
マルチモーダル対話データセットタグ付けの基準
VERSION 0.1

2023 年 5 月

京都大学大学院情報学研究科 言語メディア研究室
理化学研究所 ガーディアンロボットプロジェクト
植田 暢大
ueda@nlp.ist.i.kyoto-u.ac.jp

1 はじめに

2 者の実世界対話シーンにおいて、発話書き起こしテキストと1人称視点動画に対して、種々の意味関係を付与する際の基準について説明する。関係の付与はテキスト間照応アノテーション、物体領域アノテーション、テキスト・物体間参照アノテーションから構成される。本アノテーション基準は、京大コーパスにおける「格・省略・共参照タグ付けの基準」を基にアノテーション付与対象を画像中の物体矩形に拡張したものである。

テキスト間照応アノテーションには京大コーパスアノテーションツール^{*1}を用いる。物体領域アノテーションおよびテキスト・物体間参照アノテーションには静止画アノテーションツールを用いる。

2 データセット

アノテーションに使用するデータセットは、実世界における2者の対話シーンからテキストと画像を抽出したものである。対話は、主人とそのお手伝いロボットを想定し、リビング・ダイニング・キッチンを模した実験設備で収録した。収録内容は、両者の発話音声とロボット側の1人称視点動画である。両者の発話音声はテキストに書き起こし、1人称視点動画は1秒ごとにフレームを抽出し画像系列に変換した。

発話書き起こしテキストと画像系列に対し、図1に示すような複数種類のタグを付与する。桃色吹き出しがテキスト間照応アノテーション、橙色矩形が物体領域アノテーション、破線がテキスト・物体間参照アノテーションに対応する。

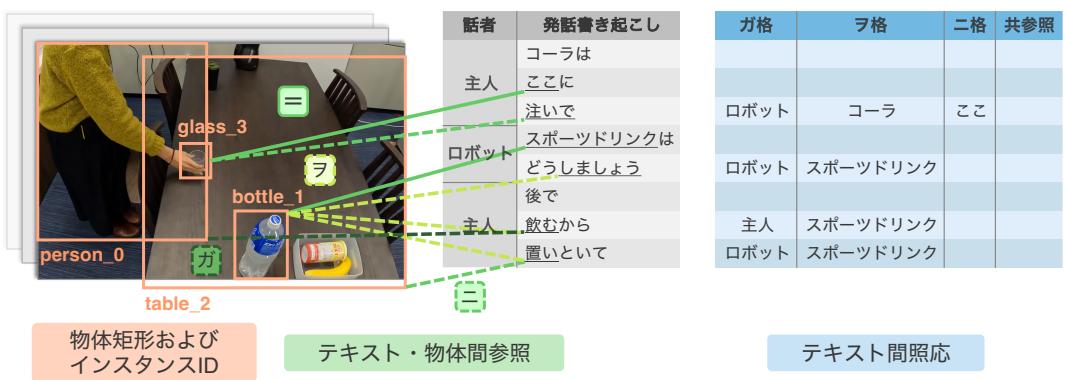


図1: 本アノテーションの概観。1人称視点画像（左）と対話書き起こしテキスト（中央）に対し、物体領域（オレンジ色矩形）、テキスト間照応関係（右）、テキスト・物体間参照関係（緑色線）が付与されている。物体領域にはクラス名とインスタンスIDが付与されている。

3 テキスト間照応アノテーション（コーパスタグ付け）

テキスト間照応アノテーションでは、基本句間の意味関係を付与する。基本句とは、自立語1語を核として、その前後に存在する付属語をまとめた言語単位であり、京都大学テキストコーパス

^{*1} <https://github.com/ku-nlp/KyotoCorpusAnnotationTool>

ス^{*2} [1] や京都大学ウェブ文書リードコーパス^{*3} [2]においてタグ付けの単位として使用されている。基本句間に付与する意味関係は、用言・サ変名詞などに対する格・省略関係、名詞間の関係、および共参照とする。詳細は格・省略・共参照タグ付けの基準^{*4}に従う。

基本句にタグ付けをするためには、基本句の境界を特定する必要がある。そのため、基本句間の意味関係の付与に先立ち、形態素・基本句・文節の境界を付与する。詳細は、形態素・構文タグ付きコーパス作成の作業基準^{*5}3.2 節「形態素・文節の区切り」に従う。この際、形態素の原形や品詞等、言語単位の境界に関係しない情報は付与しない。

本データセットに含まれるテキストは対話形式の話し言葉である。一方、格・省略・共参照タグ付けの基準はモノローグ形式の書き言葉を対象としている。本節では扱うテキストの性質の違いから生じる追加のアノテーション基準について述べる。

3.1 主人とロボットに対するタグ付け

コーパスアノテーションツールの仕様上、外界照応の照応先として「主人」や「ロボット」が選択できない。代わりに、「著者」を「主人」、「読者」を「ロボット」としてタグを付与する。

3.2 感動詞に対するタグ付け

挨拶など、相手に向けられた儀礼的な感動詞の場合はガ格に主体、ニ格に相手を付与する。

- (1) 【主人】 ただいま (ガ格: 著者, ニ格: 読者, 不特定: 人1 [?])
- (2) 【ロボ】 どういたしまして (ガ格: 読者, ニ格: 著者)
- (3) 【ロボ】 とんでもないです (ガ格: 読者, ニ格: 著者)
- (4) 【ロボ】 おかえりなさいませ (ガ格: 読者, ニ格: 著者)
- (5) 【主人】 うん (ガ格: 著者, ニ格: 読者)
- (6) 【主人】 はい (ガ格: 著者, ニ格: 読者)
- (7) 【主人】 ごめん (ガ格: 著者, ニ格: 読者)

感動詞が述語に由来する場合、その述語に準じた項を付与する。

- (8) 【主人】 いってきます (ガ格: 著者, ニ格: 会社)
- (9) 【ロボ】 いってらっしゃい (ガ格: 著者, ニ格: 会社)
- (10) 【主人】 ありがとう (ガ格: 不特定: 状況1, ニ格: 読者, ガ2格: 著者)
- (11) 【主人】 いただきます (ガ格: 著者, ヲ格: 料理)

^{*2} <https://github.com/ku-nlp/KyotoCorpus>

^{*3} <https://github.com/ku-nlp/KWDLC>

^{*4} https://github.com/ku-nlp/KWDLC/blob/master/doc/rel_guideline.pdf

^{*5} https://github.com/ku-nlp/KWDLC/blob/master/doc/syn_guideline.pdf

儀礼的な感動詞でも相手に向けた発話でない場合はガ格のみ付与する。

- (12) 【主人】 ごちそうさまでした (ガ格: 著者)

はっきりと相手に向けられた発話でなく、フィラーや独り言に近い感動詞にはタグを付与しない。

- (13) あら

- (14) へえ

- (15) ええと

- (16) ほら

- (17) よいしょ

- (18) さてと

- (19) (考え方をしながら) そうだねえ

相槌にはタグを付与しない。

- (20) うんうん

3.3 受け答えに対するタグ付け

一般的な述語と同様にタグを付与する。以下に頻出する表現に対するタグ付けの例を示す。

3.3.1 「大丈夫です」

ガ格に大丈夫か判断する対象、ガ2格に判断する主体を付与する。両者が一致する場合はガ格のみ付与する。

- (21) 【主人】 こういう細かい作業は苦手なのよね。【ロボ】いや大丈夫ですよ (ガ格: 作業, ガ2格: 読者)

- (22) 【ロボ】 顔色が悪いですが、大丈夫ですか (ガ格: 著者)

3.3.2 「分かりました」 / 「オーケーです」 / 「了解です」 / 「もちろんです」

ガ格に対象となる事柄、ガ2格に承知する主体を付与する。多くの場合、対象は何らかの動作やイベントであるため、ガ格には主に述語が付与される。ただし、述語の一部の修飾要素に焦点が当たっている場合は「？」付きでその修飾要素も付与する。

- (23) 【主人】 そこにあるプチプチも持ってきてちょうだい。【ロボ】 分かりました (ガ格: 持ってきて, ガ2格: 読者)

- (24) 【主人】 引き出しから乾電池を持ってきてちょうだい。【ロボ】 サイズは何ですか？【主人】 単

四電池が三つ必要だね。分かりました (ガ格: 持ってきて, 単四 [?], 三つ [?], ガ2格: 読者)

3.3.3 「了解しました」 / 「承知しました」

ヲ格に対象となる事柄, ガ格に了解する主体を付与する.

- (25) 【主人】ちょっと見に行ってくれる? 【ロボ】はい、了解しました (ガ格: 読者, オ格: 行ってくれる)

3.3.4 「かしこまりました」

ニツイテ格に依頼内容, ガ格に了解する主体を付与する.

- (26) 【主人】リモコンを取ってもらえる? 【ロボ】かしこまりました (ガ格: 読者, ニツイテ格: 取ってもらえる)

3.3.5 「いかがですか?」 / 「いかがでしょうか?」

ガ格にいかがか判断する対象, ガ2格に判断する主体を付与する.

- (27) 【ロボ】コーヒーはいかがですか? (ガ格: コーヒー, ガ2格: 著者)

- (28) 【ロボ】焼き加減いかがでしょうか? (ガ格: 加減, ガ2格: 著者)

3.3.6 「よろしいですか?」

ガ格やデ格に対象を付与する.

- (29) 【ロボ】テープってどこにありましたっけ? 【主人】確か玄関にあったと思う。【ロボ】こちらでよろしいですか? (ガ格: テープ, デ格: こちら)

3.3.7 「そうですね」

ガ格と=格に同意している対象を付与する.

- (30) 【主人】おもちゃが散乱しているわね。【ロボ】そうですね (ガ格: 散乱している, =格: 散乱している)

3.3.8 「どうぞ」

ガ格に主体, ニ格に相手を付与する. 何らかの物を渡す文脈では, オ格に対象となる物を付与する.

- (31) 【ロボ】本を持ってきました。どうぞ (ガ格: 読者, オ格: 本, ニ格: 著者)

ただし, 何らかの物を渡す文脈でも「はい」にオ格は付与しない.

- (32) 【ロボ】本を持ってきました。はい (ガ格: 読者, ニ格: 著者)

3.3.9 「よろしく（お願ひします）」

ガ格に依頼の主体、ヲ格に依頼の内容、ニ格に相手を付与する。

- (33) 【ロボ】 それでは、食器を洗いますね。【主人】 うんよろしくね。 (ガ格: 著者, ヲ格: 洗いますね, ニ格: 読者)

依頼の内容が漠然としている場合は、ヲ格に「不特定: 状況」を付与する。

- (34) 【主人】 今日も一日よろしくね。 (ガ格: 著者, ヲ格: 不特定: 状況, ニ格: 読者, 時間格: 今日)

「よろしくお願ひします」と言った場合、「お願ひします」にのみタグを付与し、「よろしく」にはタグを付与しない。

- (35) 【主人】 今日も一日よろしくね。【ロボ】 おはようございます。よろしくお願ひします

- (36) 【主人】 今日も一日よろしくね。【ロボ】 おはようございます。よろしくお願ひします (ガ格: 読者, ヲ格: 不特定: 状況, ニ格: 著者)

3.4 その他注意点

3.4.1 総称名詞と非総称名詞について

主に共参照のタグ付けについて、総称名詞と非総称名詞を区別する必要がある。以下の例では、「コーラ₁」は特定のコーラを指しておらず総称名詞である。一方、「コーラ₂」は手元にあるコーラを指しており、非総称名詞である。したがって、表層表現が同じでもこれらは区別し、=格ではなく≠格を付与する。

- (37) 【主人】 コーラ₁が飲みたいな。【ロボ】 こちらにコーラ₂をお持ちしました (=≠格: コーラ₁)

ヲ格についても同様である。

- (38) 【主人】 コーラが飲みたいな。【ロボ】 こちらにお持ちしました (ガ格: 読者, ヲ≠格: コーラ, ニ格: こちら)

名詞が指す物体が発話時点で目の前に現れていない場合でも、特定の対象を指していると考えられる場合は非総称名詞として扱う。以下の文中の「コーヒー」は棚の一番下にある特定のコーヒーを指しているため非総称名詞である。

- (39) 【主人】 これでお湯を沸かせるから、えーと一番下にインスタントのコーヒーがあるから、棚の一番下、取ってきてもらえるかな？

3.4.2 「元に戻す」について

物を片付けるという文脈で、「元に戻す」という表現が出現した場合、「元」は場所を表すとは考えず、「戻す」のニ格には具体的な戻す場所を付与する。以下の例では、おもちゃ箱に対応する「不特定: 物₂」が付与されている。

- (40) 【主人】あとは、床に散らかっているパズルを元に戻してちょうだい。(ガ格: 読者, ヲ格: パズル, ニ格: 不特定: 物 2)

3.4.3 その他

述語について対応する項が述語が含まれる文あるいは先行する文にない場合は「不特定: 物 1」などを使用する。該当する項が後続の文に出現した場合はその項と「不特定: 物 1」を=で結ぶ。

格のとり方について複数の可能性がある場合はより自然だと考えられる方を選ぶ。以下の例では、動詞「入れ替える」について「不特定: 物 2」の格はニ格とト格の複数の可能性がある。この例では作業者がより自然と考えられる方であればどちらのタグ付けも正しい。

- (41) 【ロボ】電池切れかもしれないですね。【ロボ】入れ替えましょうか？ (ガ格: 読者, ヲ格: 不特定: 物 1, ニ格: 不特定: 物 2)

- (42) 【ロボ】電池切れかもしれないですね。【ロボ】入れ替えましょうか？ (ガ格: 読者, ヲ格: 不特定: 物 1, ト格: 不特定: 物 2)

主人がプレイしているゲーム中の主格は「著者（主人）」ではなく「不特定: 人 n」とする。

- (43) 【主人】やった、もうすぐゴールだ。ベスト更新できるかも (ガ格: 不特定: 人 1)

4 物体領域アノテーション (BB タグ付け)

1 人称視点動画から抽出された各画像に対し、物体矩形、クラス名、インスタンス ID の 3 種類のアノテーションを付与する。本節ではそれぞれのアノテーションについて説明する。

4.1 物体矩形のアノテーション

物体矩形とは、画像中の物体が占める矩形領域のことである。物体にオクルージョンがある場合は、可視領域のみを物体矩形に含める。例えば、キッチンに立っていて上半身しか見えていない人に対しては、上半身のみが物体矩形に含まれる。

物体矩形のアノテーションでは、画像中の物体に対して物体矩形を付与する。ただし、1 画像中には物体が無数に存在するため、付与対象を対話中で参照された物体に限定する。対話中の物体への参照には、物体名や代名詞による直接的な言及のほか、「お湯を（ポットで）沸かして」といったゼロ照応に相当する間接的な言及も含まれる。

静止画アノテーションツール上では、事前に物体矩形が自動付与されている。矩形付与対象の物体に関しては、必要であれば矩形を修正する。それ以外の物体に関しては、矩形を削除するか、対話中で参照されていない物体であることを表現する何らかのタグ^{*6}を付与する。なお、矩形付与対象外の物体の物体矩形を修正しても構わない。

物体矩形は原則、一つ一つの物体それぞれに付与する。しかし、同じクラスの物体が集合として現れ、かつ集合として対話中で参照される場合は図 2 のようにまとめて 1 つの物体矩形を付与する。

^{*6} 推定基準の「自動/手動」タグなど



図 2: 重なった 4 枚の皿に対する物体矩形の付与 (20220304-56132112-1, 132.png より).

物体矩形は前後のフレームの内容を考慮して付与する. 図 3b において、「person_001」は画像単体からは判断できないが、前のフレーム（図 3a）を参照することで、左下の黒い部分が人であることが分かる.



図 3: 前後のフレームを考慮した物体矩形アノテーションの例 (20230308-57113425-1 より).

4.2 クラス名のアノテーション

それぞれの物体矩形に対して物体のクラス名を付与する. クラス名は LVIS データセット [3] において定義されている 1,203 クラスの集合から選択する. 該当するクラス名がクラス集合にない場合は最も意味的に近いと考えられるクラス名を採用する.

4.3 インスタンス ID のアノテーション

それぞれの物体矩形に対して物体を一意に識別するためにインスタンス ID を付与する. インスタンス ID は物体ごとに固有でなければならない. 例えば、同じ色の皿が 2 枚ある場合には、それぞれに異なるインスタンス ID を付与する.

インスタンス ID は動画中で一貫して付与される必要がある. すなわち、同じ物体が異なる画像フレームに出現した場合は同じインスタンス ID を付与する. ただし、似た物体が複数同時に出現するなどの理由で、動画から特定の物体の追跡が困難な場合には、可能性のあるインスタンス ID から任意に選択して付与する.

物体の集合に対して物体矩形を付与している場合、物体が 1 つでも集合から抜けた際には別のインスタンス ID を付与する（図 4）。また、集合の要素を個別に扱う場合も別のインスタンス ID を付与する（図 5）。例外として、見た目上物体が集合から抜けた場合では同じインスタンス ID を使用することができる。図 6(b) は、集合としてインスタンス ID が付与された 2 枚の大皿が重ねられ、小皿と上段の皿しか見えていない状態だが、集合に対するインスタンス ID を上段の皿に付与しても構わない（上段の皿に対するインスタンス ID が存在すればそちらを付与することも可能）。



図 4: 重なった 3 枚の皿に対する物体矩形の付与 (20220304-56132112-1, 146.png より)。

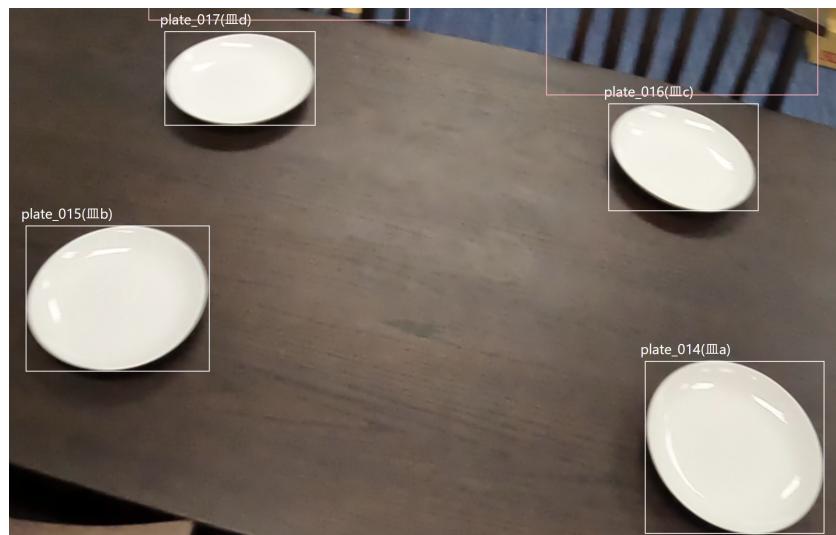


図 5: 個別の皿に対する物体矩形の付与 (20220304-56132112-1, 158.png より)

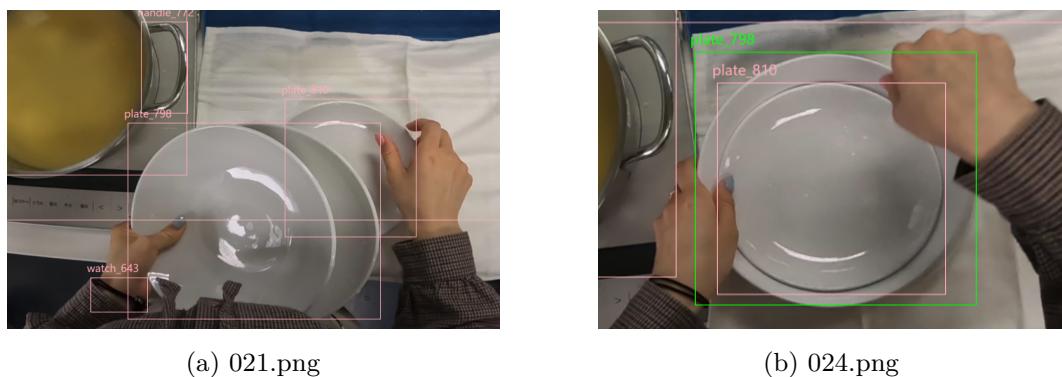


図 6: 1 要素が隠れている場合の集合に対する ID の付与 (20230519-57232051-0 より)

物体の状態が動画中で変化した場合は、インスタンス ID を変更する。例えば、目玉焼きを作る

動画では、割られていない卵と完成した目玉焼きに対して異なるインスタンス ID を付与する⁷. また、動画中で物体の構成要素が変化した場合もインスタンス ID を更新する. 図 7 は、パズル（の土台）にピースの一つである「8」がはまる様子を示している. ピースはパズルの構成要素であると考えられるため、ピースがはまつた後のインスタンス ID (16) は、はまる前のインスタンス ID (10) とは異なる.

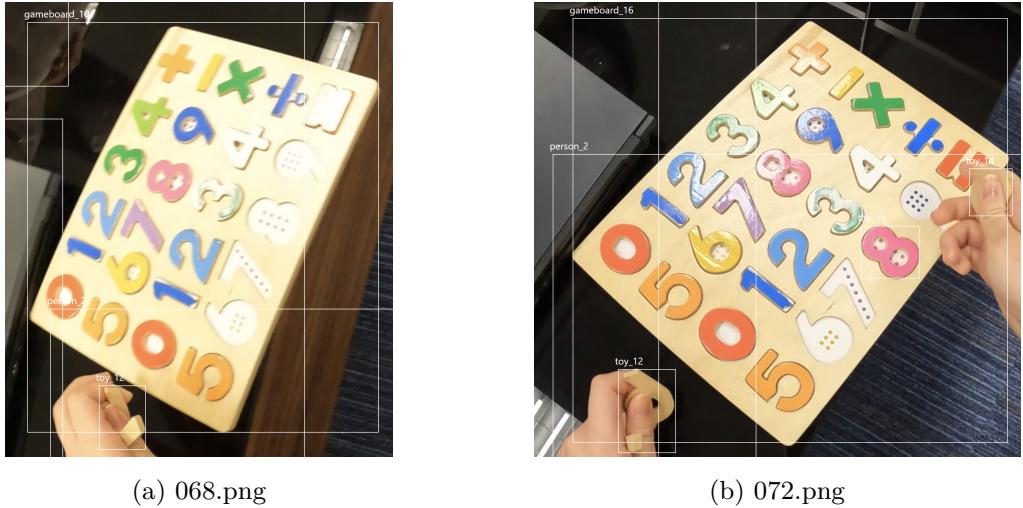


図 7: 物体の構成要素が変化する場合のインスタンス ID の付与 (20220303-56133610-0 より).

物体の状態に視覚的变化がない場合でも、対話の内容から推測されるシナリオ上で物体が変化することになっている場合は、インスタンス ID を変更する. 図 8 は、シナリオ上ではロボットがカフェオレを作り終えた場面である. しかし、収録の都合上、作る際の動作のみを行って実際にはカフェオレを作っていない. このとき、コップに視覚的变化はないが、シナリオ上はカフェオレが入っていることになっているので入れる前のコップとは異なるインスタンス ID を付与する.

物体に付随する部分が変化した場合でも、その部分が物体の構成要素ではないならインスタンス ID は変更しない. ただし、外観が変化しない、あるいは物体自体ではなく物体に付随する別の物体が変化した場合などは同一の ID を使用する. 図 9 は本棚に本を収納する場面を示しているが、本は本棚の構成要素ではないため、両フレーム中の本棚については同じインスタンス ID (7) を付与する. 他にも、リモコンの埃を拭き取る場面では、埃をリモコンの一部とは考えず、埃が拭き取られた後もリモコンのインスタンス ID を更新しない.

4.3.1 容器に対するタグ付け

何らかの容器に物が入れられた場合、容器が内容物の名称で参照されることがある. 例えば、「コーヒーはミルクと一緒にテーブルに置いてあります」という発話では、コーヒーが入ったコーヒーカップを「コーヒー」という表現で参照している. コーヒーカップのように、内容物が入る前後で参照表現が変化する物体に関しては、前後で異なるインスタンス ID を付与する. 仮に、コーヒーカップに対するインスタンス ID が同一だった場合、後述するテキスト・物体間参照アノテー

⁷ 生卵に火が通る、などの段階的な変化では変化の開始時点のフレームでインスタンス ID を変更する.

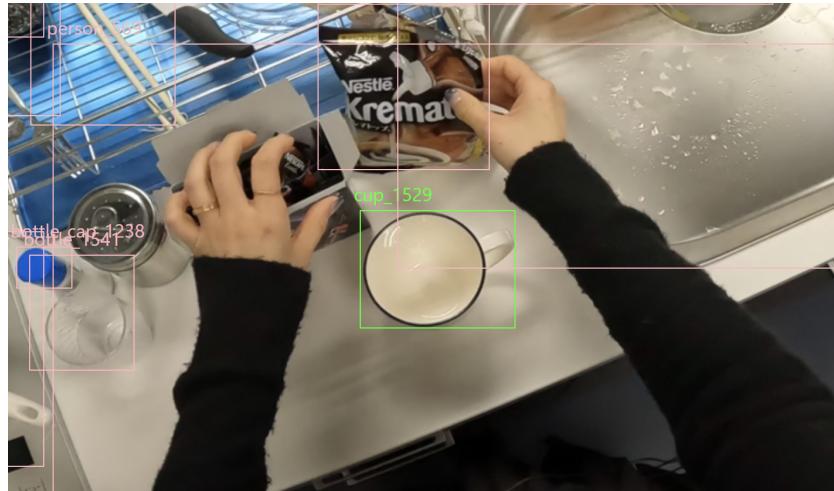


図 8: カフェオレを作る場面でのコップへのインスタンス ID の付与. 実際には作っていないためコップに視覚的变化はないが、シナリオ上では作ったことになっているためコップのインスタンス ID を変更する (20230308-57194356-0, 017.png より).



図 9: 変化した部分が物体の構成要素ではない場合のインスタンス ID の付与 (20220303-56133610-0 より)

ションにおいて、空のコーヒーカップも「コーヒー」という表現で参照されることになる^{*8}. 内容物の有無によって異なるインスタンス ID を付与することで、誤ったタグが付与されることを防ぐ.

5 テキスト・物体間参照アノテーション（関係タグ付け）

テキスト中の名詞句および述語と、画像中の物体矩形のすべての組み合わせについて参照関係を付与する。付与する際の基準は、テキスト間照応アノテーションにおける照応先を物体矩形に拡張したものに相当する。すなわち、「格・省略・共参照タグ付けの基準」に従い、名詞句についてはノ格、ノ? 格、修飾格、トイウ格、=格（共参照）を付与する。述語についてはその項に対応する物

^{*8} テキスト・物体間参照アノテーションにおけるタグ付けの省略で、コーヒーが入ったコーヒーカップが「コーヒー」という表現と関係を持つとタグ付けされれば、同じインスタンス ID を持つ空のコーヒーカップも「コーヒー」と関係を持つものとして扱われる。

体矩形を格ごとに付与する。

ただし、=関係や \neq 関係の扱いについてのみ、「格・省略・共参照タグ付けの基準」(以下、「コーパスタグ付け基準」とよぶ)とは一部異なる基準を定める。まず、タグ付けの簡略化のため「=構文」関係(5.1.2節)や「=役職」関係(5.1.3節)は使用しない。また、用言に対する共参照タグ(5.3節 共参照に関する注意点)も使用しない。

コーパスタグ付け基準では、以下のような複数を表す表現とその構成要素である単数を表す表現の関係には \neq タグを付与すると定められている(5.3節 共参照に関する注意点)。

(44) 学校、警察を含め関係機関が協力する必要がある (\neq : 学校, 警察)

図10は以下の発話を含む対話における物体矩形の例である。

(45) 【主人】この卵を一個ずつ割入れてくれる？

発話中の「卵」は図中の卵2つを指しており、「卵」という表現と図中のそれぞれの物体矩形は、複数を表す表現とその構成要素の関係にある。しかし、テキスト・物体間参照アノテーションでは、表現とそれぞれの物体矩形の間に=格を付与する。これはクラス名の異なる物体矩形に関しても同様であり、例えば「果物」という表現がりんごとバナナの物体矩形を参照している場合でも両者に=格を付与する。

コーパスタグ付け基準では、まったく異なる対象であっても同じ表現で表される要素間は「 \neq 」でタグを付けると定められている(5.2節 \neq 関係)。

(46) 国会議員に次いで、県議会議員にも呼びかける (\neq : 議員)

しかし、テキスト・物体間参照アノテーションでは対象がまったく異なる場合にタグは付与しない。

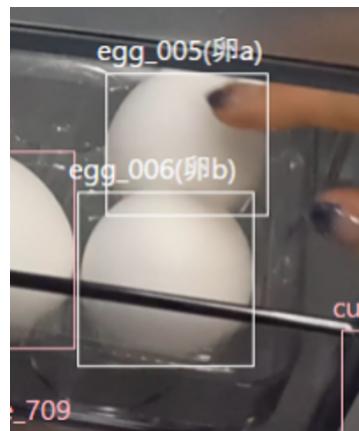


図10: 「卵」という1つの表現で参照される2つの卵の物体矩形。

ここまで、コーパスタグ付け基準では \neq 関係が付与されるが、テキスト・物体間参照アノテーションでは付与されない例を挙げてきた。一方で、テキスト・物体間参照アノテーションで \neq 関係が付与されるのは主に、総称名詞との間に関係を付与する場合と、物体のインスタンスIDが更新された場合である。前者については、テキスト間照応アノテーションにおいて総称名詞と対応する非総称名詞の間に=格を付与することに由来する。例えば例3.4.1における「コーラ1」と、実

際にロボットが持ってきたコーラに付与された物体矩形の間には \doteq 関係を付与する。後者については、以下のような例が挙げられる。

- (47) 【主人】今から目玉焼きを作ろうと思うから、手伝ってくれる？【ロボ】分かりました。卵₁は何個ありますか？（中略）【主人】この卵₂を一個ずつ割り入れてくれる？（中略）【主人】今卵₃はどんな感じ？



図 11: 物体のインスタンス ID が変化する場合のテキスト・物体間参照アノテーションの例 (20230308-57114309-0, 147.png より)。

図 11 は対話例に対応する動画の 1 フレームである。この対話では、「卵₃」は画像中の目玉焼きを指しており、目玉焼きの物体矩形（「boiled_egg_020(目玉焼き a, b)」）と \doteq 関係で結ばれる。一方、「卵₂」は火が通る前の生卵を指しており、状態変化後の目玉焼きの物体矩形とはインスタンス ID が異なる。したがって、「卵₂」と目玉焼きの物体矩形には \doteq 関係を付与する。なお、「目玉焼き」は総称名詞であるため、目玉焼きの物体矩形との間には \doteq 関係を付与する。

テキスト中の表現と物体が、総称・非総称と状態変化いずれかの関係を持つ例を挙げたが、両関係が複合する場合も存在する。例えば、「卵₁」は生卵に対する総称名詞であるため、目玉焼きの物体矩形との関係を考えると、総称・非総称の関係と状態変化の関係が複合していることが分かる。このような場合はいずれの関係も付与しない。また、2つ以上の状態変化の関係が複合する場合も関係を付与しない。以上の関係をまとめると、表 1 のようになる。なお、このときテキスト間照応アノテーションは表 2 のようになる。

表 1: 対話例中の各表現の解釈と図 11 中の物体領域との関係

表現	解釈	総称・非総称	関係	インスタンス ID
目玉焼き	一般の目玉焼き	総称	\doteq	boiled_egg_020(目玉焼き a, b)
卵 ₁	一般の卵	総称	-	-
卵 ₂	フライパンに割る前の卵	非総称	\doteq	boiled_egg_020(目玉焼き a, b)
卵 ₃	フライパン上の目玉焼き	非総称	=	boiled_egg_020(目玉焼き a, b)

表 2: 対話例中の各表現の解釈とテキスト間照応アノテーションの例

表現	解釈	総称・非総称	関係	コーパスタグ
目玉焼き	一般の目玉焼き	総称	-	-
卵 1	一般の卵	総称	≒	目玉焼き
卵 2	フライパンに割る前の卵	非総称	≒	卵 1
卵 3	フライパン上の目玉焼き	非総称	≒	目玉焼き AND 卵 2

5.1 タグ付けの省略

テキスト・物体間参照アノテーションでは、=格以外の関係についてテキスト間照応アノテーションの結果を利用してタグ付けを省略できる。例えば、以下のようなテキスト間照応アノテーションが付与済みだった場合を考える。

- (48) 【主人】そこにあるコップを、机に運んで (ガ格: 読者, ヲ格: コップ, ニ格: 机)

このとき、「コップ」とコップに対応する物体矩形に=格を付与すれば、「運んで」と「コップ」の物体矩形の関係は自動的にヲ格と推定できる(図 12)。したがって、「運んで」のヲ格のタグ付けは不要である。「机」についても同様である。ただし、「読者(ロボット)」に関しては対応する表現がテキストに出現していないため、=格が付与できず、ガ格のタグ付けが必要となる。



図 12: テキスト間照応アノテーションを利用したテキスト・物体間アノテーションの省略。

テキスト・物体間参照アノテーションではさらに、物体矩形のインスタンス ID を利用してタグ付けを省略できる。名詞句あるいは述語と、ある物体矩形の関係をタグ付けしたとき、同じ関係が同じインスタンス ID を持つ他の物体にも成立する。同じインスタンス ID を持つ他の物体との関係は、インスタンス ID から自動的に推定できるため、タグ付けが不要である。

5.2 領域に対するアノテーション

「ここ」や「あそこ」などの参照表現はしばしば物体ではなく空間あるいは平面上の領域を指示する。このような表現を領域参照表現とよぶ。領域参照表現に対しては、適当と思われる領域に矩形を付与し、参照表現と適切な関係タグを付与する。このとき、矩形を付与するフレームは対応する領域参照表現を含む発話区間内のフレーム全てとし、矩形クラス名は「region」とする。また、クラス名が「region」の矩形を領域矩形とよぶ。図 13 に以下の発話についての領域矩形アノテーションの例を示す。発話中の「そこ」という表現はいずれの物体にも対応せず、机の上的一部の領域を参照しているため、その領域を領域矩形として対応するフレームにタグ付けする。同時に、こ

の領域矩形と「そこ」に=格を付与する。

(49) 【主人】ついでにそこにあるコップとカップも片付けてちょうだい。



図 13: 領域矩形アノテーションの例 (20220304-56132132-0, 069.png より).

領域矩形は物体に依らないため、その位置・大きさはある程度作業者の主観に依存する。領域矩形アノテーションでは、主観による影響を小さくするため、できる限り言語的および視覚的文脈を考慮する。図 14, 図 15 はそれぞれ以下の発話に対応する領域矩形アノテーションの例である。

(50) 【主人】そこらへんに置いてくれるかしら？

(51) 【ロボ】こちらに置いておきますね。

図 14 では、主人の手が指示示す方向や、ものを置くことができる範囲を考慮して領域矩形を付与している。図 15 では、文脈からものを置く場所が確定したため、領域矩形をより狭い範囲に設定している。



図 14: 主人の発話「そこらへんに置いてくれるかしら。」についての領域矩形アノテーションの例 (20220302-56132034-1, 105.png より).

領域参照表現には、「ここ」や「あそこ」などの指示代名詞だけでなく、「上・下・右・左・中・

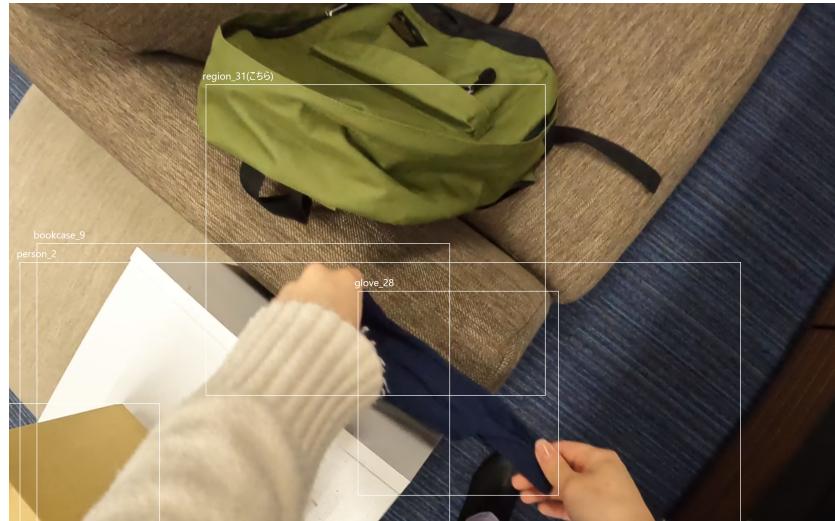


図 15: ロボットの発話「こちらに置いておきますね。」についての領域矩形アノテーションの例 (20220302-56132034-1, 110.png より).

横・そば・端」などの別の物体との相対的な位置関係を示す名詞も含まれる。また、「そこらへん」や「机の端の方」など、他の名詞と複合する場合もある。指示代名詞の中でも「こっち」「そっち」「あっち」などの方角を表す表現は、領域が対応しないため、領域参照表現に含まない。ただし、いずれの場合も領域参照表現であるかは文脈に依存する。以下の例において、前者の「こちら」は領域参照表現ではないが、後者の文脈における「こちら」は領域参照表現である。

(52) こちらを向いてください。

(53) こちらに置いてください。

「冷蔵庫の中」「ソファーの下」など領域が物体で隠れて見えない場合は矩形を付与しない。領域参照表現にはゼロ代名詞も含まれる。すなわち、用言やサ変名詞に対しても領域矩形と共に適切な関係タグを付与する。図 16 のフレームにおける発話「それでは、運ぶのを手伝ってもらってもいいですか？」では「運ぶ」の二格の項が省略されており、それはキッチンのまな板周辺の領域に対応する。したがって、適切な領域に領域矩形を付与し、「運ぶ」と二格の関係を付与する。



図 16: ロボットの発話「それでは、運ぶのを手伝ってもらってもいいですか？」についての領域矩形アノテーションの例 (20230301-57113858-2, 026.png より).

5.3 複数の解釈が可能な場合

以下の例における「床」という表現は、物体としては部屋全体の床に相当するが、文脈を考慮するとおもちゃが散乱している一部の床を指していると考えられる。この場合、部屋全体の床とregionとしての床両方に物体矩形を付与する（図17）。

(54) 【主人】おもちゃが床に散乱しているわね

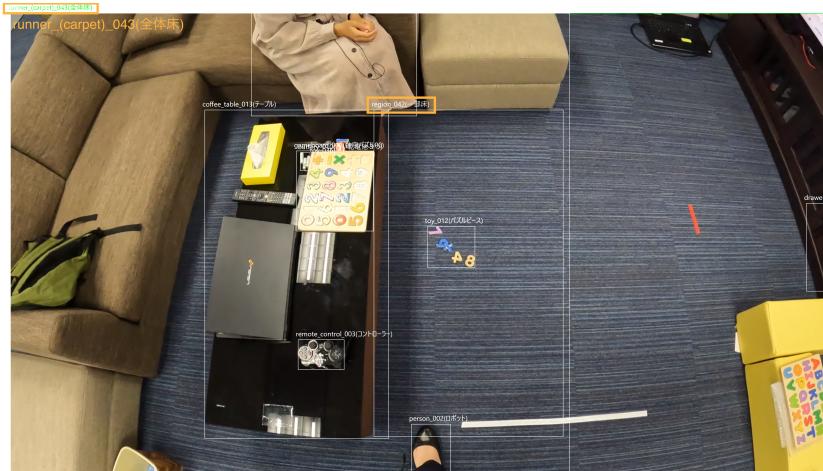


図17: 「床」に対する物体矩形の付与 (20220303-56133752-2, 176.png より)

6 未解決の問題

6.1 繰り返し表現の分割基準について

対話中では「はいはい」、「うんうん」、「そっかそっかそっか」、「ごめんごめん」等の繰り返し表現がしばしば出現する。これらの表現を1つの形態素として扱うのか、繰り返し単位ごとに形態素および基本句に分割するのかは未定である。現状、「はいはい」、「うんうん」は1形態素とし、「そっかそっかそっか」、「ごめんごめん」は形態素および基本句として分割している。

参考文献

- [1] Sadao Kurohashi and Makoto Nagao. Building a Japanese Parsed Corpus while Improving the Parsing System. In *Proceedings of the NLPRS*, pp. 719–724, 1998.
- [2] 萩行正嗣, 河原大輔, 黒橋禎夫. 多様な文書の書き始めに対する意味関係タグ付きコーパスの構築とその分析. 自然言語処理, Vol. 21, No. 2, pp. 213–247, 2014.
- [3] Agrim Gupta, Piotr Dollar, and Ross Girshick. LVIS: A dataset for large vocabulary instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.