Natural Language Knowledge Representation

Gabi Stanovsky

About me

- Third year PhD student at Bar Ilan University

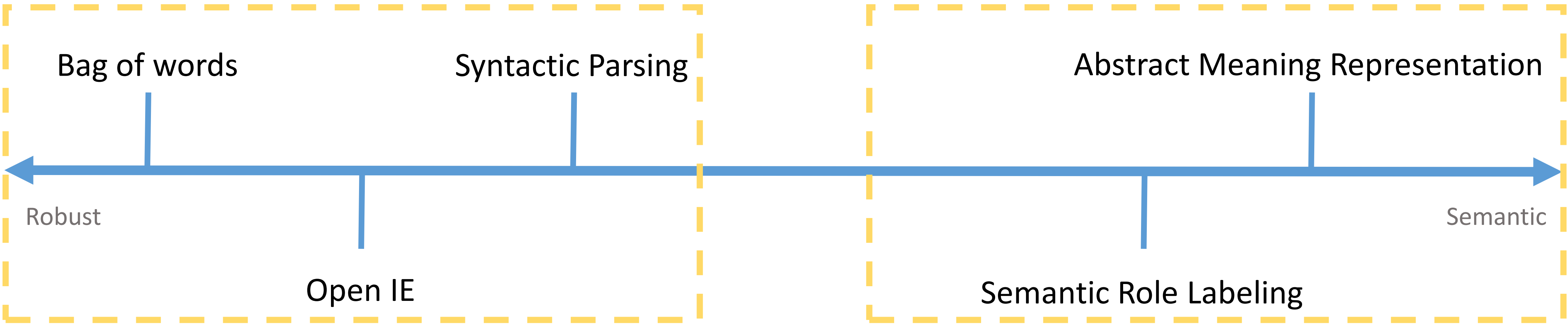
- Advised by Prof. Ido Dagan

- This summer: Intern at IBM Research
- Last Summer: Intern at AI2

Language Representations
A semantic scale

Robust, Scalable

Queryable, Formal



Redundant, Not readily usable

Small Domains, Low accuracy

Inspired by slides from Yoav Artzi

In This Talk

- Explorations of applicability
  - Using Open IE as an intermediate structure

- Finding a better tradeoff
  - PropS
  - Identifying non-restrictive modification

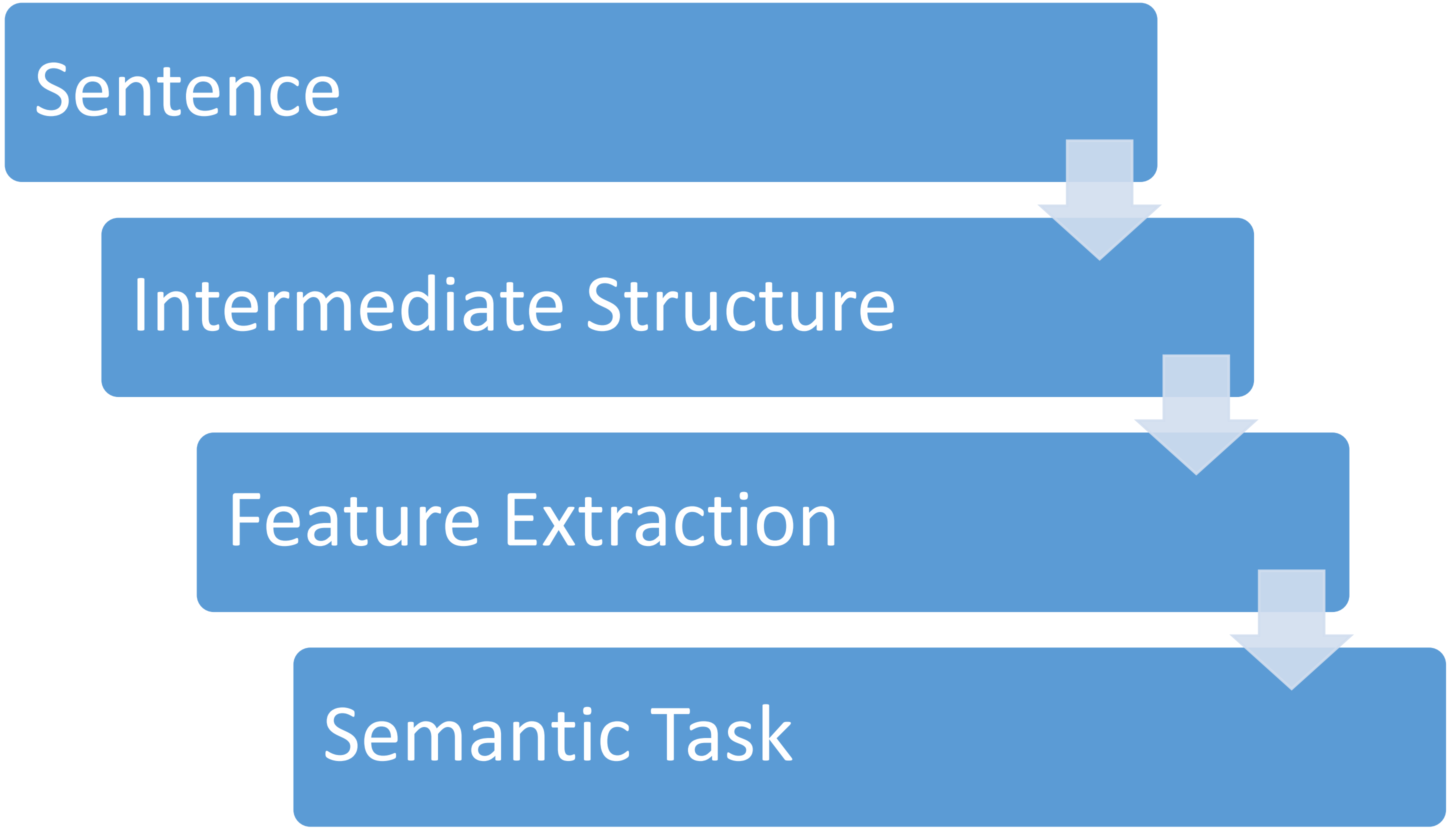
- Evaluations
  - Creating a large benchmark for Open Information Extraction

Open IE as an Intermediate Structure for Semantic Tasks

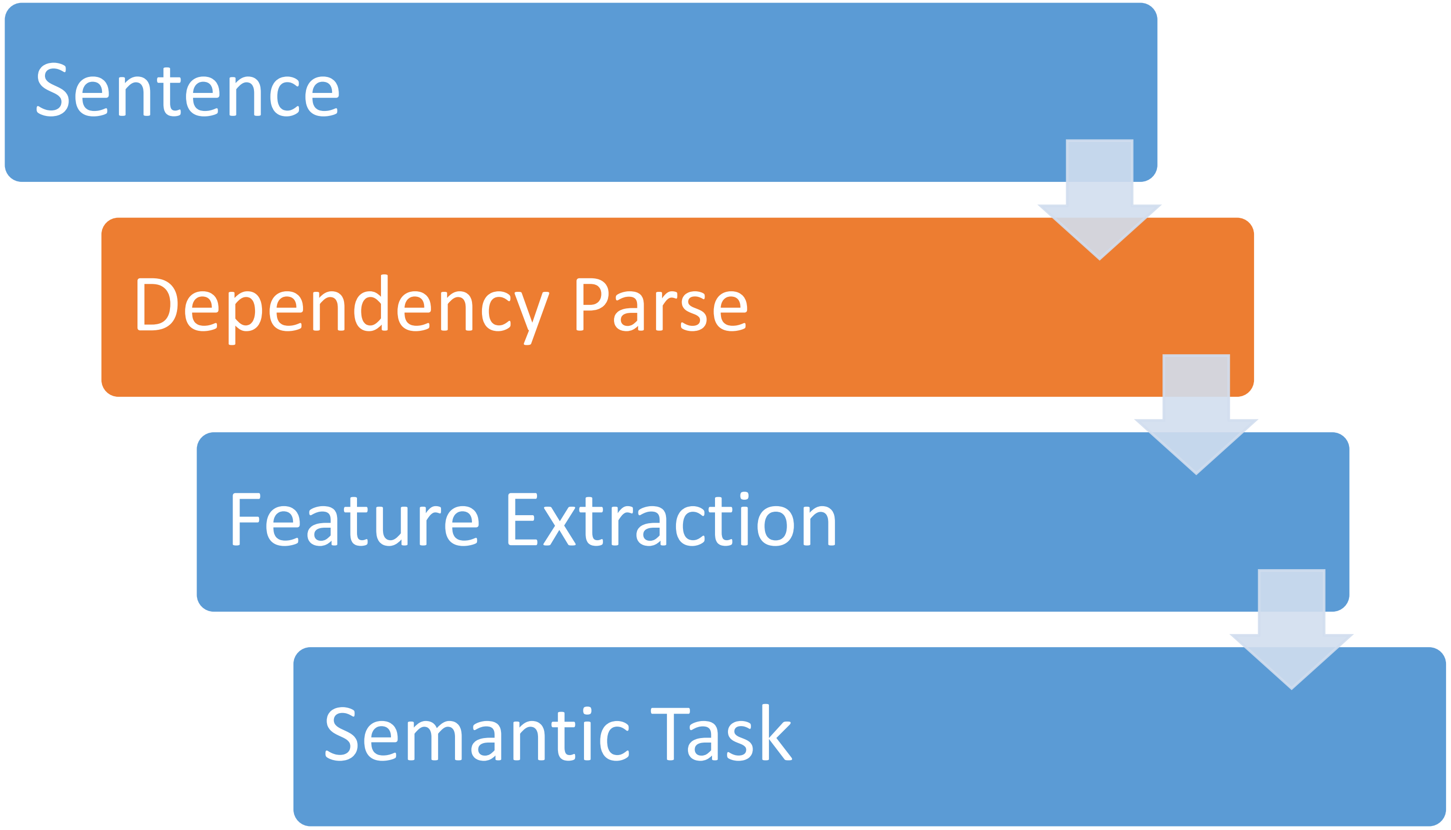
Gabriel Stanovsky, Ido Dagan and Mausam

ACL 2015

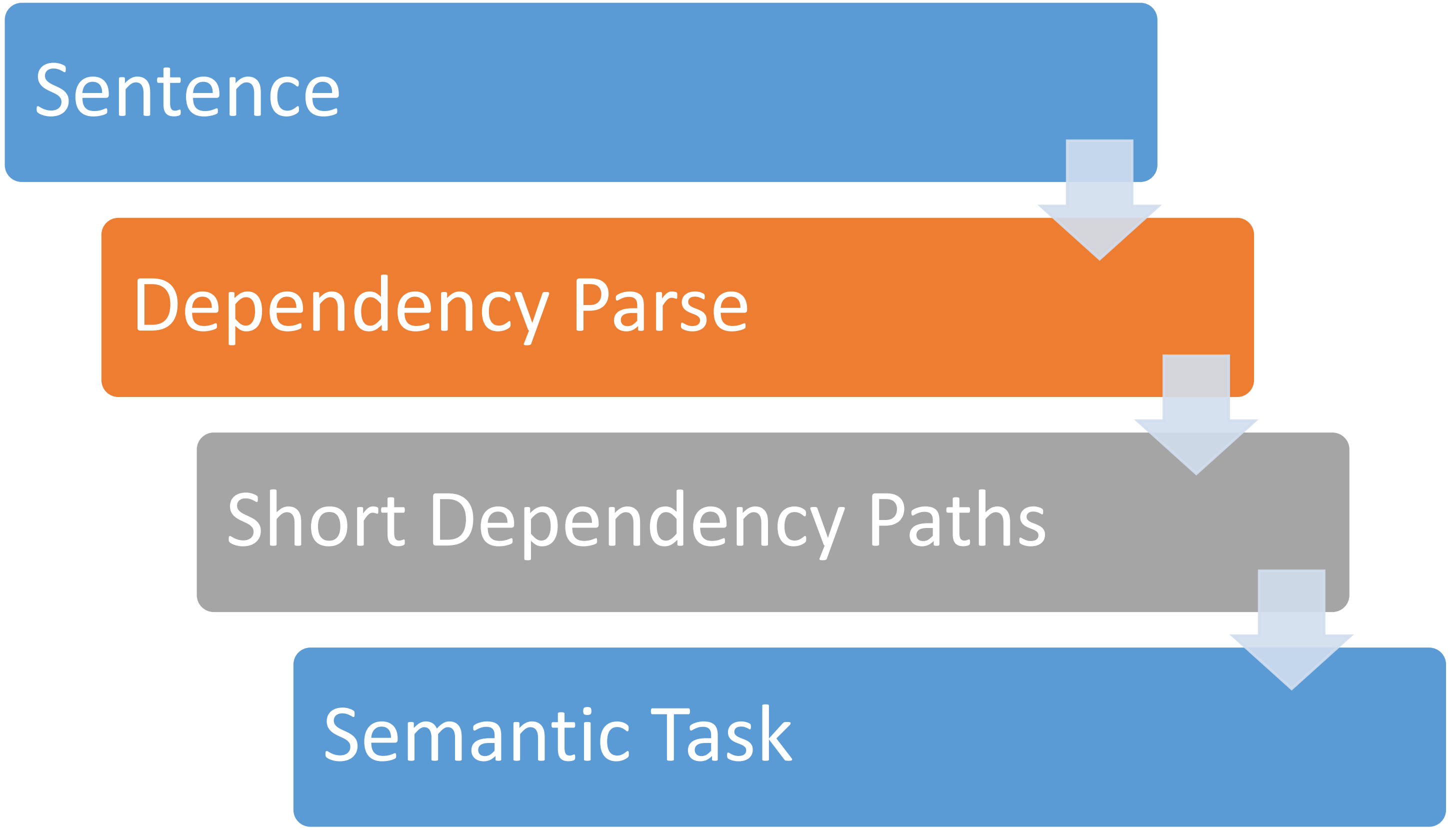
Sentence Level Semantic Application



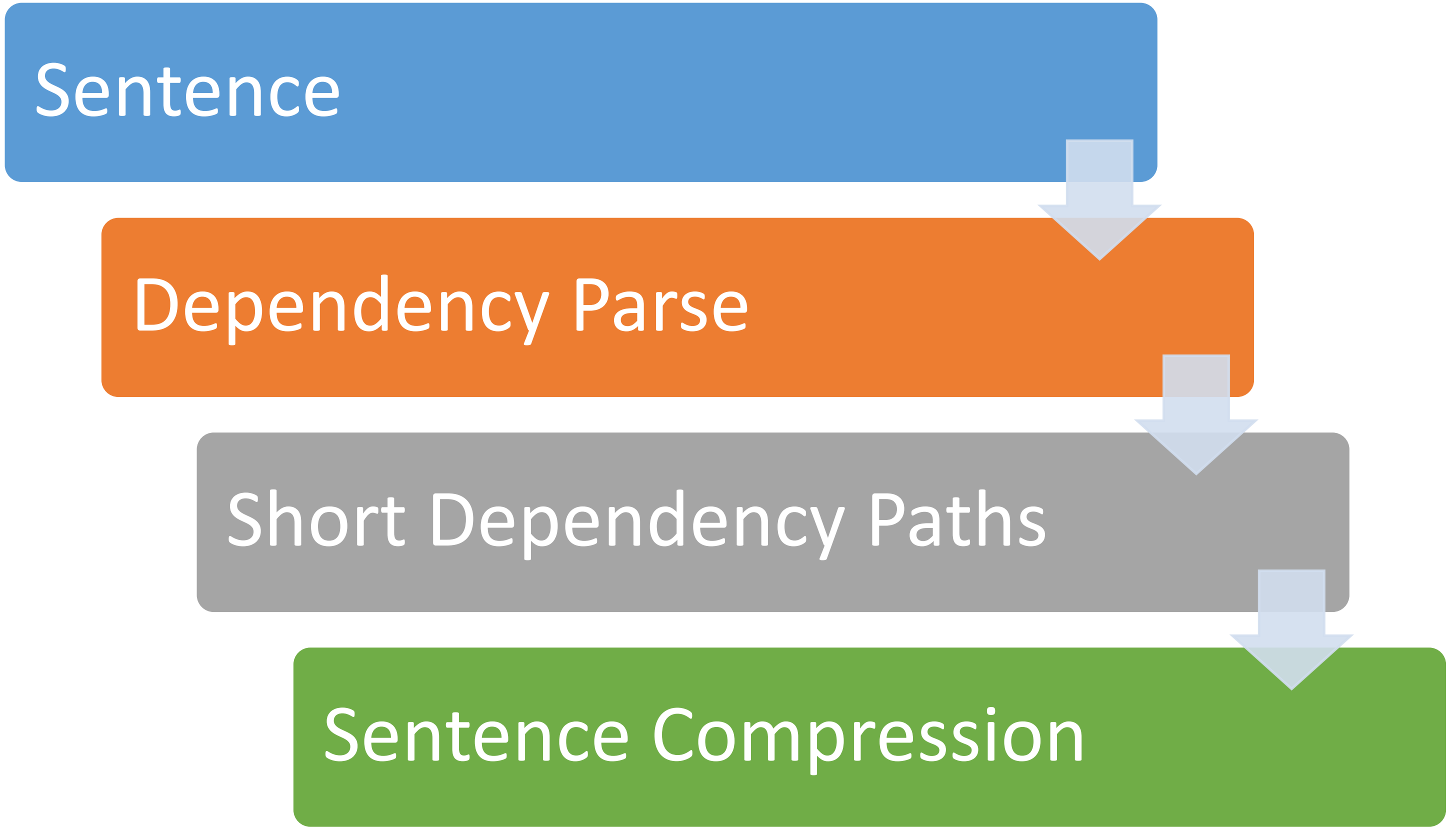
Example: Sentence Compression



Example: Sentence Compression



Example: Sentence Compression



Research Question

- **Open Information Extraction** was developed as an end-goal on itself
- …Yet it makes **structural decisions**

Can Open IE serve as a useful ***intermediate representation****?*

Open Information Extraction

(John, **married**, Yoko)

(John, **wanted to leave**, the band)

(The Beatles, **broke up**)

Open Information Extraction

(John, **wanted to leave**, the band)

argument | predicate | argument

Open IE as Intermediate Representation

- Infinitives and multi word predicates

(John, **wanted to leave**, the band)

(The Beatles, **broke up**)

Open IE as Intermediate Representation

- Coordinative constructions

  "John decided to compose and perform solo albums"

(John, **decided to compose**, solo albums)

(John, **decided to perform**, solo albums)

Open IE as Intermediate Representation

- Appositions

  “Paul McCartney, founder of the Beatles, wasn’t surprised”

(Paul McCartney, **wasn’t surprised**)

(Paul McCartney, **[is] founder of**, the Beatles)

Open IE as Intermediate Representation

- Test Open IE versus:

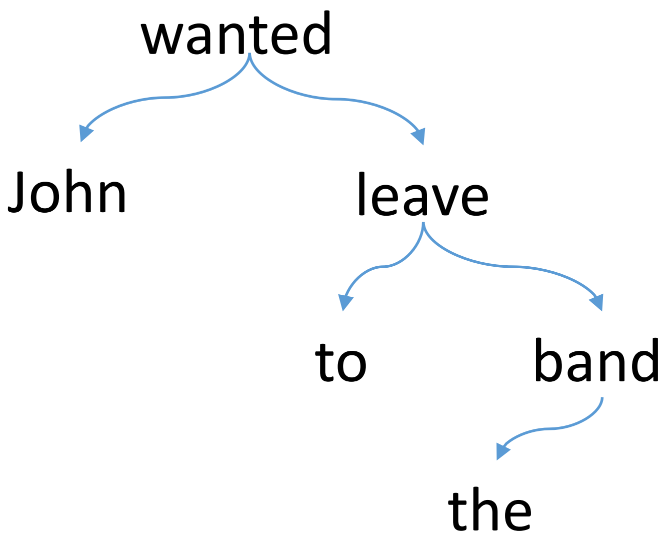
Open IE as Intermediate Representation

- Test Open IE versus:
  - Bag of words

John wanted to leave the band

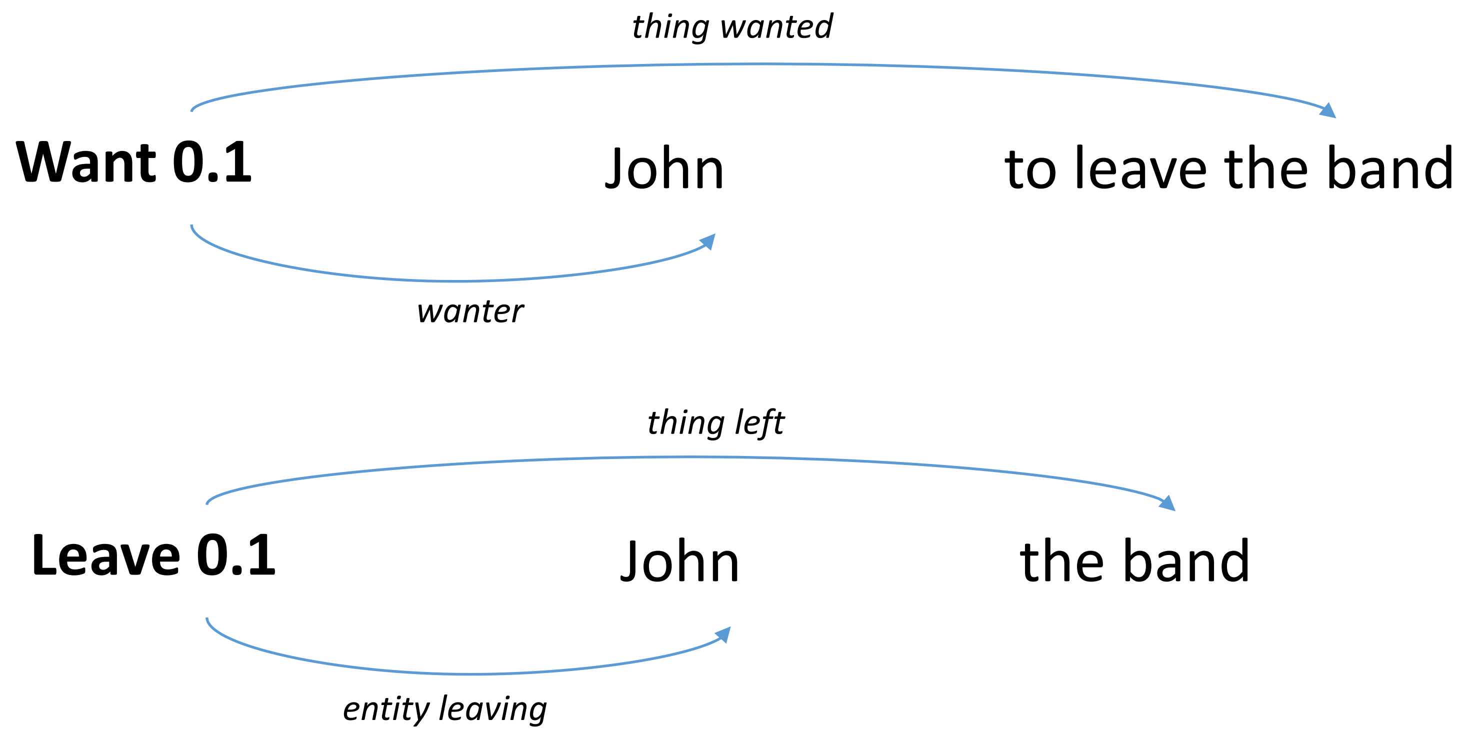
Open IE as Intermediate Representation

- Test Open IE versus:
  - Dependency parsing

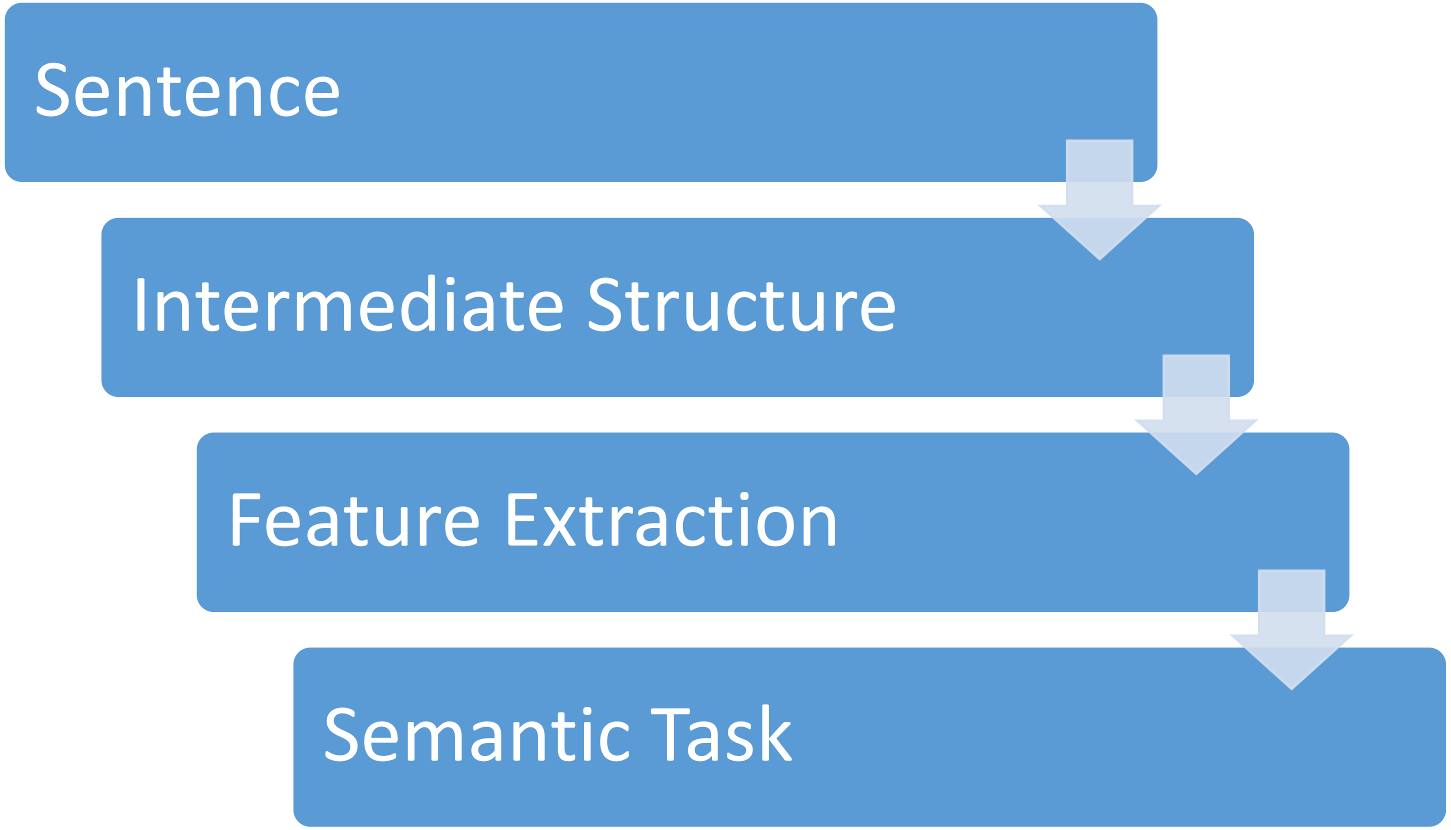


Open IE as Intermediate Representation

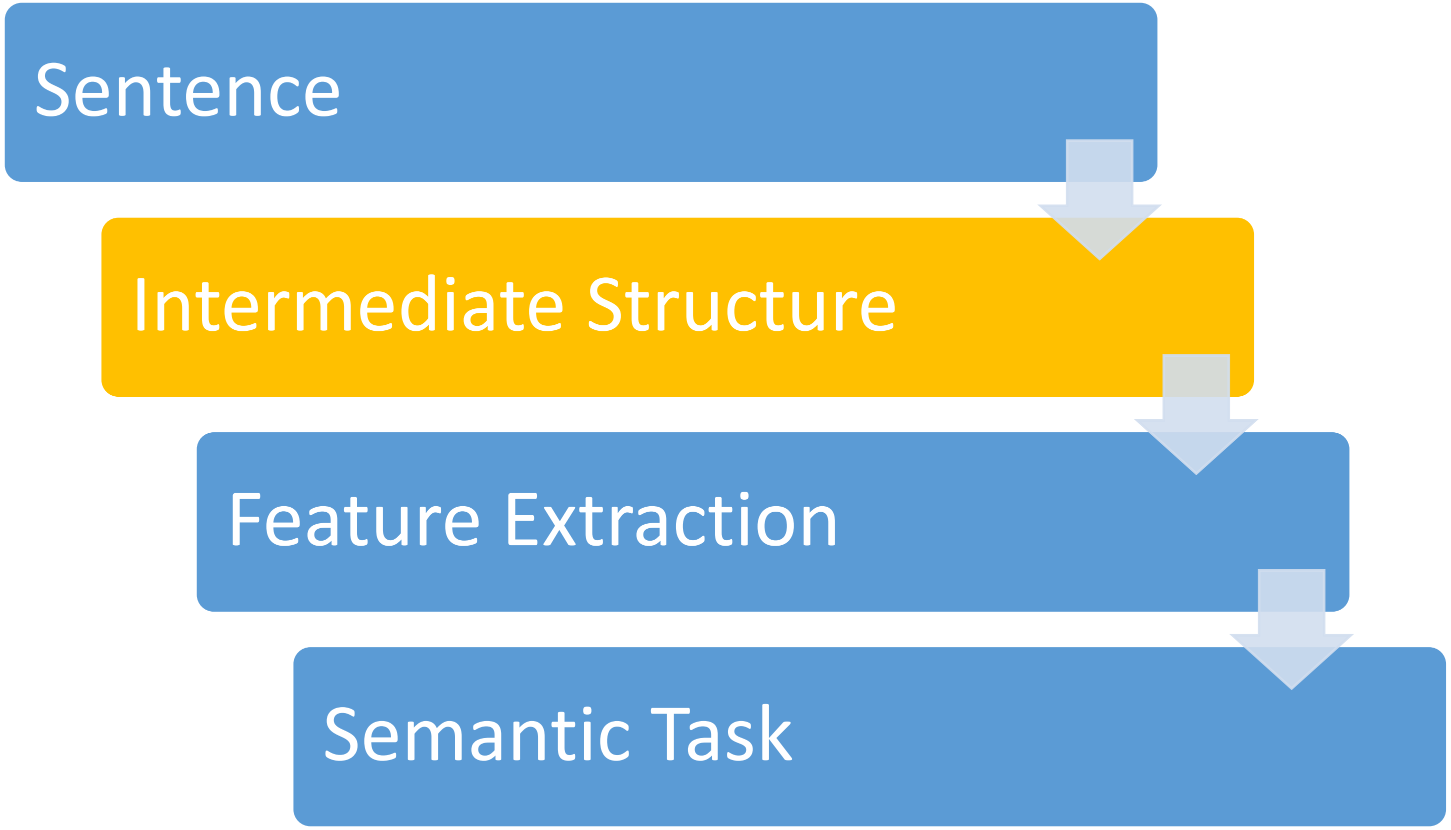
- Test Open IE versus:
  - Semantic Role Labeling



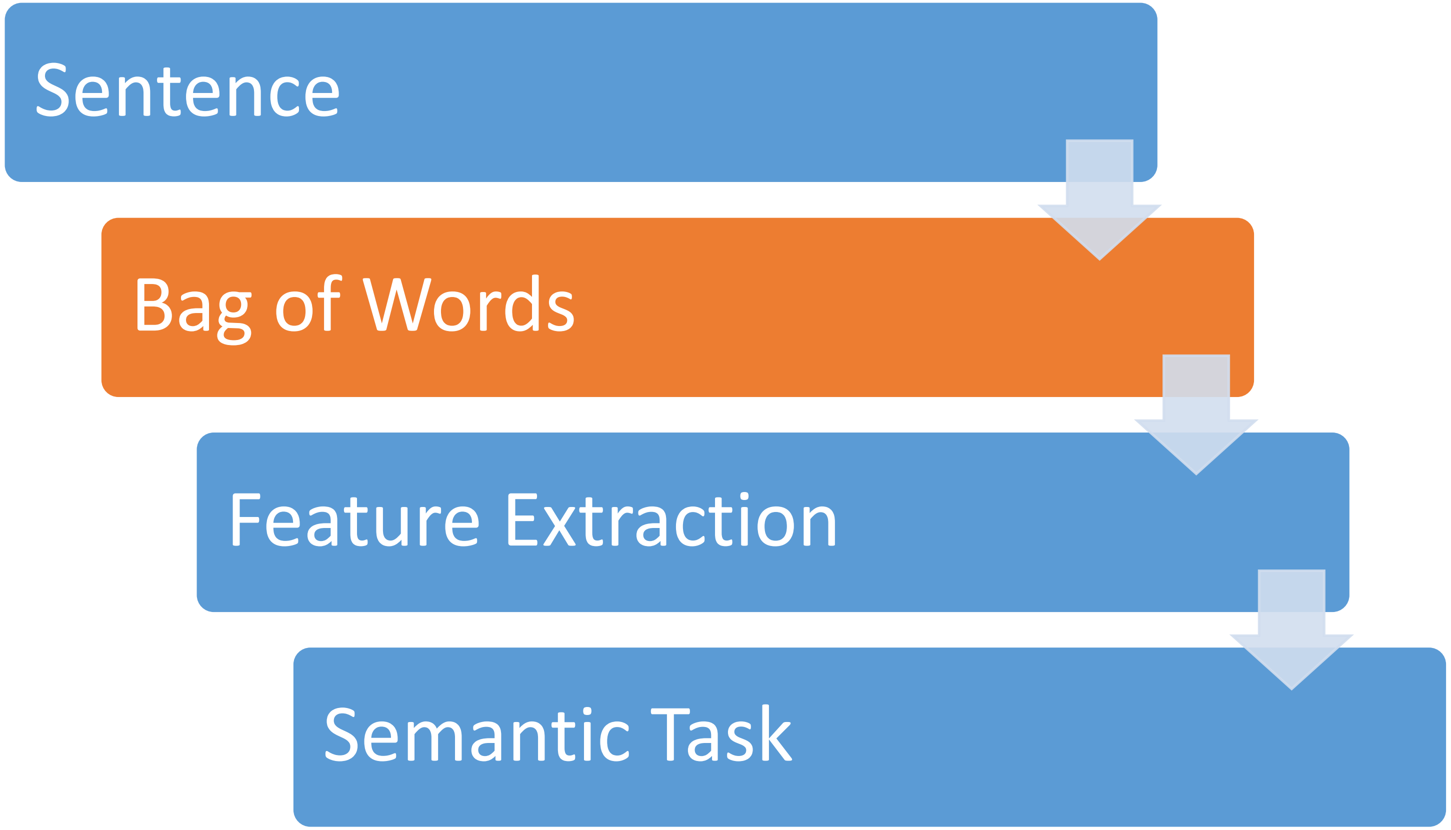
Extrinsic Analysis



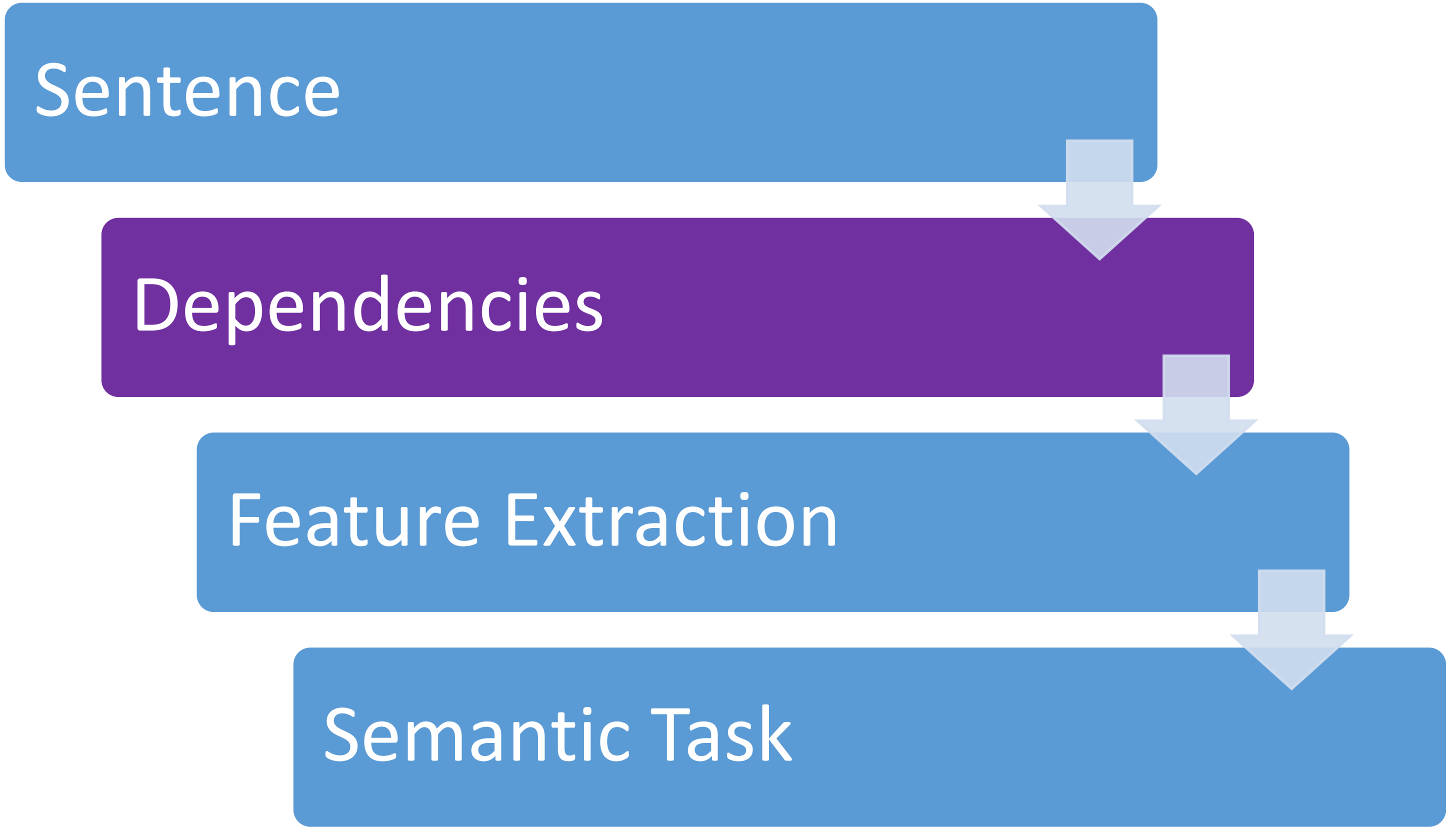
Extrinsic Analysis



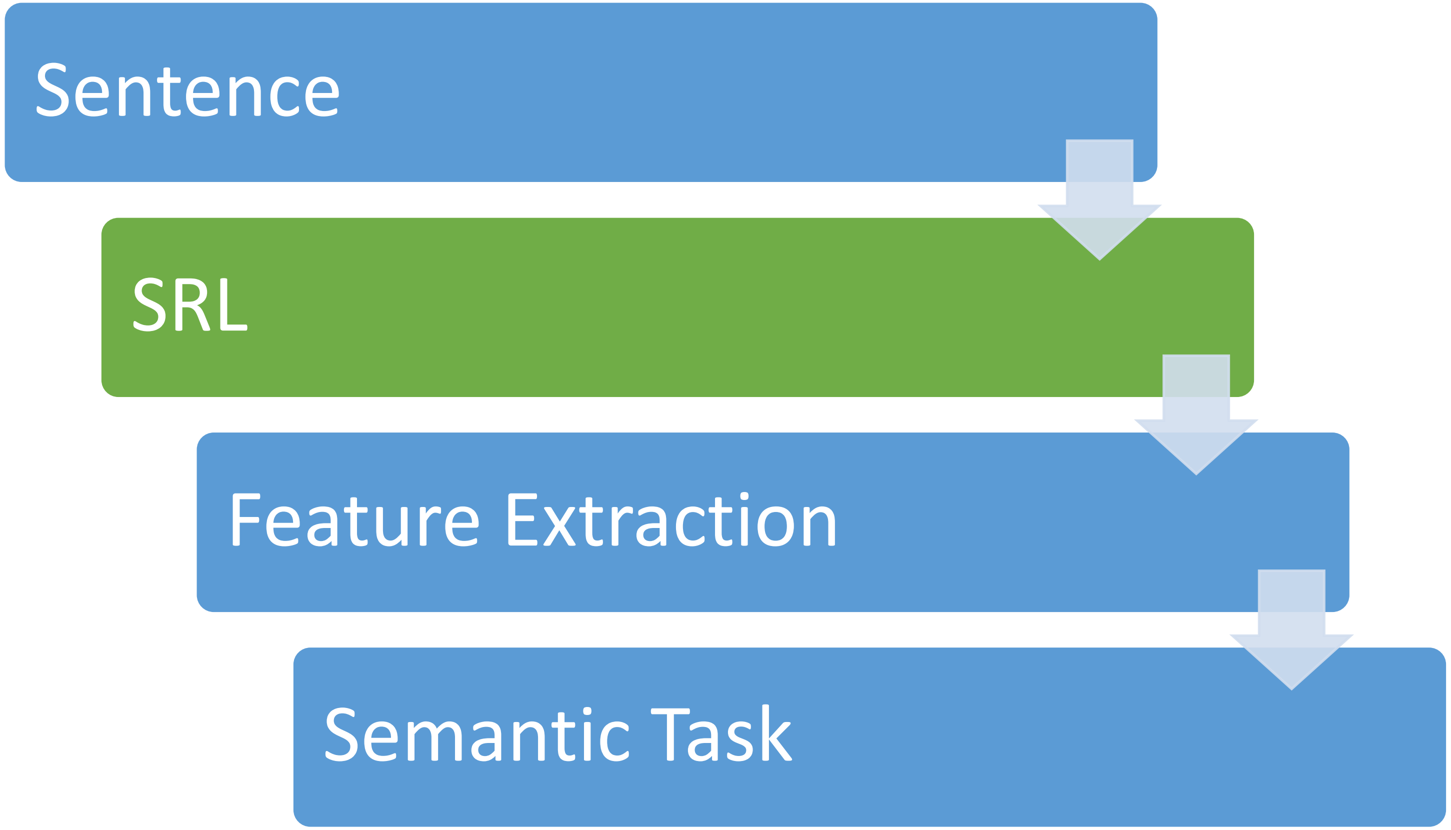
Extrinsic Analysis



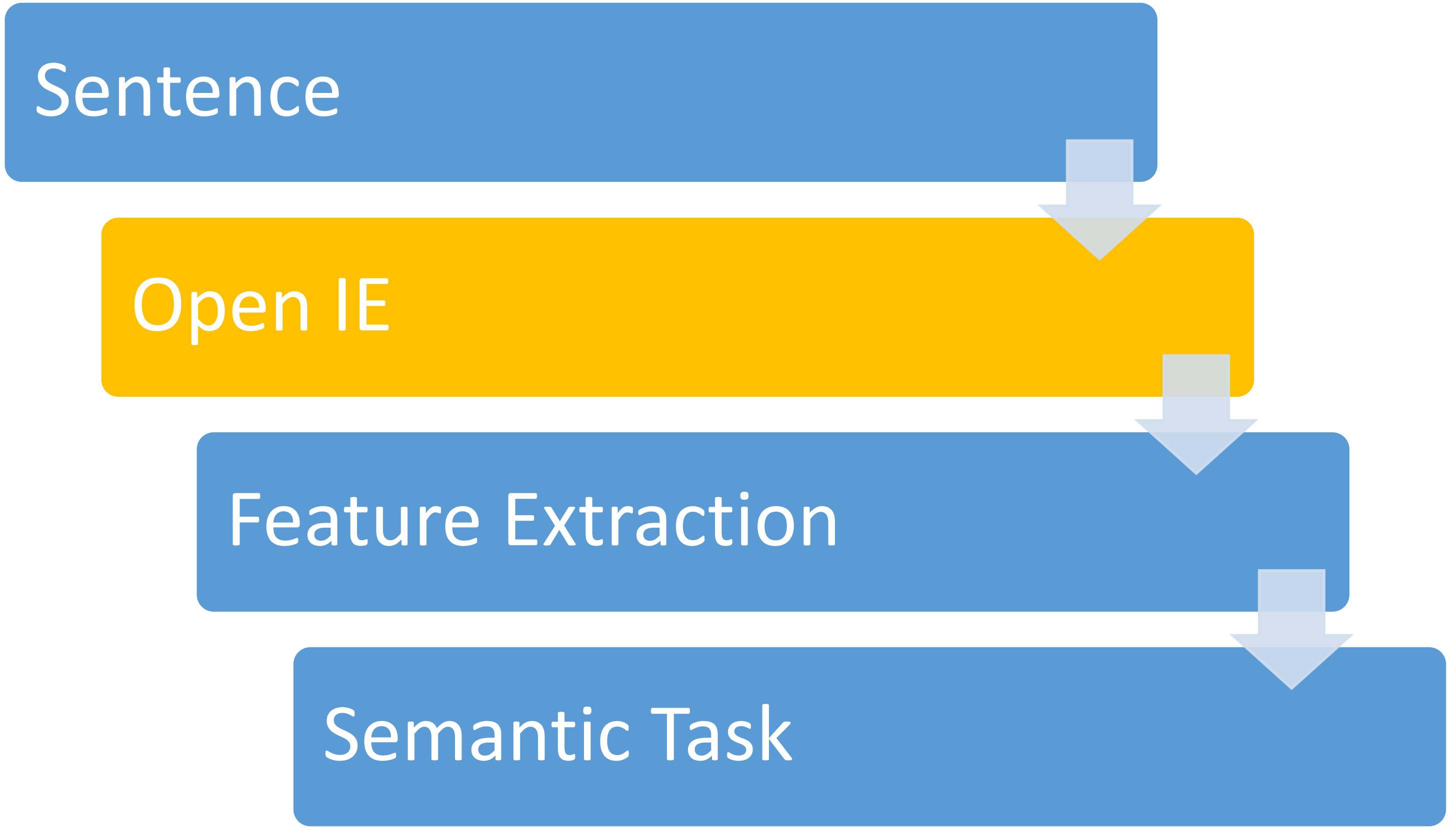
Extrinsic Analysis



Extrinsic Analysis



Extrinsic Analysis



Textual Similarity

- Domain Similarity
  - *Carpenter* ←→ *hammer* **[Domain similarity]**
  - Various test sets:
    - Bruni (2012), Luong (2013), Radinsky (2011), and ws353 (Finkelstein et al., 2001)
    - ~5.5K instances

- Functional Simlarity
  - *Carpenter* ←→ *Shoemaker* **[Functional similarity]**
  - Dedicated test set:
    - Simlex999 (Hill et al, 2014)
    - ~1K instances

Word Analogies

- (**man** : **king**), (**woman** : **?**)

Word Analogies

- (man : king), (woman : queen)

Word Analogies

- (man : king), (woman : queen)
- (Athens : Greece), (Cairo : ?)

Word Analogies

- (man : king), (woman : queen)
- (Athens : Greece), (Cairo : Egypt)

Word Analogies

- (**man** : **king**), (**woman** : **queen**)
- (**Athens** : **Greece**), (**Cairo** : **Egypt**)
- Test sets:
  - Google (~195K instances)
  - MSR (~8K instances)

Reading Comprehension

- MCTest, (Richardson et. al., 2013)

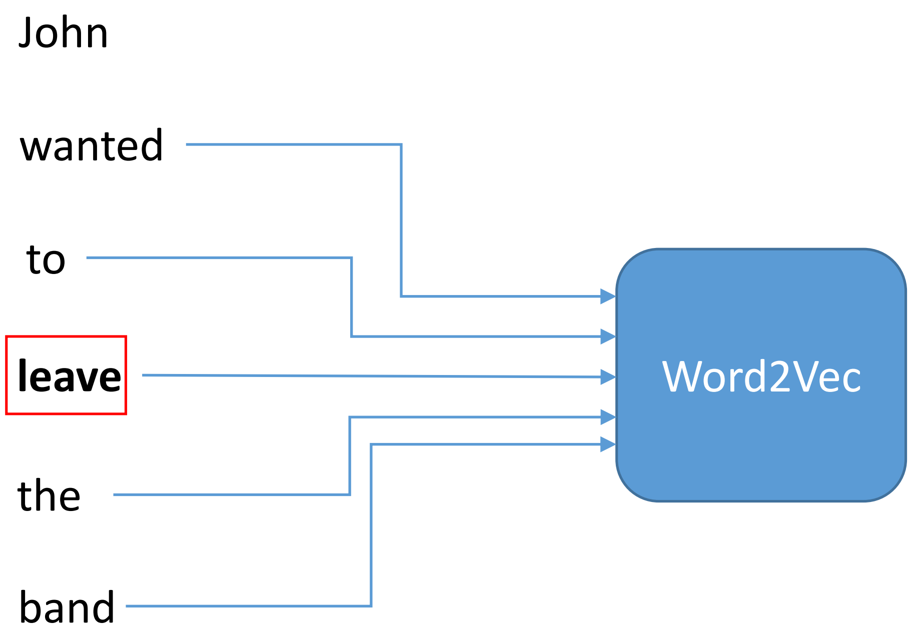
- Details in the paper!

Textual Similarity and Analogies

- Previous approaches used distance metrics over word embedding:
  - (Mikolov et al, 2013) - **lexical contexts**
  - (Levy and Goldberg, 2014) - **syntactic contexts**
- We compute embeddings for **Open IE** and **SRL** contexts
- Using the same training data for all embeddings (1.5B tokens Wikipedia dump)

Computing Embeddings

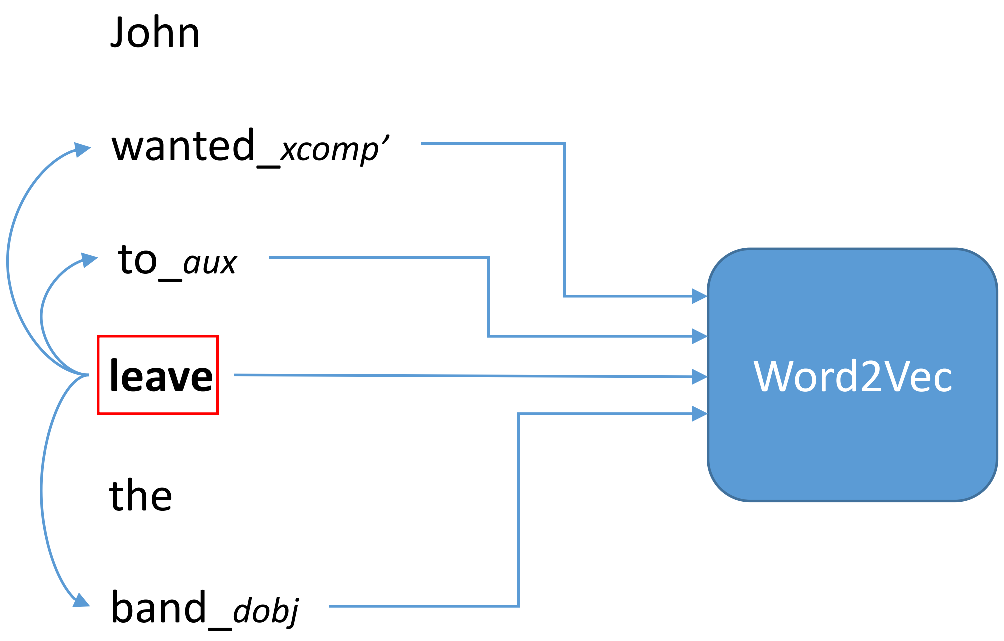
- **Lexical contexts**
  (for word **leave**)



(Mikolov et al., 2013)

Computing Embeddings

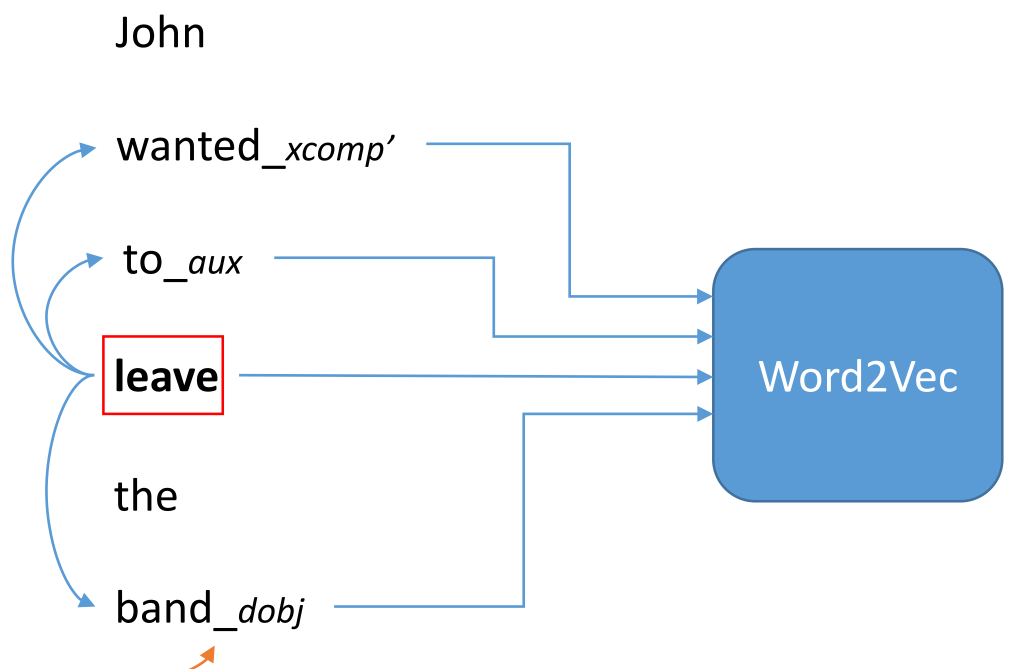
- **Syntactic contexts**
  (for word **leave**)



(Levy and Goldberg, 2014)

Computing Embeddings

- **Syntactic contexts**
  (for word **leave**)

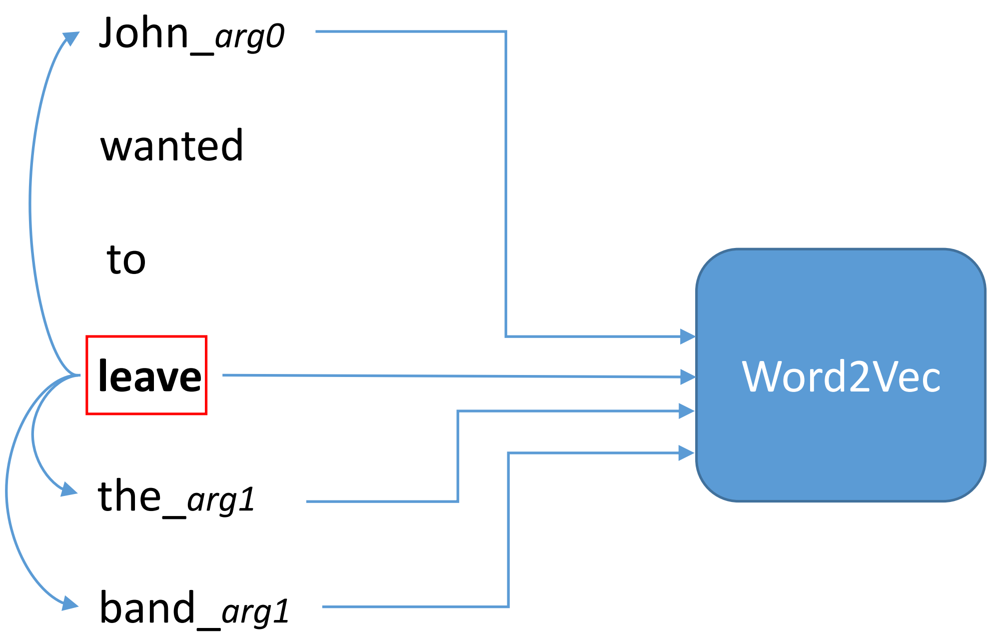


(Levy and Goldberg, 2014)

A context is formed of word + syntactic relation

Computing Embeddings

- **SRL contexts**
  (for word **leave**)

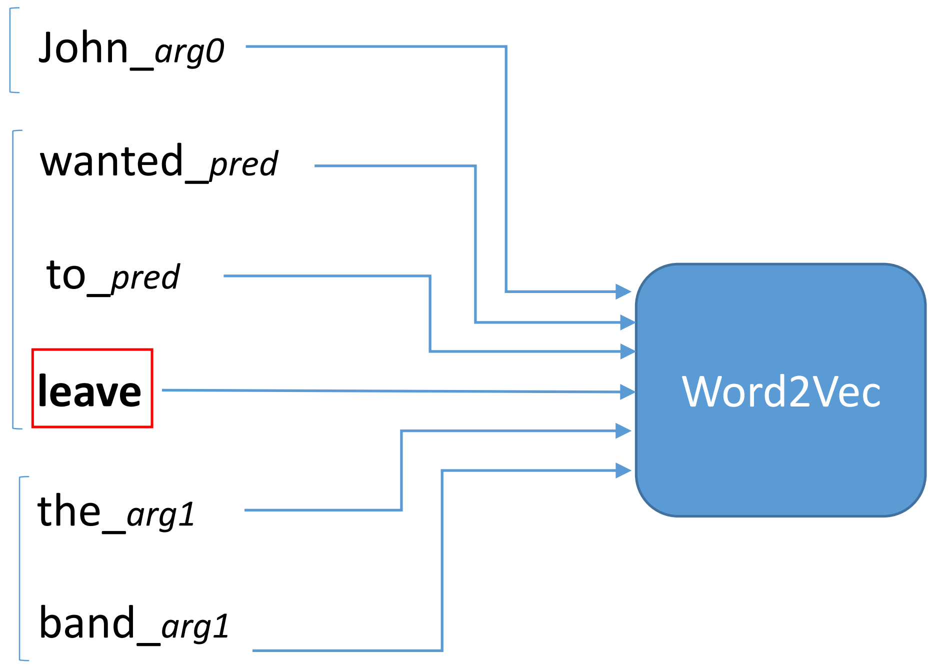


Available at author's website

Computing Embeddings

- **Open IE contexts**
  (for word **leave**)

(John, **wanted to leave**, the band)



Available at author's website

Results on Textual Similarity

	Open IE	Lexical	Deps	SRL
bruni	**.757**	.735	.618	.491
luong	**.288**	.229	.197	.171
radinsky	**.681**	.674	.592	.433
simlex	.39	.365	**.447**	.306
ws353-rel	**.647**	.64	.492	.551
ws353-sym	**.77**	.763	.759	.439
ws353-full	**.711**	.703	.629	.693

Results on Textual Similarity

	Open IE	Lexical	Deps	SRL
bruni	**.757**	.735	.618	.491
luong	**.288**	.229	.197	.171
radinsky	**.681**	.674	.592	.433
simlex	.39	.365	**.447**	.306
ws353-rel	**.647**	.64	.492	.551
ws353-sym	**.77**	.763	.759	.439
ws353-full	**.711**	.703	.629	.693

Syntactic does better on functional similarity

Results on Analogies

	Google		MSR	
	Add	Mul	Add	Mul
Open IE	**.714**	**.719**	**.529**	**.55**
Lexical	.651	.656	.438	.455
Deps	.34	.367	.4	.434
SRL	.352	.362	.389	.406

Additive

$$\arg\max_{b^*\in V} \left(\cos\left(b^*, b\right) - \cos\left(b^*, a\right) + \cos\left(b^*, a^*\right)\right)$$

Multiplicative

$$\arg\max_{b^*\in V} \frac{\cos\left(b^*, b\right)\cos\left(b^*, a^*\right)}{\cos\left(b^*, a\right) + \varepsilon}$$

Results on Analogies

	Google		MSR	
	Add	Mul	Add	Mul
Open IE	**.714**	**.719**	**.529**	**.55**
Lexical	.651	.656	.438	.455
Deps	.34	.367	.4	.434
SRL	.352	.362	.389	.406

State of the art with this amount of data

Additive

$$\arg\max_{b^* \in V} \left(\cos\left(b^*, b\right) - \cos\left(b^*, a\right) + \cos\left(b^*, a^*\right)\right)$$

Multiplicative

$$\arg\max_{b^* \in V} \frac{\cos\left(b^*, b\right) \cos\left(b^*, a^*\right)}{\cos\left(b^*, a\right) + \varepsilon}$$

PropS
Generic Proposition Extraction

Gabriel Satanovsky   Jessica Ficler   Ido Dagan   Yoav Goldberg

http://u.cs.biu.ac.il/~stanovg/propextraction.html

What's missing in Open IE?

Structure!

- *Intra-proposition* structure
  - NL propositions are more than SVO tuples
    - E.g., **The president thanked the speaker of the house who congratulated him**
- *Inter-proposition* structure
  - Globally consolidating and structuring the extracted information

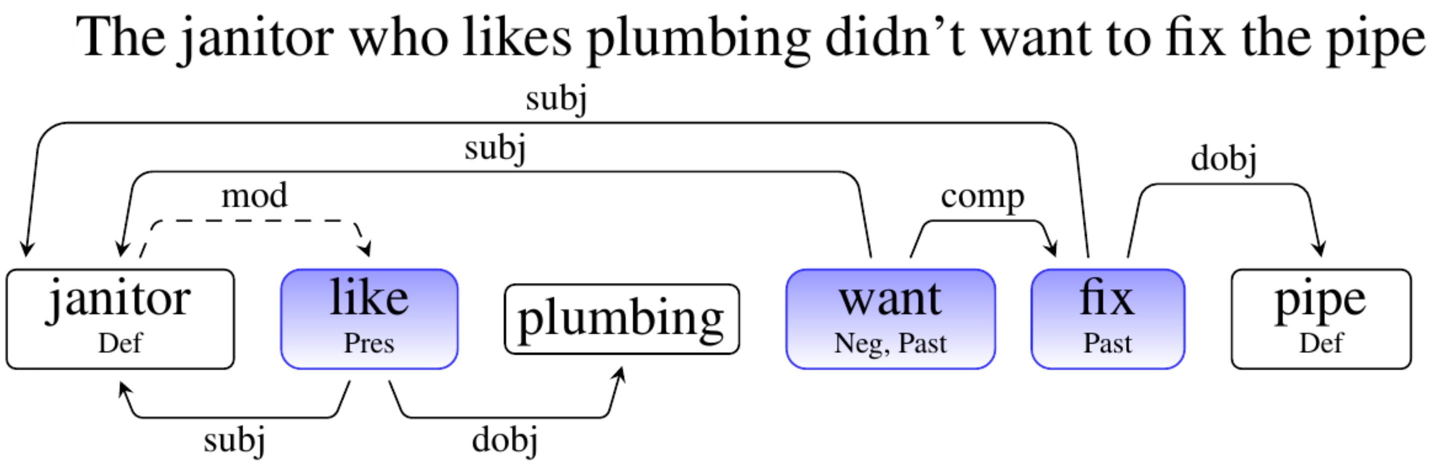
E.g. **aspirin relieve headache** = **aspirin treat headache**

PropS motivation

- Semantic applications are primarily interested in the predicate-argument structure conveyed in texts
- Commonly extracted from dependency trees
  - Yet it is often a non-trivial and cumbersome process, due to syntactic over-specification, and the lack of abstraction & canonicalization
- Our goal:
  - Accurately get as much semantics as given by syntax
  - Stems from a technical standpoint
  - Yet raises some theoretic issues regarding the syntax – semantics interface
    - Over generalizing might result in losing important semantic nuances

PropS

- A simple, abstract and canonicalized sentence representation scheme
  - **Nodes** represent atomic elements of the proposition
    - Predicates, arguments or modifiers
  - **Edges** encode argument (solid) or modifier (dashed) relations



PropS Properties

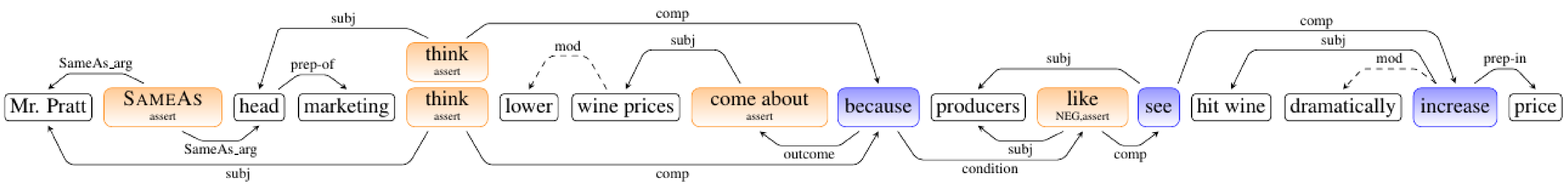
- **Abstracts** away syntactic variations
  - Tense, passive vs. active voice, negation variants, etc.

- **Unifies** semantically similar constructions
  - Various types of predications:
    - Verbal
    - Adjectival
    - Conditional
    - ….

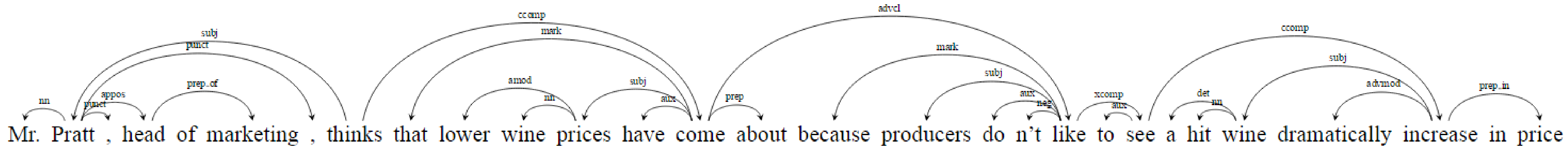
- **Differentiates** over semantically different propositions
  - E.g. restrictive vs. non-restrictive modification

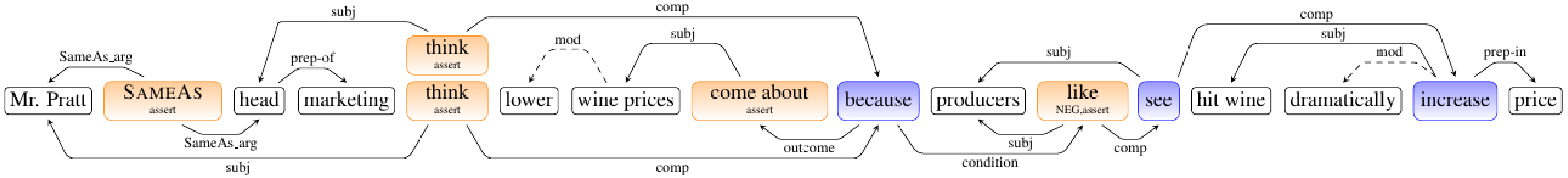
“Mr. Pratt, head of marketing, thinks that lower wine prices have come about because producers don’t like to see a hit wine dramatically increase in price.”

Props (17 nodes and 19 edges)



Dependency parsing (27 nodes and edges)





"Mr. Pratt, head of marketing, thinks that lower wine prices have come about because producers don't like to see a hit wine dramatically increase in price."

- Extracted propositions:

(1) lower wine prices have come about	[asserted]
(2) hit wine dramatically increase in price	
(3) producers see (2)	
(4) producers don't like (3)	[asserted]
(5) Mr Pratt is the head of marketing	[asserted]
(6) (1) happens because of (4)	
(7) Mr Pratt thinks that (6)	[asserted]
(8) the head of marketing thinks that (6)	[asserted]

PropS Methodology

- Corpus based analysis
  - Taking semantic applications perspective
  - Focusing on the most commonly occurring phenomena

- Feasibility criterion
  - **High accuracy** would be feasibly derivable **from available manual annotations**
  - **Reasonable accuracy** for baseline parser on top of **automatic dependency parsing**

PropS Handled Phenomena

- Certain syntactic details are abstracted into node features
  - Modality
  - Negation
  - Definiteness
  - Tense
  - Passive or active voice

- Restrictive vs. non restrictive modification
  - Implies different argument boundaries:
    - *[The boy who was born in Hawaii]* ***went*** *home* [restrictive]
    - *[Barack Obama] who was born in Hawaii* ***went*** *home* [non-restrictive]

PropS Handled Phenomena (cont.)

- Distinguishing between **asserted** and **attributed** propositions
  - *John **passed** the test*
  - *the teacher **denied** that John **passed** the test*

- Distinguishing the different types of appositives and copulas
  - ***The company, Random House,** didn't report its earnings* [**appositive**]
  - ***Bill Clinton, a former U.S president,** will join the board* [**predicative**]

PropS Handled Phenomena (cont.)

- … and more:
  - Conditionals
  - Raising vs. control constructions
  - Non-lexical predications (expletives, possessives, etc.)
  - Temporal expressions

PropS Provided Resources

- Human annotated gold-standard
  - 100 sentences from the PTB annotated with our gold structures
- High-accuracy conversion of the WSJ
  - Computed (rule-based) on top of integration of several manual annotations
    - PTB Constituency
    - Propbank
    - Vadas et al(2007)'s NP structure
- Baseline parser
  - Rule based converter over automatically generated dependency parse trees

PropS Conversion Accuracy

Traditional LAS was modified to account for non 1-1 correspondence between words and nodes

	Feature Computation			Modified LAS		
	P	R	F1	P	R	F1
WSJ	.95	.97	.96	.9	.92	.91
PROPS	.88(.88)	.89(.84)	.89(.86)	.83(.8)	.81(.81)	.82(.8)

Table 2: Conversion accuracy, WSJ is compared against gold standard, PROPS against the gold standard and WSJ (in parentheses).

PropS Empirical Demonstration: Reading Comprehenstion

Rule-based methods for answering questions from [MCTest]
Simple similarity metrics. Applied once over dependency and ***PropS***

Method	Correct
PROPS	**66.34%**
dependencies	64.58%
lexical	60.44%

Table 3: Results on MCTest corpus

PropS Future Work

- Nominalizations
  - *“Instagram’s acquisition by Facebook”*
- Improved restrictiveness annotations
  - Work in ACL 16
- Conjunctions
  - Improving conjunctions underlying parsing and representation
- Quantifications

Annotating and Predicting Non-Restrictive Modification

Stanovsky and Dagan,
ACL 2016

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

- Restrictive modification
  - The content of the modifier is an **integral part** of the meaning of the containing clause
  - AKA: integrated (Huddleston)

- Non-restrictive modification
  - The modifier presents an **separate or additional unit of information**
  - AKA: supplementary (Huddleston), appositive, parenthetical

	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
- **Automatic prediction** of non-restrictive modifiers
  - Using lexical-syntactic features

Previous work

- Rebanking CCGbank for improved NP interpretation
  (Honnibal, Curran and Bos, ACL ‘10)
  - Added automatic non-restrictive annotations to the CCGbank
  - Simple punctuation implementation
    - Non restrictive modification $\leftarrow\rightarrow$ The modifier is preceded by a comma
  - No intrinsic evaluation

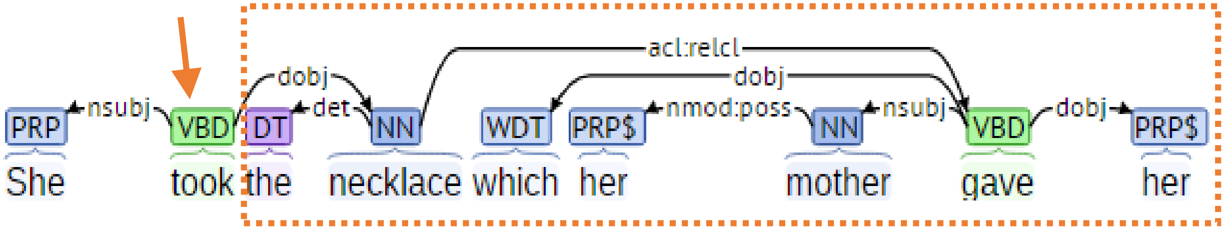
Previous work

- Relative clause extraction for syntactic simplification
  (Dornescu et al., COLING '14)
  - Trained annotators marked spans as restrictive or non-restrictive
  - Conflated argument span with non-restrictive annotation
    - This led to low inter-annotator-agreement
      - **Pairwise F1 score of 54.9%**
    - Develop rule based and ML baselines (CRF with chunking feat.)
      - **Both performing around ~47% F1**

Our Approach

Consistent corpus with QA based classification

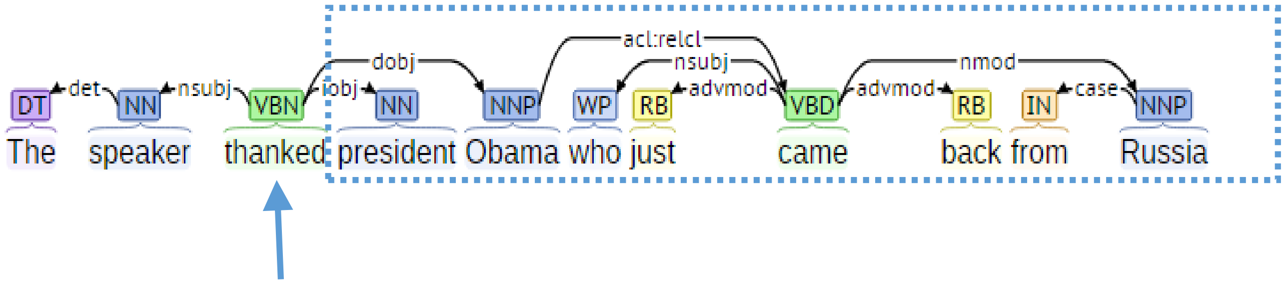
1. Traverse the syntactic tree from predicate to NP arguments
2. Phrase an argument role question, which is answered by the NP (*what? who? to whom? Etc.*)
3. For each candidate modifier (= syntactic arc) - check whether when omitting it the NP still provides the same answer to the argument role question



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

Crowdsourcing

- This seems fit for crowdsourcing:
  - **Intuitive** - Question answering doesn’t require linguistic training
  - **Binary decision** – Each decision directly annotates a modifier

Corpus

- CoNLL 2009 dependency corpus
  - Recently annotated by QA-SRL -- we can borrow most of their role questions
- Each NP is annotated on Mechanical Turk
  - Five annotators for 5c each
- Final annotation by majority vote

Expert annotation

- Reusing our previous expert anntation, we can assess if crowdsourcing captures non-restrictiveness
- Agreement
  - Kappa = 73.79 (substantial agreement)
  - F1 =85.6

Candidate Type Distribution

	#instances	%Non-Restrictive	Agreement (K)
Prepositive adjectival modifiers	677	41%	74.7
Prepositions	693	36%	61.65
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 annotation covered 1930 NPs in 1241 sentences

Candidate Type Distribution

	#instances	%Non-Restrictive	Agreement (K)
Prepositive adjectival modifiers	677	41%	74.7
Prepositions	693	36%	61.65
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

Candidate Type Distribution

	#instances	%Non-Restrictive	Agreement (K)
Prepositive adjectival modifiers	677	41%	74.7
Prepositions	693	36%	61.65
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 balanced between the two classes

Predicting non-restrictive modification

- CRF features:
  - Dependency relation
  - NER
    - Modification of named entity tend to be non-restrictive
  - Word embeddings
    - Contextually similar words will have similar restricteness value
  - Linguistically motivated features
    - The word introducing the modifier,
      - “that” indicates restrictive, while a wh-pronoun as indicates non-restrictive (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**

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**

Prepositions and adjectives are harder to predict

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**

Commas are good in precision but poor for recall

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**

Dornescu et al. performs better on our dataset

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**

Our system highly improves recall

To Conclude this part...

- A large non-restrictive gold standard
  - Directly augments dependency trees
- Automatic classifier
  - Improves over state of the art results

Creating a Gold Benchmark for Open IE

Stanovsky and Dagan,
EMLP 2016

Open Information Extraction

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

- Used in various applications for populating large databases from raw open domain texts
  - A scalable and open variant of the Information Extraction task

Open IE Evaluation

- Open IE task formulation has been lacking formal rigor
  - No common guidelines → **No large corpus for evaluation**
- Annotators examine *a small sample* of their system's output and judge it according to *some* guidelines

→ **Precision oriented** metrics

→ Numbers are **not comparable**

→ Experiments are **hard to reproduce**

Goal

- In this work we -
  - Analyze common evaluation principles in prominent recent work

  - Create a **large gold standard corpus** which follows these principles
    - Uses previous annotation efforts
    - Provides both precision **and recall** metrics

  - Automatically evaluate the performance of the most prominent OIE systems on our corpus
    - First automatic & comparable OIE evaluation
    - Future systems can easily compare themselves

Converting QA-SRL to Open IE

- Intuition:
  - All of the QA pairs over a single predicate in QA-SRL correspond to a single Open IE extraction
- Example:
  - "**Barack Obama**, the newly elected president, **flew** **to Moscow** **on Tuesday**"
  - QA-SRL:
    - Who **flew** somewhere? **Barack Obama**
    - Where did someone **fly**? **to Moscow**
    - When did someone **fly**? **on Tuesday**

→ (Barack Obama, **flew**, to Moscow, on Tuesday)

Example

- John Bryce, Microsoft's head of marketing **refused** to **greet** Arthur Black
  - Who refused something? John Bryce
  - Who refused something? Microsoft's head of marketing
  - What did someone refuse to do? **greet** Arthur Black
  - Who was not greeted? Arthur Black
  - Who did not greet someone? John Bryce

→

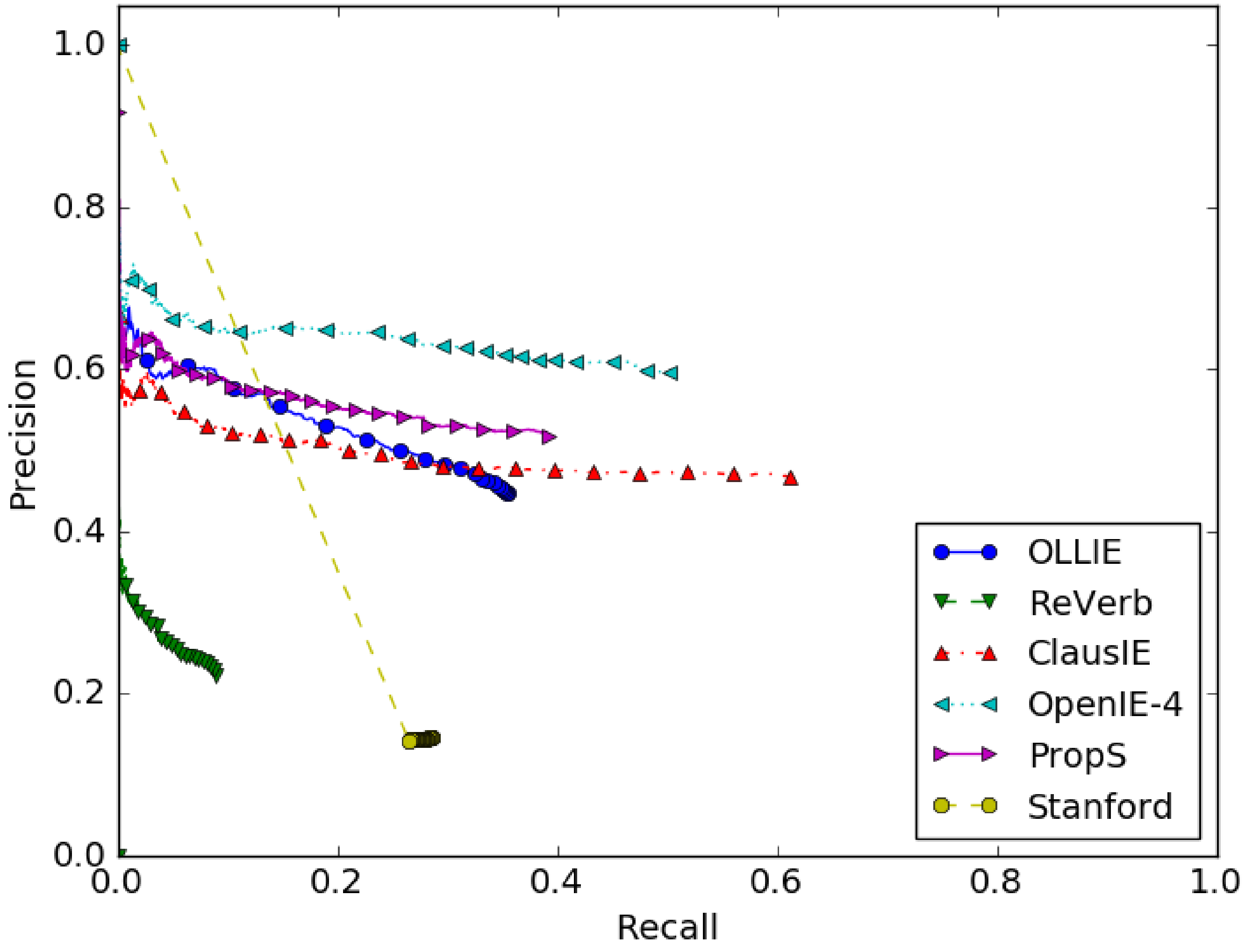
(John Bryce, **refused to greet**, Arthur Black),
(Microsoft's head of Marketing , **refused to greet**, Arthur Black)

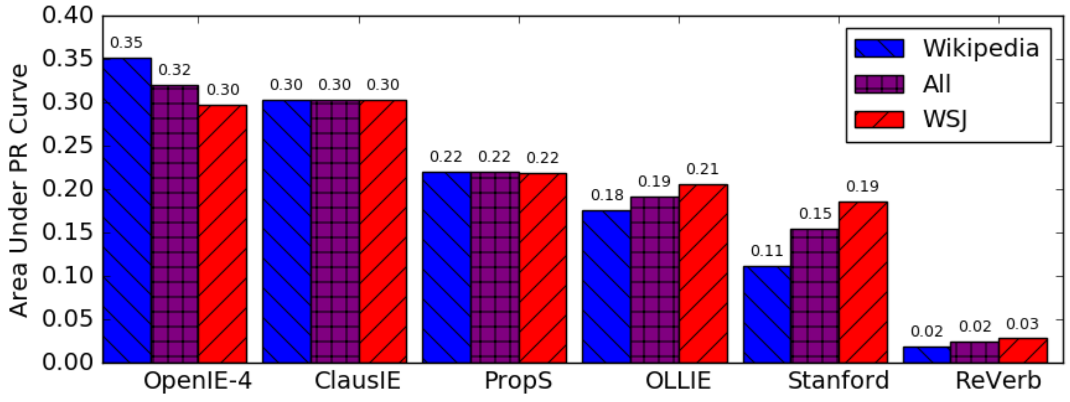
Resulting Corpus

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

- 13 times bigger than largest previous corpus (ReVerb)

Evaluations: PR-Curve





- Stanford – Assigns a probability of 1 to most of its extractions (94%)
- Low Recall
  - Most missed extractions seem to come from questions with multiple answers (usually **long range dependencies**)
- Low Precision
  - Allowing for softer matching functions (lowering threshold), raises precision and **keeps the same trends**

Conclusions

- We discussed a framework for argument annotation:
  - Formal Definition
  - Expert and crowdsource annotation
  - Automatic prediction
  - Automatic conversion from quality annotations

Conclusions

- We discussed a framework for argument annotation:
  - Formal Definition
  - Expert and crowdsource annotation
  - Automatic prediction
  - Automatic conversion from quality annotations

Thanks For Listening!