

プランニングの知識を用いた照応・省略解決

志賀 聡子 徳永 健伸 田中 穂積

東京工業大学 大学院情報理工学研究所

{satoko,take,tanaka}@cl.cs.titech.ac.jp

1 はじめに

人間とコンピュータのインターフェースの手段として、最も望ましいのは自然言語を用いることである。しかし、コンピュータに自然言語を理解させるには多くの困難な問題があり、その中の一つに照応・省略解決がある。

これまでの自然言語理解の研究において、話者の意図理解にプランやゴールといった概念が扱われてきた [3]。本稿では、このような概念が対話中に現れる照応・省略の解決にも役立つことを示し、プランニングの知識を用いた照応・省略解決手法を提案する。

自然言語で仮想世界内のロボットを操作する対話システムとして、「積木の世界」で知られる Winograd の SHRDLU がある [5]。また、近年の音声認識技術の発展に伴い、新山は音声による指令を理解するシステム傀儡を開発した [4]。提案手法は、このような対話システムにおける照応・省略解決を想定している。

2 関連研究

従来の照応・省略解決の研究は、テキスト（書き言葉）を対象としたものが多く、主に代名詞などが出現した際に、先行文脈中の名詞句と結び付けることを目的としている。Grosz らの Centering 理論 [1] が代表的な手法であり、この理論をもとに多くの研究がなされている。たとえば、亀山は Centering 理論を日本語の省略解決に適用した [2]。これらの手法は、文の表層に現れる情報を解析して先行詞を同定している。

一方、SHRDLU や傀儡など、ユーザの指令に対して行動する対話システムでは、ロボットの行動手続きを生成するために、「誰が何をどうする」という情報をすべて理解することが必要で、省略の検出が必須になる。また、テキストとは異なり、先行詞がそれまでの発話文中に現れるとは限らないという問題点がある。さらに、先行詞は世界内の実体と結び付ける必要がある。たとえば「球」という表現に対し、世界内に球が複数あれば、どの球かを同定しなければならない。

傀儡における照応・省略解決アルゴリズムは、まず、発話を格フレームに変換し、その後、照応・省略解決

済みの履歴発話と比較して先行詞を同定する。このために、発話スレッドと呼ばれる履歴データベースに発話を分類して保存しておく。発話の分類には、主語・述部の一致や手がかりとなる語を用いる。そして、照応・省略表現を含む発話が入力された場合には、まず発話スレッドを同定し、発話スレッド内の履歴発話の格と、照応のある発話の格を比較することで、照応・省略を解決する。もし、参照すべき発話が見つからなければ、仮想世界内を探索する。

3 問題点

従来の照応・省略解決手法では、たとえば、次のような場合に問題が出てくる。

対話例 1: 次の行動が予測できる場合

#1-1 「ロボット A は赤い球を押して」

#1-2 「青い球の前に行って」

#1-3 「それを押して」

発話 #1-3 の目的格「それ」は「青い球」と解決するべきである。しかし、傀儡のアルゴリズムでは、述部「押す」が一致していることから発話 #1-1 を参照し、「それ」の先行詞は「赤い球」となる。さらに、発話 #1-2 を参照したとしても、発話 #1-2 には目的格がないため、発話 #1-3 の目的格を補完できない。この場合、人間が「それ」は「青い球」とであると判断するのは、わざわざ青い球の前に行ったのだから、次は青い球を押すのであろうと推論しているからだと考えられる。つまり、ある行動が次の行動のための準備であり、次の行動が予測できる場合に、ユーザは暗黙のうちに焦点化を行うと考えられる。

対話例 2: カメラの移動でユーザの視線が変化する場合

#2-1 「ロボット A は青い球を押して」

#2-2 「カメラは右を映して」

(→その結果、画面に赤い球が映る)

#2-3 「それを押して」

発話 #2-3 の目的格「それ」は発話 #2-1 から「青い球」と解決するのではなく、文脈にはないが発話 #2-2 の

動作によって映った「赤い球」と解決すべきである。カメラの向きが変わり、ユーザの目の前には新しい物体が現れると、ユーザの焦点は暗黙のうちにその新しい物体に移動すると考えられる。

このように、表層に現れた情報の解析だけでは不十分な照応・省略現象が存在する。人間がロボットに指令を行う際、頭の中に大きなゴールがあり、図1のように人間自身がプランニングを行って、そのプランを分割して順次指示していくと考えられる。そして、その間で冗長な言い回しを避けるために照応・省略表現を使うので、このような照応・省略を解決するためには、システムもプランニングの知識を使って推論する必要がある。

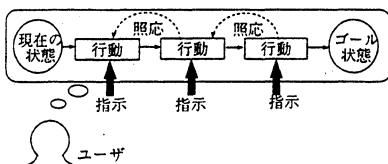


図1: ユーザのプランニングと指示

4 プランベースの照応・省略解決

4.1 プランへのマッピング

本稿における対話システムのアーキテクチャは図2のようにになっている。システムは、ユーザの指令を、ロボットの行動を生成するプランナに渡すためのゴールに変換しなければならない。そこで、発話を格フレームに変換し、それをもとに述部と対応するプランライブラリ中の行動を取り出す(図3)。

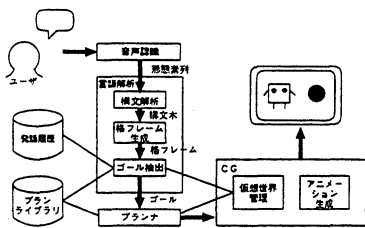


図2: 対話システムのアーキテクチャ

行動には必須格が引数として付与されている。これらの引数を発話の格フレームと対応させて埋めるが、代名詞であったり、発話の中に存在しない格は変数として残る。また、表層から世界内の実体と一意に対応付けられないものも変数となる。こうして変数のまま

残された部分が、次に照応・省略を解決すべき箇所であるとわかる。

また、各行動にはそれぞれ、その行動を行うために必要な前提条件と、その行動を行った結果得られる効果が付与されている。行動を取り出すことができれば、その行動の効果がユーザのゴールであると推測する。

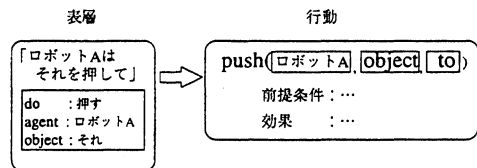


図3: プランへのマッピング

ここで、発話の述部が省略されている場合には、プランライブラリとの対応が取れないので、まず述部の省略を解決する。たとえば、対話例3のようにゴールとなる状態が直接指定されている場合には、その状態と等しい効果から検索する。

対話例3: ゴール状態が指定される場合

#3-1 「カメラは赤い球が中心に見えるように(映して)」

その他に、従来の傀儡における省略解決手法と同様に、先行文脈を参照する手がかり語なども利用する。

4.2 照応・省略解決アルゴリズム

4.2.1 プランベースの照応・省略解決

抽出したゴールに含まれている照応・省略を解決するために、まず世界内の実体に対し、表層に現れている情報によって制約をかけて、先行詞の候補を絞り込む。たとえば、「それ」「その」という指示詞がある場合、候補はユーザから見えているものに限定される。ここで得られた候補に対し、本手法では次の二つの点にプランニングの知識を利用して、先行詞を同定する。

1. ユーザのプランを認識して参照先の発話を同定

ある行動の効果が、別の行動の前提条件となっていた場合、この二つの行動間には意味的な関連性があり、同一ゴールを目指した発話を分割したものであると考えられる。そこで、行動の効果と前提条件の連鎖からユーザのプランを認識する(図4)。照応・省略を含む発話が入力された場合に、その発話が属するプランがわかれば、参照すべき発話を同一プラン内のものに同定できる。

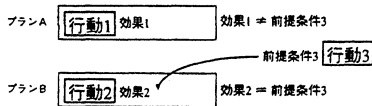


図 4: プランの同定

2. プラン内の情報を利用して参照先の格を同定

プラン認識時に、参照先発話の効果と、照応・省略のある発話の前提条件が一致しているのを、これらのマッチングを行うことで、発話中の参照すべき部分を同定する(図5)。対話例1の発話#1-2と発話#1-3の場合、照応詞「それ」と先行詞「青い球」の格は一致していないが、本手法ではこのような場合にも参照先の格を同定できる。

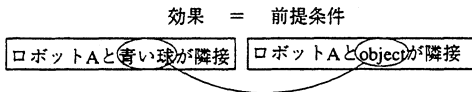


図 5: 参照格の同定

4.2.2 効果と前提条件の部分マッチング

一般に、行動には複数の前提条件と効果があるので、マッチングの際にはそのうち一つでもマッチすればよいとする。これは、ユーザがプランニングした結果の行動を逐一指示するとは限らず、途中を省略してロボットに推論させる場合があることを考慮している。たとえば、次の対話例4に示す場合である。

対話例4: 指示が省略される場合

#4-1 「赤い球の前に行って」

#4-2 「それを右に押して」

発話#4-2が指示された時点から赤い球を右に押すためには、その前にまず球の左に移動して、それから右を向くという行動をしなければならない。図6に示すように、ユーザはまずロボットに対象を同定させるためにプラン中の一番目の行動を指示し、次に四番目の行動を指示しており、途中の二番目と三番目の行動の指示は省略している。この場合には、発話#4-2の指示を実行するのに必要な前提条件を全ては満たしておらず、赤い球とロボットが隣接しているという条件だけが成立している。

4.2.3 効果と仮想世界の状態

これまで述べてきたように、行動の効果と前提条件のマッチングを行うことで、照応・省略を解決するこ

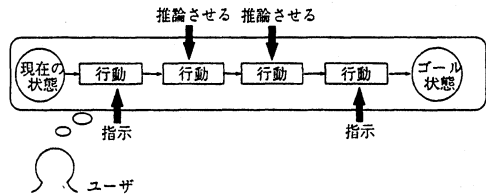


図 6: ユーザのプランニングと指示の省略

とができる。しかし、ある行動による効果は、他の行動によって失われな限り持続する。そのため、同一プラン内の直前の行動の効果だけではなく、現在持続しているすべての効果とのマッチングをとるべきである。現在の仮想世界の状態には、先行するすべての行動の効果が反映されて、持続している効果のみが残っている。したがって、図7のように前提条件と現在の仮想世界の状態とのマッチングを取ればよい。これにより、文脈に応じて起こる、プランライブラリに明示されていない効果も捉えることができる。

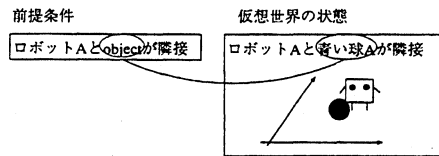


図 7: 仮想世界と前提条件のマッチングによる解決

また、直前の行動の効果によるものではなく、もともと前提条件が満たされていることもある。その場合には、いきなり「それ」という指示を与えても、同様の推測が可能である。行動の前提条件と現在の仮想世界の状態のマッチングをすることで、このような直示表現も扱うことができる。

4.2.4 注目度スコアによる先行詞の決定

先行詞の候補を絞り込んだ結果、候補が複数残ったり、一つも残らない場合が出てくる。そこで、候補が複数残った場合は、その各候補にスコアを与える。候補が一つも残らなかった場合は、表層から選出した各候補にスコアを与える。そして、最もスコアの高いものを先行詞とする。本手法では、次の式でスコアを計算する。

$$score = \alpha S1 + \beta S2 + \gamma S3$$

α, β, γ は人手で付けた重みであり、 $S1, S2, S3$ は次の要素である。

$S1$: 先行文脈からの注目度

S2: 実行するロボットの視線による注目度

S3: ユーザ (カメラ) の視線による注目度

S1 は、その候補について言及された頻度と、どれだけ最近言及されたかの履歴をもとに計算する。また、S2, S3 は、現在の仮想世界の状態を見て、距離の近さと、視界の中心線との角度をもとに計算する。

5 結果と考察

提案手法を適用したモジュールを実装した結果、本手法は従来の手法で解決できなかった主格、目的格の照応・省略 (対話例 1,2) に特に有効であった。また、ゴール状態を指定される際に起こる述部の省略 (対話例 3) の解決も可能になった。

次に、解決できなかったものについて述べる。状態を指定する発話でも、状態の継続が指定される発話における述部の省略 (対話例 5) は解決できない。

対話例 5: 状態の継続が指定される場合

#5-1 「ロボット A が画面からはみ出すことのないように (映して)」

これは、まず状態を継続する行動を扱えるように、動作の途中の状態を監視する動的なプランニングを導入する必要がある。そして、照応・省略を解決するために、他の制約などの情報を考慮することが必要である。

また、並列なプランを複数指示するときに起こり得る、複数の格が関係した省略 (対話例 6) は解決できない。この場合、述部の省略をまず「行く」と解決するため、「の右に」という情報が欠如する。これは、述部を単純に動詞のみで扱うのではなく、その他の格も考慮してプランにマッピングすべきであることを示している。

対話例 6: 複数の格が関係している場合

#6-1 「ロボット A は赤い球の右に行って」

#6-2 「ロボット B は青い球 (の右に行って)」

さらに、状態を表す格内に照応・省略がある場合 (対話例 7) にも解決できない。これは、格解析が不十分なためであり、状態を表す格内をさらに主格や目的格に分解できれば、提案手法が適用できると思われる。

対話例 7: 状態を表す格内に照応・省略がある場合

#7-1 「ロボット A は赤い球の前に行って」

#7-2 「それが青い球の隣に来るように押して」

また、今回は世界内の実体を指示する照応・省略のみを対象としたが、イベントなど実体以外のものを指示する照応・省略 (対話例 8) もあり得る。この場合、

プランの概念でイベントのまとまりを扱うことが考えられる。

対話例 8: 実体以外を指示する場合

#8-1 「それ、もう一度やりなおして」

6 おわりに

本稿では、ロボットを自然言語で操作するタスクにおける、プランニングの知識を用いた照応・省略解決手法を提案した。これにより、従来のような表層に現れる情報に基づく手法では解決できなかった、関連性のある行動間での非明示的な焦点の移動に追従することが可能になった。本手法では、行動の持つ前提条件と仮想世界の状態のマッチングを行うことにより、先行詞を同定する。

今回は、限定された単純な行動のみを対象としているが、今後、複数のロボットでの協調作業や、動的なプランニングが必要な行動など、複雑な行動を扱えるように拡張することが考えられる。提案手法では、照応・省略解決に用いる知識をロボットが行動するためのプランニングに用いる知識と共有しているため、拡張性に優れている。しかし、プランへのマッピング時の曖昧性など、新たなアルゴリズムの拡張が必要となることも考えられる。また、人間は言葉で指示を与えながら、同時に視線や指差などのジェスチャーによって対象を指し示しており、照応・省略の解決を行う際に、パラ言語学的な情報も重要な要素となる。たとえば、対話例 2 のように視線の移動によっても焦点化が起こることから、現状のようにカメラの向きでユーザの視線を考慮するのではなく、より詳細に情報を取得できることが望まれる。

参考文献

- [1] Barbara J. Grosz, Arvind K. Joshi, and Scott Weinstein. Providing a Unified Account of Definite Noun Phrases in Discourse. In *proceedings of ACL*, pp. 44-49, 1983.
- [2] Megumi Kameyama. A Property-Sharing Constraint in Centering. In *proceedings of ACL*, pp. 200-206, 1986.
- [3] Roger C. Schank and Robert P. Abelson. *Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures*. L. Erlbaum, 1977.
- [4] 新山祐介, 徳永健伸, 田中穂積. 自然言語を理解するソフトウェアロボット: 傀儡. 情報処理学会論文誌, Vol. 42, No. 6, pp. 1359-1367, 2001.
- [5] T. Winograd. *Understanding Natural Language*. Academic Press, 1972.