

仮想空間中のエージェントとの対話による行動制御

岡崎 篤[†] 船越 孝太郎[‡] 徳永 健伸[§] 田中 穂積[§]

東京工業大学 大学院情報理工学研究科

1. はじめに

音声対話システムでは、自然言語による発話を理解する必要がある。その際、冗長性を避けるために、照応・省略表現が頻繁に使われる。したがって、音声対話システムにおいて、照応・省略の解決が不可欠である。

新山によって開発された音声対話システム傀儡[1]では、表層的な情報や過去の対話履歴を用いた照応・省略解決手法を実装している。

しかし、このような情報を利用するだけでは解決できない事例があることが分かっており、それに対して志賀は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ている[2]。

本研究では、プランニングの知識を用いた照応・省略解決手法を実装し、仮想世界中のソフトウェアエージェントと対話を行なうシステムを構築した。

2. システムの概要

システムの動作画面を図 1 に示す。仮想世界中に複数のエージェントが存在し、エージェントはユーザの発話からその意図を解釈して行動する。エージェントの動きは、CG アニメーションとしてユーザに提示される。

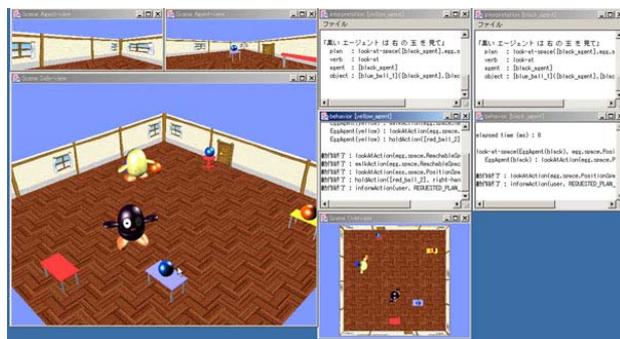


図 1 : システムの動作画面

2.1 システムのアーキテクチャ

システムのアーキテクチャの概念図を、図 2 に示す。

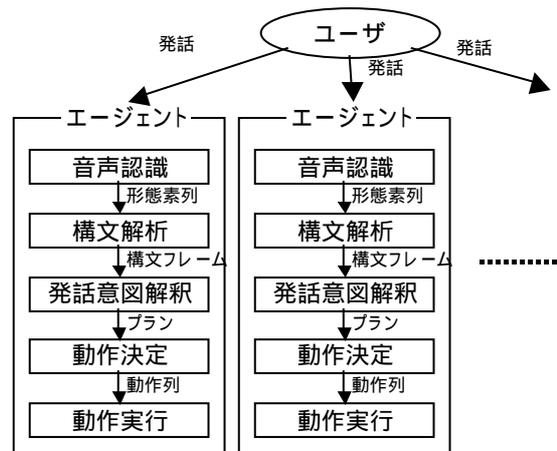


図 2 : システムのアーキテクチャ

各エージェントは、自分の状態などに依存して、ユーザの発話をそれぞれ独立して解釈し、振る舞いを決定する。そのため、例えば「青いボールを持って」と言われた場合に、エージェントがそれぞれ自分にする発話だと解釈する可能性があり、さらにエージェントごとに別の「青いボール」と解釈する可能性もある。

実際のシステムでは、音声認識を複数動作させることは現実的でないため、音声認識と構文解析の後でエージェントごとに解析結果を振り分けている。

構文解析については、言い直し・言いよどみに対応した構文解析器[3]を利用している。

2.2 処理の流れ

例えば「右の机の上にあのボールを置いて」という発話があると、構文フレームの段階では、「あのボール」や「右の机」が具体的に仮想世界中のどの実体を指すのか決まっていない。また、この発話は主語が省略されているため、誰に対する命令かも分からない。

それに対して、これを元に発話の意図を解釈すると、例えば「発話者は agent-2 に ball-1 を table-3 の上に置いて欲しい」というような、仮

Controlling agents through speech dialogue

[†]Okazaki Atsushi

[‡]Funakoshi Kotaro

[§]Tokunaga Takenobu

[§]Tanaka Hozumi

Department of Computer Science, Tokyo Institute of Technology

想世界中の実体を特定した意図として理解できる。

この意図を理解した上で、例えばこの発話者の意図を満たすために何をすべきかをプランニングし、まず ball-1 を取るために ball-1 の近くまで行ってそれを取って(そのボールを持っていない場合)、その後で table-3 の近くまで行ってそのボールをその机の上に置く、というような実行すべき動作列を生成する。

本研究で実装したシステムは、エージェントに対する命令文のみに対応している。したがって、ユーザの発話に含まれる意図はエージェントに何かをする要求の意図であると仮定しているので、発話の意図を解釈した結果はプランの形で表すことができ、発話の意図を解釈することは構文フレームからプランにする処理として捉えることができる。

3. 照応・省略解決手法

3.1 手法の概略

自然言語を理解するには、照応・省略表現を解決する必要がある。特に本研究が対象としている音声対話では、冗長な表現を避けるために照応・省略が使われやすい。例えば「それを赤い机の上に置いて」という発話では、主語が省略されていて、「それ」が照応表現となっている。

書き言葉を対象とする照応・省略解決と異なり、音声対話システムとして実際に実装し、その解釈をアニメーションとして提示しようとする、参照先の名詞句を単に表層的な先行文脈から探すのではなく、仮想世界の中の実体から探す必要がある。同様に、「青いボール」のように明示的に言及されている場合でも、仮想世界の中に青いボールが複数存在すればどちらの実体を指しているのかを区別する必要がある。このような実体の特定も、広義の照応解決とみなすことができる。

また、過去の対話履歴から適切な対象を探す手法は、照応・省略解決によく用いられるが、このような手法は過去に言及されていない対象を参照する表現に対応できない。これは書き言葉ではあまり問題にならないが、音声対話システムの場合は、仮想世界中のある物体の方を見ながら、もしくは指差しながら、「それを取って」というような直示を含む発話があり得る。したがって、照応・省略の参照先が過去の対話で言及されているはずであるという前提は必ずしも成り立たない。

さらに、照応・省略の解決において、例えば「それ」という名詞句だけに着目して解決しようとしても限界がある。「それを机の上に置いて」の「それ」が何を指すかは、「机」をどの机と解釈するかによって適切な解釈が変わり、例えばその机に既に置かれている物は「それ」が指す対象として適切でない。これは、「青いボールを机の上に置いて」の「青いボール」が具体的にどの実体を指しているのか解釈する場合も、同様のことが言える。つまり、意図を解釈するためには、発話の部分ごとに着目してそれに対する最良の解釈を探し出してそれらをつなぎ合わせても、全体として最良の解釈が得られるとは限らず、「誰がどの物体をどこに置くのか」というような全体の解釈に対して最良なものを探さなければならない。

以上に述べた手法の要件を満たし、かつ以下のような特徴を持つ照応・省略解決手法を実装した。

- 広義の照応解決と一般的な意味での照応・省略解決を区別せず、同じように扱う
- 可能性のある実体すべてを候補にする
- 可能性のある解釈すべてを組合せ的に候補として挙げておいて、その中から全体の解釈として最良のものを選ぶ

省略や「それ」のような場合は仮想世界中のすべての物体を候補とし、「そのボール」、「ボール」、「机の上にある青いボール」などの場合はその表現に合うすべての物体を候補とする。

そして、まず可能性のある解釈すべてを組合せ的に候補として挙げ、その後でその中から最良の解釈を選ぶ手法をとっている。本研究では前述の通り、発話の意図をプランとして扱うので、前者は構文フレームを 1 個以上のプラン候補にマッピングする処理となり、後者はそのようにして得られた候補の中から最良のものを選択する処理となる。

3.2 構文フレームからプランへのマッピング

発話から得られた構文フレームをどのプランにマッピングするかは、基本的には動詞を手がかりに決定する。例えば「机の前まで歩いて」と言われた場合には、「エージェントがある場所まで歩く」プランに対応付けられる。

また、プランごとに、プランを特徴付けるパラメータを持つ。例えば「エージェントがある場所まで歩く」プランならば、(1)誰が(2)どこまで歩くのかという 2 つのパラメータが必要となる。

以上を踏まえ、構文フレームからプランへの

マッピングでは、対応関係を記述した知識とシソーラスをあらかじめ記述しておいて利用する。

対応関係を記述した知識には、どの動詞を持つ時にどのプランに対応する可能性があるか、さらにプランの各パラメータが構文フレームのどの格と対応するかが、記述されている。

また、必要な格が省略されている場合にどのようにそれを補えばよいかも記述されていて、仮想世界中のオブジェクトすべてが候補になる場合もあれば、「歩いて」と言われた場合に行き先を表すパラメータに動作主の近くの空間を補う(「近くの場所まで歩く」解釈)ように、別のパラメータを利用して補う場合もある。

シソーラスには、仮想世界中に存在するオブジェクトの意味クラスについて上位下位関係を記述し、これを用いて各プランの各パラメータに対して意味的制約を記述する。例えば「それを机の上に置いて」という発話を解釈する際に、単語「それ」だけに着目するとすべての実体が候補となってしまうため、別のエージェントを机の上に置くような誤った解釈を防ぐのにこの制約は不可欠である。

しかし一方で、このような制約が厳密に守られるとは限らない。例えば、「赤いボールを青い机の上に置いて」と同様の意味で、「赤いボールを青い机に置いて」と言うことも多い。この広義の比喩と呼べる問題に対処するため、シソーラスの各意味クラスに対して、置く先の場所に「机」は可能だがその机の上という意味に置き換えなければならない、というような規則も記述しておくことにした。

そのほか、「青いボールも」のように手がかり語がある場合には、過去の対話履歴から参照先の発話を探し、足りない格をその発話の格で補ないながら、構文フレームからプランへのマッピングを行なう。対話の流れによっては、「青いボール」をどう解釈するかに依存して、参照先とみなされる発話が異なる可能性がある。

また、動詞が省略された場合には、過去の発話の中にそれが参照している発話があるはずであり、まず参照先の発話を探し、その参照先の発話(手がかり語がない場合には直前の発話)の動詞が省略されたものとみなす。

動詞によってプランの種類が1つに決まらない場合もある。例えば「行って」の場合には、「歩く」「走る」などの可能性がある。どれを選ぶべきかは、例えば移動先が近くであって特に急いでいなければ「歩く」と解釈すべきなど、プランのパラメータの組合せ全体で決まるため、構文フレームをプランにマッピングする時点で

は、可能性のあるものすべてを候補とする。動詞からプランへマッピングする際にその曖昧性を許して、後でプランのパラメータなどとの関係において適切なものを選択することで、表層的な解釈にとどまらない意味を考慮した解釈を可能にしている。

3.3 プランの選択

プランの候補の中から最良と思われる解釈を選択するため、以下の基準で適切度が最も高い解釈のみを残して絞り込む。残りの候補が1つになるまで、この順序で各基準を適用する。

- プランの前提条件を満たす解釈ほど優先
- 最近プランの前提条件が満たされた解釈ほど優先
- 既にプランの効果が満たされていない解釈ほど優先
- 文内照応しやすさに関する表層的な特徴に沿う解釈ほど優先
- 最近言及されたかどうかに関して表層的な特徴に沿う解釈ほど優先

「青いボールを赤い机の上に置いて」と言う代わりに、「青いボールを持って」「それを赤い机の上に置いて」のように分けて発話することがある。そのような過去の発話で意図された行動の結果は、意図が正しく解釈されていれば仮想世界に反映されているはずで、物を置くためにはその前に持っていなければならないというように、仮想世界の中でプランの前提条件を満たしている解釈ほど、文脈に沿った自然な解釈であると言える。

また、例えば「それを赤い机の上に置いて」と言った場合に、既に赤い机の上にある物体を置く解釈は適切ではないため、プランの効果が既に満たされている解釈ほど、適切でない解釈とみなす。

プランニングの知識を使うことで、コンピュータによる自然言語処理が不得手にしがちである、人間同士の対話では前提となっている意味および常識的知識を考慮して、解釈を行なうことができる。

構文フレームをプランにマッピングする際の意味制約が、仮想世界中のオブジェクトの意味クラスで決まる静的な制約であるのに対し、プランニングの知識を利用することは、発話のあった時点でボールを持っていたかどうかなど、仮想世界で動的に変化する状況に関する制約となる。

実際の対話システムとして実装してみると、プランニングに関する知識が照応・省略の解決

に有用である一方、表層的な情報や過去の対話履歴が不要になるわけではないことが分かった。

その代表例が「青いボールをその右にある机の上に置いて」の「その」のような文内照応である。発話の中で照応・省略表現で現れた場合、同じ発話内の別な名詞句と同じ実体を参照している解釈の方がより適切である。逆に「ボールを赤いボールの右にある机の上に置いて」の「赤いボール」のように照応・省略表現を使わずに現れた場合には、同じ発話内の別の名詞句と同じ実体を参照していない解釈の方が適切である。

また、対話中の発話はそれぞれ独立したものでなく関係を持っているので、発話の中で照応・省略表現で現れたということは、最近言及された実体を指しているのが冗長性を避けた可能性が高いと言える。そのため、発話の中で照応・省略表現が現れた場合、過去の対話履歴を参照して、より最近言及されている実体に解釈しているものを優先させる。これは、過去の対話履歴の中から参照先を探すような照応・省略解決手法での対話履歴の利用の仕方と大きく異なり、表層的な傾向を利用しているだけであり、しかもプランニングの知識を利用して差がつかなかった場合にのみ使われる。

解釈の候補を絞り込んでなお複数の候補が残っている場合、動作の主体が自分である解釈を選ぶ。これは、主語が省略された場合に、文脈から判断できるエージェント1体に対して命令する以外に、すべてのエージェントに対して命令する場合を考慮しているからである。また、実際のシステムでは、自分に対して命令された発話を自分に対する命令でないと解釈してしまうのを防ぐためでもある。

自分に対して命令されている解釈が複数残っている場合、空間指示表現のもっともらしさや動作主であるエージェントの視界を考慮して解釈を1つ選ぶ。例えば「右にある机に青いボールを置いて」と言った場合、「右にある机」としてもっともらしい位置にある机を使った解釈が採用される。また、「青いボールを取って」と言った場合、動作主であるエージェントの目の前にある青いボールを使った解釈を優先する。

3.4 手法の評価

本研究の照応・省略解決手法を評価するため、小規模な実験を行なった。

実験の方法は、2体のエージェントとテーブル5つとボール4つが置かれた仮想世界を用意し、ランダムに生成した目標となる配置図を被験者

に見せて、その配置を実現するように、エージェントに対する命令をユーザに発話してもらう、というものである。

実験で被験者が行なった実際の発話の例として、以下のようなものがあった。

- 1) 青いボールを取って
- 2) 後ろを向いて
- 3) 前の机にそれを置いて

「後ろを向いて」はボールを置く机を視界に入れるための発話であるが、表層的には対話の焦点が「青いボール」から外れてしまったように見える。

一般に、焦点が他に移ったように見える発話や、間違った解釈に導くような名詞句を含んだ発話が間に入ることで、対話履歴から参照先を探すような従来の照応・省略解決手法は正しく解決できなくなってしまう。

しかし、本研究の手法で用いたプランニングに関する知識によって現在の状況を把握することで、最近言及された対象に過度に惑わされずに、発話の合った時点で持っているボールを置く解釈を優先することができた。

4. おわりに

本研究では、音声対話システムに不可欠な照応・省略解決にプランニングに関する知識を使った手法を用い、音声対話システムを実装した。実際にシステムとして動作させて必要な改良を加えることにより、多くの言語現象を正しく扱えるシステムを実現できた。

今後の課題として、本研究の手法を評価するために、より大規模な実験で定量的な評価を行なうことが挙げられる。

また、現在のシステムではエージェントに対する命令を表す発話しか扱えないなど、扱える発話のバリエーションには制限がある。質問や情報伝達を表す発話など扱えるようにシステムを拡張することも、今後の課題である。

参考文献

- [1] 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
- [2] 志賀聡子, プランニングの知識を用いた照応・省略解決, 言語処理学会第8回年次大会, pp.599-602, 2002
- [3] 船越孝太郎, 単語の意味役割を用いた自己修復表現の処理. 言語処理学会第8回年次大会, pp.655-658, 2002