Q7-13

実世界対話における参照関係の統合的解析

稲積 駿 [1,2], 植田 暢大 [3], 吉野 幸一郎 [4,2,1]

1. 奈良先端科学技術大学院大学
2. 理化学研究所 ガーディアンロボットプロジェクト, 3. 京都大学, 4. 東京科学大学

実世界対話の意味理解

テキストの照応・述語項構造解析

語句間の**参照関係**を特定するタスク

- **直接参照**: 共参照と対応　　コーヒーカップ = これ
- **間接参照**: 述語項構造や橋渡し照応と対応　　取って →ヲ コーヒーカップ

マルチモーダル参照解析 [Ueda+, 2024]

語句が指す**物体**の参照関係を特定するタスク　　取って →ヲ

フレーズグラウンディング:
語句-物体間の直接参照のみを特定する場合　　コーヒーカップ =

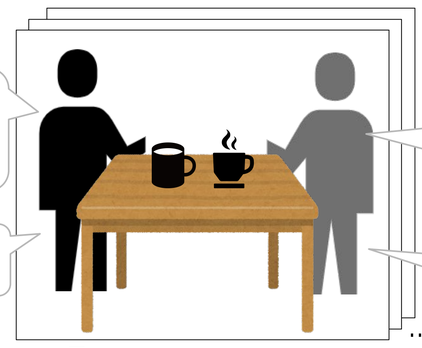
テキストに含まれる**イベント**「**誰**が **誰**に **何**を **どうする**」を
物体と紐付けて理解するシステムが実現可能

2者の実世界対話をシステムが解析する例

話者1の発話

[Φ_ガ] コーヒーカップを
[Φ_ヲ] 取って頂けますか？

ありがとう！



話者2の発話

はい。
これですよね。

どういたしまして。

システムの視点

直接参照 (共参照): これ = コーヒーカップ

間接参照 (ガ格): 取って → 話者2

間接参照 (ヲ格): 取って → コーヒーカップ

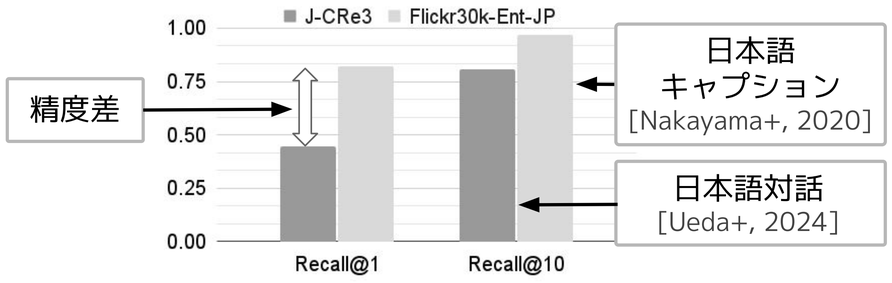
間接参照 (ニ格): 取って → 話者1

既存モデルの課題

GLIP [Li+, CVPR2022]: 語句-物体間の直接参照を
大量の画像とそのキャプション・クラスラベルで学習

対話の解析においてGLIPは:

1. **指示詞**に対する**直接参照**の解析が困難　　これ =
2. **間接参照**の解析を扱えず**省略**された語句の解析が困難　　取って →



GLIPによる直接参照の解析結果

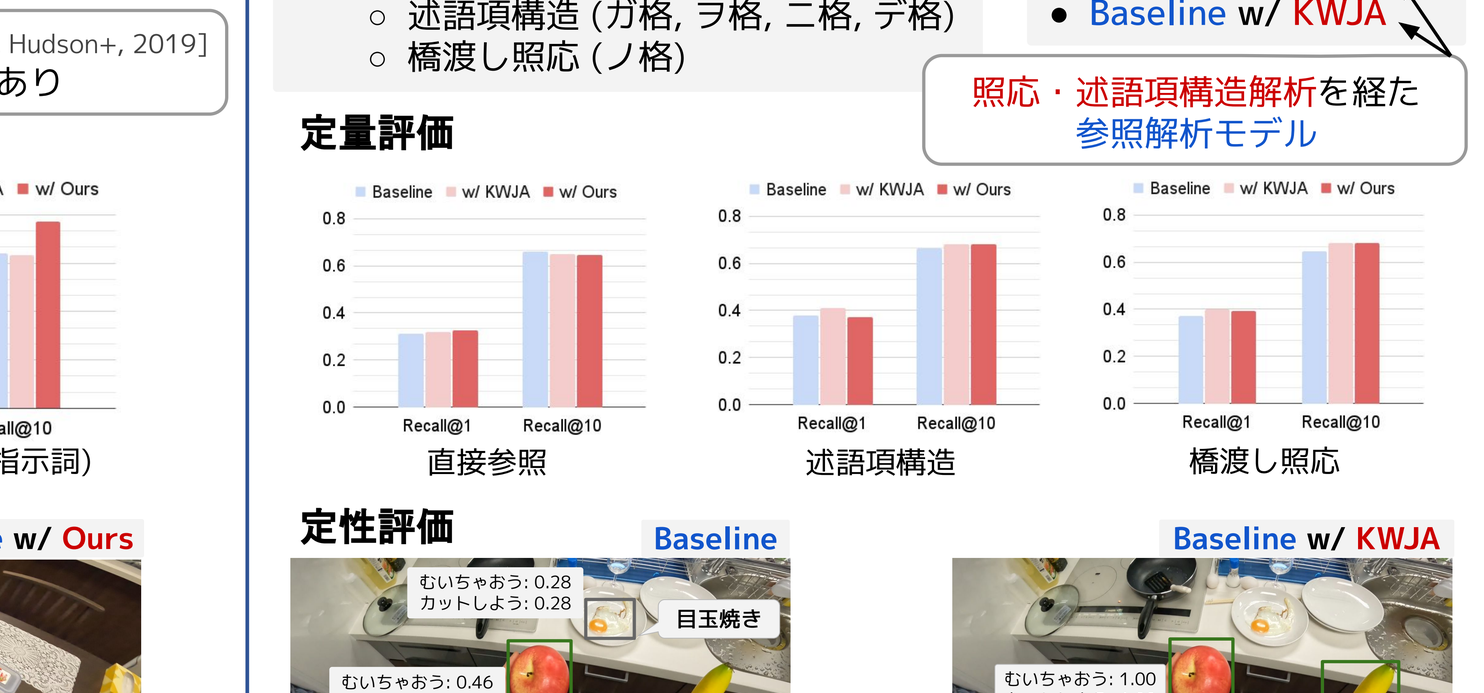
1 と 2 の曖昧性解消を通して
対話への**マルチモーダル参照解析の性能向上**を目指す

提案手法

照応・述語項構造の知識を考慮して 語句-物体間の解析精度向上を図る

e.g.) コーヒーカップ / これ　が既知であれば　これ = 　が一意に特定可能

照応・述語項構造解析 と **マルチモーダル参照解析**を統合的に扱う枠組みの提案



フレーズグラウンディングの結果

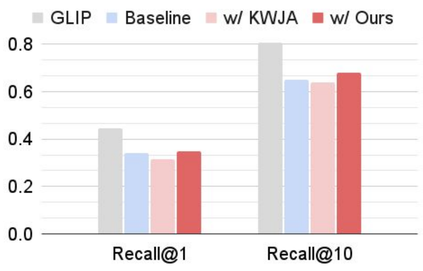
比較モデル

- Baseline
- Baseline w/ Ours
- Baseline w/ KWJA [Ueda+, 2023]
- GLIP [Li+, CVPR2022]

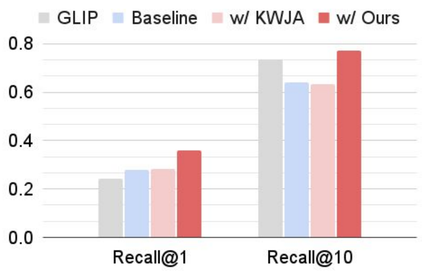
1. 共参照関係でテキスト解析モデルを学習
2. 参照解析モデルを追加で学習

英語データ [Krishna+, 2017, Hudson+, 2019] による事前学習あり

定量評価



J-CRe3の評価結果 (全件)



J-CRe3の評価結果 (指示詞)

定性評価

Baseline

Baseline w/ Ours

お湯が沸いたら、ここに入れてくれる?

共参照関係の学習により:

- **指示詞に対する解析精度が向上**
- **指示詞から物体への予測の確信度が強まる**

マルチモーダル参照解析の結果

解析対象

- 直接参照
- 間接参照
  - 述語項構造 (ガ格, ヲ格, 二格, デ格)
  - 橋渡し照応 (ノ格)

比較モデル

- Baseline
- Baseline w/ Ours
- Baseline w/ KWJA

照応・述語項構造解析を経た
参照解析モデル

定量評価

直接参照

述語項構造

橋渡し照応

定性評価

Baseline

Baseline w/ KWJA

[バナナとリンゴを] 両方むいちゃおうか。
で、3人分にカットしよう。

照応関係・述語項構造の学習により:

- **語句-物体間の間接参照の解析精度が向上**
- **述語から物体への予測の確信度が強まる**