Proposition Extraction
Formulation, Crowdsourcing and Prediction

Gabi Stanovsky

Introduction

What, How and Why

Propositions

- Statements for which a truth value can be assigned
  - Bob loves Alice
  - Bob gave a note to Alice

- A single predicate operating over arbitrary number of arguments
  - **loves**: (Bob, Alice)
  - **gave**: (Bob, a note, to Alice)

- Primary (atomic) unit of information conveyed in texts

Proposition Extraction

Barack Obama, the 44th U.S. president, was born in Hawaii

- Barack Obama **is** the 44th U.S. president
- Barack Obama **was born** in Hawaii
- The 44th U.S. president **was born** in Hawaii

Representations

SRL

Barack Obama, the 44th U.S. president, was born in Hawaii

ARG0 | Born-01 | LOC

Open IE

(Barack Obama, **is**, the 44th U.S. president)
(Barack Obama, **was born**, in Hawaii)
(the 44th U.S. president, **was born**, in Hawaii)

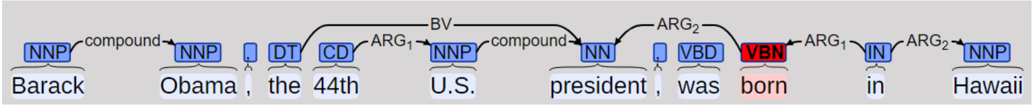
AMR

```
(b1 / born-01
    :ARG0 (p / person
        :name (n / name
            :op1 "Barack"
            :op2 "Obama")
        :ARG0-of (p / preside-01
            :ARG1 (s / state   :wiki "U.S.")
            :NUM  (q / quant :value "44th")
    :LOC (s / state
        :wiki "Hawaii")
```

Neo-Davidsonian

∃e **born**(e1) & Agent(e1, **Barack Obama**))
& LOC(e1, **Hawaii**)
∃e2 **preside**(e2) & Agent(e2, **Barack Obama**)
& Theme(e2, **U.S.**) & Count(e2, **44th**)

MRS



Why?

Useful in a variety of applications

- **Summarization**
  Toward Abstractive Summarization Using Semantic Representations
  Liu et al., NAACL 2015

- **Knowledge Base Completion**
  Leveraging Linguistic Structure For Open Domain Information Extraction
  Angeli et al., ACL 2015

- **Question Answering**
  Using Semantic Roles to Improve Question Answering
  Shen and Lapata, EMNLP 2007

But…

“I train an end-to-end deep bi-LSTM directly over word embeddings”

And yet…

Structured knowledge can help neural architectures

- **Lexical Semantics**
  Improving Hypernymy Detection with an Integrated Path-based and Distributional Method
  Shwartz et al., ACL 2016

- **Semantic Role Labeling**
  Neural semantic role labeling with dependency path embeddings
  Roth and Lapata, ACL 2016

- **Machine Translation**
  Towards String-to-Tree Neural Machine Translation
  Aharoni and Goldberg, ACL 2017

My Research Questions

1. **Foundations**
   What are the desired requirements from proposition extraction?
   - *Specifying and Annotating Reduced Argument Span Via QA-SRL, ACL 2016*
   - *Getting More Out Of Syntax with PropS*

2. **Annotation**
   Can we scale annotations through crowdsourcing?
   - *Annotating and Predicting Non-Restrictive Noun Phrase Modifications, ACL 2016*
   - *Creating a Large Benchmark for Open Information Extraction, EMNLP 2016*

3. **Applications**
   How can we effectively predict proposition structures?
   - *Recognizing Mentions of Adverse Drug Reaction in Social Media Using Knowledge-Infused Recurrent Models, EACL 2017*
   - *Porting an Open Information Extraction System from English to German, EMNLP 2016*
   - *Open IE as an Intermediate Structure for Semantic Tasks, ACL 2015*

Outline

- **Non-restrictive modification**
  - Crowdsourcing
  - Prediction with CRF

- **Supervised Open Information Extraction**
  - Formalizing
  - Automatic creation of large gold corpus
  - Modeling with bi-LSTMs

- **Next steps**

Non-Restrictive Modification

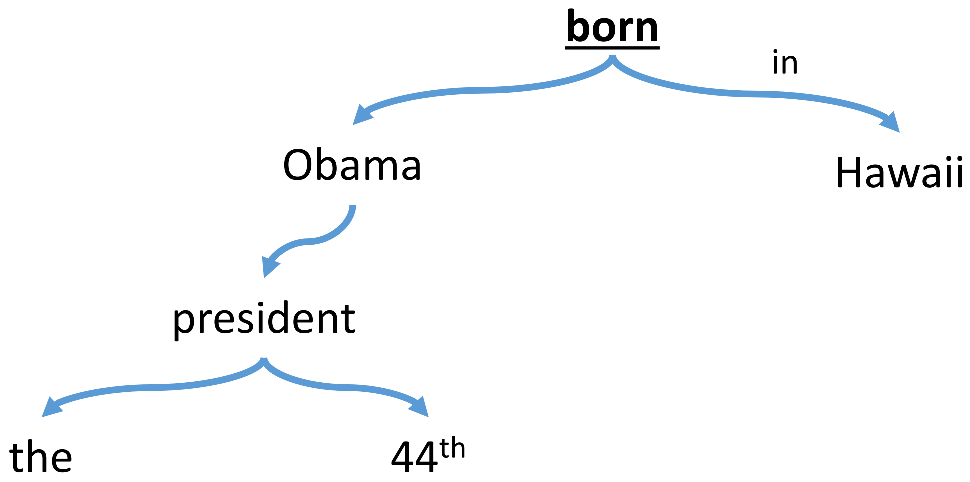
Argument Span

Obama, the 44^{th} president, was <u>born</u> in Hawaii

- Arguments are typically perceived as answering **role questions**
  - Who was *<u>born</u>* somewhere?
  - Where was someone *<u>born</u>*?
- Implicit in most annotations
- QA-SRL annotates with explicit role questions

Argument Span: The Inclusive Approach

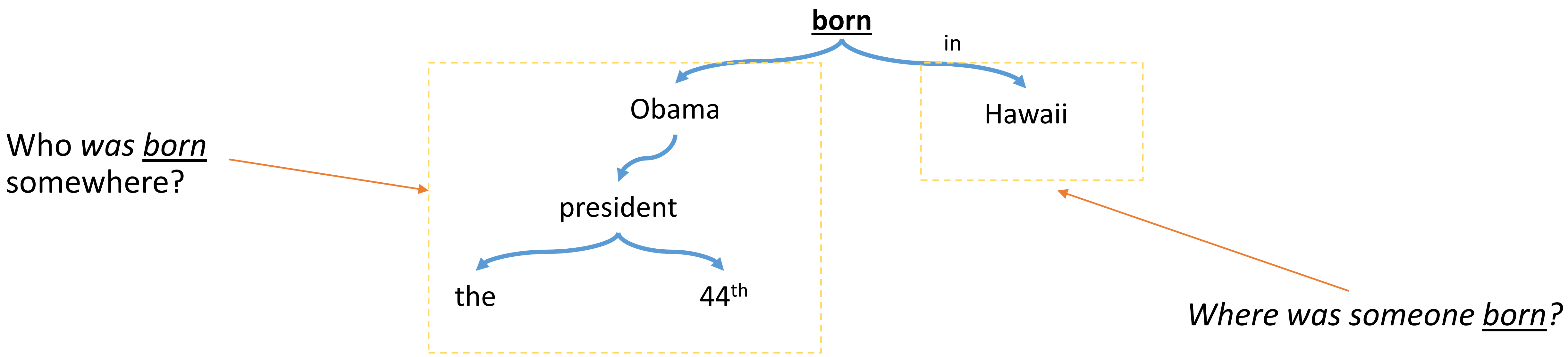
- Arguments are full syntactic constituents



- PropBank
- FrameNet
- AMR

Argument Span: The Inclusive Approach

- Arguments are full syntactic constituents



- PropBank
- FrameNet
- AMR

Can we go shorter?

$\boxed{\textit{Obama,}}$ *the 44th president, was <u>born</u> in Hawaii*

Who *was <u>born</u>* somewhere?

- More concise, yet sufficient answer

Motivation: Applications

- Sentence Simplification

Barack Obama~~, the 44th president,~~ thanked ~~vice president~~ Joe Biden and Hillary Clinton~~, the secretary of state~~

- Knowledge Base Completion

  Angeli et al. , ACL 2015

- Text Comprehension

  Stanovsky et al, ACL 2015

Different types of NP modifications
(from Huddleston et.al)

- Restrictive modification
  - An **integral part** of the meaning of the containing clause
- Non-restrictive modification
  - Presents **separate or additional information**

	Restrictive	Non-Restrictive
Relative Clause	She took the necklace **that her mother gave her**	The speaker thanked president Obama **who just came back from Russia**
Infinitives	People **living near the site** will have to be evacuated	Assistant Chief Constable Robin Searle, **sitting across from the defendant**, said that the police had suspected his involvement since 1997.
Appositives		Keeping the Japanese happy will be one of the most important tasks facing **conservative leader** Ernesto Ruffo
Prepositional modifiers	the kid **from New York** rose to fame	Franz Ferdinand **from Austria** was assassinated om Sarajevo
Postpositive adjectives	George Bush's **younger brother** lost the primary	Pierre Vinken, **61 years old**, was elected vice president
Prenominal adjectives	The **bad** boys won again	The water rose a **good** 12 inches

Goals

- Create a **large corpus** annotated with non-restrictive NP modification
  - Consistent with gold dependency parses
  - Crowdsourceable with good agreement levels

- **Automatic prediction** of non-restrictive modifiers
  - Enabled by the new corpus

Previous work

- Rebanking CCGbank for Improved NP Interpretation
  (Honnibal, Curran and Bos, 2010)
  - Added automatic non-restrictive annotations to the CCGbank
  - Simple implementation
    - Non restrictive modification ←→ The modifier is preceded by a comma
  - No intrinsic evaluation

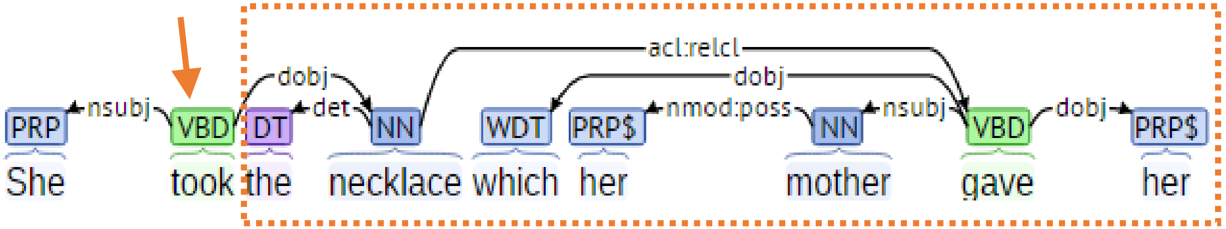
Previous work

- Relative Clause Extraction for Syntactic Simplification
  (Dornescu et al., 2014)
  - Conflated argument span and non-restrictive annotation
    - Span agreement - **54.9% F1**
    - Restrictiveness agreement - **0.51 kappa (moderate)**
  - Develop rule based and ML baselines (CRF with chunking feat.)
    - **Both performing around ~47% F1**

Our Approach

Syntax-consistent QA based classification

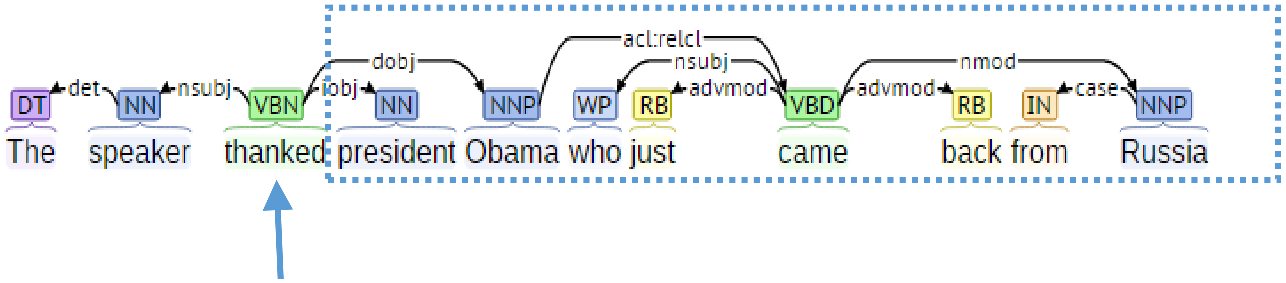
1. Traverse from predicate to NP argument
2. Phrase an argument role question answered by the NP (*what? who? to whom?*)
3. Omitting the modifier still provides the same answer?



What did someone take?

X The necklace ~~which her mother gave her~~

Restrictive



Who was thanked by someone?

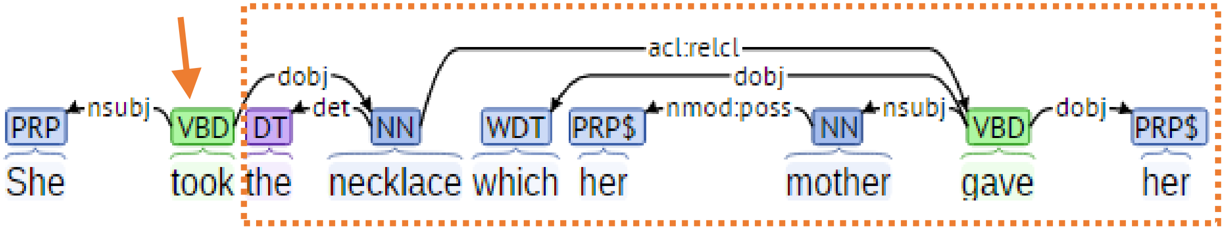
V President Obama ~~who just came back from Russia~~

Non-restrictive

Our Approach

Syntax-consistent QA based classification

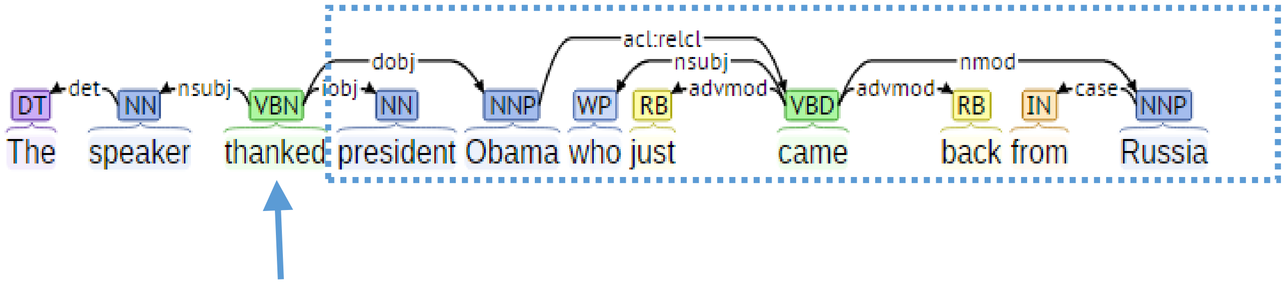
1. Traverse from predicate to NP argument
2. Phrase an argument role question answered by the NP (*what? who? to whom?*)
3. Omitting the modifier still provides the same answer?



What did someone take?

X The necklace ~~which her mother gave her~~

Restrictive



Who was thanked by someone?

V President Obama ~~who just came back from Russia~~

Non-restrictive

Our Approach

1. Can be effectively annotated by non-experts
   - Doesn’t require any linguistic knowledge
   - Language independent (hopefully)

1. Focuses on restrictiveness
   - Doesn’t require span annotation

Corpus

- CoNLL 2009 dependency corpus
  - We can borrow most role questions from QA-SRL
- Each NP is annotated on Mechanical Turk
  - Five annotators for 5c each
  - Consolidation by majority vote

Corpus Analysis

	#instances	%Non-Restrictive	Agreement (K)
Prepositions	693	36%	61.65
Prepositive adjectival modifiers	677	41%	74.7
Appositions	342	73%	60.29
Non-Finite modifiers	279	68%	71.04
Prepositive verbal modifiers	150	69%	100
Relative Clauses	43	79%	100
Postpositive adjectival modifiers	7	100%	100
Total	**2191**	**51.12%**	**73.79**

Corpus Analysis

	#instances	%Non-Restrictive	Agreement (K)
Prepositions	**693**	**36%**	**61.65**
Prepositive adjectival modifiers	677	41%	74.7
Appositions	**342**	**73%**	**60.29**
Non-Finite modifiers	279	68%	71.04
Prepositive verbal modifiers	150	69%	100
Relative Clauses	43	79%	100
Postpositive adjectival modifiers	7	100%	100
Total	**2191**	**51.12%**	**73.79**

→ Prepositions and appositions are harder to annotate

Corpus Analysis

	#instances	%Non-Restrictive	Agreement (K)
Prepositions	693	36%	61.65
Prepositive adjectival modifiers	677	41%	74.7
Appositions	342	73%	60.29
Non-Finite modifiers	279	68%	71.04
Prepositive verbal modifiers	150	69%	100
Relative Clauses	43	79%	100
Postpositive adjectival modifiers	7	100%	100
Total	**2191**	**51.12%**	**73.79**

→ The corpus is fairly balanced between the two classes

Predicting non-restrictive modification

- CRF features:
  - Dependency relation
  - NER
    - Named entity modification tends to be non-restrictive
  - Word embeddings
    - Contextually similar words ←→ similar restrictiveness value
  - Linguistically motivated features
    - The word preceding the modifier (Huddleston)

Results

Modifier Type	#	Precision			Recall			F1		
		Honnibal	Dornescu	Our	Honnibal	Dornescu	Our	Honnibal	Dornescu	Our
Prepositional	135	.83	.67	.69	.1	.16	.41	.18	.26	.51
Adjectival	111	.33	.38	.59	.06	.06	.21	.11	.11	.31
Appositive	78	.77	.81	.82	.34	.93	.98	.47	.87	.89
Non-Finite	55	.77	.63	.64	.29	.97	.97	.42	.76	.77
Verbal	20	0	.75	.75	0	1	1	0	.86	.86
Relative clause	13	1	.85	.85	.27	1	1	.43	.92	.92
Total	412	.72	.72	**.73**	.19	.58	**.68**	.3	.64	**.72**

Error Analysis

Modifier Type	#	Precision			Recall			F1		
		Honnibal	Dornescu	Our	Honnibal	Dornescu	Our	Honnibal	Dornescu	Our
Prepositional	135	.83	.67	.69	.1	.16	.41	.18	.26	.51
Adjectival	111	.33	.38	.59	.06	.06	.21	.11	.11	.31
Appositive	78	.77	.81	.82	.34	.93	.98	.47	.87	.89
Non-Finite	55	.77	.63	.64	.29	.97	.97	.42	.76	.77
Verbal	20	0	.75	.75	0	1	1	0	.86	.86
Relative clause	13	1	.85	.85	.27	1	1	.43	.92	.92
Total	412	.72	.72	**.73**	.19	.58	**.68**	.3	.64	**.72**

Prepositions and adjectives are harder to predict

Error Analysis

Modifier Type	#	Precision			Recall			F1		
		Honnibal	Dornescu	Our	Honnibal	Dornescu	Our	Honnibal	Dornescu	Our
Prepositional	135	.83	.67	.69	.1	.16	.41	.18	.26	.51
Adjectival	111	.33	.38	.59	.06	.06	.21	.11	.11	.31
Appositive	78	.77	.81	.82	.34	.93	.98	.47	.87	.89
Non-Finite	55	.77	.63	.64	.29	.97	.97	.42	.76	.77
Verbal	20	0	.75	.75	0	1	1	0	.86	.86
Relative clause	13	1	.85	.85	.27	1	1	.43	.92	.92
Total	412	.72	.72	**.73**	.19	.58	**.68**	.3	.64	**.72**

Commas are good in precision but poor for recall

Error Analysis

Modifier Type	#	Precision			Recall			F1		
		Honnibal	Dornescu	Our	Honnibal	Dornescu	Our	Honnibal	Dornescu	Our
Prepositional	135	.83	.67	.69	.1	.16	.41	.18	.26	.51
Adjectival	111	.33	.38	.59	.06	.06	.21	.11	.11	.31
Appositive	78	.77	.81	.82	.34	.93	.98	.47	.87	.89
Non-Finite	55	.77	.63	.64	.29	.97	.97	.42	.76	.77
Verbal	20	0	.75	.75	0	1	1	0	.86	.86
Relative clause	13	1	.85	.85	.27	1	1	.43	.92	.92
Total	412	.72	.72	**.73**	.19	.58	**.68**	.3	.64	**.72**

Dornescu et al. performs better on our dataset

Error Analysis

Modifier Type	#	Precision			Recall			F1		
		Honnibal	Dornescu	Our	Honnibal	Dornescu	Our	Honnibal	Dornescu	Our
Prepositional	135	.83	.67	.69	.1	.16	.41	.18	.26	.51
Adjectival	111	.33	.38	.59	.06	.06	.21	.11	.11	.31
Appositive	78	.77	.81	.82	.34	.93	.98	.47	.87	.89
Non-Finite	55	.77	.63	.64	.29	.97	.97	.42	.76	.77
Verbal	20	0	.75	.75	0	1	1	0	.86	.86
Relative clause	13	1	.85	.85	.27	1	1	.43	.92	.92
Total	412	.72	.72	**.73**	.19	.58	**.68**	.3	.64	**.72**

Our system highly improves recall

To conclude this part…

- Large non-restrictive gold standard
  - Directly augmenting dependency trees
- Automatic classifier
  - Improves over state of the art results

Supervised Open Information Extraction

Supervised Open Information Extraction

- **Problem**: No large benchmark for Open IE evaluation!

- **Approach:**
  - Identify common extraction principles
  - Extract a large Open IE corpus from QA-SRL
  - Train a transducer Bi-LSTM

Open Information Extraction

- Extracts SVO tuples from texts
  - Barack Obama, the U.S president, was **born in** Hawaii
    → (Barack Obama, **born in**, Hawaii)

  - Obama and Bush were **born in** America
    → (Obama, **born in**, America), (Bush, **born in**, America)

- Useful for populating large databases
  - A scalable open variant of Information Extraction

Open IE: Many parsers developed

- TextRunner (Banko et al., NAACL 2007)
- WOE (Wu and Weld, ACL 2010)
- ReVerb (Fader et al., 2011)
- OLLIE (Mausam et al., EMNLP 2012)
- KrakeN (Akbik and Luser, ACL 2012)
- ClausIE (Del Corro and Gemulla, WWW 2013)
- Stanford Open Information Extraction (Angeli et al., ACL 2015)
- DEFIE (Bovi et al., TACL 2015)
- Open-IE 4 (Mausam et al., ongoing work)
- PropS-DE (Falke et al., EMNLP 2016)
- NestIE (Bhutani et al., EMNLP 2016)

Problem: Open IE evaluation

- Open IE task formulation has been lacking formal rigor
  - No common guidelines → **No large corpus for evaluation**
- Post-hoc evaluation:
  - Annotators judge *a small sample* of their output

→ **Precision oriented** metrics

→ Figures are **not comparable**

→ Experiments are **hard to reproduce**

Previous evaluations

System	#Sentences	Genre	Metric	#Annot.	Agreement
TextRunner	400	Web	% Correct	3	-
WOE	300	Web, Wiki, News	Precision / Recall	5	-
ReVerb	500	Web	Precision / AUC	2	86%, .68 k
KrakeN	500	Web	% Correct	2	87%
Ollie	300	News, Wiki, Biology	Precision/Yield AUC	2	96%
ClauseIE	300	Web, Wiki, News	Precision/Yield	2	57% / 68% / 63%

→ **Hard to draw general conclusions!**

Solution:

Common Extraction Principles

Large Open IE Benchmark

Supervised Model

Common principles

1. **Open lexicon**

2. **Soundness**
   "Cruz refused to endorse Trump"
   ReVerb: (Cruz; **endorse**; Trump)
   OLLIE: (Cruz; **refused to endorse**; Trump)

3. **Minimal argument span**
   *"Hillary **promised** better education, social plans and healthcare coverage"*
   *ClausIE: (Hillary, **promised**, better education), (Hillary, **promised**, better social plans), (Hillary, **promised**, better healthcare coverage)*

Solution:

Common Extraction Principles

Large Open IE Benchmark

QA-SRL → Open IE

Supervised Model

Open IE vs. SRL vs. QA-SRL

QA-SRL isn't limited to a lexicon

	Open IE	Traditional SRL	QA-SRL
Open lexicon	V	X	V
Consistency	V	V	V
Reduced arguments	V	X	V

QA-SRL format solicits reduced arguments
(Stanovsky et al., ACL 2016)

Converting QA-SRL to Open IE

- Intuition: generate all independent extractions

- Example:
  - "Barack Obama, the newly elected president, **flew** to Moscow on Tuesday"
  - QA-SRL:
    - Who **flew** somewhere? — Barack Obama / the newly elected president
    - Where did someone **fly**? — to Moscow
    - When did someone **fly**? — on Tuesday

    → OIE: (Barack Obama, **flew**, to Moscow, on Tuesday)
    (the newly elected president, **flew**, to Moscow, on Tuesday)

➔ Cartesian product over all answer combinations
  - Special cases for nested predicates, modals, preposition and auxiliaries

Resulting Corpus

Corpus	WSJ	WIKI	All
#Sentences	1241	1959	3200
#Predicates	2020	5690	7710
#Questions	8112	10798	18910
#Extractions	**4481**	**5878**	**10359**

- Validated against an expert annotation of 100 sentences (95% F1)
- 13 times bigger than largest previous OIE corpus (ReVerb)

Follow-up Work

- Analysing Errors of Open Information Extraction Systems
- RelVis: Benchmarking OpenIE Systems
  (Schneider et al., 2017)

- MinIE: Minimizing Facts in Open Information Extraction
  (Gashteovski et al, 2017)

- A Large-scale Evaluation of PredPatt against PropBank
  (Zhang et al, 2017)

Solution:

Common Extraction Principles

Large Open IE Benchmark

Supervised Model

BIO Encoding

May, the British PM, **plans for** Brexit on which the UK has voted for last June

BIO Encoding

Multiple extractions by repeating labels

May$_{\text{A0-B}}$, the$_{\text{A0-B}}$ British$_{\text{A0-I}}$ PM$_{\text{A0-I}}$, **plans**$_{\text{P-B}}$ **for**$_{\text{P-I}}$ Brexit$_{\text{A1-B}}$ on which the UK has voted for last June

→ (May; **plans for**; Brexit)

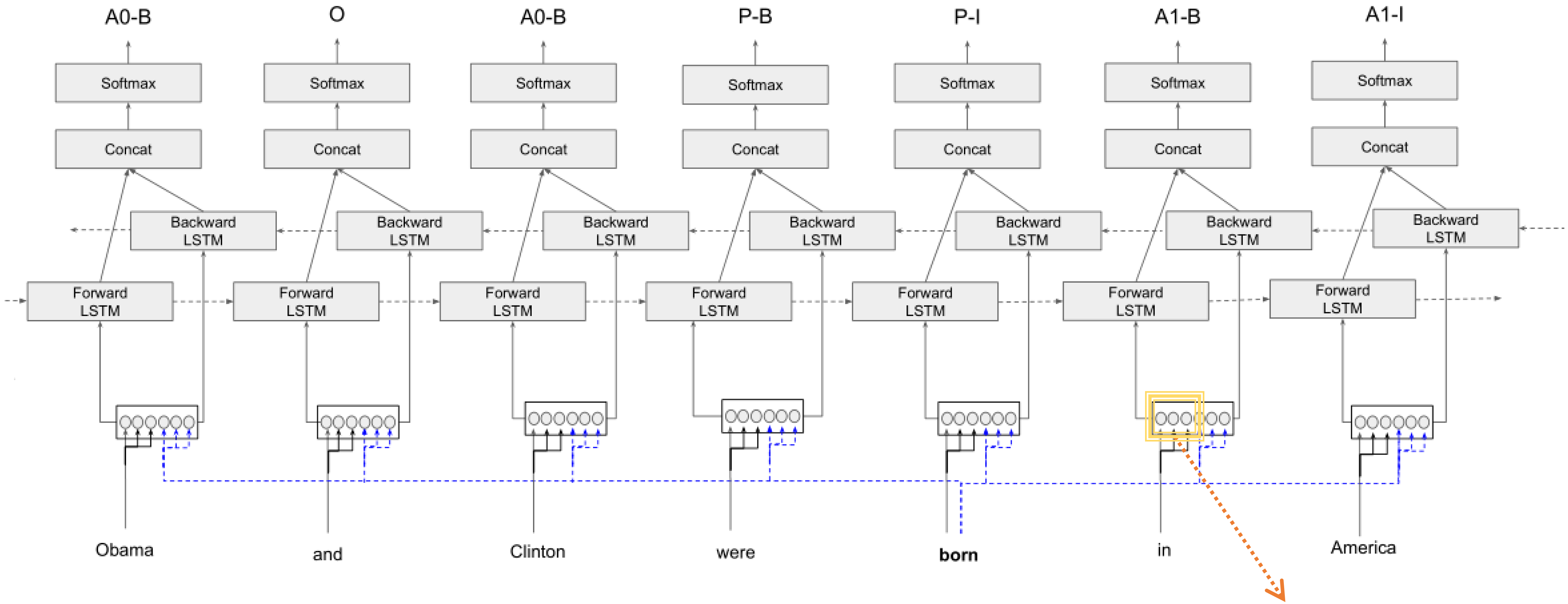
→ (The British PM; **plans for**; Brexit)

the British PM, plans for Brexit$_{\text{A1-B}}$ on which the$_{\text{A0-B}}$ UK$_{\text{A0-I}}$ **has**$_{\text{P-B}}$ **voted**$_{\text{P-I}}$ **for**$_{\text{P-I}}$ last$_{\text{A2-B}}$ June$_{\text{A2-I}}$

→(the UK; **has voted for**; Brexit; last June)

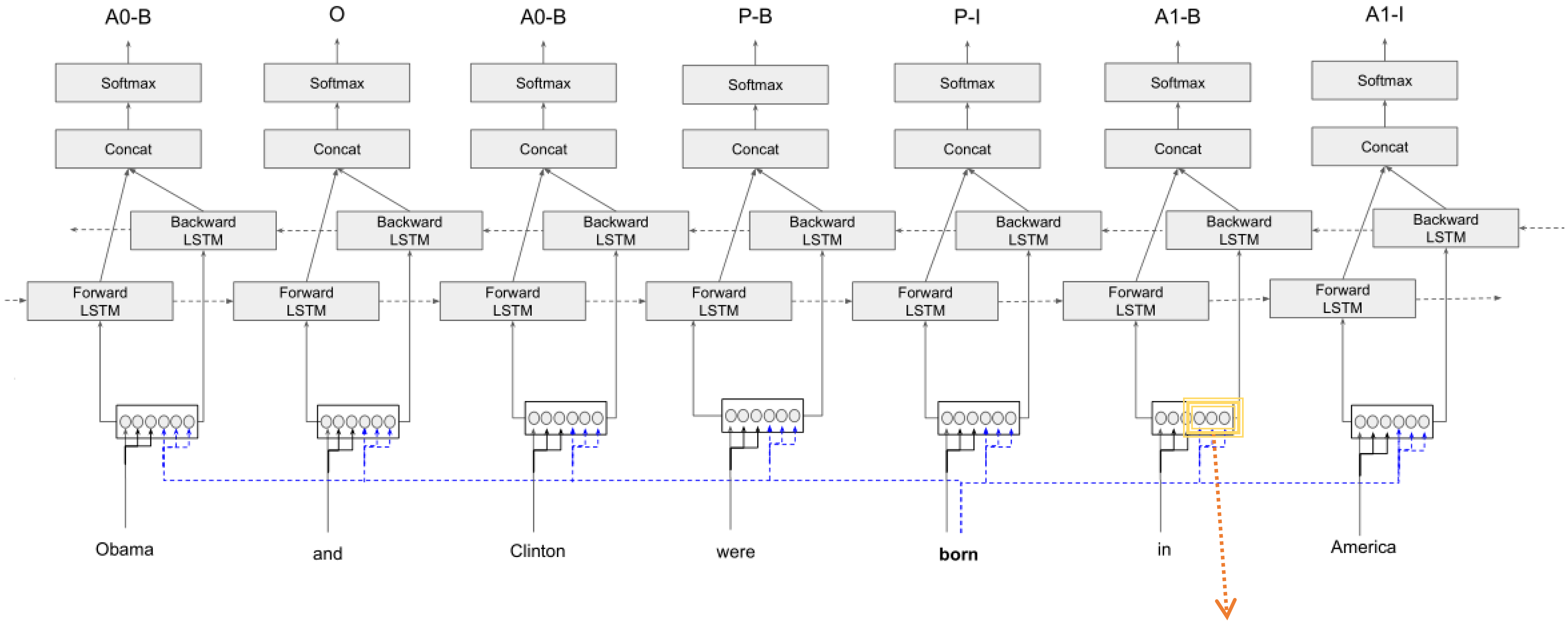
Argument label ≈ Argument role

End to End Model



POS and pretrained word embeddings

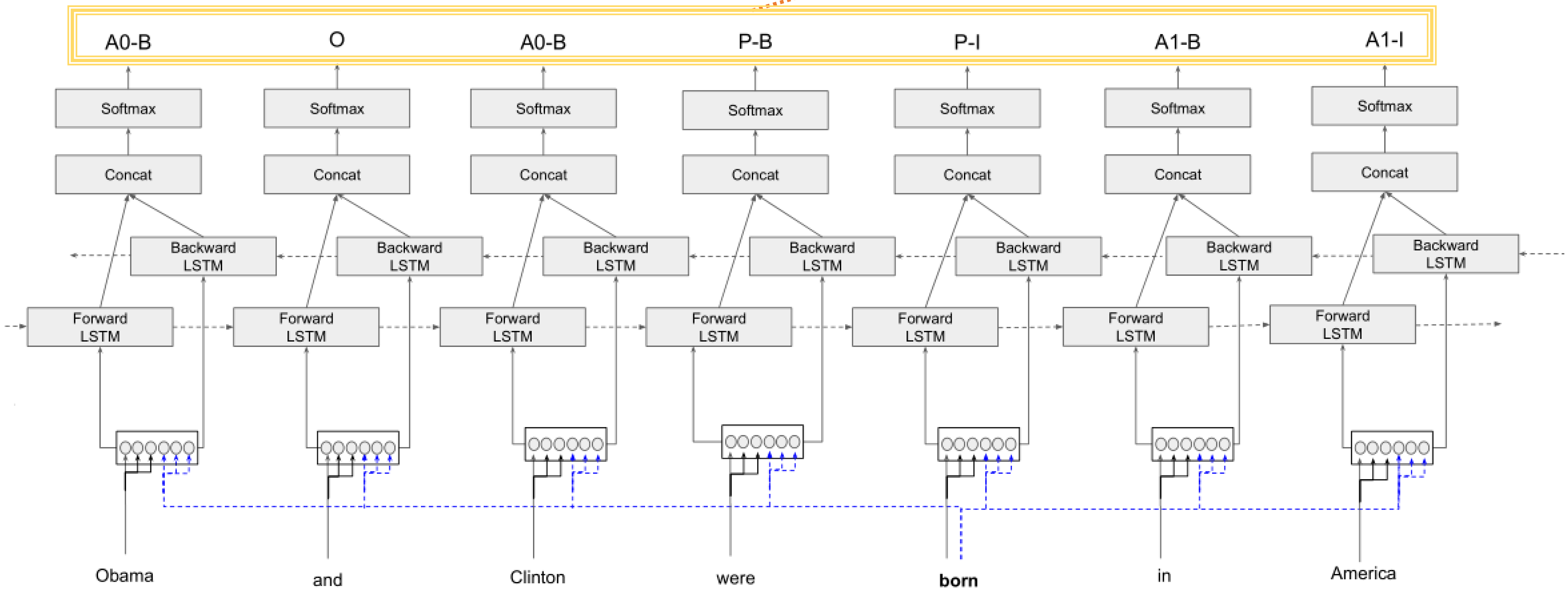
End to End Model



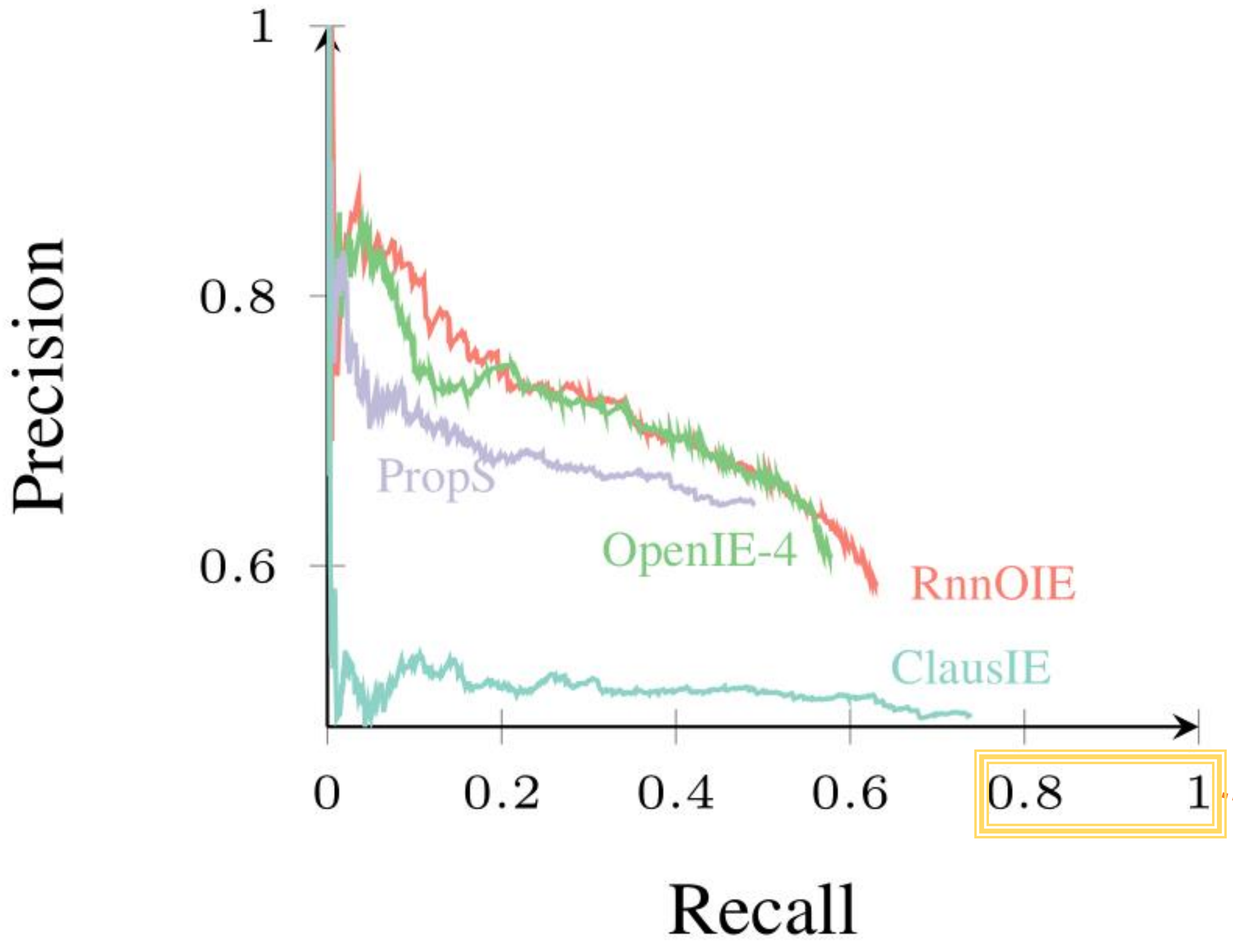
Predicate head concatenated to all word feats

End to End Model

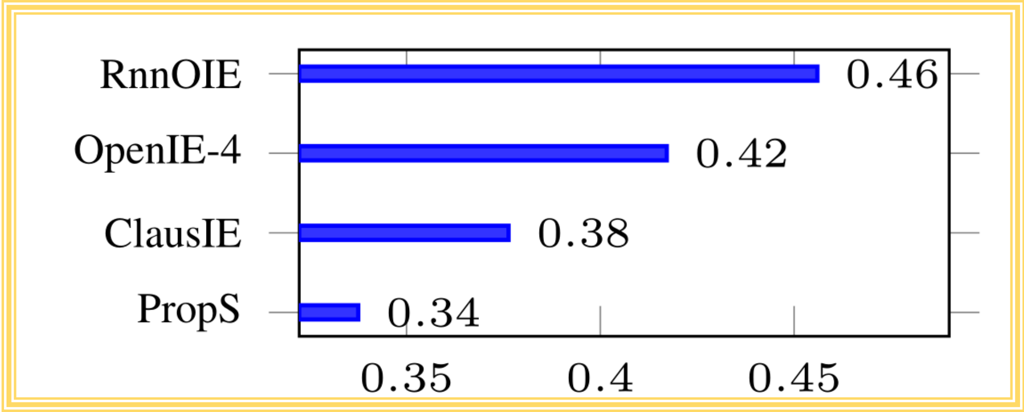
Confidence = Π (word prob)



Evaluation



Area Under the Curve



4 points over previous state of the art

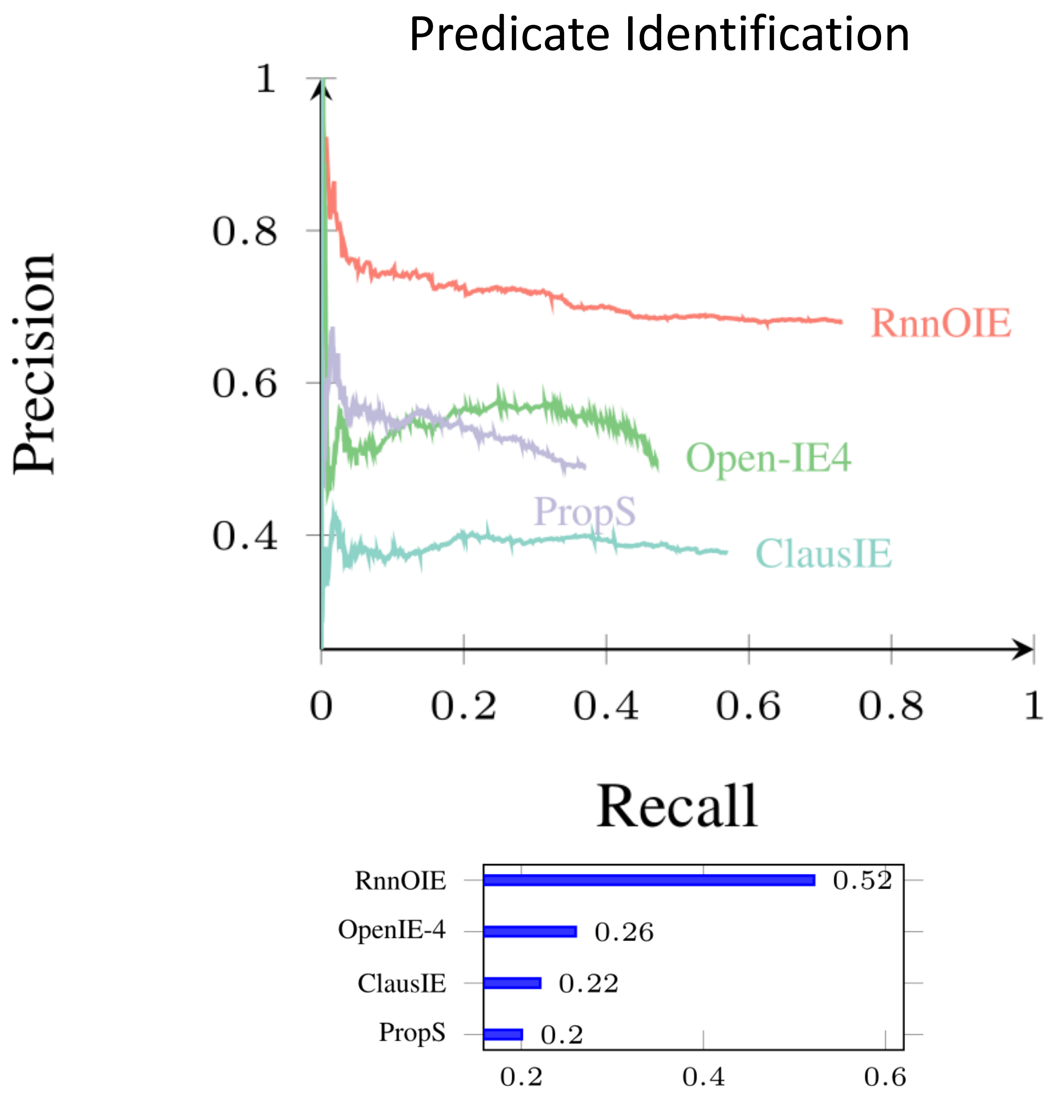
Low recall: Missed long-range dep, pronoun resolution

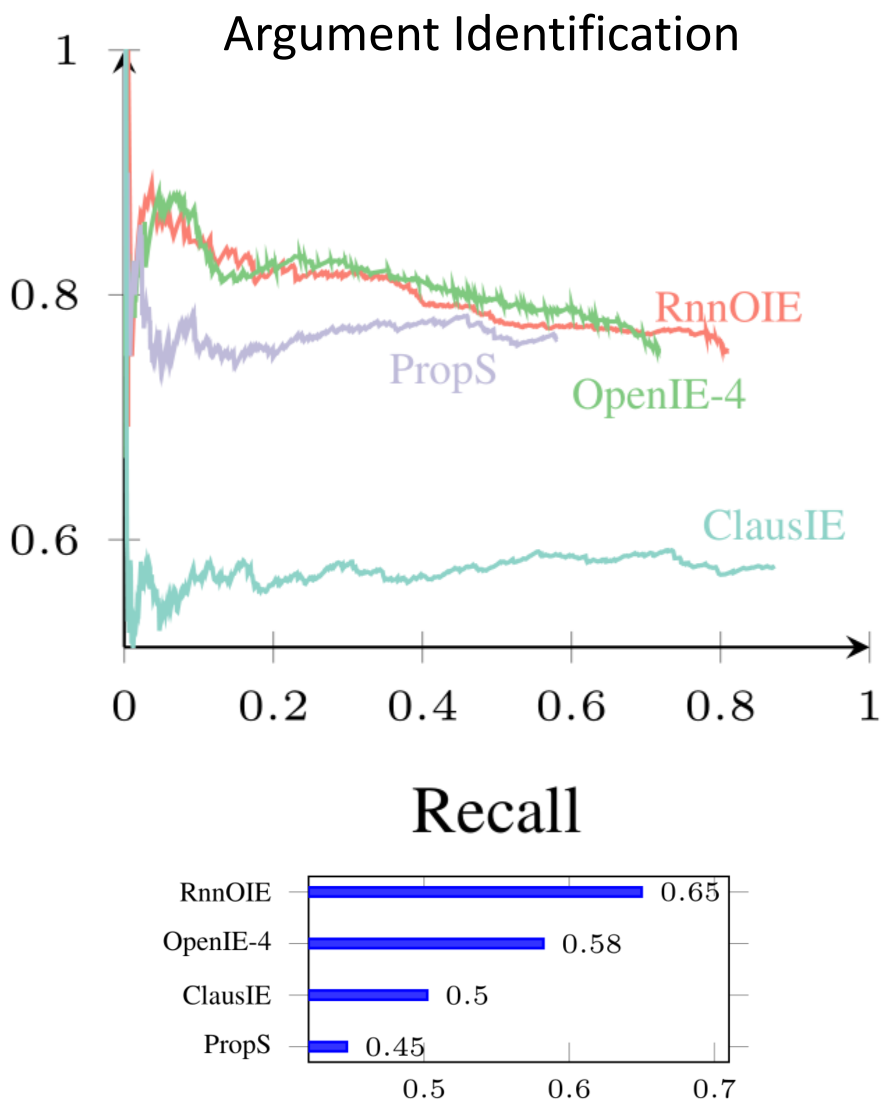
Analysis

System	# Extractions	Arg. per Prop.	Words per Arg
Gold	1730	2.45	26.91
ClausIE	2768	2.00	28.89
Open IE4	1793	3.07	22.75
PropS	1551	2.68	29.00
RnnOIE	1993	3.19	23.40

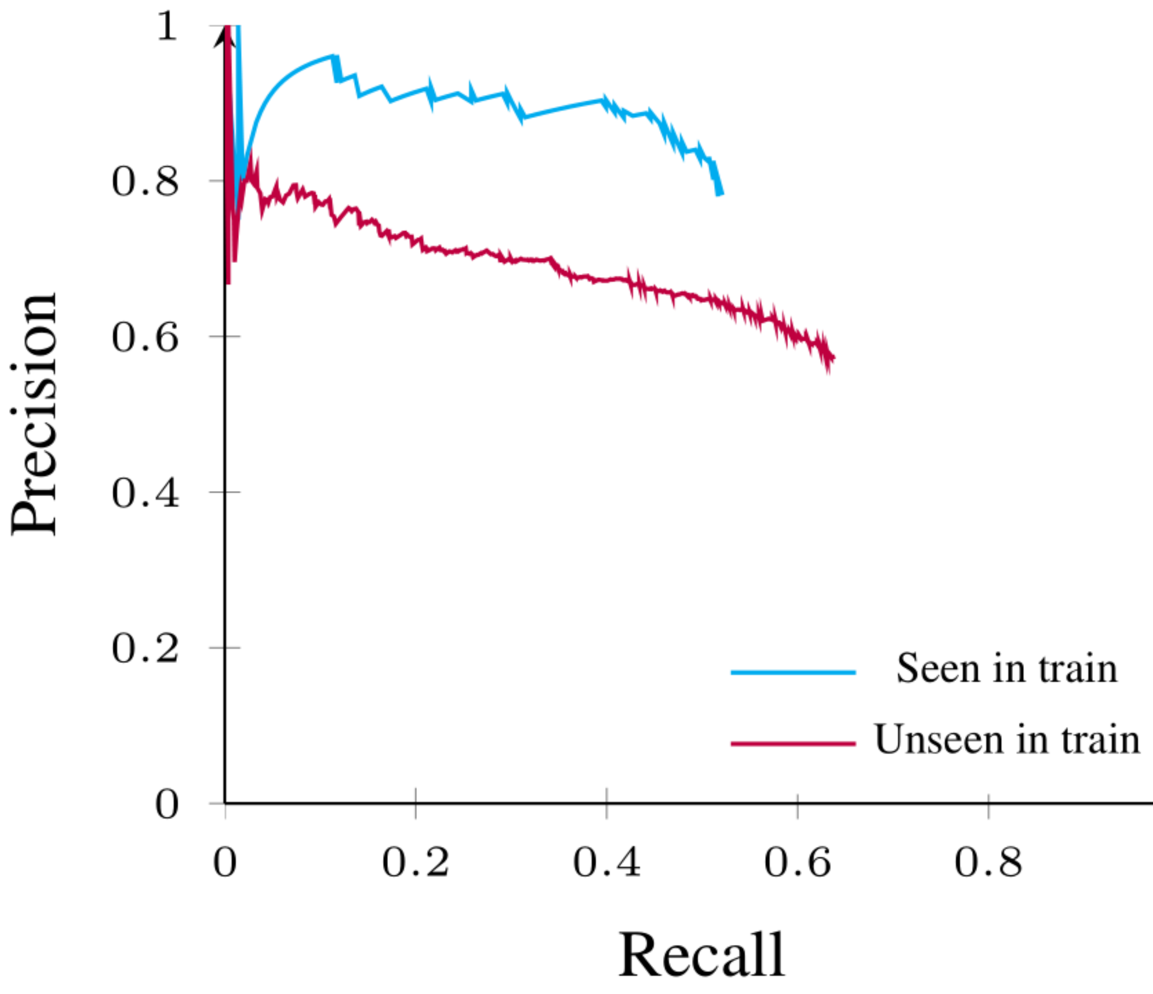
- RnnOIE overproduces and over-shortens arguments

Analysis


Predicate Identification
Precision
Recall
RnnOIE
Open-IE4
PropS
ClausIE
1
0.8
0.6
0.4
0
0.2
0.4
0.6
0.8
1
RnnOIE 0.52
OpenIE-4 0.26
ClausIE 0.22
PropS 0.2
0.2
0.4
0.6


Argument Identification
Recall
RnnOIE
PropS
OpenIE-4
ClausIE
1
0.8
0.6
0
0.2
0.4
0.6
0.8
1
RnnOIE 0.65
OpenIE-4 0.58
ClausIE 0.5
PropS 0.45
0.5
0.6
0.7

Analysis



We generalize for unseen predicates

- 24% of predicates unseen in test

Conclusions

We’ve seen..

- Non-Restrictive modification
  - Crowdsourcing annotations
  - Modeling with CRF
  - Future work:
    - Distributive coordination
- Supervised Open IE
  - Automatically converted corpus
  - Transducer Bi-LSTMs
  - Future work:
    - Better confidence estimation
    - Model improvements

Future Work

- Layered structured representation
  - Integrating various levels of semantic annotations
  - Deep Multitask Learning for Semantic Dependency Parsing, (Peng et al., 2017)
- Crowdsourcing
  - Learning from partial annotations
- Multi-sentence
  - Collapsing co-referring nodes
- Multilingual

Thanks for Listening!