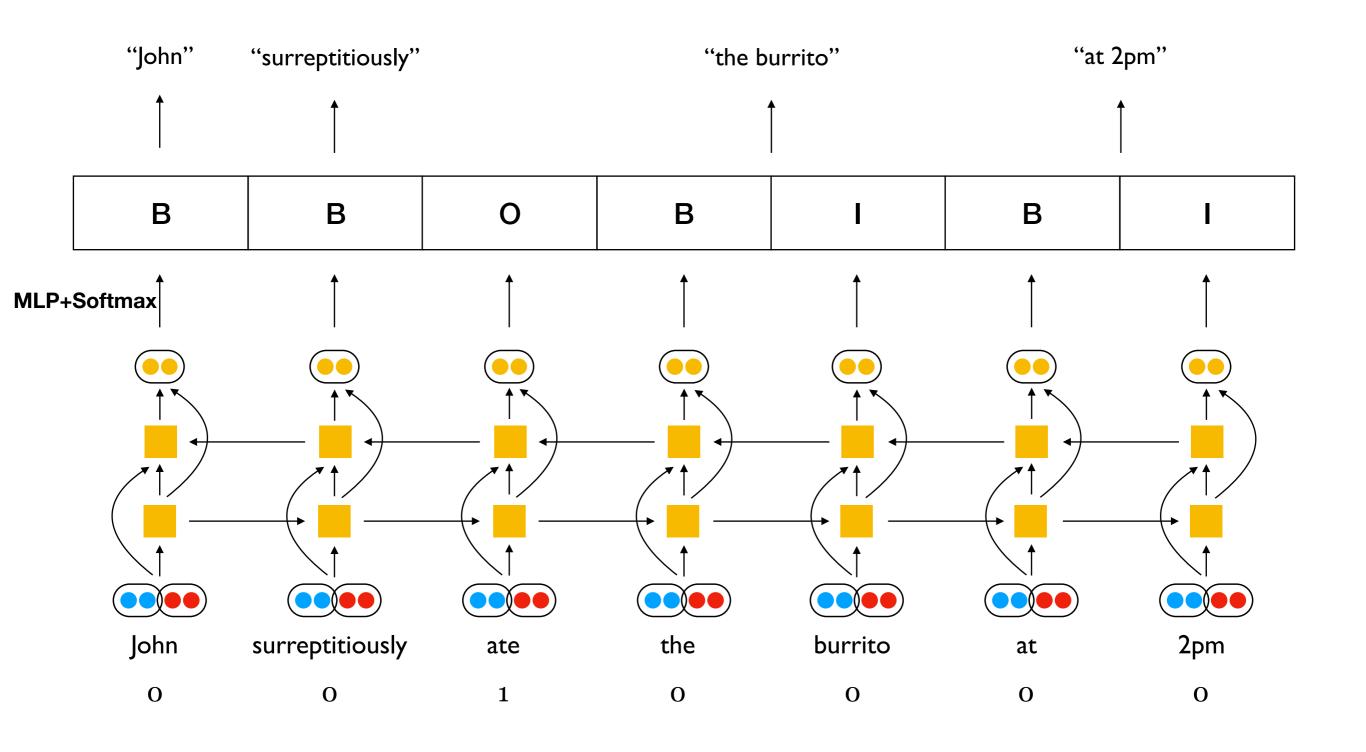
- Alternating Bi-LSTM with Highway Connections and Recurrent Dropout [He et al 2017]
- Input includes predicate indicator

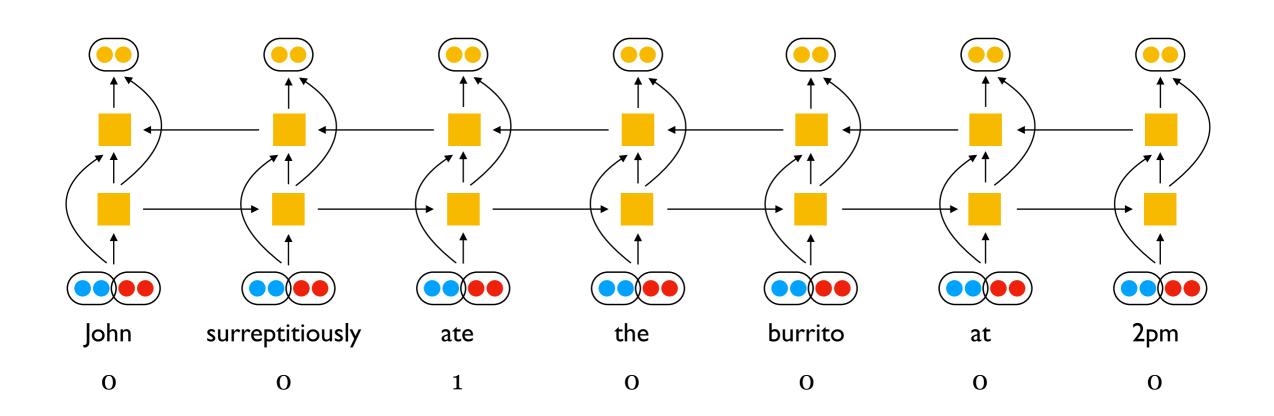


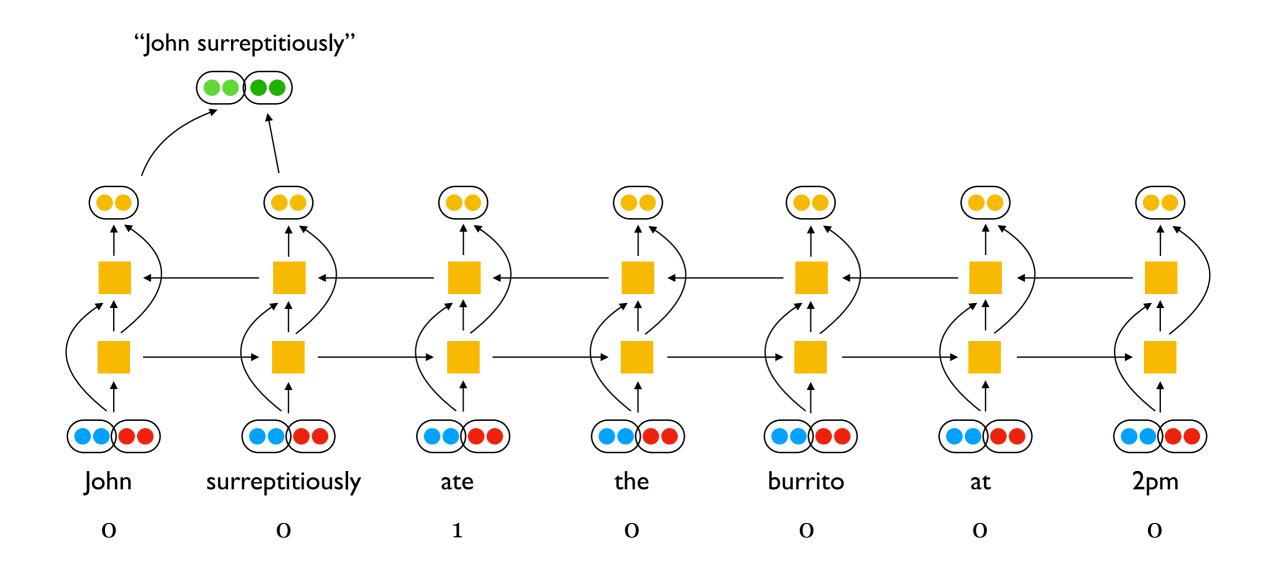
#### **Argument Detection - BIO Model**

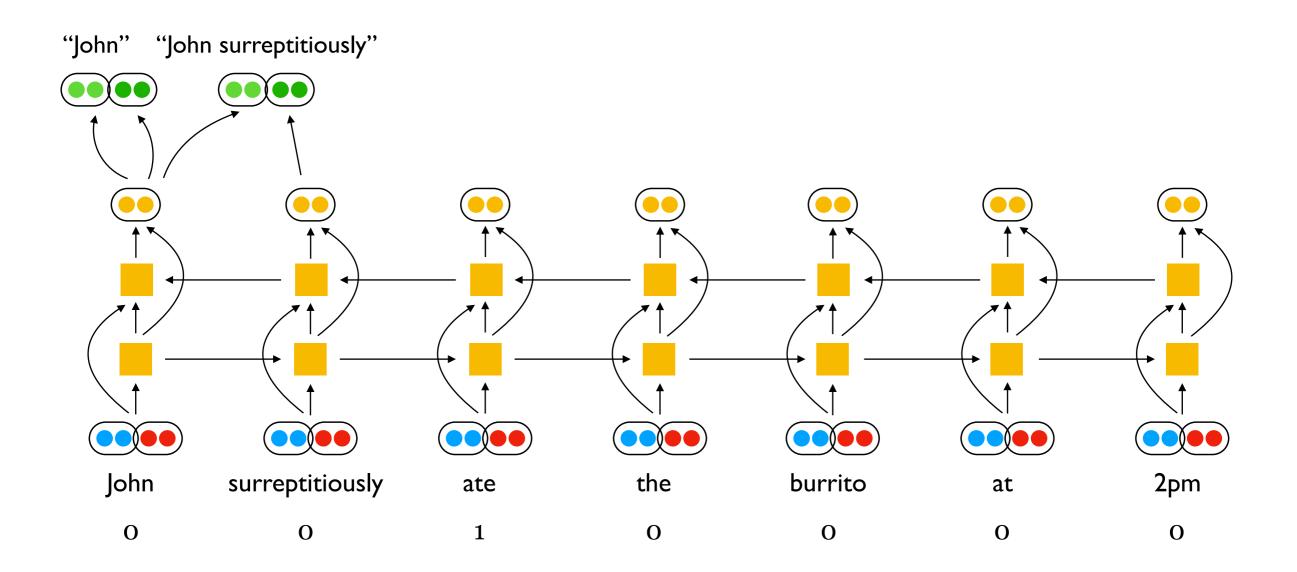


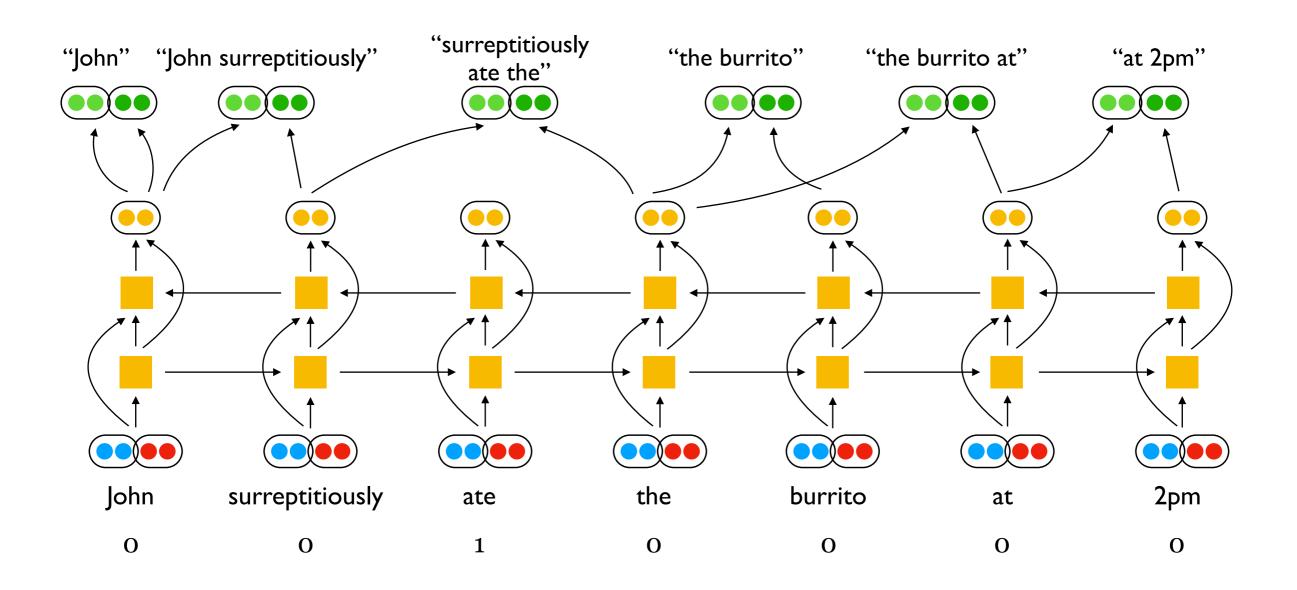
#### **Argument Detection - BIO Model**

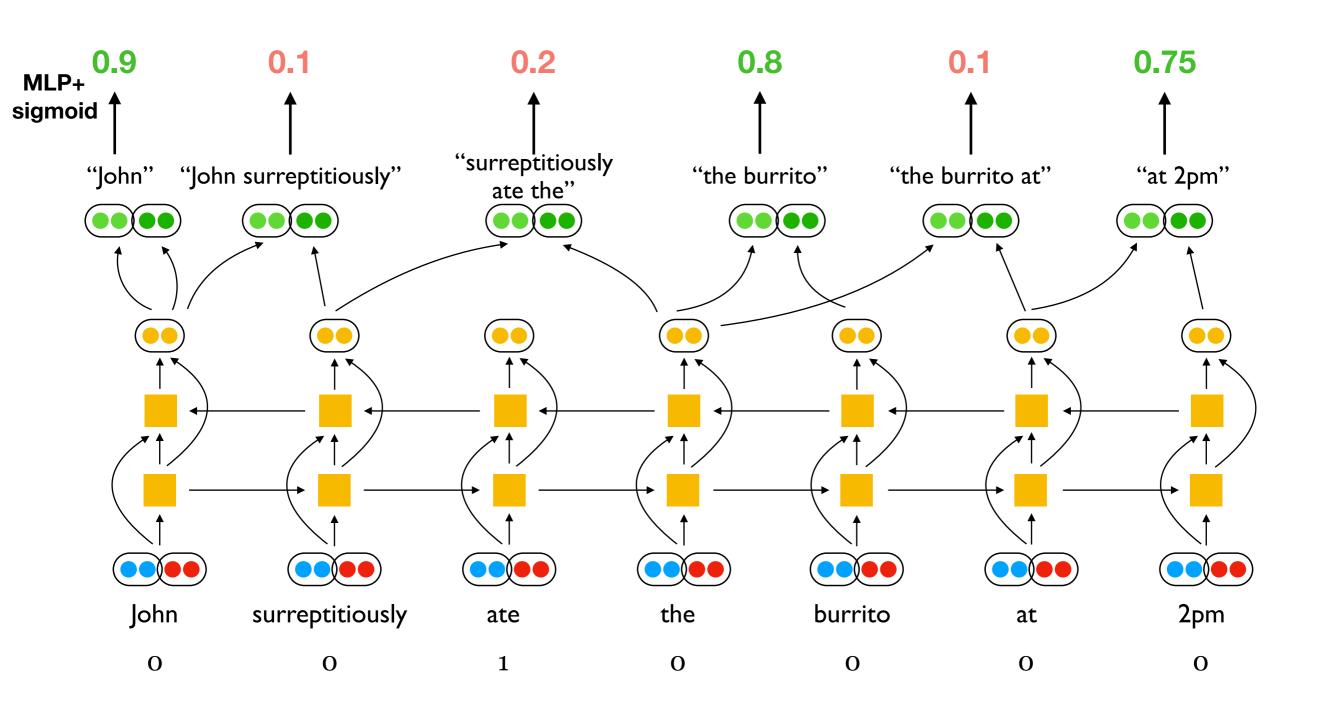


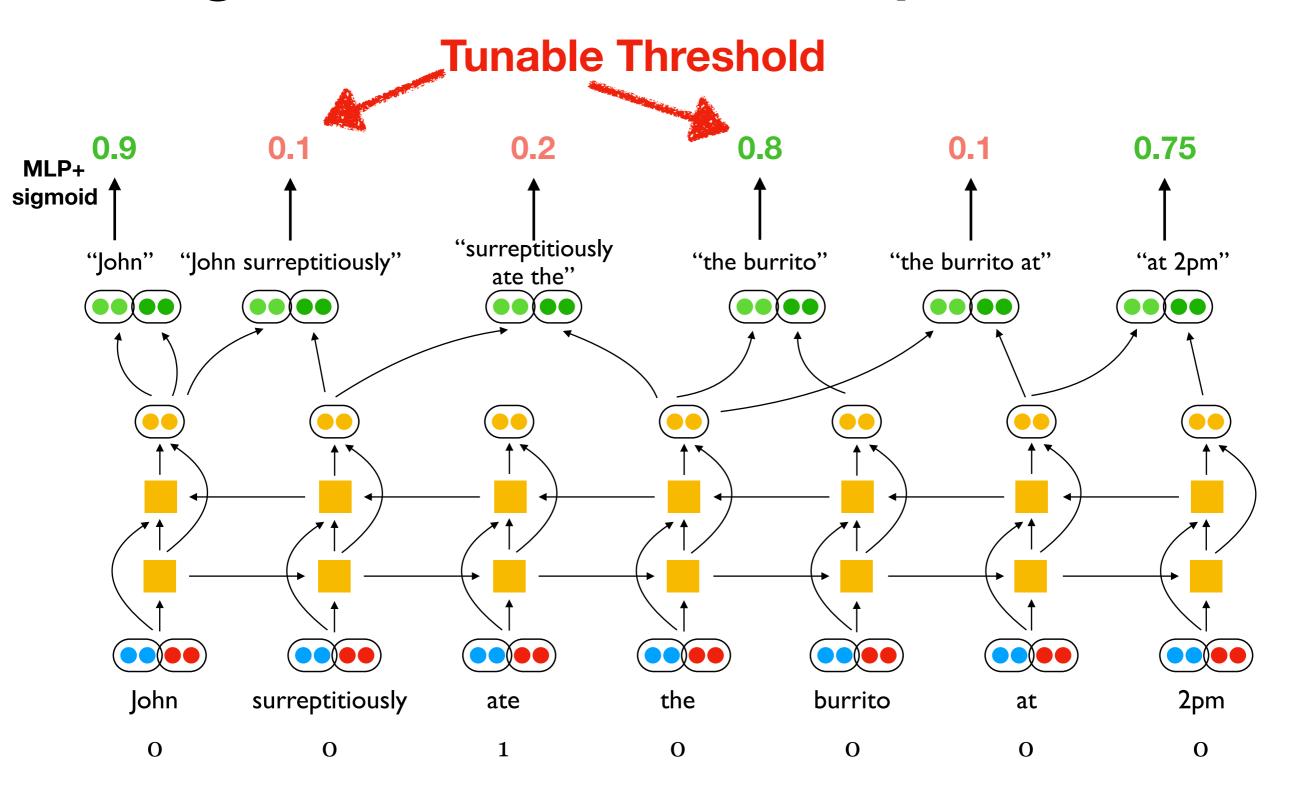








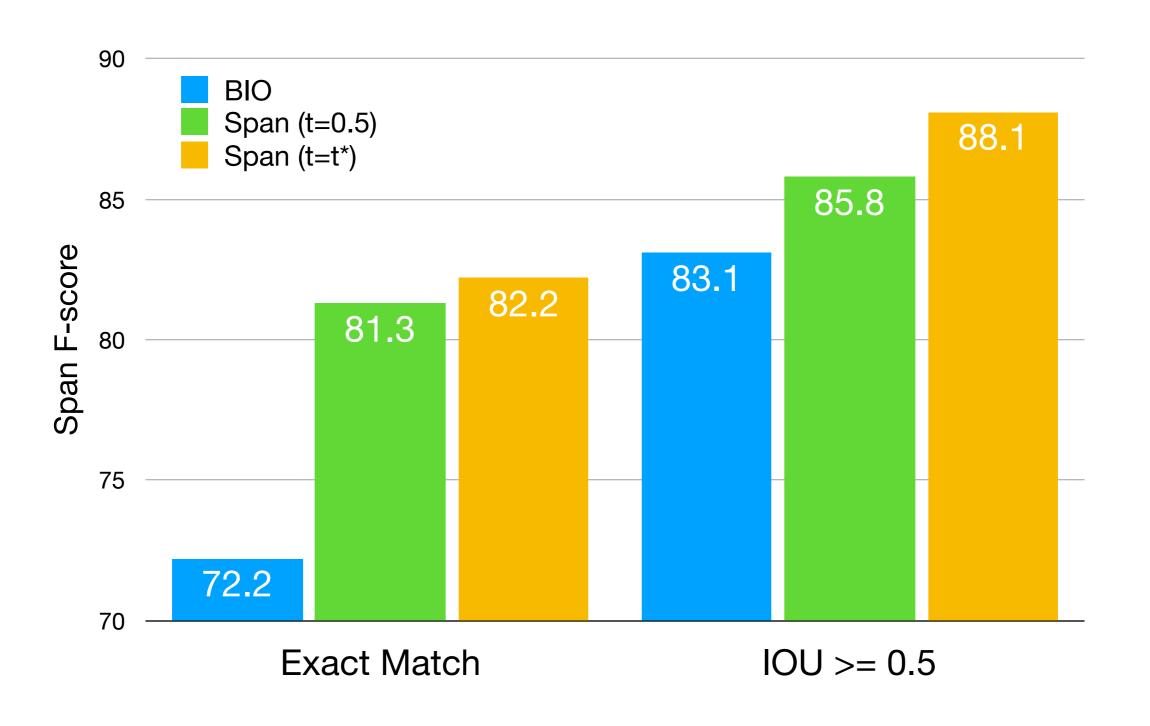




### **Argument Detection**

- 4 layer Alternating Bi-LSTM with Highway Connections and Recurrent Dropout [He et al 2017]
- Trained to maximize log-likelihood

## **Argument Detection**



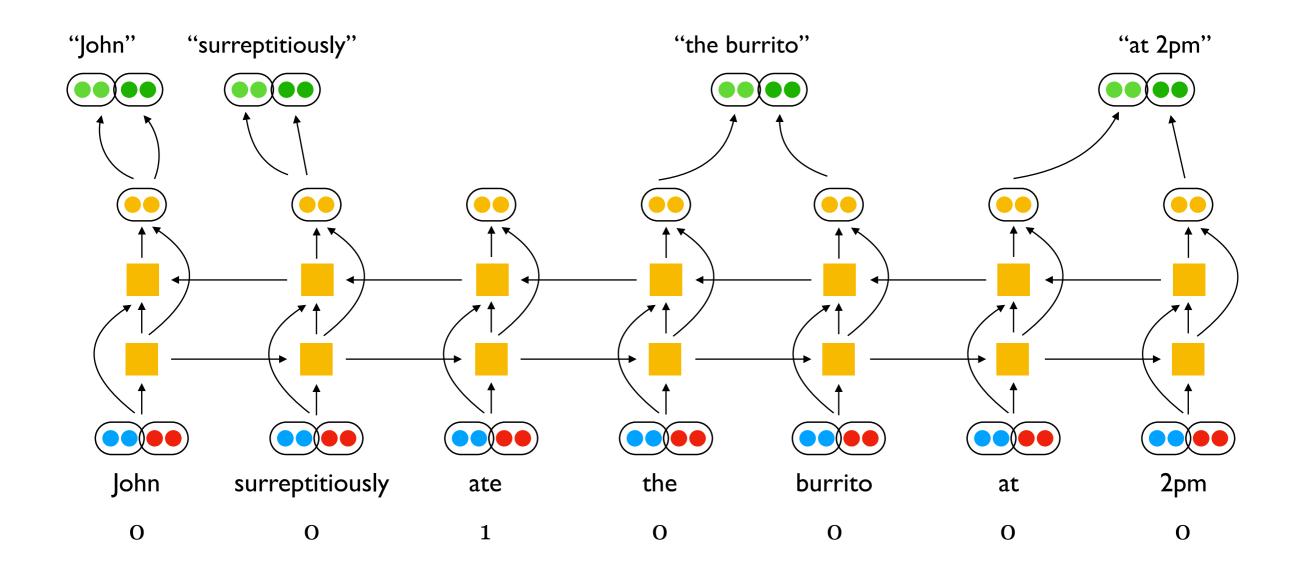
### Question Generation

Wh	Aux	Subj	Verb	Obj	Prep	Obj2
Who	Ø	Ø	Ø	Ø	Ø	Ø
What	did	someone	stem	someone	on	someone
Where	didn't	something	past	something	to	something
When	might		past participle		by	
Why	will		present		from	
How						

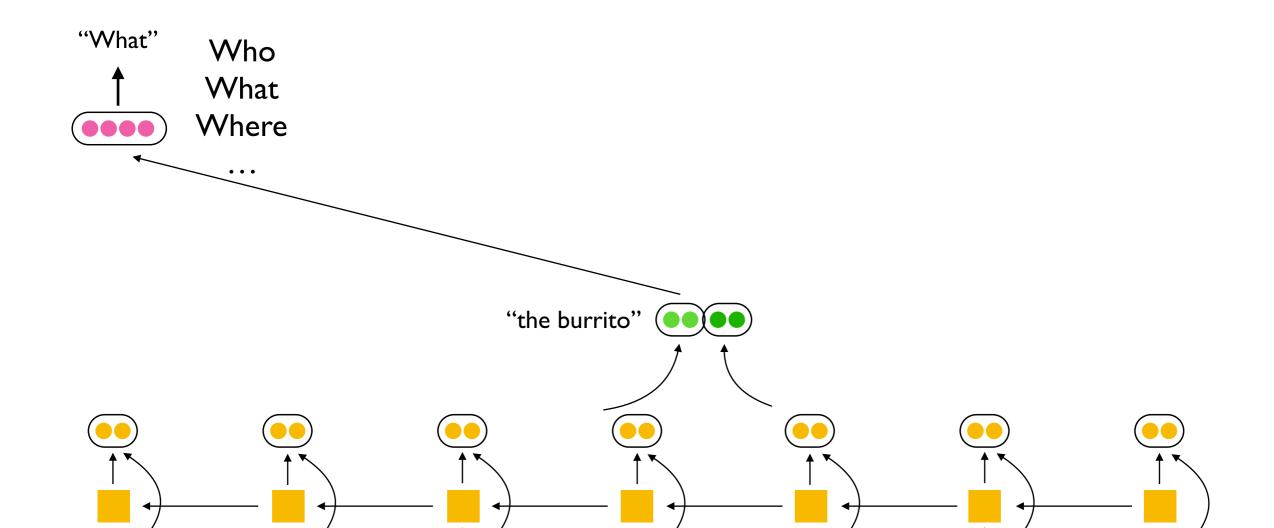
### Question Generation

Wh	Aux	Subj	Verb	Obj	Prep	Obj2
Who	Ø	Ø	Ø	Ø	Ø	Ø
What	did	someone	stem	someone	on	someone
Where	didn't	something	past	something	to	something
When	might		past participle		by	
Why	will		present		from	
How	•••				• • •	
	1 1 1 1 1					

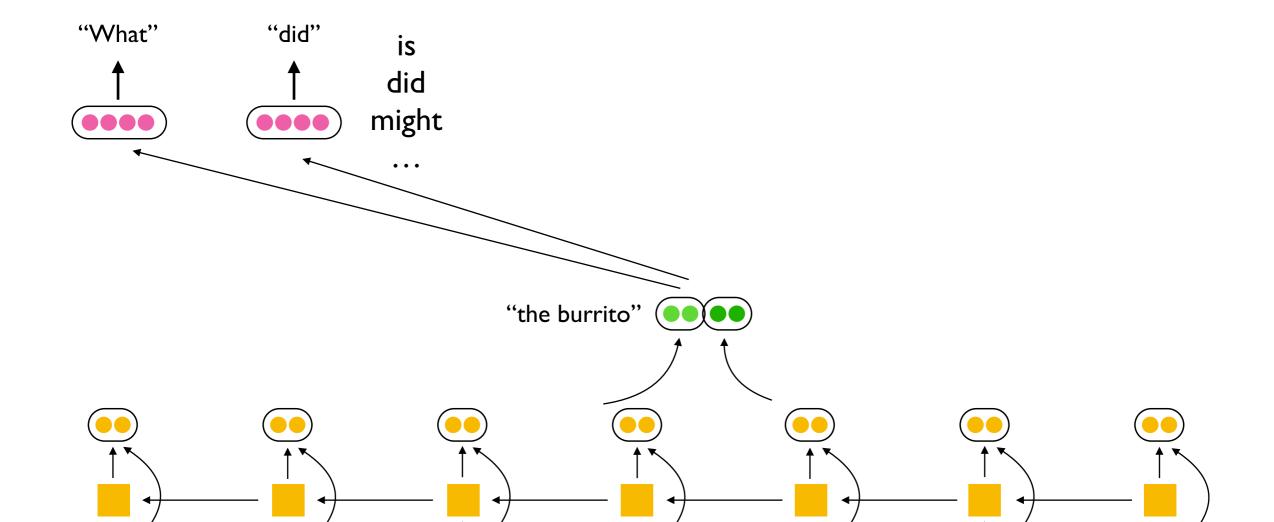
- Two models:
  - Local slot prediction
  - Sequential Model (LSTM)



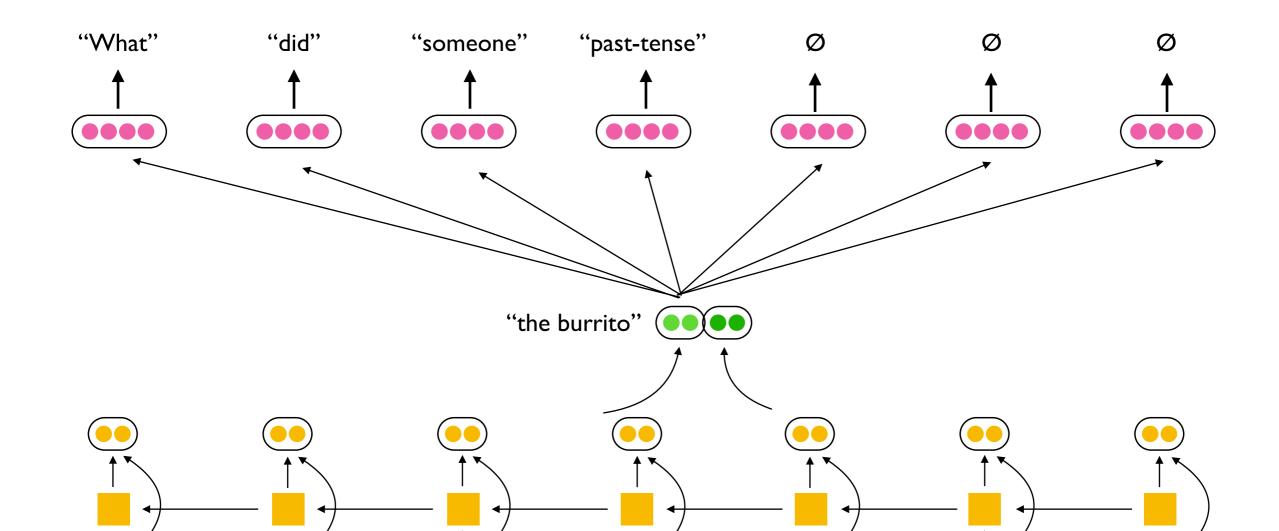
Wh	Aux	Subj	Verb	Obj1	Prep	Obj2
		•		_	•	_



Wh	Aux	Subj	Verb	Obj1	Prep	Obj2
		•		_	•	_

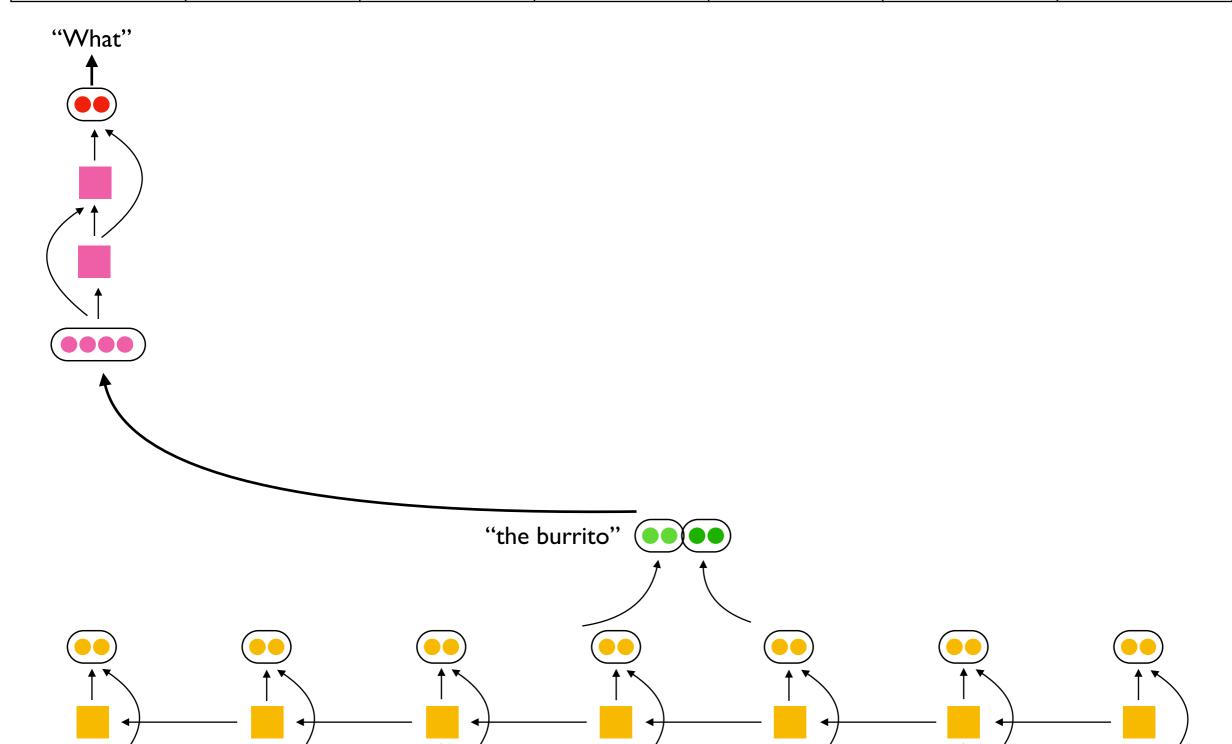


Wh Aux Subj Verb Obj1 Prep Obj2
---------------------------------



#### **Question Generation - Sequential**

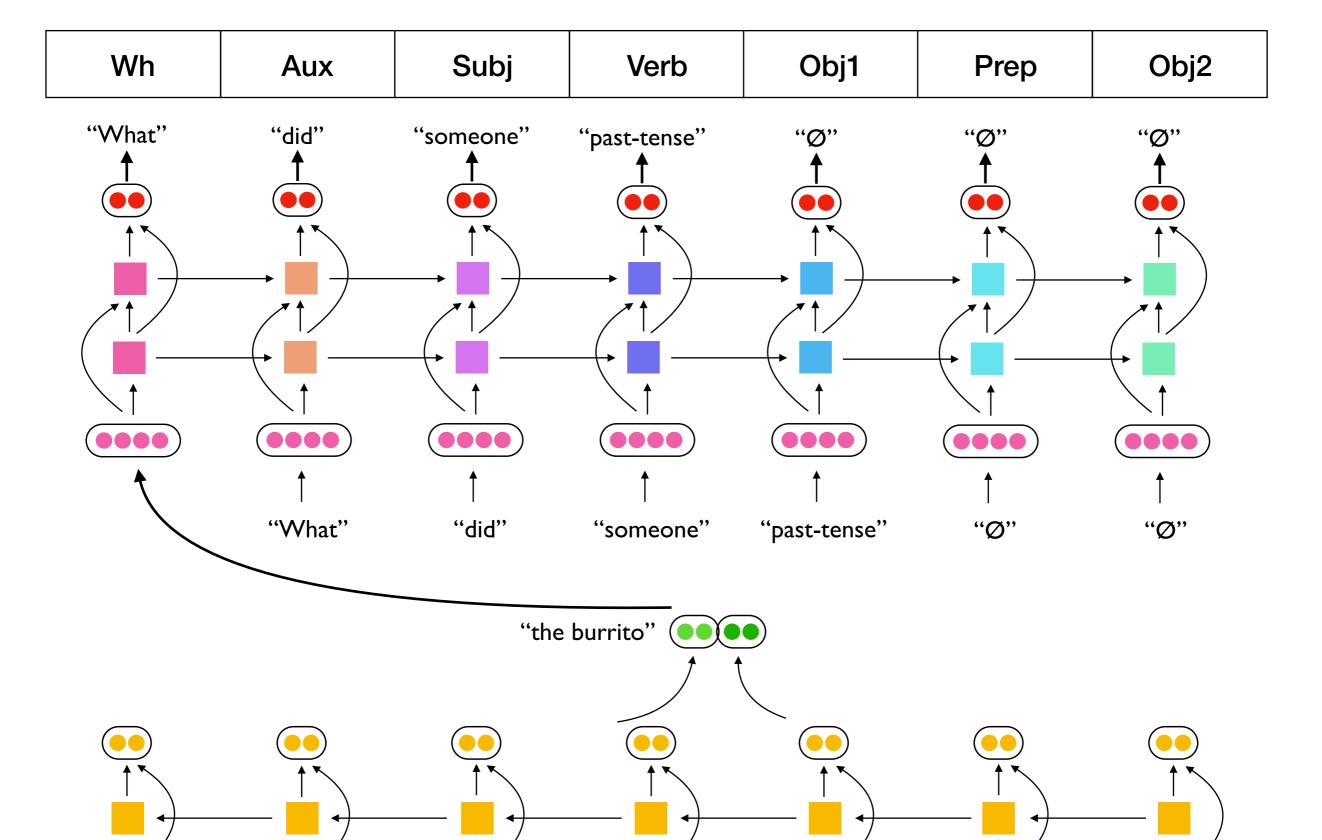
Wh Aux Subj Verb Obj1 Prep Obj2



#### **Question Generation - Sequential**

Subj Wh Verb Obj1 Obj2 Aux Prep "What" "did" "What" "the burrito"

#### **Question Generation - Sequential**



### **Question Generation**

- 4 layer Alternating Bi-LSTM with Highway Connections and Recurrent Dropout [He et al 2017]
- Trained to maximize log-likelihood

### **Evaluation Questions**

Paraphrasing

"Who ate something?

"Who was something eaten by?

### **Evaluation Questions**

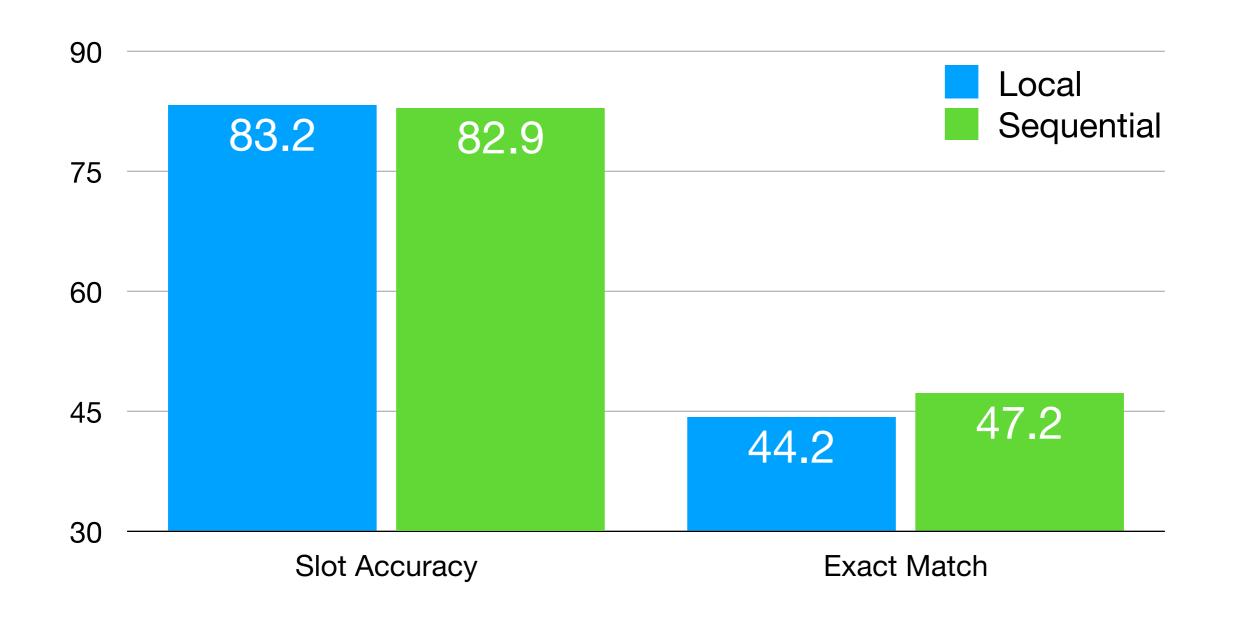
Paraphrasing

"Who ate something?

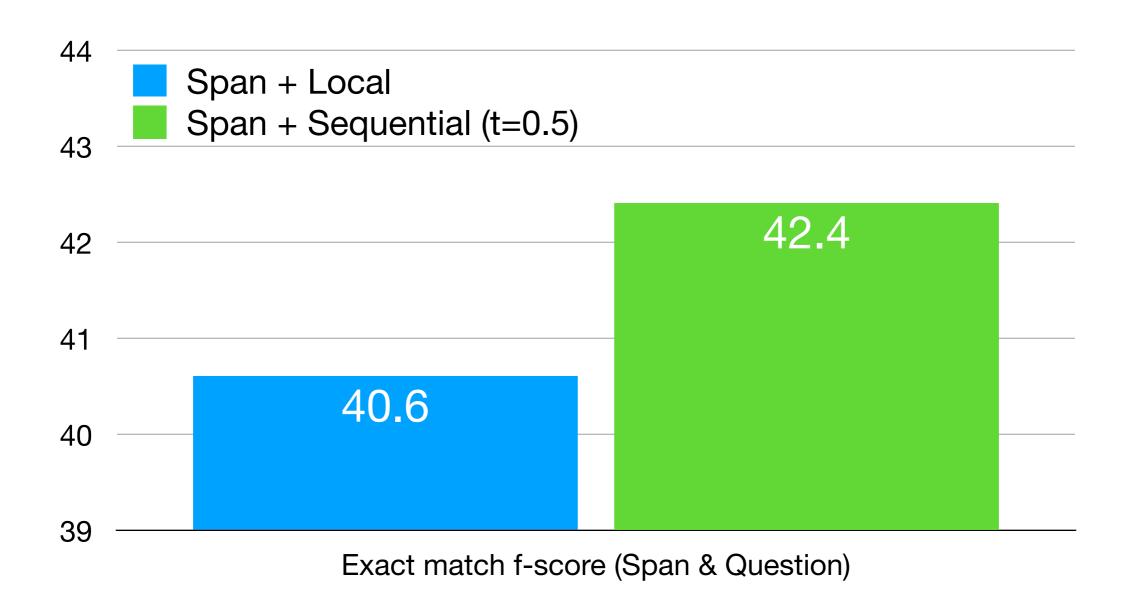
"Who was something eaten by?

- Given gold span:
  - Slot-level accuracy
  - Exact Match (full question)

### **Question Generation**



## Full Parsing Accuracy



# Large-scale QA-SRL Parsing

- 1. Scale up QA-SRL data annotation
- 2. Train a QA-SRL Parser
- 3. Improve Recall

## In 1950 Alan M. Turing *published* "Computing machinery and intelligence" in Mind, in which he *proposed* that machines could be *tested* for intelligence *using* questions and answers.

	Who published something?	Alan M. Turing
published	What was published?	"Computing Machinery and Intelligence"
	When was something published?	In 1950
	Where was something published?	in Mind
proposed	Who proposed something?	Alan M. Turing
	What did someone propose?	that machines could be tested for intelligent using questions and answers
	When did someone propose something?	In 1950
	What did someone propose something in?	"Computing Machinery and Intelligence"
tested	What can be tested?	machines
	What can something be tested for?	intelligence
	How can something be tested?	using questions and answers
using	What was being used?	questions and answers
	Why was something being used?	tested for intelligence

## In 1950 Alan M. Turing *published* "Computing machinery and intelligence" in Mind, in which he *proposed* that machines could be *tested* for intelligence *using* questions and answers.

published	Who published something?	Alan M. Turing
	What was published?	"Computing Machinery and Intelligence"
	When was something published?	In 1950
	Where was something published?	in Mind
proposed	Who proposed something?	Alan M. Turing
	npacts training an	that machines could be tested for intelligent using questions and answers
		In 1950
	When did someone propose something?  CVAIUATION What did someone propose something in?	"Computing Machinery and Intelligence"
tested	What can be tested?	machines
	What can something be tested for?	intelligence
	How can something be tested?	using questions and answers
using	What was being used?	questions and answers
	Why was something being used?	tested for intelligence

## In 1950 Alan M. Turing *published* "Computing machinery and intelligence" in Mind, in which he *proposed* that machines could be *tested* for intelligence *using* questions and answers.

published	Who published something?	Alan M. Turing
	What was published?	"Computing Machinery and Intelligence"
	When was something published?	In 1950
	Where was something published?	in Mind
proposed	Who proposed something?	Alan M. Turing
	What did someone propose?	that machines could be tested for intelligent using questions and answers
	Filling Gaps	In 1950
	What did someone propose something in?	"Computing Machinery and Intelligence"
tested	What can be tested?	machines
	What can something be tested for?	intelligence
	How can something be tested?	using questions and answers
using	What was being used?	questions and answers
	Why was something being used?	tested for intelligence

## Annotation Pipeline

 $\mathcal{X}$ 

John surreptitiously ate the burrito at 2pm

**Predicate** detection

Identify verbs with POS + heuristics

Question generation

One worker writes as many QA-SRL questions as possible, and provides the answer

Workers are shows questions, provide answers or mark as invalid

**Validation** 

## Annotation Pipeline

 ${\mathcal X}$  John surreptitiously ate the burrito at 2pm

**Predicate** detection

Question generation

**QA-SRL** Parser

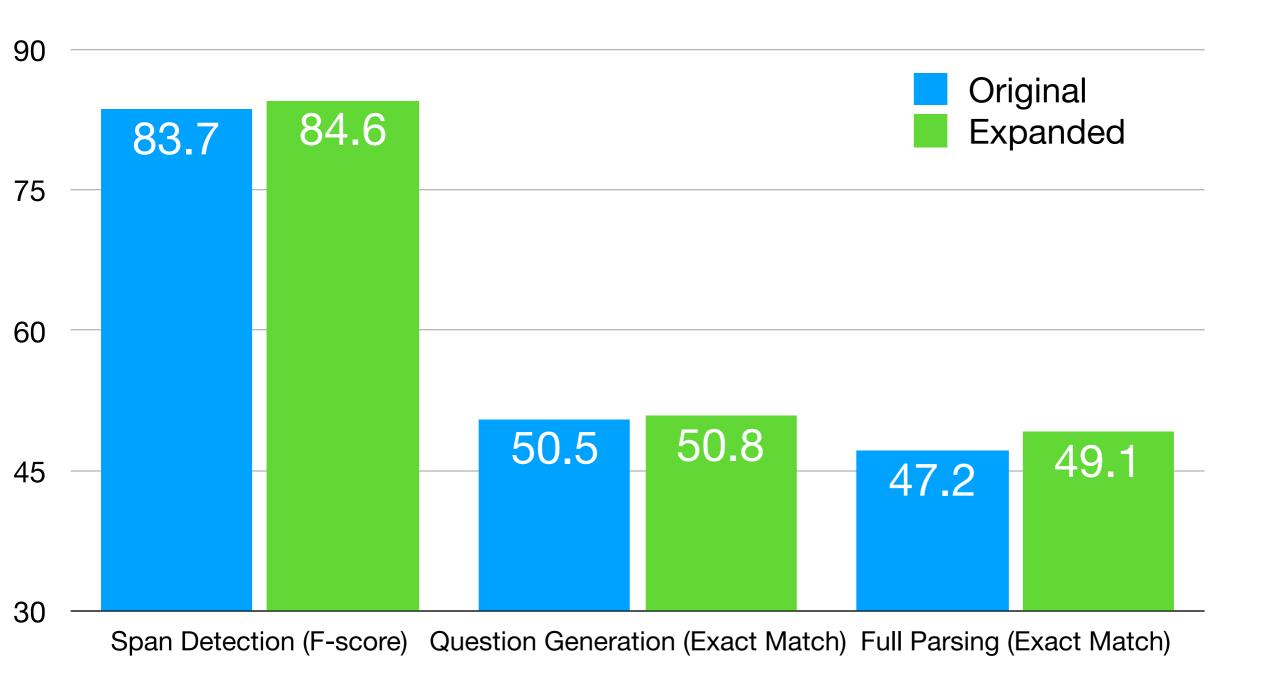
**Validation** 

Workers are shows questions, provide answers or mark as invalid

## Data Expansion

- Overgenerate questions with low span threshold
- Validate
- +46,715 valid QA-pairs (~11% after filtering)

## Data Expansion



#### Evaluation

- Exact Match for Question Generation is overly harsh
  - Paraphrasing

"Who ate something?

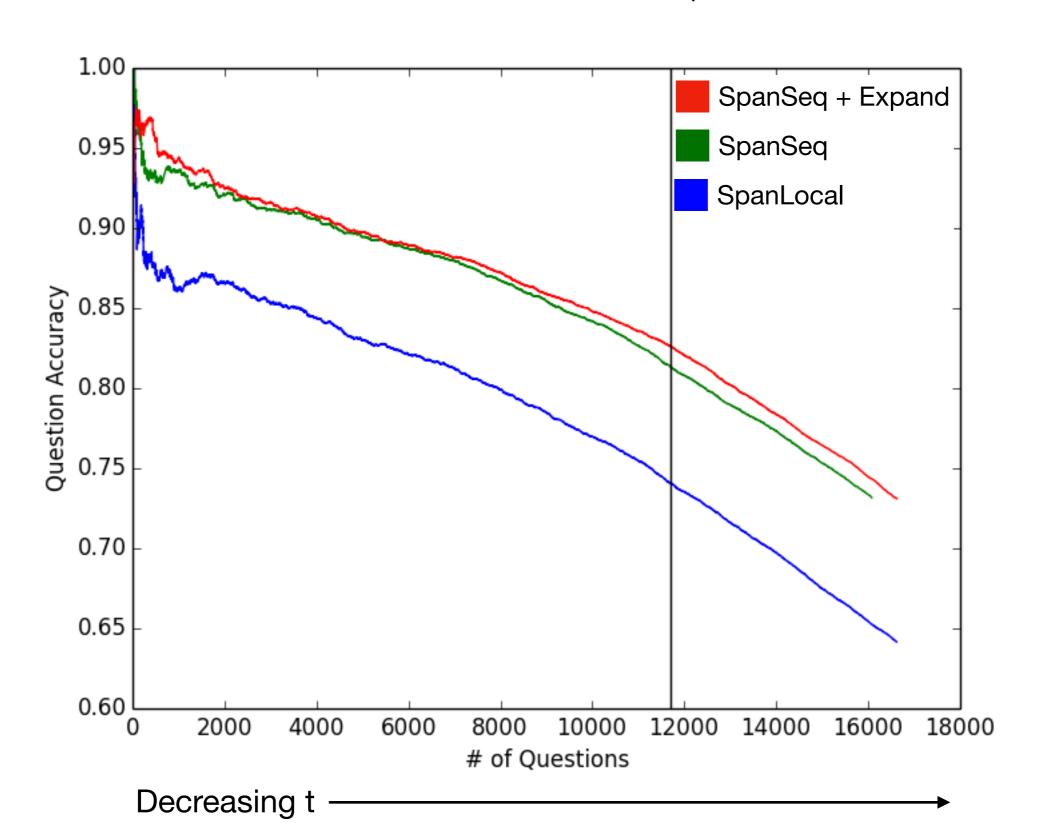
"Who was something eaten by?

- Missing Questions
  - Penalizes correct predictions missing from data

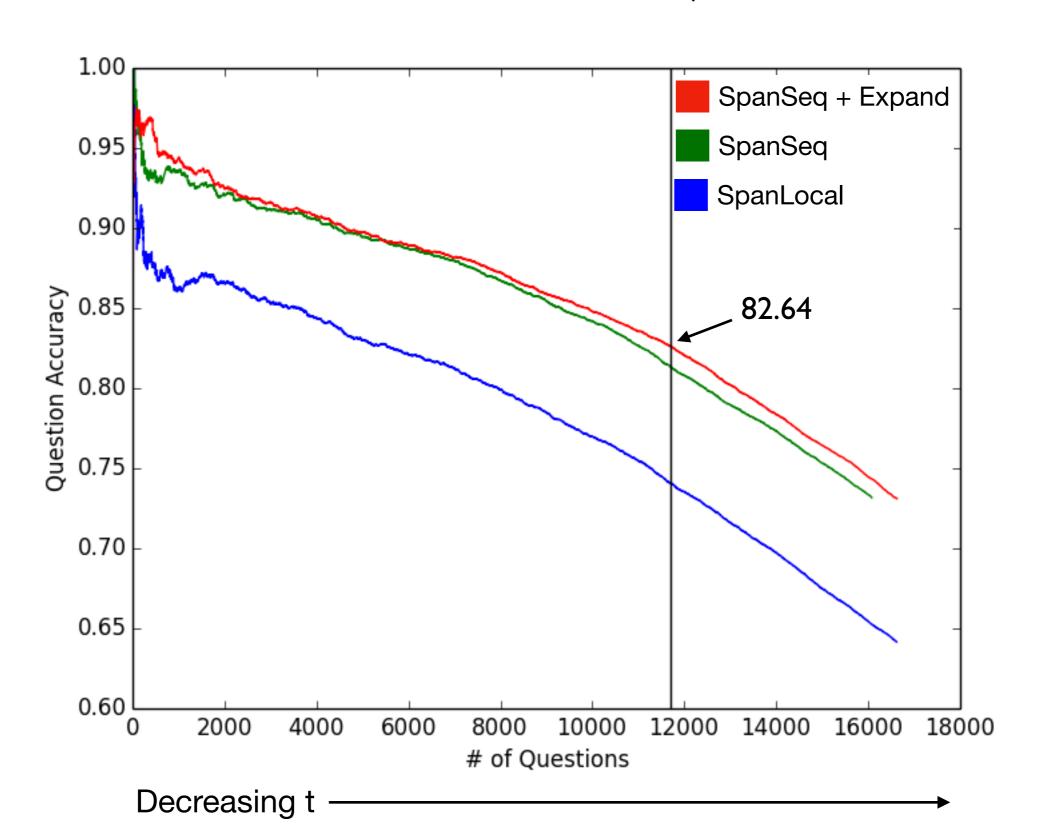
#### Human Evaluation

- Validate model predictions with 6 annotators
- Generated Question valid if 5 out of 6 annotators provided answers
- Predicted span correct if exactly matches any annotator's answer

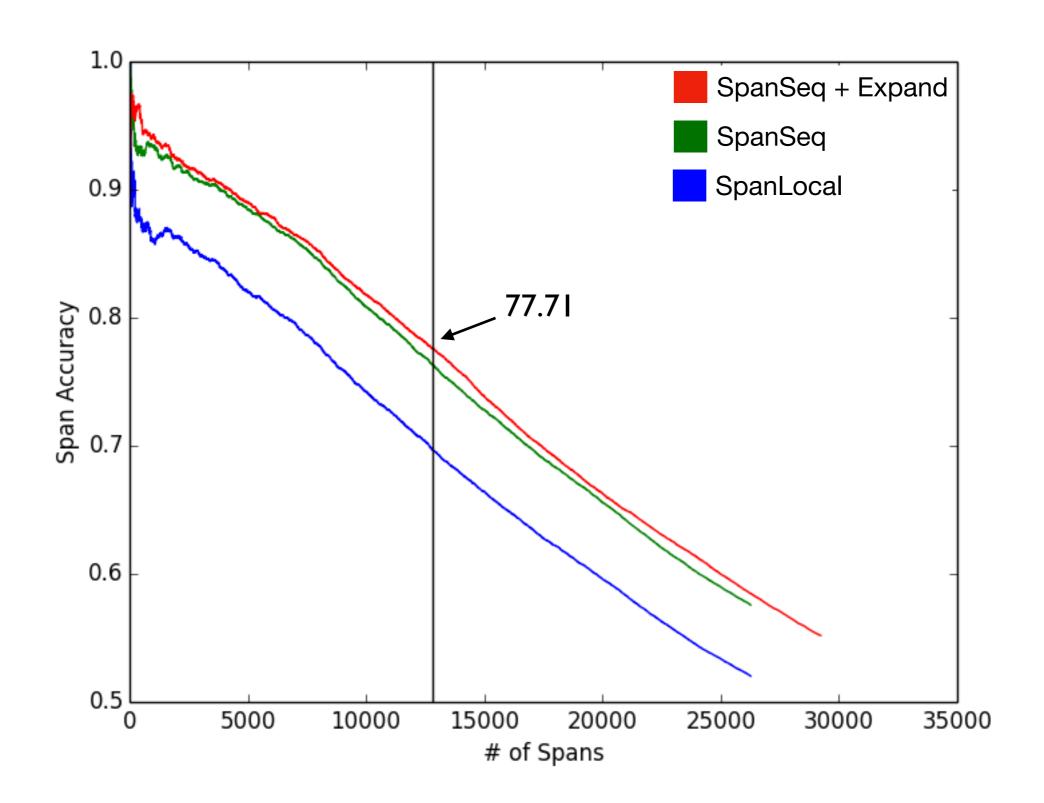
#### Human Eval - Questions



#### Human Eval - Questions



### Human Eval - Arguments



### Example Output

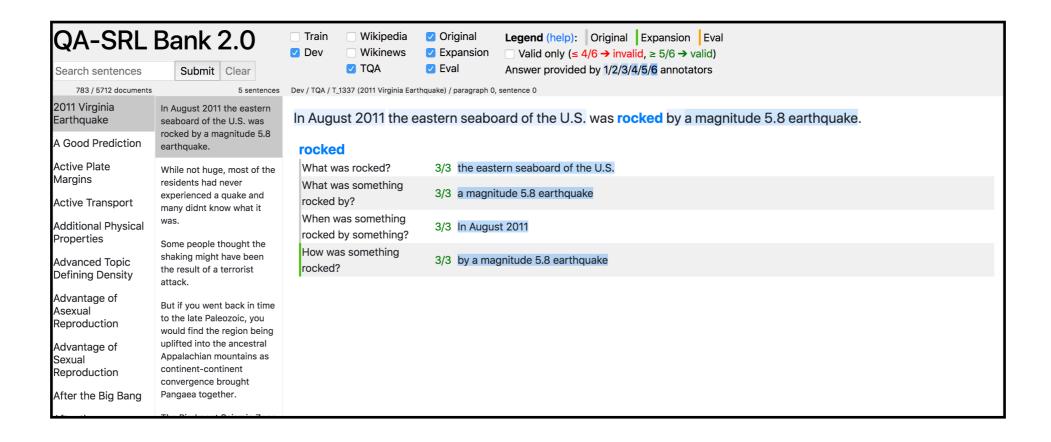
Some of the vegetarians he **met** were members of the Theosophical Society, which had been **founded** in 1875 to further universal brotherhood, and which was **devoted** to the study of Buddhist and Hindu literature.

	Who met someone?	Some of the vegetarians	
met	Who met?	he	
	What did someone meet?	members of the Theosophical Society	
founded	What had been founded?	members of the Theosophical Society	
	What had been founded?	the Theosophical Society	
	When was something founded?	in 1875	
	Why has something been founded?	to further universal brotherhood	
devoted	What was devoted to something?	members of the Theosophical Society	
	What was something devoted to?	the study of Buddhist and Hindu literature	

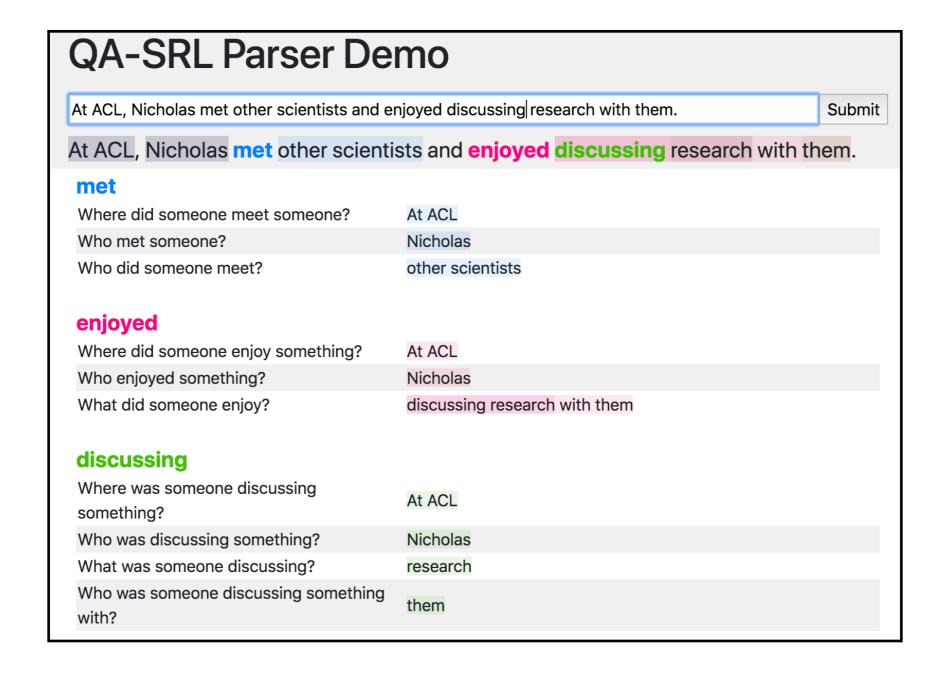
#### Conclusion

- Large crowdsourced dataset of QA-SRL annotations
- High quality QA-SRL Parser
- Techniques for data expansion

# Code and Data available at: <a href="http://qasrl.org">http://qasrl.org</a>



# Code and Data available at: <a href="http://demo.qasrl.org">http://demo.qasrl.org</a>



#### Thank You!









http://qasrl.org/



