Pseudo Zero Pronoun Resolution Improves Zero Anaphora Resolution

Ryuto Konno[1], Shun Kiyono[2,3], Yuichiroh Matsubayashi[3,2], Hiroki Ouchi[4,2], Kentaro Inui[3,2]

[1]Recruit Co., Ltd. [2]RIKEN [3]Tohoku University
[4]Nara Institute of Science and Technology

Summary

- Task: **Zero Anaphora Resolution (ZAR)** in Japanese

  *The **criminal**'s weapon was found in the victim's room.*

  *It seems that **ϕ** used a hammer.*

- We proposed
  - A new **pretraining task** for ZAR
  - A new **finetuning method** for ZAR

- Results:
  - The two proposals **boost the SoTA** performance of ZAR
  - Our analysis provides new insights on the remaining challenges

Task: Zero Anaphora Resolution (ZAR)

Let's take a look at the following example:

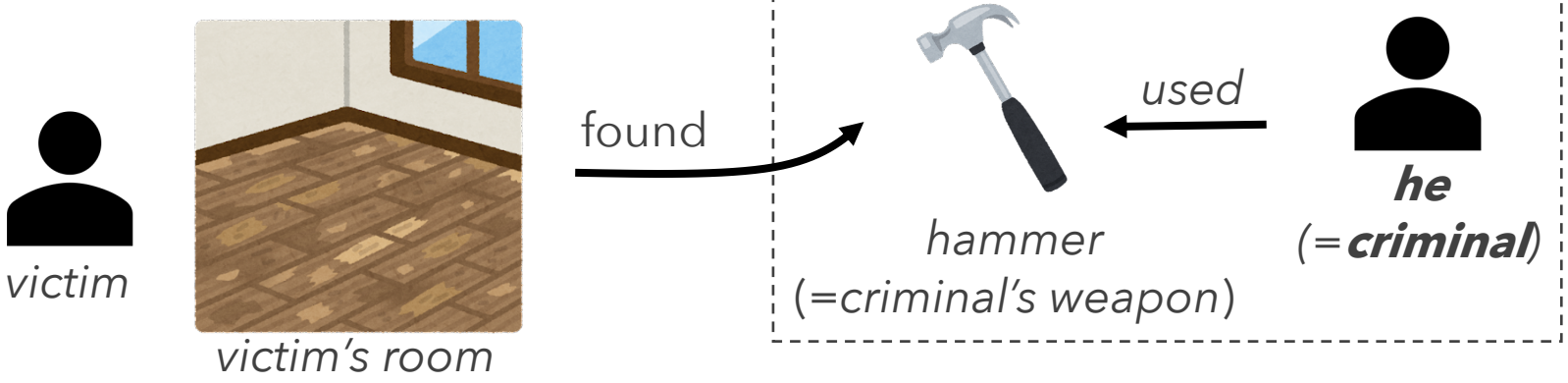
In English

*The **criminal**'s weapon was found in the victim's room.*

refer to

*It seems that **he** used a hammer.*

We can guess what "***he***" means



Task: Zero Anaphora Resolution (ZAR)

- What is "Zero Anaphora"?

In English

*The **criminal**'s weapon was found in the victim's room.*

refer to

*It seems that **he** used a hammer.*

pronoun

We can guess what "***he***" means
→ "***he***" is omitted

In Japanese, Chinese, Korean, Italian, Spanish, …

*The **criminal**'s weapon was found in the victim's room.*

zero anaphora

It seems that ϕ used a hammer.

zero pronoun

- What is "Zero Anaphora Resolution (ZAR)" ?
  - Recognizing the antecedents of zero pronouns

Japanese Zero Anaphora Resolution

*The **criminal**'s weapon was found in the victim's room.*

zero anaphora

*It seems that (**ϕ**) used a hammer.*

In Japanese

被害者の	部屋	から	**犯人**の	凶器が	見つかった。
victim-GEN	*room*	*from*	***criminal***-GEN	*weapon-NOM*	*was found.*

zero anaphora

(**ϕ**-NOM)	ハンマーを	使用した	模様。
(**ϕ**-NOM)	*hammer-ACC*	*used*	*seem.*

Not easy to find where the zero pronoun is in the sentences

Cast to predicate-argument structure analysis

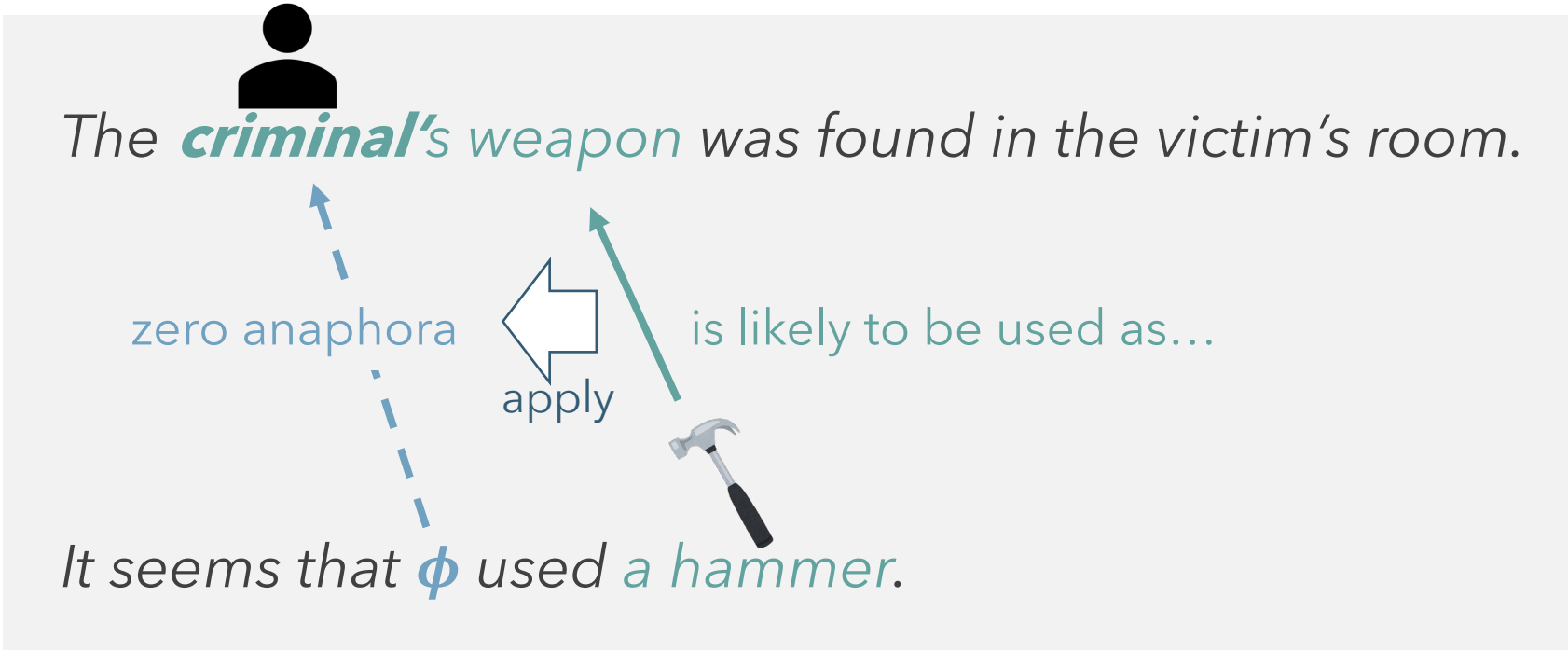
The semantic arguments of "*used*" are…

- Nominative (subject) : *the ciriminal*
- Accusative (direct object) : *a hammer*

Two Research Questions for ZAR Task

The model needs …

(1) To acquire a large amount of anaphoric relational knowledge

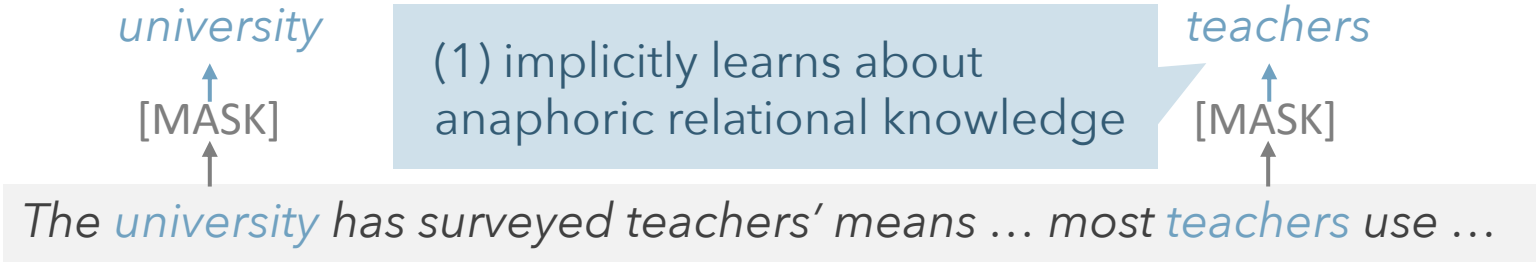


(2) To apply the acquired knowledge to the ZAR task

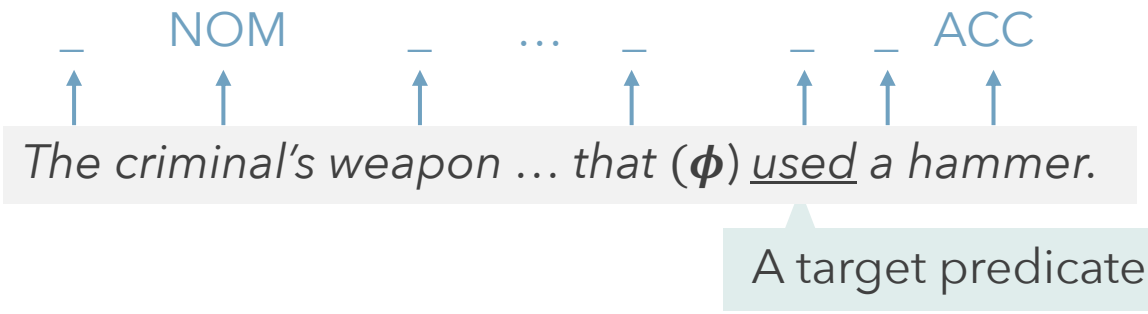
Recent Approaches: Masked Language Models

[devlin+'19]

- Pretraining: **Cloze Task**



- Finetuning: **Argument Selection with Label Probability (AS)**
  - Identifying Nominative (NOM), Accusative (ACC), and Dative (DAT)

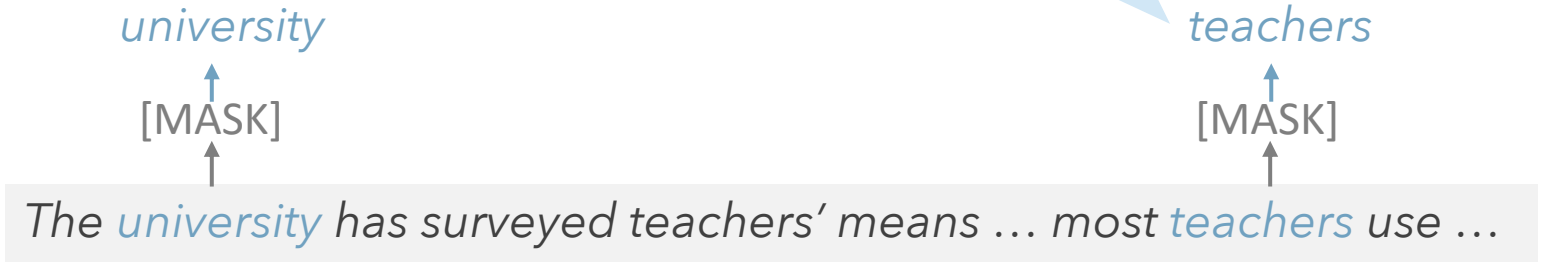


※ We use Japanese sentences in our experiments

Two Problems on Previous Approaches

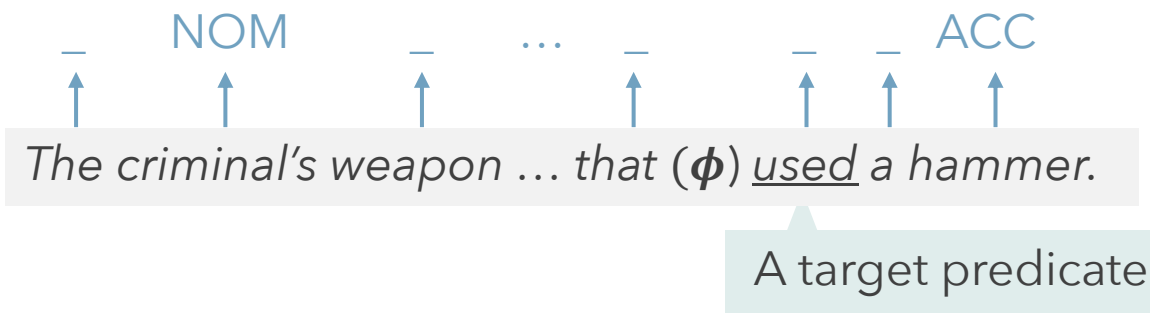
- Pretraining: **Cloze Task**

☹ No supervision on anaphoric relations



☹ Pretrain-finetune discrepancy [Yang+'19]
- [MASK] is not used
- Last Layer is replaced/added

- Finetuning: **Argument Selection with Label Probability (AS)**
  - Identifying Nominative (NOM), Accusative (ACC), and Dative (DAT)



※ We use Japanese sentences in our experiments

Our Approaches

- Pretraining: **Pseudo Zero Pronoun Resolution (PZERO)**

Explicit supervision on anaphoric relations

[MASK]

The university has surveyed teachers' means … most teachers use …

Smoother adaptation
- Both predict antecedents
- We can use the same network structure

- Finetuning: **Argument Selection as PZERO (AS-PZERO)**
  - Identifying Nominative (NOM), Accusative (ACC), and Dative (DAT)

*The criminal's weapon … that (**ϕ**) used a hammer.* [MASK] *NOM used*

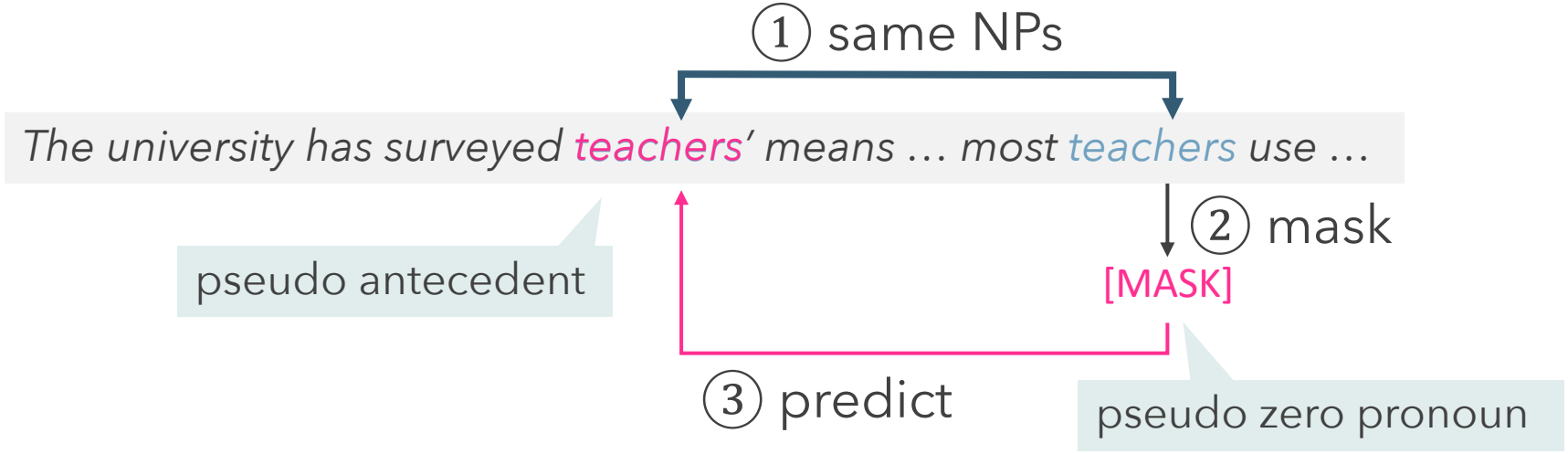
A target predicate

Query chunk

※ We use Japanese sentences in our experiments

Pretraining: Pseudo Zero Pronoun Resolution (PZERO)

1. We assume that same noun phrases (NPs) are coreferent
2. One of them is masked as a **pseudo zero pronoun**
3. The model predicts the other NPs as its **pseudo antecedents**

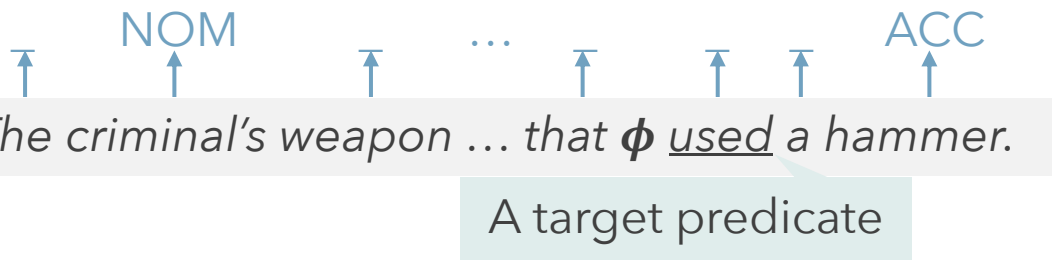


Pseudo Zero Pronoun Resolution (PZERO):

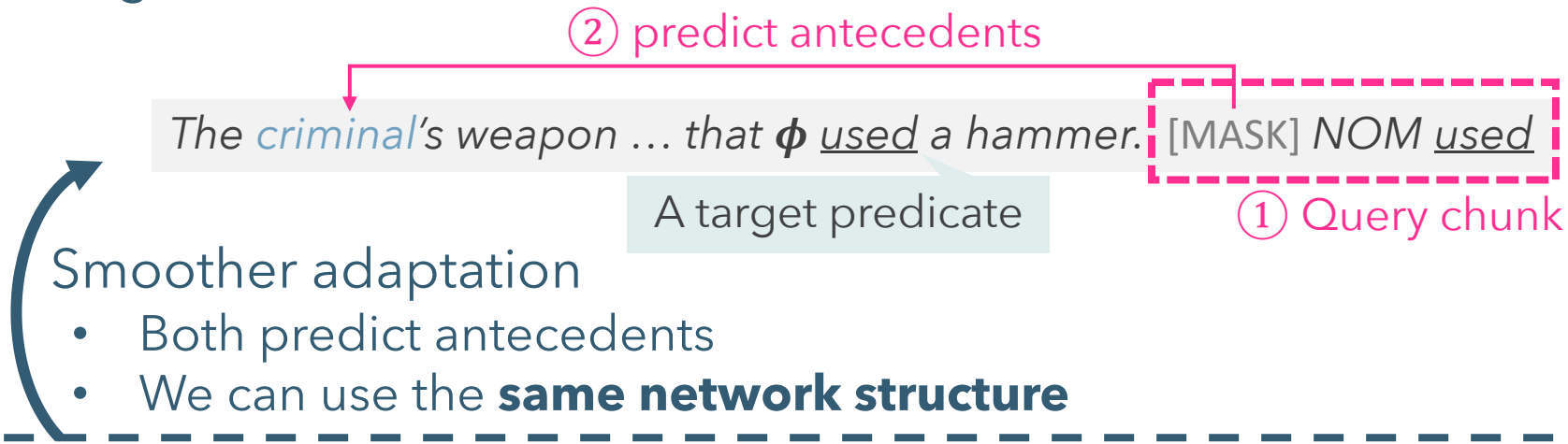
- provides **explicit supervision** on anaphoric relations
- is too strong assumption but
  can provide a **large-scale** dataset from raw corpora

Finetuing: Argument Selection as PZERO (AS-PZERO)

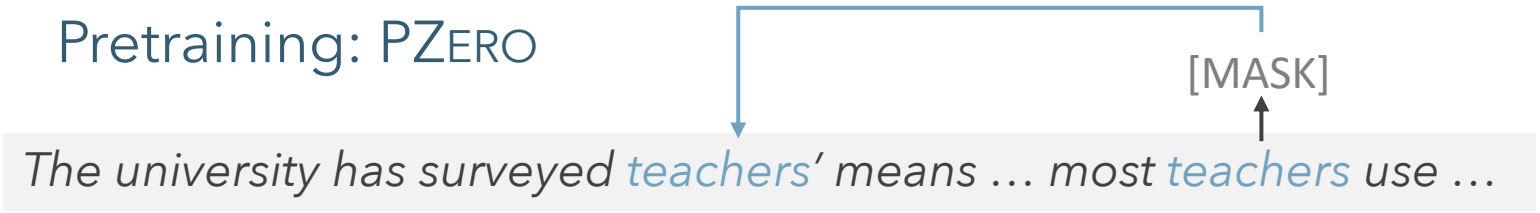
- Argument Selection with Label Probability (AS)



- Argument Selection as PZERO (AS-PZERO)



- Pretraining: PZERO



Experiments

- We initialized the model parameters with the pretrained masked language model
  - bert-base-japanese model (transformers library)

- Further Pretraining on Japanese Wikipedia corpus
  - Cloze Task 30K updates
  - PZERO Task 30K updates

Can this combination improve the performance of ZAR?

- Finetuning on NAIST Text Corpus [iida+'17]
  - Baseline Model (AS)
  - Proposed Model (AS-PZERO)

- **Evaluation**
  - Data: NAIST Text Corpus Test set
  - Metrics: F_1 score

Results

- ① The effective of PZERO task
- ② The effective of AS-PZERO model

Pretraining	Further Pretraining		Finetuning		ZAR F_1
Cloze	Cloze	PZERO	AS	AS-PZERO	All
✓	✓		✓		62.54 ± 0.47
✓	✓			✓	62.85 ± 0.19
✓		✓	✓		63.06 ± 0.19
✓		✓		✓	**64.18** ± 0.23

① ②

Results

- ① The effective of PZERO task
- ② The effective of AS-PZERO model

Pretraining	Further Pretraining		Finetuning		ZAR F_1
Cloze	Cloze	PZERO	AS	AS-PZERO	All
✓	✓		✓		62.54 ± 0.47
✓	✓			✓	62.85 ± 0.19
✓		✓	✓		63.06 ± 0.19
✓		✓		✓	**64.18** ± 0.23

① ②

The model affectively learns anaphoric relational knowledge

Results

- ① The effective of PZERO task
- ② The effective of AS-PZERO model

Pretraining	Further Pretraining		Finetuning		ZAR F_1
Cloze	Cloze	PZERO	AS	AS-PZERO	All
✓	✓		✓		62.54 ± 0.47
✓	✓			✓	62.85 ± 0.19
✓		✓	✓		63.06 ± 0.19
✓		✓		✓	**64.18** ± 0.23

① ②

The model successfully address the pretrain-finetune discrepancy

Results

- ① The effective of PZERO task
- ② The effective of AS-PZERO model

Pretraining	Further Pretraining		Finetuning		ZAR F_1
Cloze	Cloze	PZERO	AS	AS-PZERO	All
✓	✓		✓		62.54 ± 0.47
✓	✓			✓	62.85 ± 0.19
✓		✓	✓		63.06 ± 0.19
✓		✓		✓	**64.18** ± 0.23

① ②

Both the proposals improve the performance and achieve SoTA

※ Please see the paper for detailed results

Analysis

- The following cases are hard to predict
  - The arguments far from a target predicate

  - The arguments of predicate in the passive voice

active

The man used ϕ

more
difficult

passive

ϕ was used by the man

※ Please see the paper for detailed analysis

Conclusion

- Task: **Zero Anaphora Resolution (ZAR)** in Japanese

- We proposed
  - A new **pretraining** task for ZAR, **PZERO** task
  - A new **finetuning** method for ZAR, **AS-PZERO** model

- Results:
  - The two proposals **boost the SoTA** performance of ZAR
  - Our analysis provides new insights on the remaining challenges

Thank you for listening!

ryuto_konno@r.recruit.co.jp