Question-Answer Driven Semantic Role Labeling

Using Natural Language to Annotate Natural Language

Luheng He, Mike Lewis, Luke Zettlemoyer

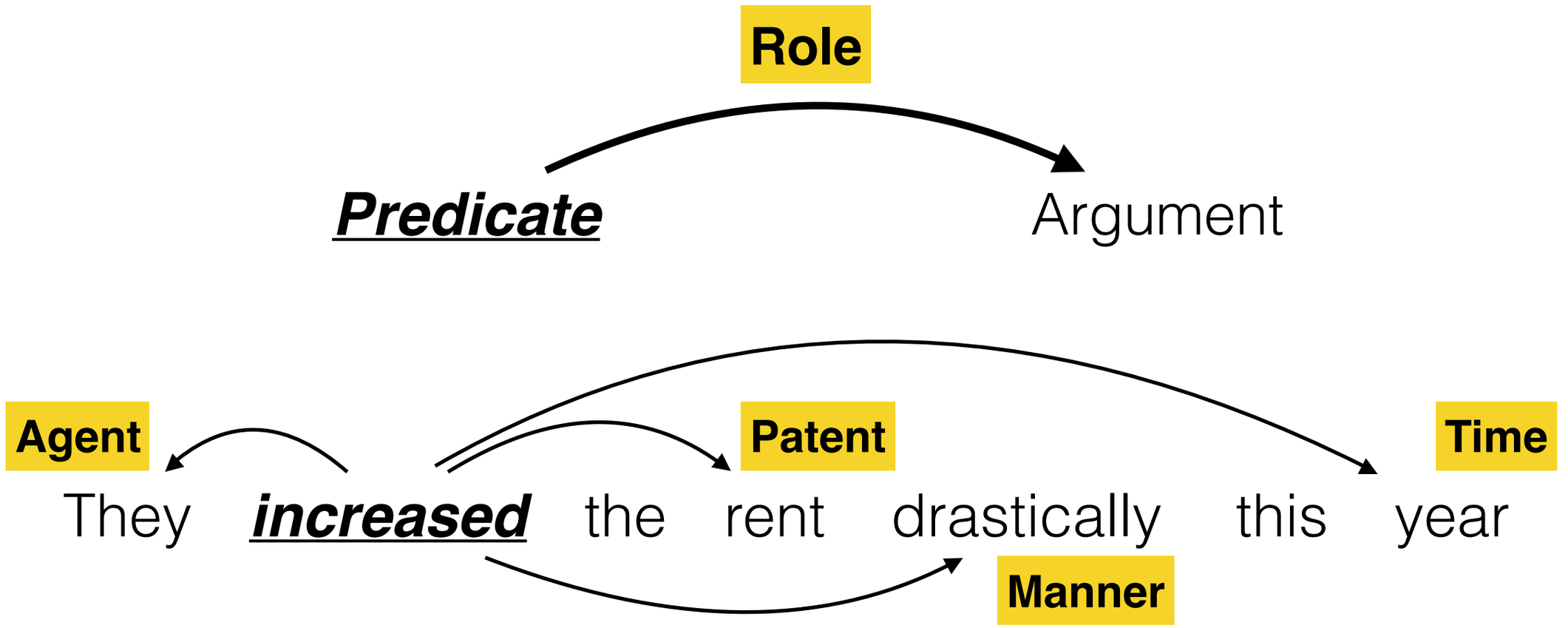
University of Washington

EMNLP 2015

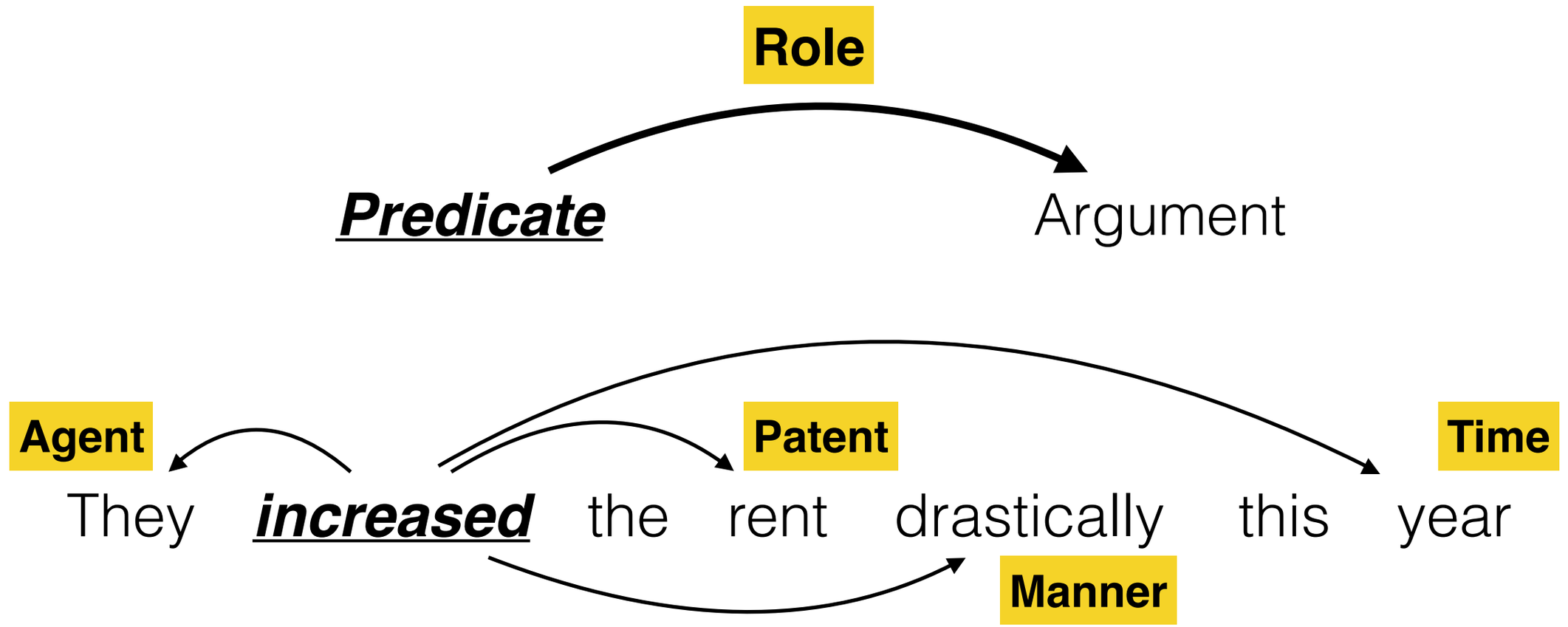
Semantic Role Labeling (SRL)

who did **what** to **whom**, **when** and **where**?

Semantic Role Labeling (SRL)



Semantic Role Labeling (SRL)



- Defining a set of roles can be difficult
- Existing formulations have used different sets

Existing SRL Formulations and Their Frame Inventories

FrameNet

1000+ semantic frames,
10,000+ frame elements (roles)

Frame: Change_position_on_a_scale
This frame consists of words that indicate the change of an Item's position on a scale (the Attribute) from a starting point (Initial_value) to an end point (Final_value). The direction (Path) …
Lexical Units:
…, reach.v, rise.n, ***rise.v,*** *rocket.v, shift.n, …*

PropBank

10,000+ frame files
with predicate-specific roles

Roleset Id: rise.01 , ***go up***

Arg1-: *Logical subject, patient, thing rising*
Arg2-EXT: *EXT, amount risen*
Arg3-DIR: *start point*
Arg4-LOC: *end point*
Argm-LOC: *medium*

Unified Verb Index, University of Colorado http://verbs.colorado.edu/verb-index/
PropBank Annotation Guidelines, Bonial et al., 2010
FrameNet II: Extended theory and practice, Ruppenhofer et al., 2006
FrameNet: https://framenet.icsi.berkeley.edu/

This Talk: QA-SRL

- Introduce a **new SRL** formulation with **no frame or role inventory**
- Use **question-answer pairs** to model verbal predicate-argument relations
- Annotated **over 3,000 sentences in weeks** with **non-expert**, part-time annotators
- Showed that this data is **high-quality** and **learnable**

Our Annotation Scheme

Given sentence and a verb:

They ***increased*** the rent this year .

Our Annotation Scheme

Given sentence and a verb:

They ***increased*** the rent this year .

<u>Step 1:</u> Ask a question about the verb:

Who increased something ?

Our Annotation Scheme

Given sentence and a verb:

They ***increased*** the rent this year .

<u>Step 1:</u> Ask a question about the verb:

Who increased something ?

<u>Step 2:</u> Answer with words in the sentence:

They

Our Annotation Scheme

Given sentence and a verb:

They ***increased*** the rent this year .

<u>Step 1:</u> Ask a question about the verb:

Who increased something ?

<u>Step 2:</u> Answer with words in the sentence:

They

<u>Step 3:</u> Repeat, write as many QA pairs as possible …

Our Annotation Scheme

Given sentence and a verb:

They ***increased*** the rent this year .

<u>Step 1:</u> Ask a question about the verb:

Who increased something ?

<u>Step 2:</u> Answer with words in the sentence:

They

<u>Step 3:</u> Repeat, write as many QA pairs as possible …

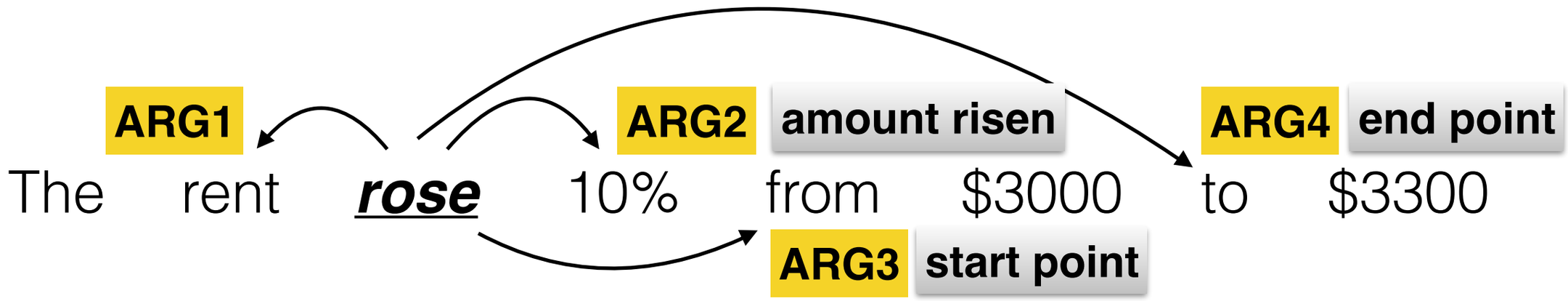
What is increased ?

the rent

When is something increased ?

this year

Previous Method: Annotation with Frames



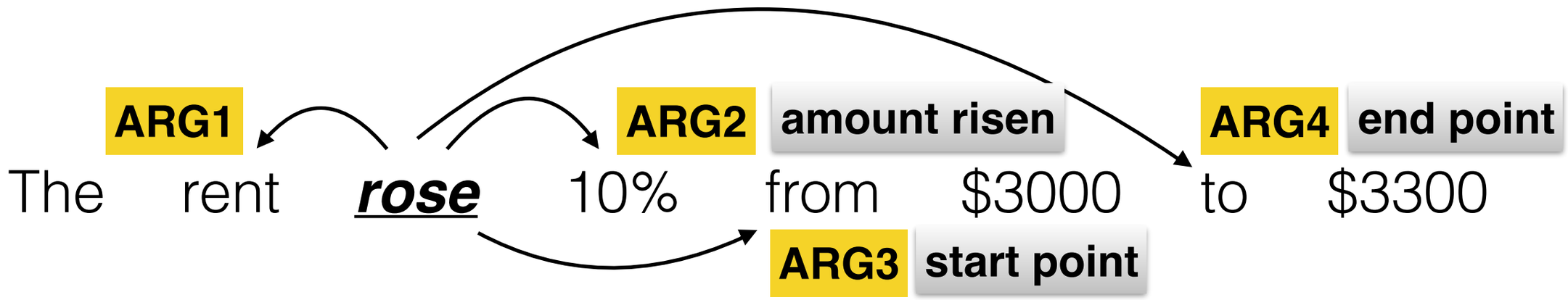
Frameset: rise.01 , *go up*

Arg1-: *Logical subject, patient, thing rising*
Arg2-EXT: *EXT, amount risen*
Arg3-DIR: *start point*
Arg4-LOC: *end point*
Argm-LOC: *medium*

- Depends on pre-defined frame inventory
- Annotators need to:
  1) Identify the Frameset
  2) Find arguments in the sentence
  3) Assign labels accordingly
- If frame doesn't exist, create new

The Proposition Bank: An Annotated Corpus of Semantic Roles, Palmer et al., 2005
http://verbs.colorado.edu/propbank/framesets-english/rise-v.html

Our Method: Q/A Pairs for Semantic Relations



Wh-Question	Answer
What ***rose*** ?	the rent
How much did something ***rise*** ?	10%
What did something ***rise*** from ?	$3000
What did something ***rise*** to ?	$3300

Comparing to Existing SRL Formulations

No Role Inventory!

Question

(Verbal) Predicate → Answer

Question-Answer Driven SRL
(QA-SRL)

Large Role Inventory

Role

Predicate → Argument

SRL

Advantages

- Easily explained
- No pre-defined roles, few syntactic assumption
- Can capture implicit arguments
- Generalizable across domains

Advantages

- Easily explained
- No pre-defined roles, few syntactic assumption
- Can capture implicit arguments
- Generalizable across domains

Limitations

- Only modeling verbs (for now)
- Not annotating verb senses directly
- Can have multiple equivalent questions

Advantages

- Easily explained
- No pre-defined roles, few syntactic assumption
- Can capture implicit arguments
- Generalizable across domains

Limitations

- Only modeling verbs (for now)
- Not annotating verb senses directly
- Can have multiple equivalent questions

Challenges

- What questions to ask?
- Quality - Can we get good Q/A pairs?
- Coverage - Can we get all the Q/A pairs?

Outline

Motivation and Intuition

- Semantic Role Labeling
- Our Method: QA-SRL

Data Collection and Analysis

- Annotation Task Design
- Dataset Statistics
- Quality Analysis

Learning Tasks and Baselines

Future Work and Conclusion

Question-Answer Driven SRL

Given sentence ***s***, target verb ***v***

Annotate all possible question-answer pairs ***<q,a>***

Question-Answer Driven SRL

Given sentence ***s***, target verb ***v***

Annotate all possible question-answer pairs **<*q*,*a*>**

- Question ***q*** should start with a **wh-word** and contain the target verb ***v***
- Answer ***a*** should be a phrase from the sentence ***s***. Multiple correct answers are allowed.

Writing Questions

$$q \in \mathbf{WH} \times \mathbf{AUX} \times \mathbf{SBJ} \times \mathbf{TRG} \times \mathbf{OBJ1} \times \mathbf{PP} \times \mathbf{OBJ2}$$

Writing Questions

$$q \in \mathbf{WH} \times \mathbf{AUX} \times \mathbf{SBJ} \times \mathbf{TRG} \times \mathbf{OBJ1} \times \mathbf{PP} \times \mathbf{OBJ2}$$

WH: Who, What, When, Where, Why, How, How much

Writing Questions

$$q \in \mathbf{WH} \times \mathbf{AUX} \times \mathbf{SBJ} \times \mathbf{TRG} \times \mathbf{OBJ1} \times \mathbf{PP} \times \mathbf{OBJ2}$$

WH: Who, What, When, Where, Why, How, How much

AUX: Auxiliary verbs, including negations. i.e. is, might, wo n’t

Writing Questions

$$q \in \mathbf{WH} \times \mathbf{AUX} \times \mathbf{SBJ} \times \mathbf{TRG} \times \mathbf{OBJ1} \times \mathbf{PP} \times \mathbf{OBJ2}$$

WH: Who, What, When, Where, Why, How, How much

AUX: Auxiliary verbs, including negations. i.e. is, might, wo n't

SBJ, OBJ1, OBJ2: someone, something, do something, etc.

Writing Questions

$$q \in \mathbf{WH} \times \mathbf{AUX} \times \mathbf{SBJ} \times \mathbf{TRG} \times \mathbf{OBJ1} \times \mathbf{PP} \times \mathbf{OBJ2}$$

WH: Who, What, When, Where, Why, How, How much

AUX: Auxiliary verbs, including negations. i.e. is, might, wo n't

SBJ, OBJ1, OBJ2: someone, something, do something, etc.

TRG: Target verb, including inflected forms.

Writing Questions

$$q \in \mathbf{WH} \times \mathbf{AUX} \times \mathbf{SBJ} \times \mathbf{TRG} \times \mathbf{OBJ1} \times \mathbf{PP} \times \mathbf{OBJ2}$$

WH: Who, What, When, Where, Why, How, How much

AUX: Auxiliary verbs, including negations. i.e. is, might, wo n't

SBJ, OBJ1, OBJ2: someone, something, do something, etc.

TRG: Target verb, including inflected forms.

PP: Preposition. i.e. to, for, from, about, etc.

Writing Questions

WH*	AUX	SBJ	TRG*	OBJ1	PP	OBJ2
Who			***built***	something		
What	had	someone	***said***			
When	was	someone	***expected***		to	do something
Where	might	something	***rise***		from	

Annotation Interface

Revenue **rose** 33 % to $ 378.1 million from $ 283.8 million .

rose

WH	AUX	PH1	TRG	PH2	PP	PH3	?	Answer1	Answer2

Who
What
When
Where
How
How much
Why

But Western Union has said it must **lower** the interest rate on its debt to regain full financial health .

lower

WH	AUX	PH1	TRG	PH2	PP	PH3	?	Answer1	Answer2

Annotation Interface

Revenue **rose** 33 % to $ 378.1 million from $ 283.8 million .

rose

WH	AUX	PH1	TRG	PH2	PP	PH3	?	Answer1	Answer2

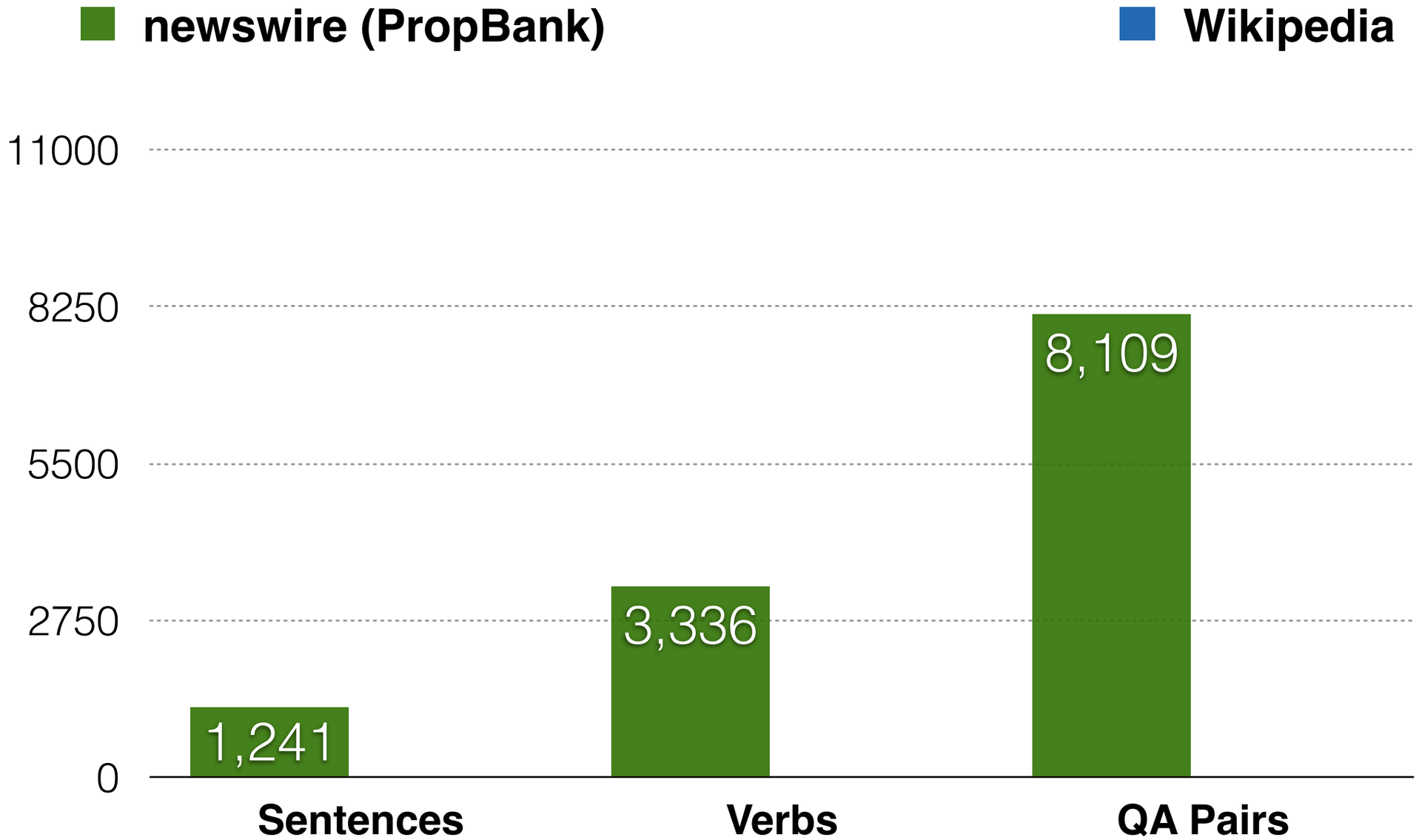
Who
What
When
Where
How
How much
Why

But Western Union has said it must **lower** the interest rate on its debt to regain full financial health .

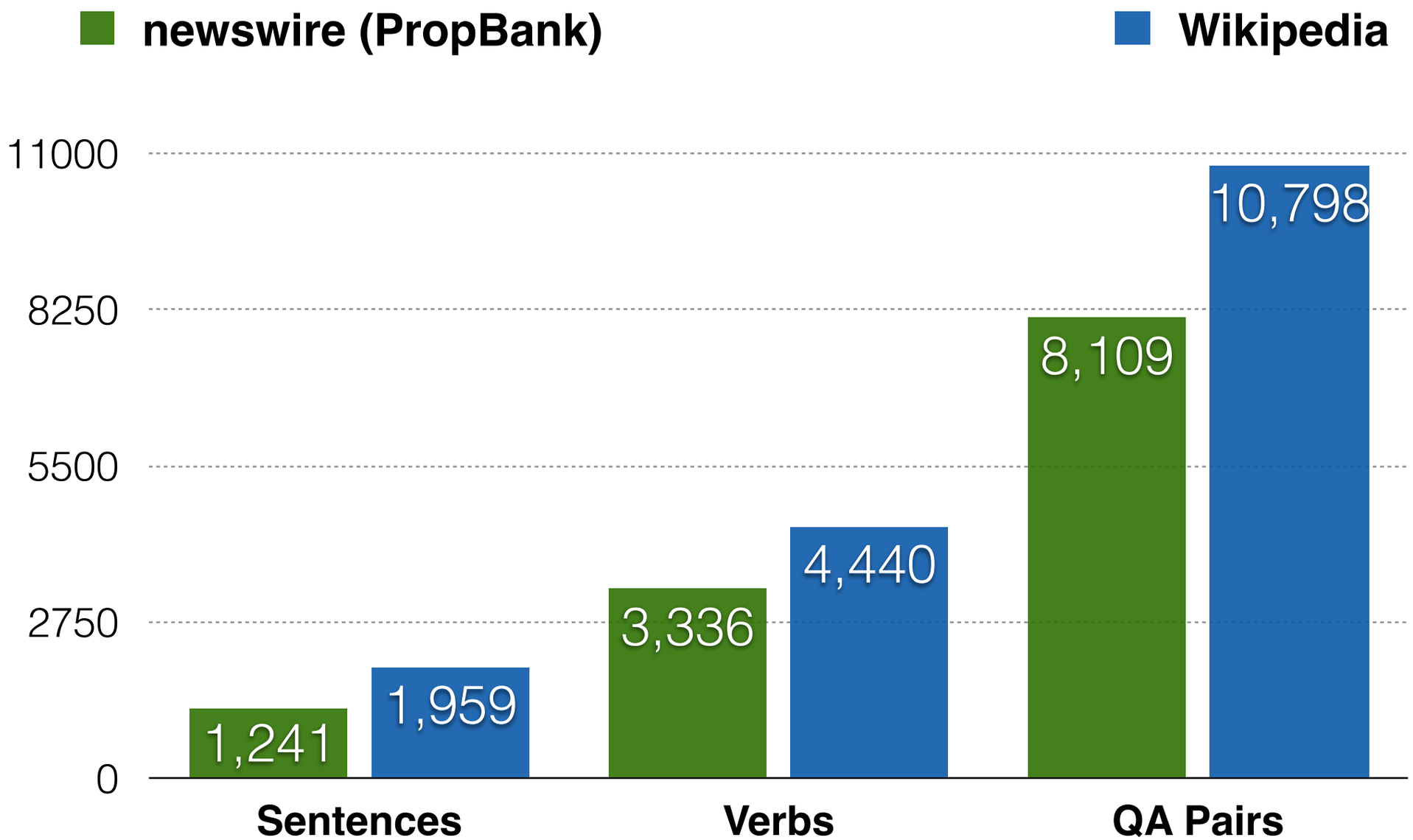
lower

WH	AUX	PH1	TRG	PH2	PP	PH3	?	Answer1	Answer2

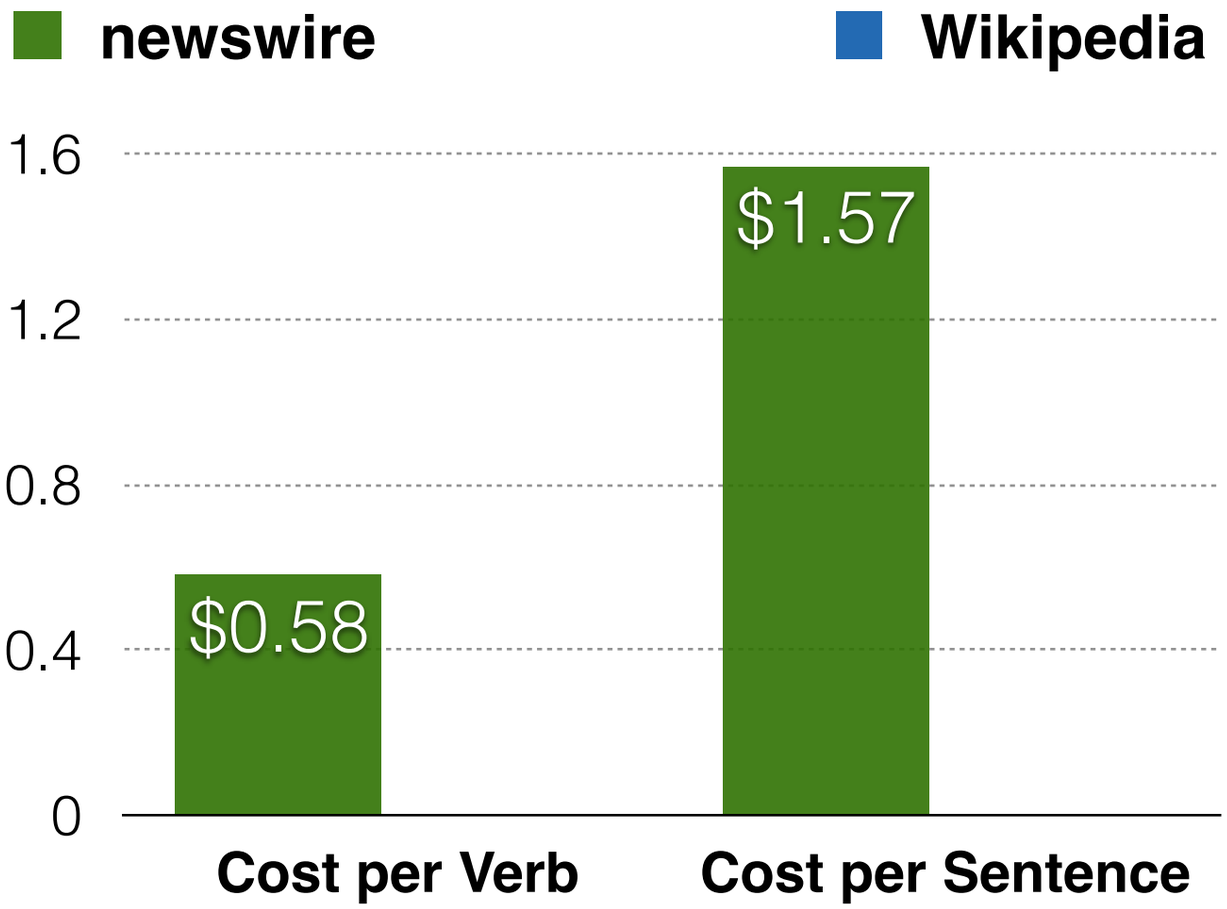
Dataset Statistics



Dataset Statistics

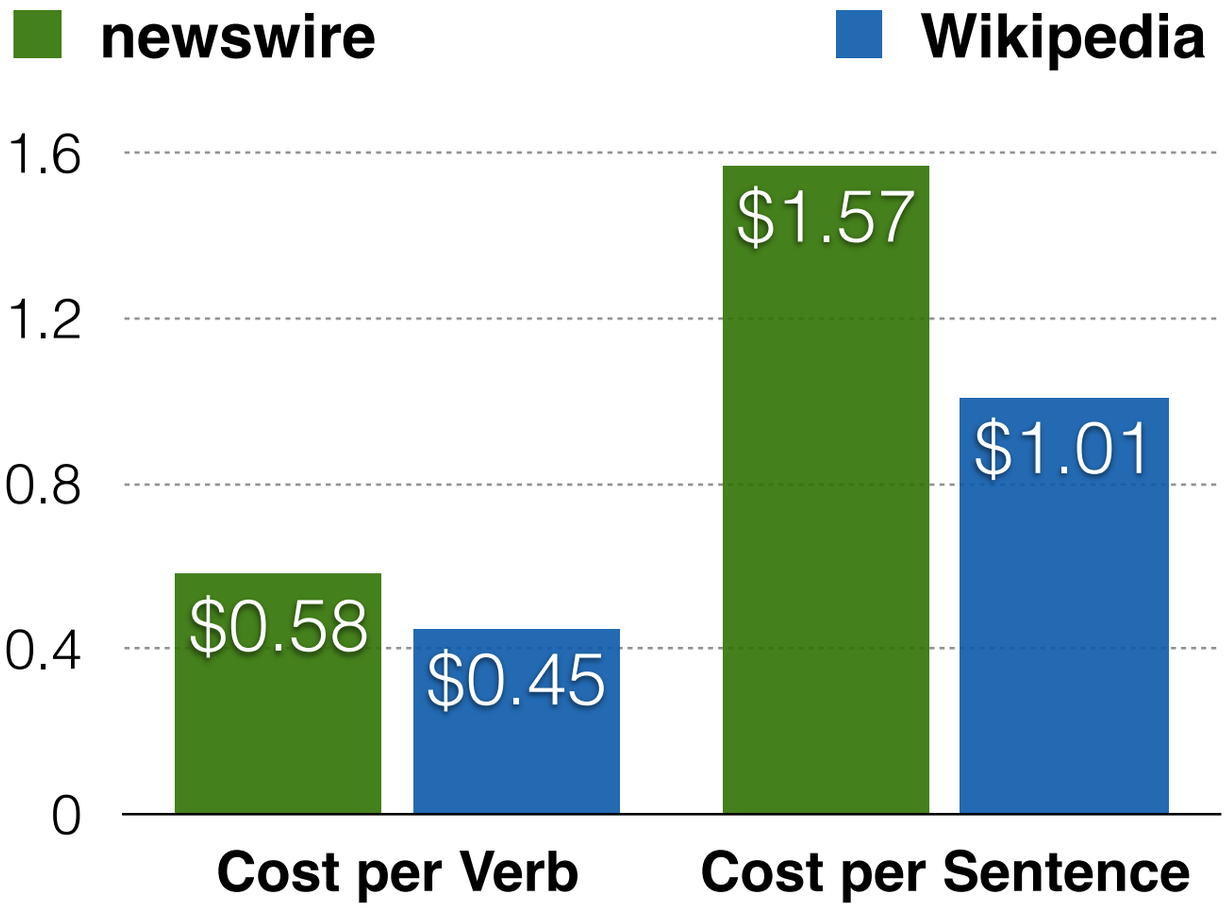


Cost and Speed



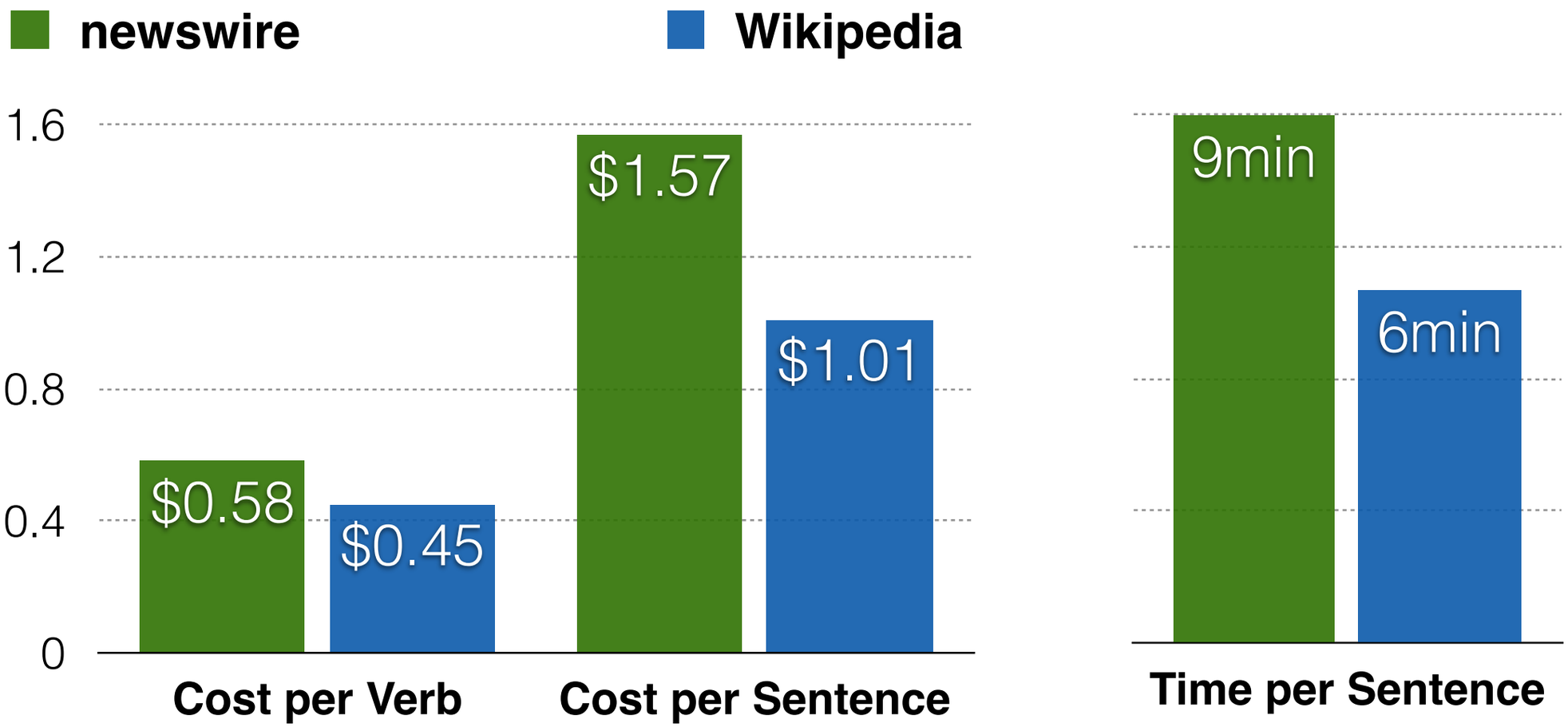
- Part-time freelancers from upwork.com (hourly rate: $10)
- ~2h screening process for native English proficiency

Cost and Speed



- Part-time freelancers from upwork.com (hourly rate: $10)
- ~2h screening process for native English proficiency

Cost and Speed



- Part-time freelancers from upwork.com (hourly rate: $10)
- ~2h screening process for native English proficiency

Sample Annotation

Sentence: Clad in his trademark black velvet suit , the soft-spoken clarinetist announced that . . . and that it was his mother 's birthday , so he was going to ***play*** her favorite tune from the record .

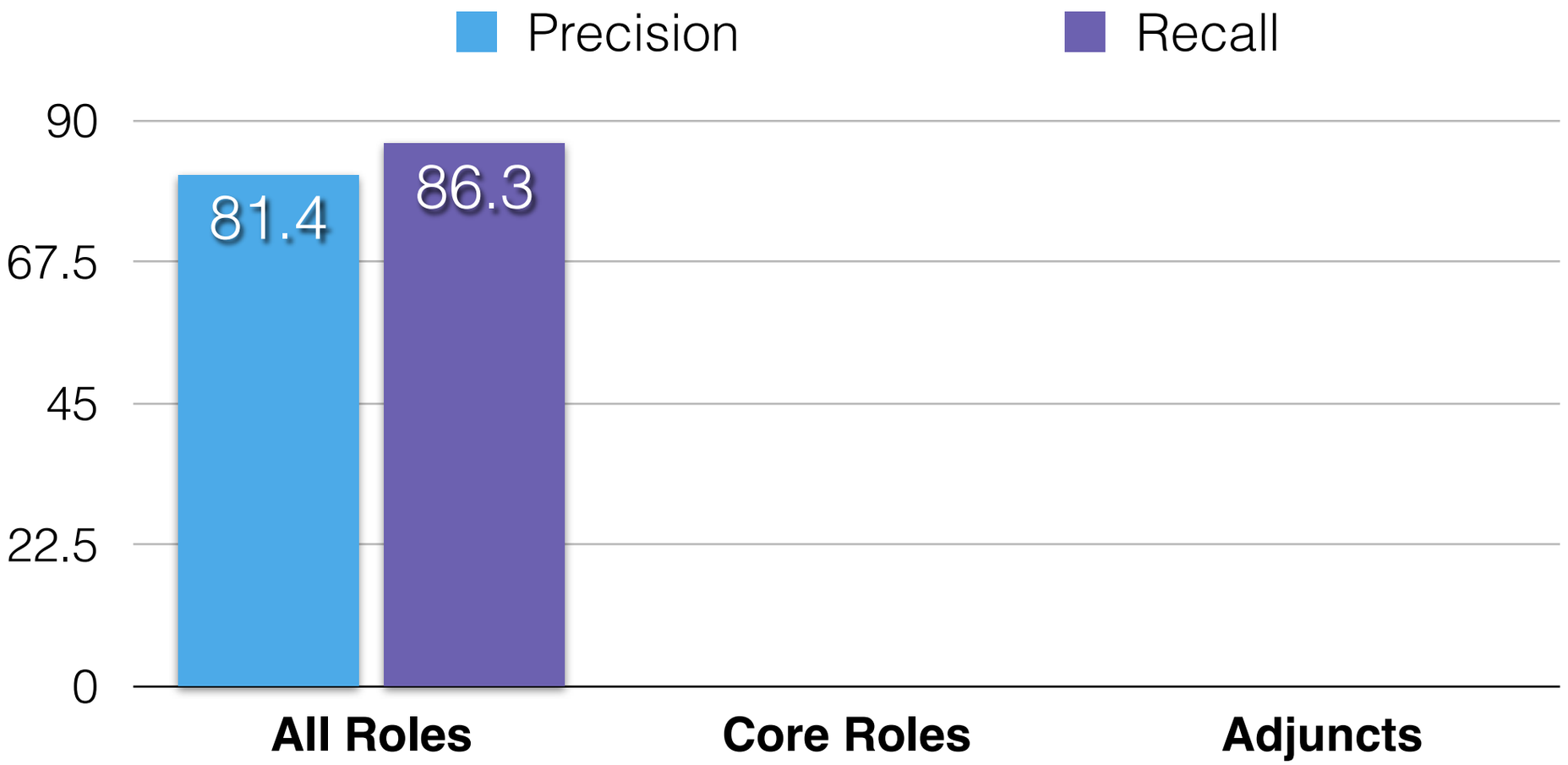
QA-SRL	**PropBank (CoNLL-2009)**
Who would play something ? the soft-spoken clarinetist / **he**	ARG0: **he**
What would be played ? her favorite **tune** from the record	ARG1: **tune**
When would someone play something? his mother 's birthday	/

Sample Annotation

Sentence: Clad in his trademark black velvet suit , the soft-spoken clarinetist announced that . . . and that it was his mother 's birthday , so he was going to ***play*** her favorite tune from the record .

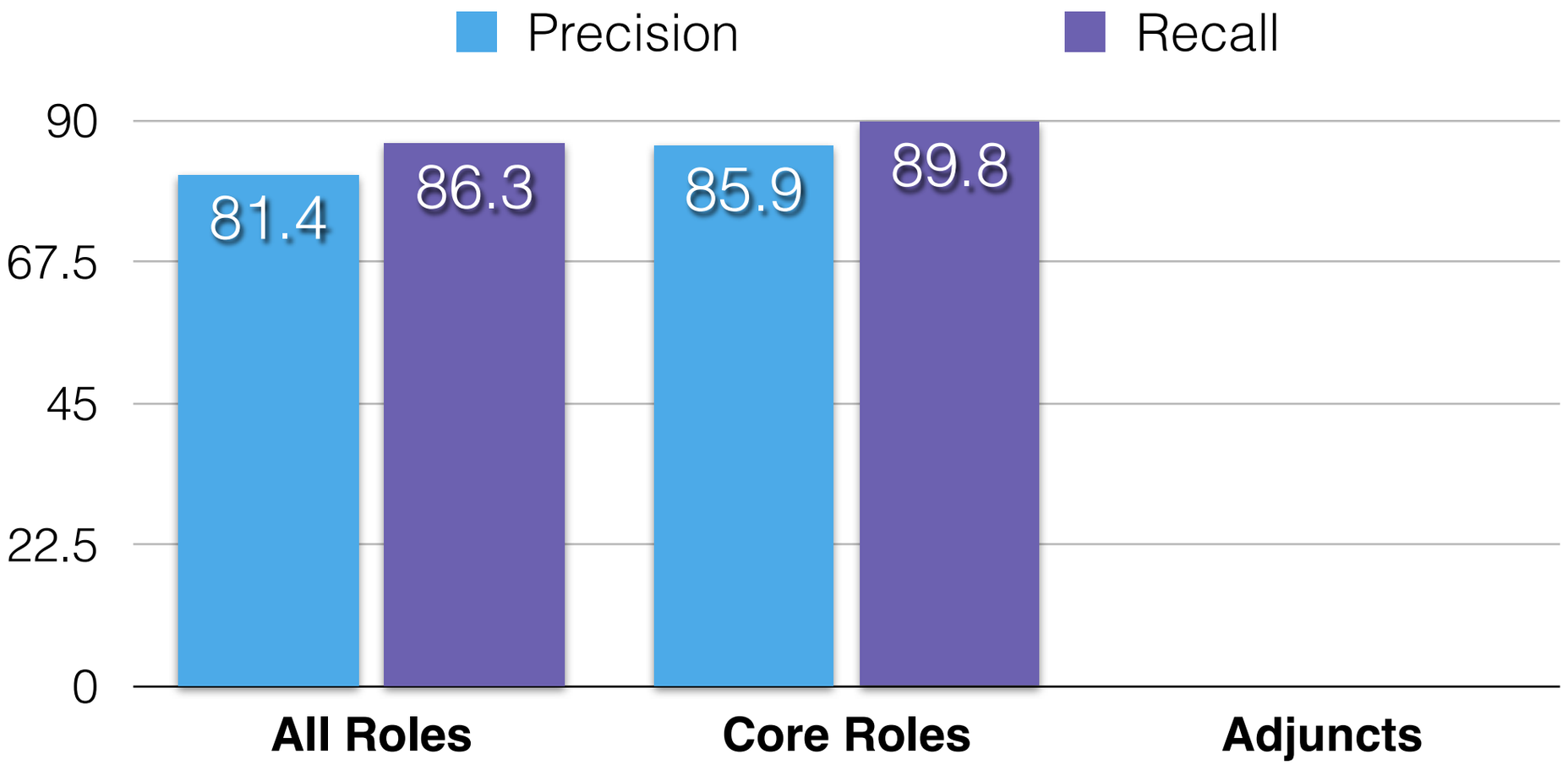
	QA-SRL	**PropBank (CoNLL-2009)**
match	Who would play something ? the soft-spoken clarinetist / **he**	ARG0: **he**
match	What would be played ? her favorite **tune** from the record	ARG1: **tune**
precision loss	When would someone play something? his mother 's birthday	/

Agreement with PropBank: Results



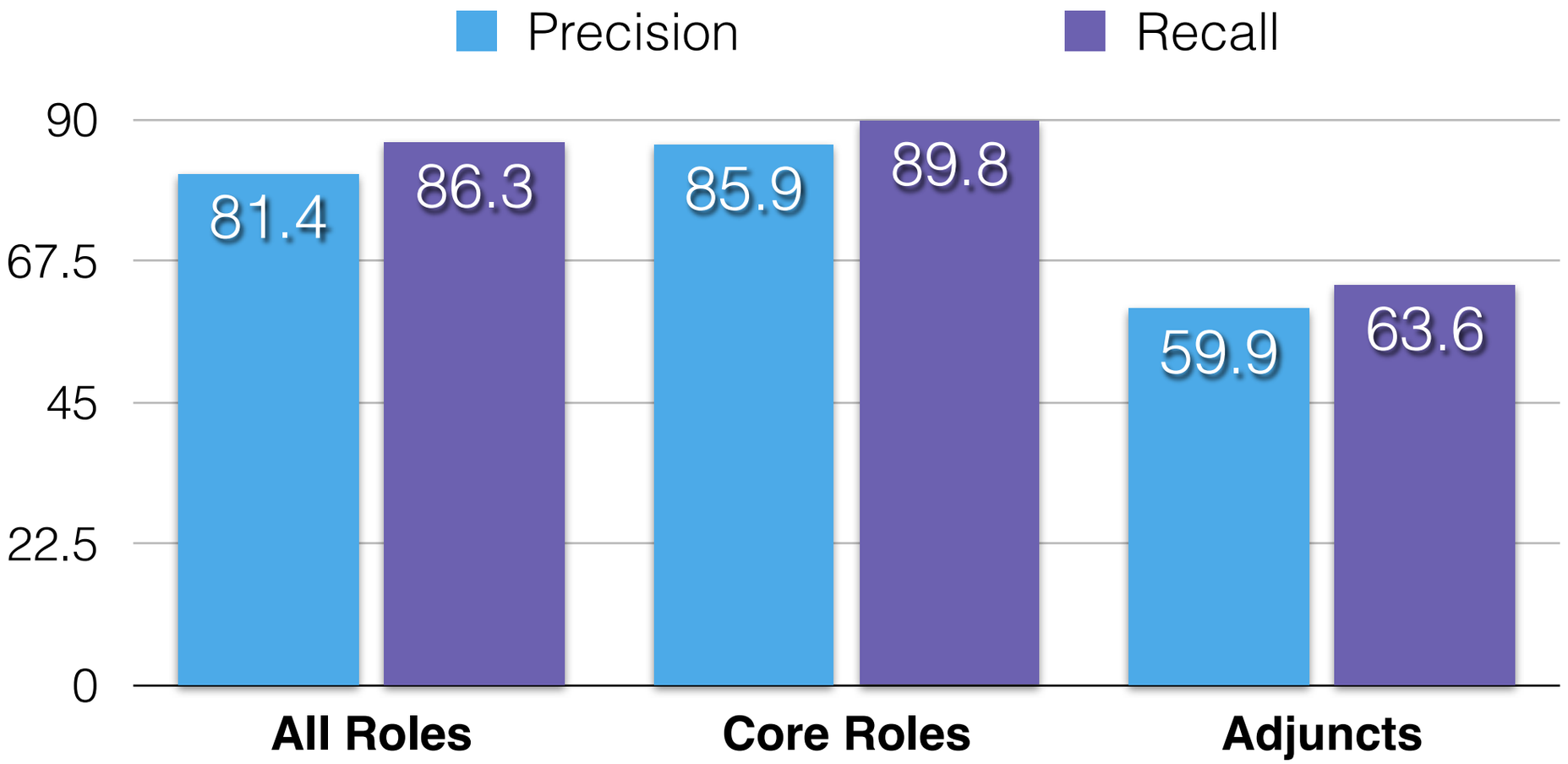
Core Roles: A0-A5
Adjuncts: ADV, CAU,DIR, EXT, LOC, MNR, PNC, PRD, TMP

Agreement with PropBank: Results



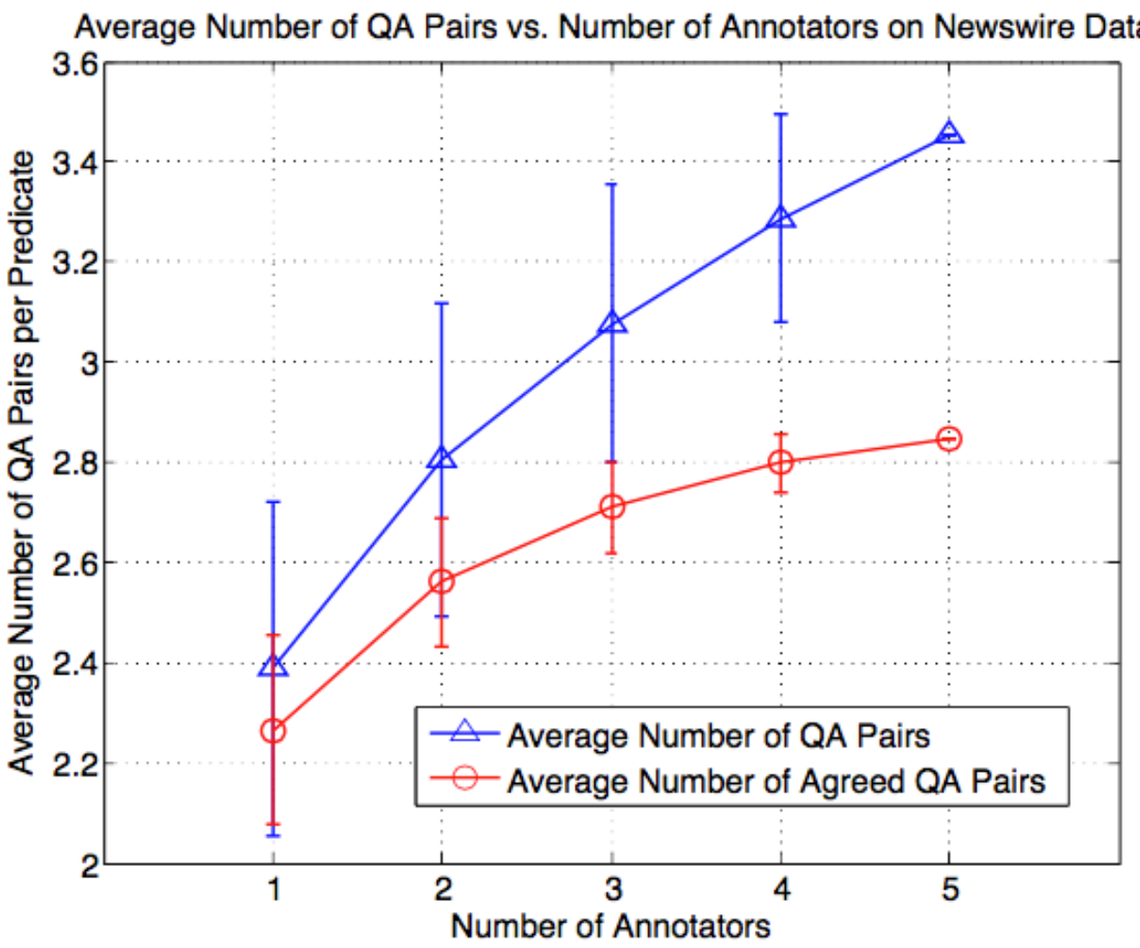
Core Roles: A0-A5
Adjuncts: ADV, CAU,DIR, EXT, LOC, MNR, PNC, PRD, TMP

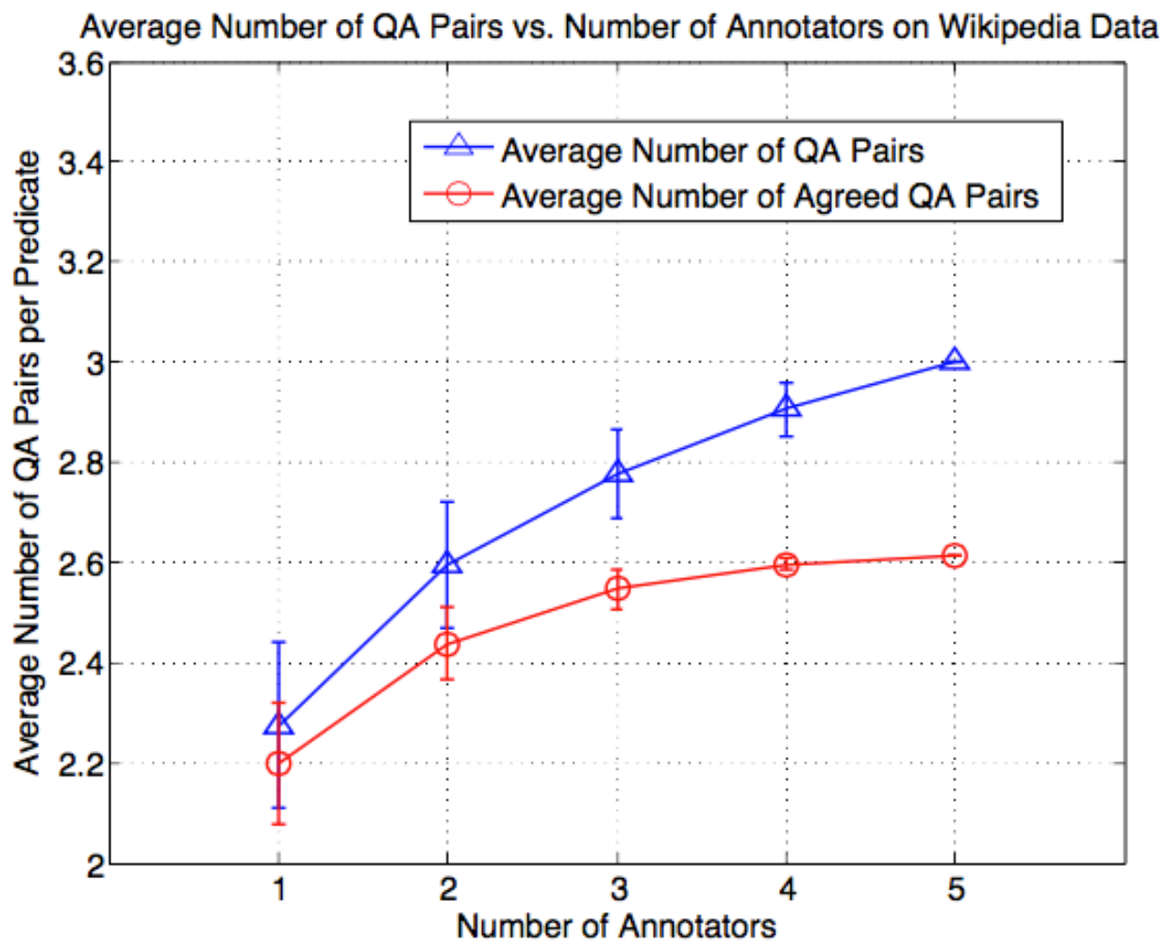
Agreement with PropBank: Results



Core Roles: A0-A5
Adjuncts: ADV, CAU,DIR, EXT, LOC, MNR, PNC, PRD, TMP

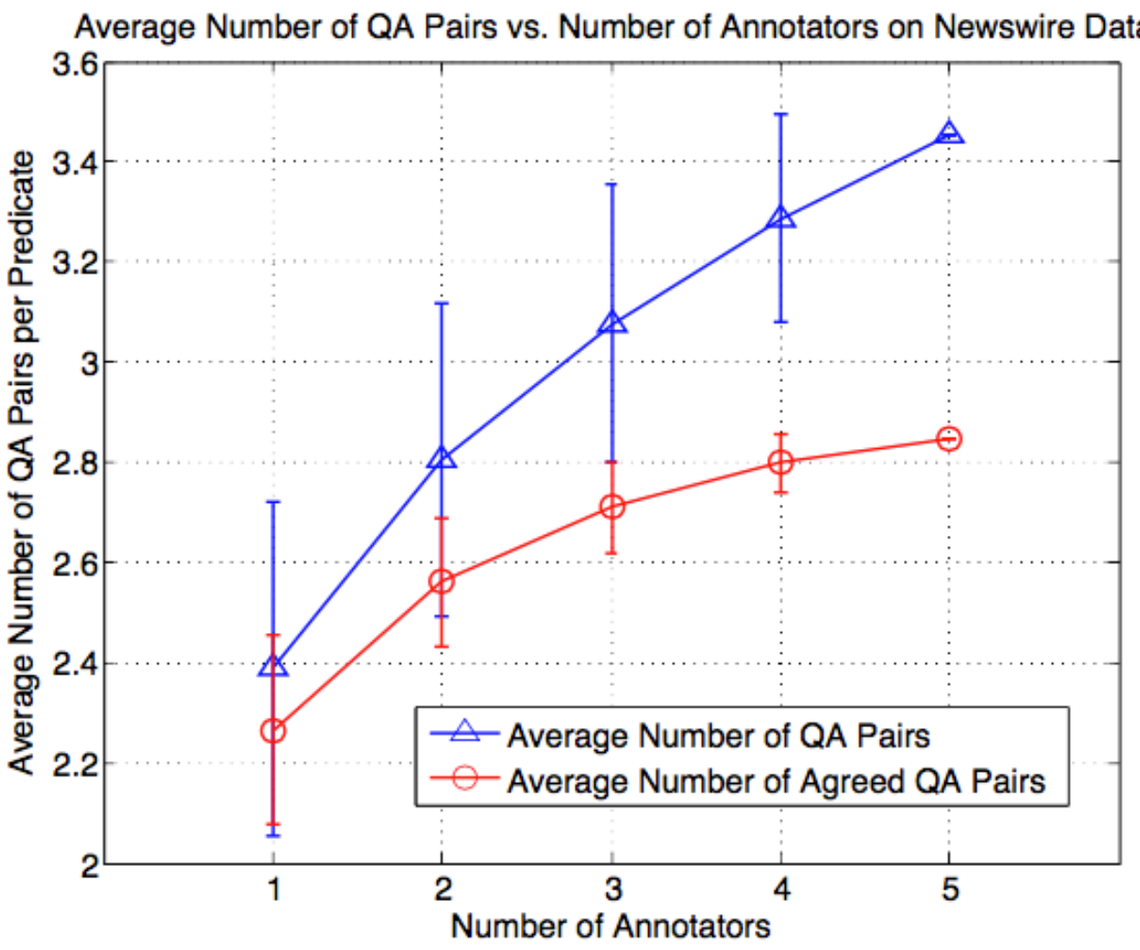
Inter-Annotator Agreement

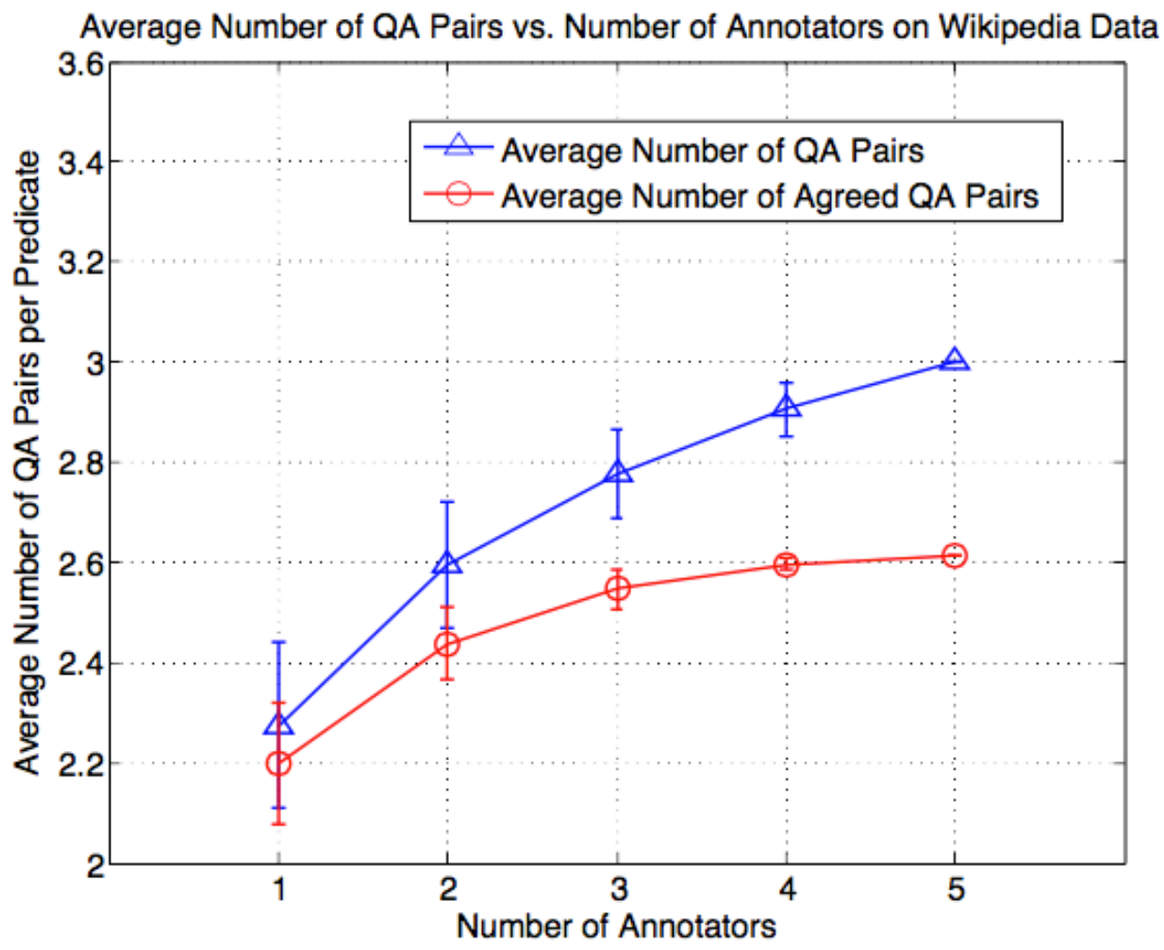




- **QA Equivalence:** Same wh-word + Overlapping answers
- **Agreed QA Pairs:** Proposed by at least 2 of the 5 annotators

Inter-Annotator Agreement





- Agreed QA pairs by five annotators: **2.6-2.8 QA/verb**
- One annotator can recover: **2.2-2.3 QA/verb (80%)**

Wh-words vs. PropBank Roles

	Who	What	When	Where	Why	How	HowMuch
ARG0	1575	414	3	5	17	28	2
ARG1	285	2481	4	25	20	23	95
ARG2	85	364	2	49	17	51	74
ARG3	11	62	7	8	4	16	31
ARG4	2	30	5	11	2	4	30
ARG5	0	0	0	1	0	2	0
AM-ADV	5	44	9	2	25	27	6
AM-CAU	0	3	1	0	23	1	0
AM-DIR	0	6	1	13	0	4	0
AM-EXT	0	4	0	0	0	5	5
AM-LOC	1	35	10	89	0	13	11
AM-MNR	5	47	2	8	4	108	14
AM-PNC	2	21	0	1	39	7	2
AM-PRD	1	1	0	0	0	1	0
AM-TMP	2	51	341	2	11	20	10

Outline

Motivation and Intuition

- Semantic Role Labeling
- Our Method: QA-SRL

Data Collection and Analysis

- Annotation Task Design
- Dataset Statistics
- Quality Analysis

Learning Tasks and Baselines

- Question Generation
- Answer Identification

Future Work and Conclusion

Question Generation

Task	Given sentence **s** and target verb **v**, predict a set of questions that are *grammatical* and *answerable*.
Motivation	In the future, automate part of the annotation process, further reduce cost and speed up annotation.

Question Generation: Basic Idea

- Pick a role in the sentence
- Predict the right pronoun.
- Fill in the rest of the question.

$s =$ They ***increased*** the rent this year .

Question Generation: Basic Idea

- Pick a role in the sentence
- Predict the right pronoun.
- Fill in the rest of the question.

$s =$ They ***increased*** the rent this year .

✓ Who increased something ?

✗ Why was something increased ?

role not present

Question Generation: Basic Idea

- Pick a role in the sentence
- Predict the right pronoun.
- Fill in the rest of the question.

$s =$ They ***increased*** the rent this year .

✓ Who increased something ?

✗ Why was something increased ?

role not present

✗ What increased someone ?

wrong pronoun

Question Generation: Basic Idea

- Pick a role in the sentence
- Predict the right pronoun.
- Fill in the rest of the question.

$s =$ They ***increased*** the rent this year .

✓ Who increased something ?

✗ Why was something increased ? **role not present**

✗ What increased someone ? **wrong pronoun**

✗ When increased someone something ? **wrong template**

Question Generation: 2-Step Method

Step 1: Role/Pronoun Prediction as Multi-label Learning

$$\mathcal{R} = \{\mathrm{R0}, \mathrm{R1}, \mathrm{R2}, \mathrm{R2}[pp], wh, wh[pp]\}$$
$$wh \in \{\mathbf{Where}, \mathbf{When}, \mathbf{Why}, \mathbf{How}, \mathbf{HowMuch}\}$$
$$\mathcal{L} = \{role{:}pronoun\_val \mid role \in \mathcal{R}\}$$

*Details can be found in paper

Question Generation: 2-Step Method

Step 1: Role/Pronoun Prediction as Multi-label Learning

$$\mathcal{R} = \{\mathrm{R0}, \mathrm{R1}, \mathrm{R2}, \mathrm{R2}[pp], wh, wh[pp]\}$$
$$wh \in \{\mathbf{Where}, \mathbf{When}, \mathbf{Why}, \mathbf{How}, \mathbf{HowMuch}\}$$
$$\mathcal{L} = \{role{:}pronoun\_val \mid role \in \mathcal{R}\}$$

Step 2: Template-based Generation with Abstract Questions

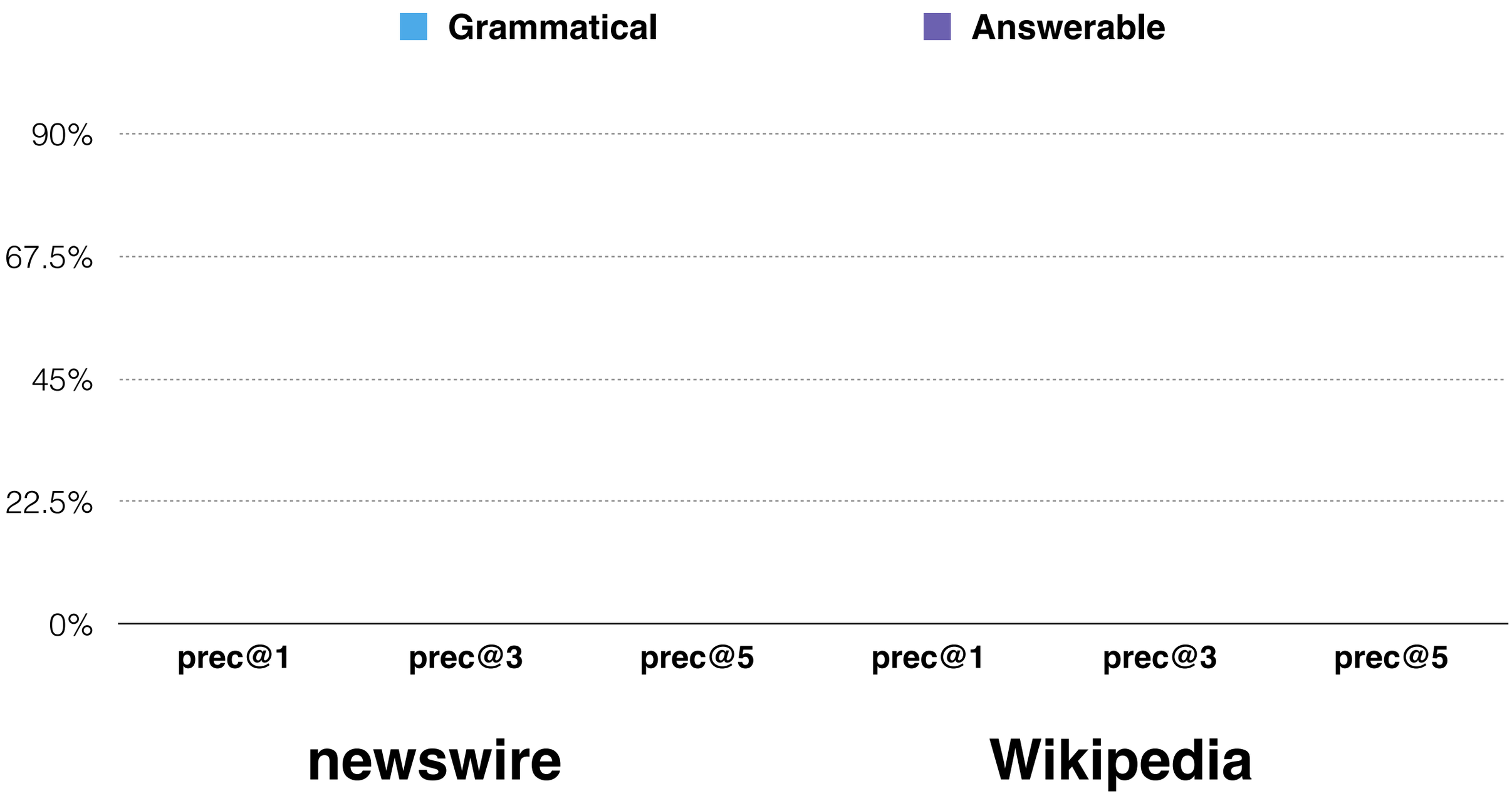
R0:someone R1:something

WH	**SBJ**	**Voice**	**OBJ1**	**OBJ2**
R0	/	active	R1	/

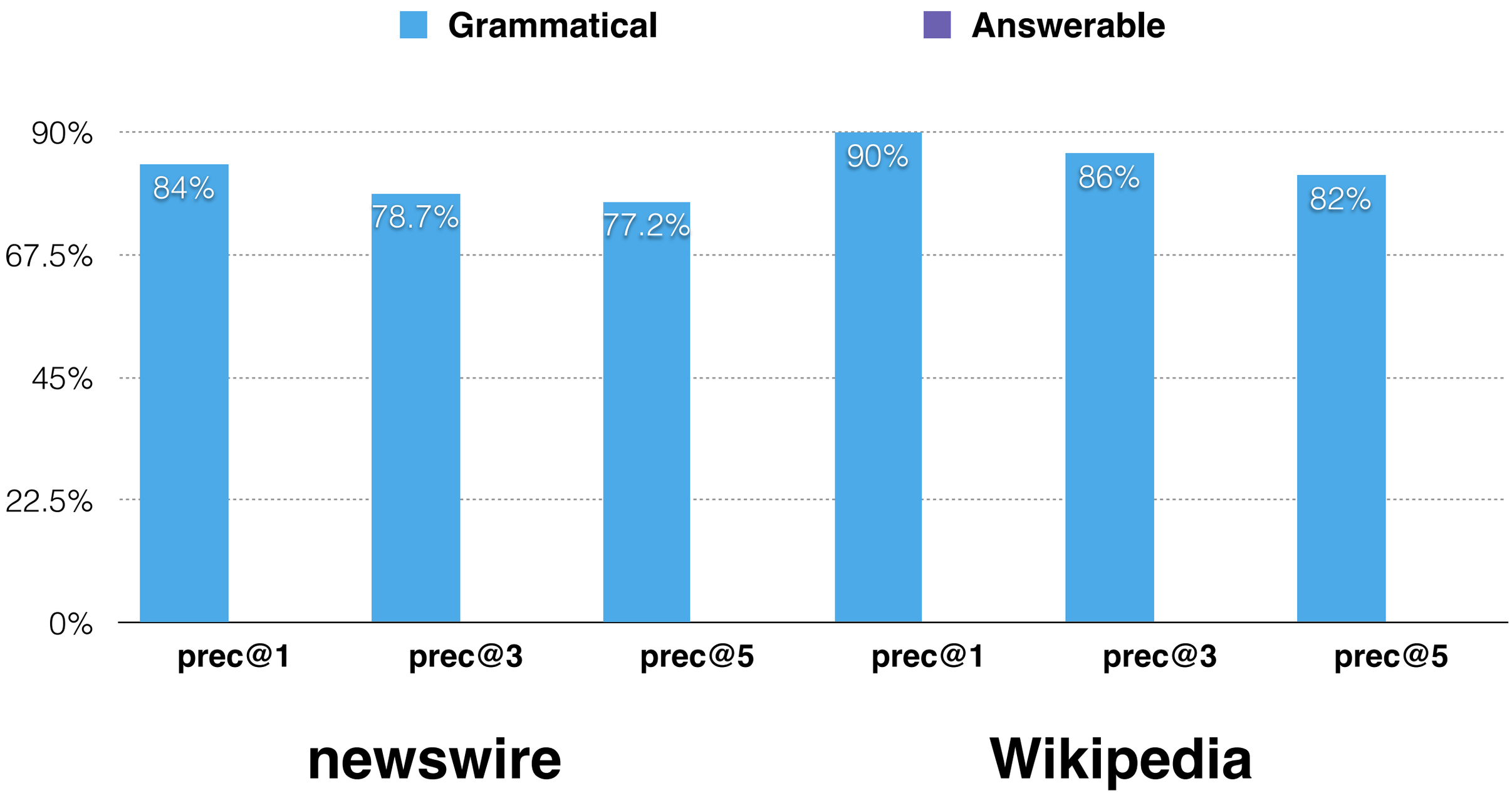
→ Who increased something ?

*Details can be found in paper

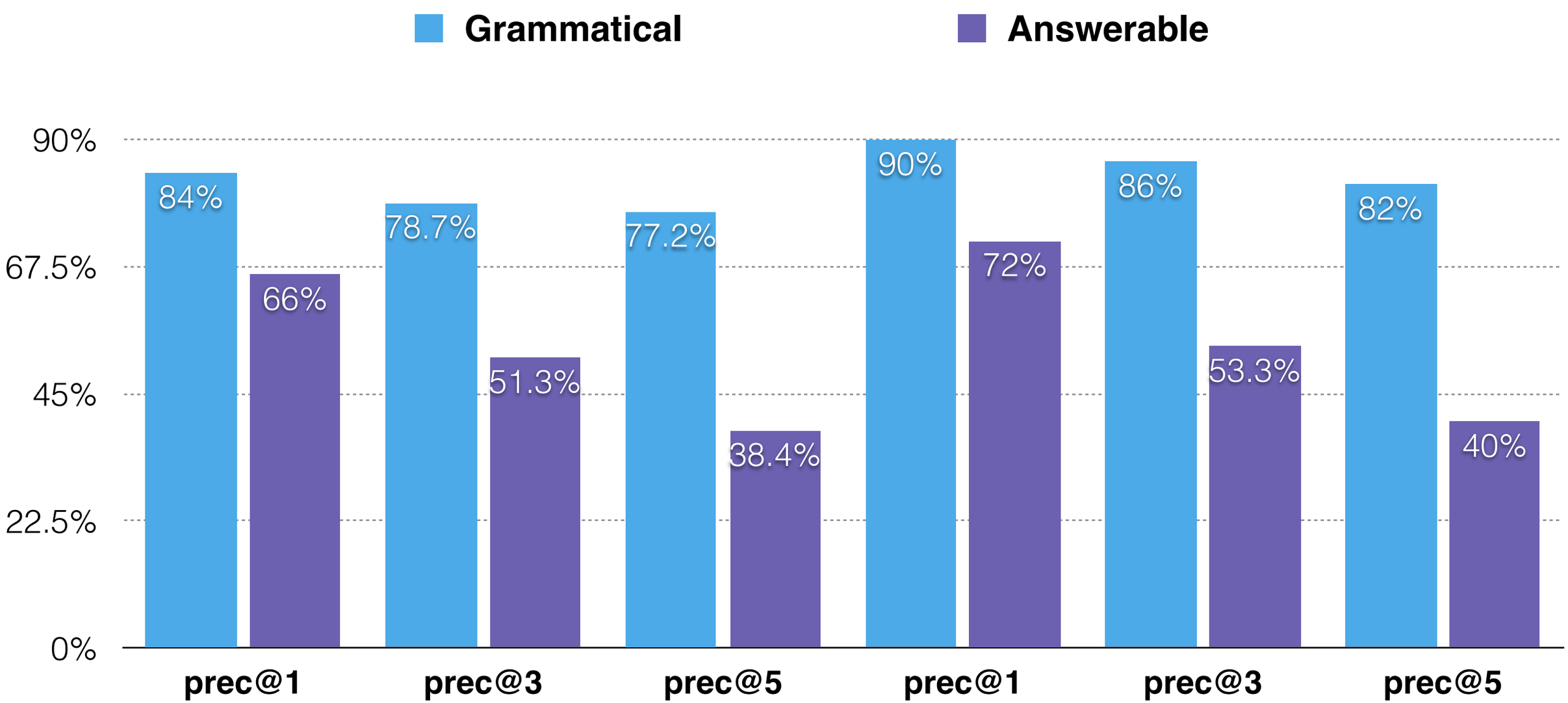
Question Generation: Results



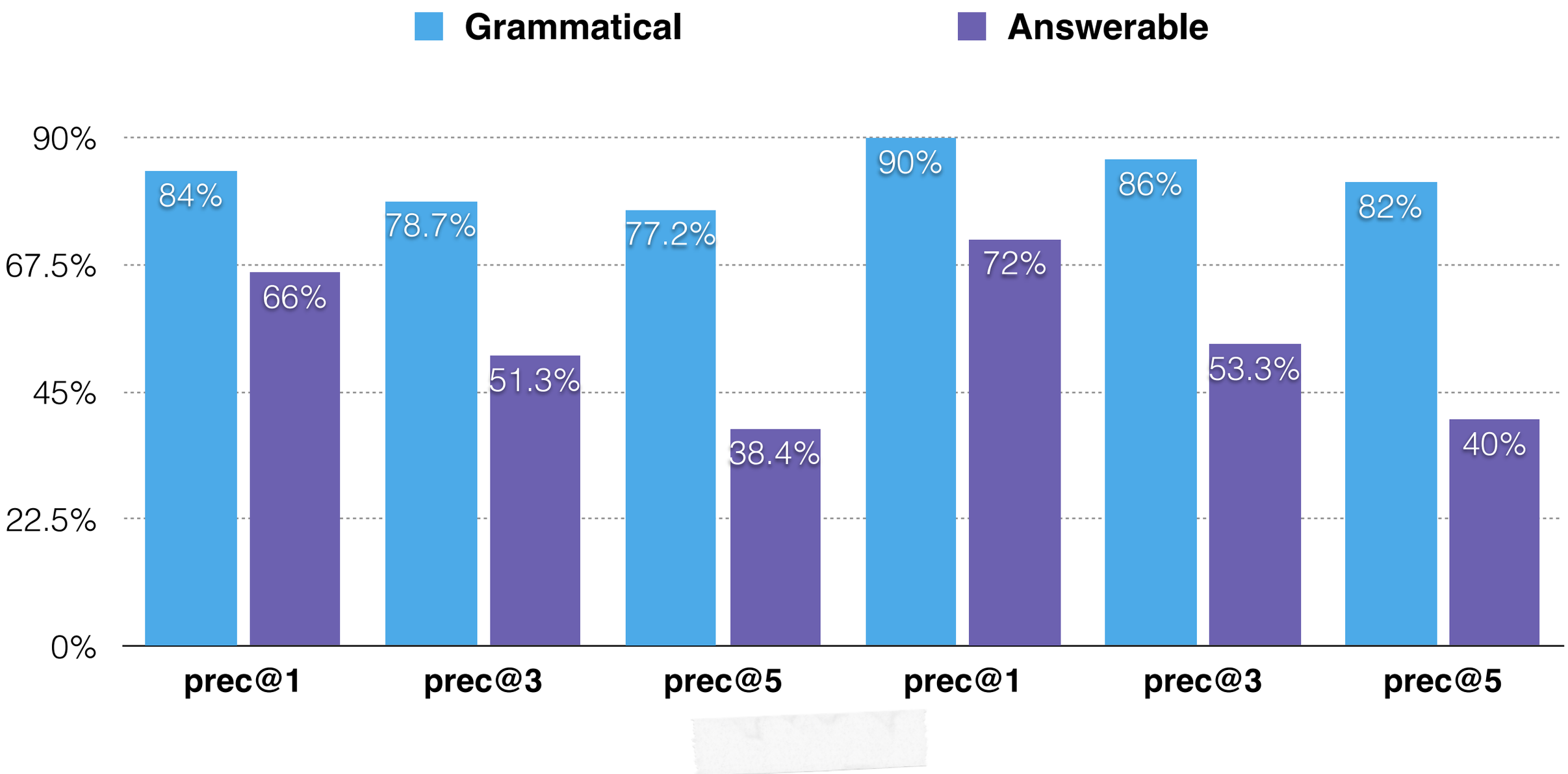
Question Generation: Results



Question Generation: Results



Question Generation: Results



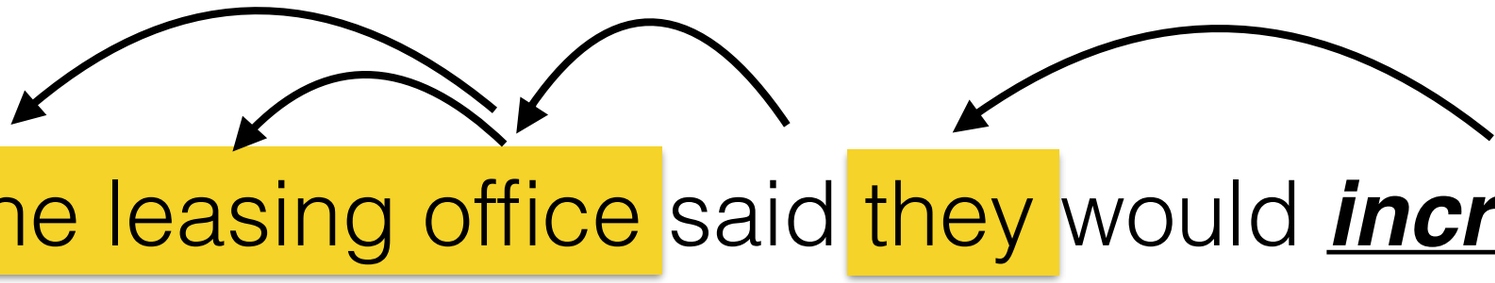
- In question prediction: **2 Question/verb** answerable
- In annotated data: **2.6-2.8 QA/verb**

Answer Identification

Task	Given sentence **s**, target verb **v**, and question **q**, predict a word in the sentence that answers the question **q**.

Motivation	In the future, build an end-to-end SRL system trained by QA-SRL data. (Analogy to SRL - questions:roles, answers:arguments).

Answer Identification: Basic Idea



$s =$ The leasing office said they would ***increase*** the rent .

$v =$ ***increase*** $q =$ Who would increase something ?

Arcs from k-best dependency trees

Annotated answer spans. Space: $2^{|s|}$

Answer Identification: Basic Idea

$s =$ The leasing office said they would ***increase*** the rent .

$v =$ ***increase*** $q =$ Who would increase something ?

Arcs from k-best dependency trees

Annotated answer spans. Space: $2^{|s|}$

Training samples: $\langle s, v, q, \text{office} \rangle, \langle s, v, q, \text{they} \rangle$ Space: $|s|$

Answer Identification: Basic Idea

$s =$ The leasing office said they would ***increase*** the rent .

$v =$ ***increase*** $q =$ Who would increase something ?

Arcs from k-best dependency trees

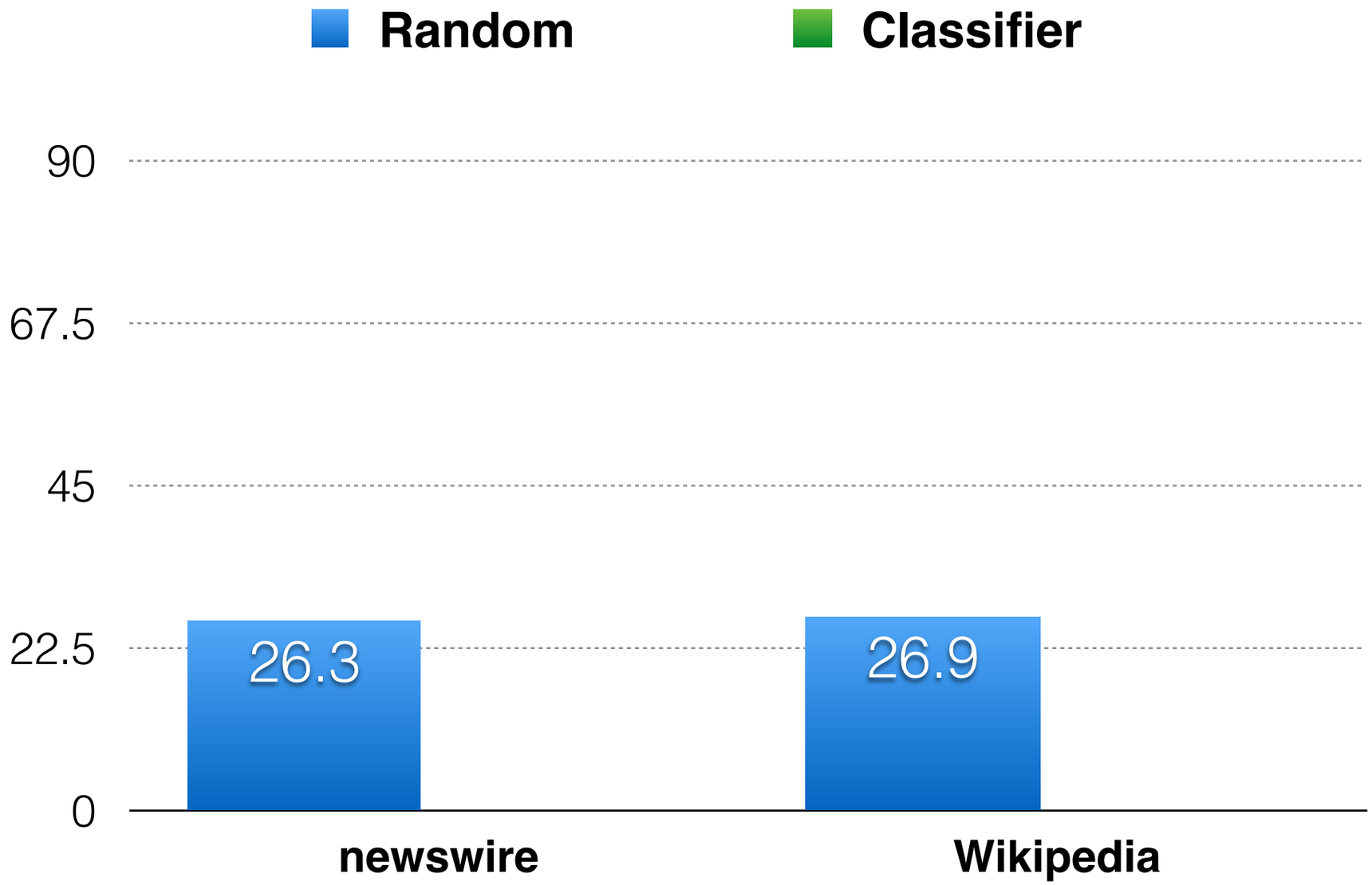
Annotated answer spans. Space: $2^{|s|}$

Training samples: $\langle s, v, q, \text{office} \rangle, \langle s, v, q, \text{they} \rangle$ Space: $|s|$

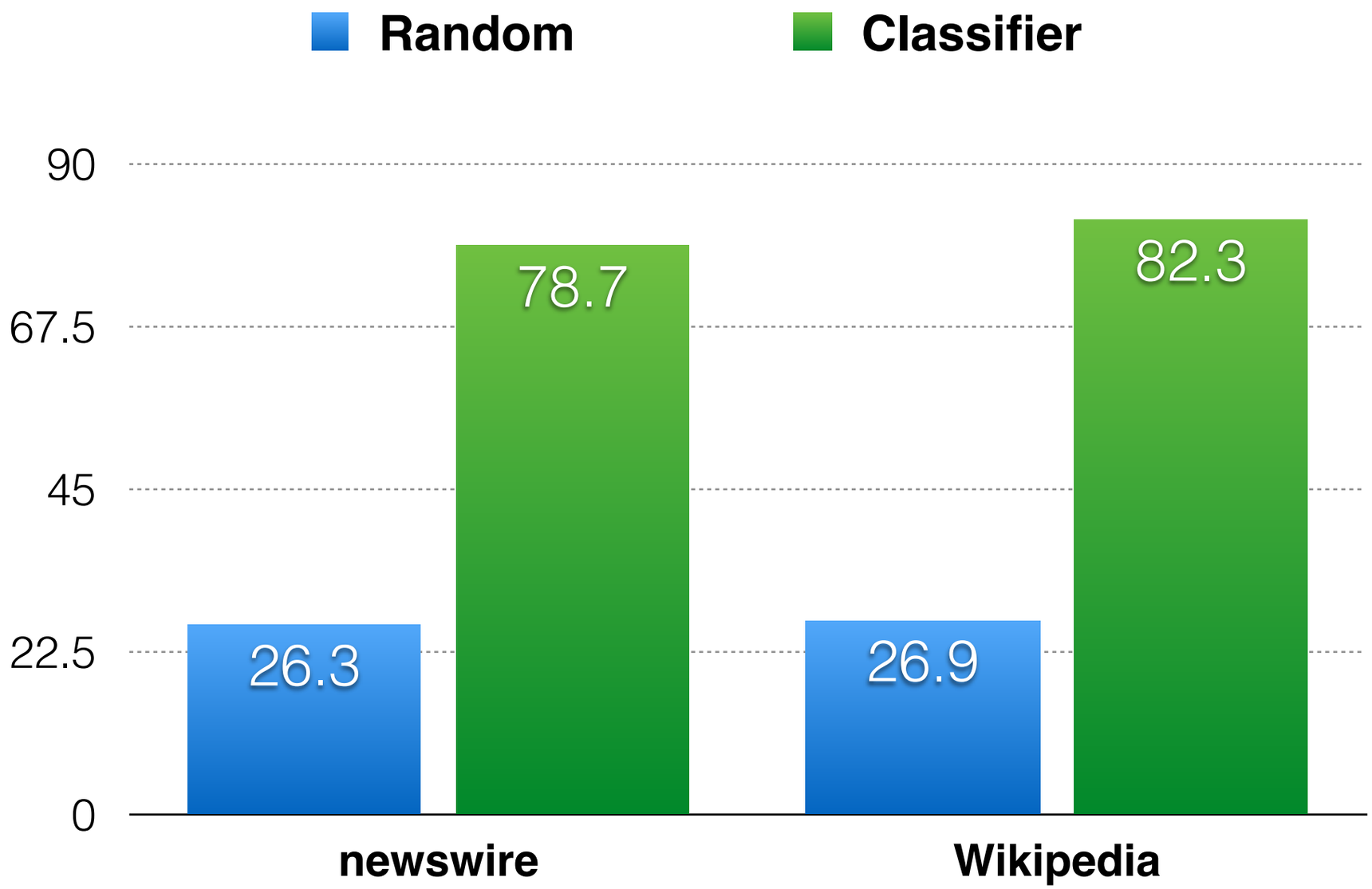
✓ Ex. of correct predictions: “**office**”, “**leasing**”, “**they**”

✗ Ex. of wrong predictions: “**rent**”

Answer Identification: Results



Answer Identification: Results



Outline

Motivation and Intuition

- Semantic Role Labeling
- Our Method: QA-SRL

Data Collection and Analysis

- Annotation Task Design
- Dataset Statistics
- Quality Analysis

Learning Tasks and Baselines

- Question Generation
- Answer Identification

Future Work and Conclusion

- Generalization
- Question Suggestion
- Training a Joint Parser

Future Work: Generalization

- Generalize to non-verbal predicates:

S: The rent increase came as a ***shock*** to us .

Q: Who was ***shocked*** ? **A:** us

- Generalize to other languages:

他们 今年 涨了 房租 。

they this year increased the rent

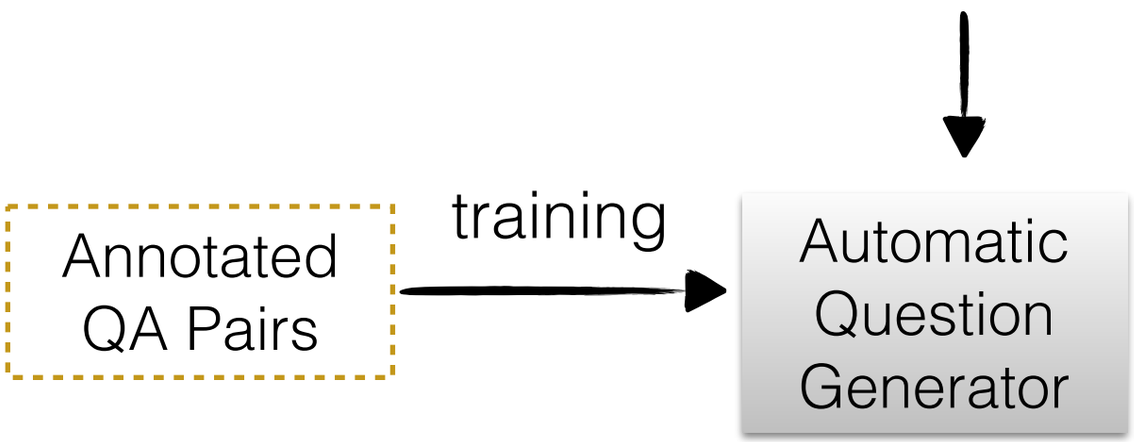
Q: 房租 什么 时候 涨了 ？ **A:** 今年

rent when increased this year

Future Work:
Automatic Question Suggestion

Given new sentence and verb:

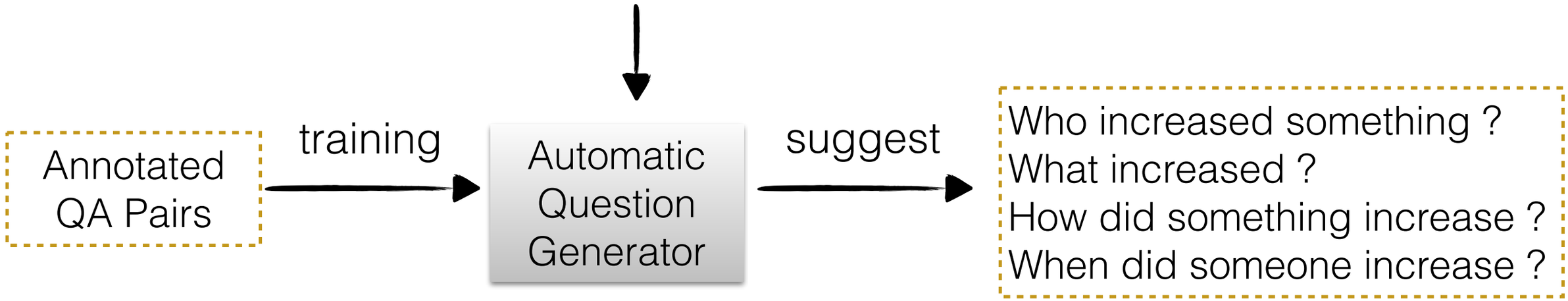
"I can 't believe they ***increased*** the rent by so much ."



Future Work:
Automatic Question Suggestion

Given new sentence and verb:

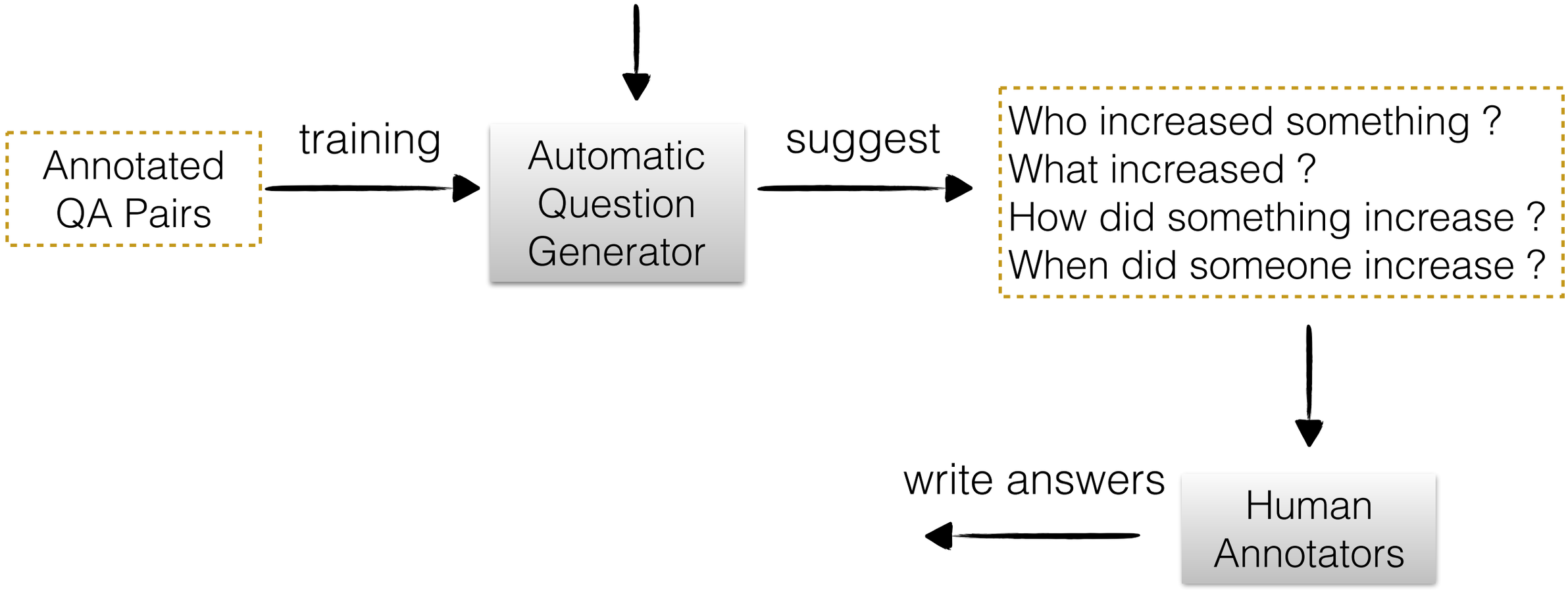
"I can 't believe they ***increased*** the rent by so much ."



Future Work:
Automatic Question Suggestion

Given new sentence and verb:

"I can 't believe they ***increased*** the rent by so much ."



Future Work: Automatic Question Suggestion

Given new sentence and verb:

"I can 't believe they ***increased*** the rent by so much ."

Annotated QA Pairs

training

Automatic Question Generator

suggest

Who increased something ?
What increased ?
How did something increase ?
When did someone increase ?

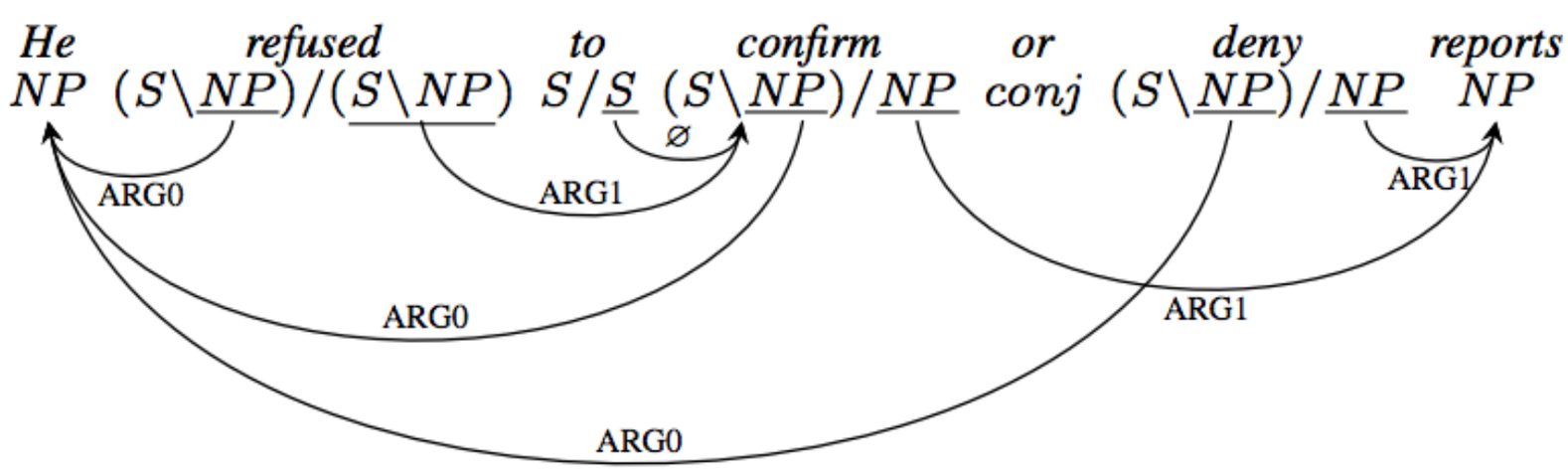
Human Annotators

write answers

Who increased something? - **they**
What increased? - **the rent**
How did something increase? - **by so much**
When did someone increase? - **unanswerable**

Future Work: Training a Joint Parser

- Use question-answer pairs to train a joint parser, to improve on both syntax and semantics
- Combine with other SRL data, i.e. PropBank, FrameNet

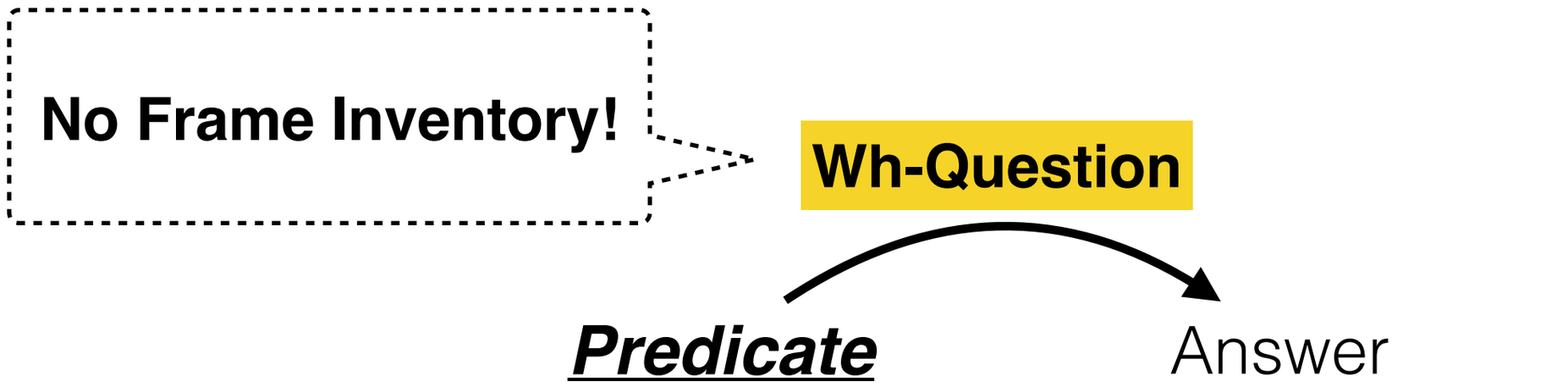


Joint A* CCG Parsing and Semantic Role Labeling, Lewis et al., EMNLP-2015. **(Presentation: Sunday 6B)**

Contributions

- Introduced question-answer driven semantic role labeling (QA-SRL).
- High quality QA annotation with a lightweight template-based scheme.
- Two new QA-SRL learning baselines: question generation and answer identification.
- Releasing data and annotation tool - https://dada.cs.washington.edu/qasrl/

Thank You!
Questions?



QA-SRL Project Page:
https://dada.cs.washington.edu/qasrl/