対話システムにおけるプランベースの照応解析*

徳永 健伸† 関谷 幸恵‡ 田中 穂積§
東京工業大学 大学院情報理工学研究科¶

概 要

本稿では，音声対話システムにおいて，ユーザのプランを考慮して照応解消をおこなう手法について述べる．ユーザは発話を通して自分が意図するゴールを達成する．したがって，それぞれの発話はその背後にあるゴールと密接に関係している．この関係を前提として，ユーザは発話中で省略や照応を用い，冗長な表現を避けることができる．本稿で述べる手法では，各発話に対応するプラン演算子の前提条件と効果の関係を手がかりに，同じゴールを持つプラン演算子 (発話) のグループを同定する．照応や省略の先行詞は同一のゴールを持つ発話グループの中から探索する．この手法を音声対話コーパスを用いて評価し，その有効性を確認した．

1 はじめに

照応解消に関する研究は盛んにおこなわれてきたが，その多くはテキスト中の文脈照応を，主に表層に表われる言語的手がかりによって解消するものであった．たとえば，センタリング理論では，談話の話題がテキスト中の「センター」と呼ばれる要素によって維持され，センターが照応や省略の先行詞になりやすいと仮定している [5, 6, 10]．各要素がセンターになる可能性は各要素の文法的な役割によって優先順位が付けられ，センターは談話の進行とともに，いくつかの制約によって管理される．

センタリング理論は対話にも適用されているが，音声対話では発話の境界が明確でないこと，いわゆる「文」を構成しない断片的な発話が多いことなどが問題となっている [2]．特に断片的な発話では多くの省略が生じるので，表層的な手がかりだけでは不十分なことが多い．日本語ではテキストにおいても多くの省略が生じることが知られており，亀山らはセンタリング理論を日本語の省略解消にも適用しているが，音声対話における省略とテキストにおける省略を同等に扱えるかどうかについては十分な研究がされていない．

Winograd の SHRDLU は対話における照応解決の草分け的研究である [12]．SHRDLU では，キーボードを介したシステムとの対話によって，グラフィック画面上に表示されたさまざまな積木を操作するシステムである．SHRDLU は対話中の照応解消をすることができたが，照応解消のための特別な理論を実装していたわけではなく，その時点での積木の世界の状態を参照して，経験則によって照応を解消していた．また，画面上に積木が提示されているので直示的な照応も扱うことができた．新山らは SHRDLU を音声対話に拡張し，SHRDLU のような状況では，世界の状況に関する知識が照応解消に有用であると指摘している [8]．

本稿では，SHRDLU のように，ユーザとの対話によって変化する世界をユーザに映像として提示するような状況において，ユーザのプランを考慮して照応や省略を解消する手法について述べる [13]．談話におけるプランの重要性は，すでに多くの研究者によって指摘されている [7, 3]．また，小規模な音声対話のコーパスを用いて，センタリング理論とプランに基づく照応解消手法を評価・比較した結果についても述べる．

2 問題設定

本稿でとりあげるシステムは，日本語の音声入力によって仮想世界中のエージェント (にわとり，馬，雪だるま，カメラ) に命令を出し，赤，青それぞれ 2 つの球を操作するものである [8]．エージェントがユーザの発話を理解し，行動する様子はグラフィック画面上にアニメーションとして表示される．

以下はこのシステムの対話例である．

(1-1) にわとりは球を左から押して．
(1-2) もっと．

---

*Plan-based anaphora resolution in speech dialog
†Tokunaga Takenobu
‡Sekiya Satie
§Tanaka Hozumi
¶Department of Computer Science, Tokyo Institute of Technology

(1-3) 雪だるまもそれを押して．

(1-4) カメラはその球を正面から写して．

この例からもわかるとおり，発話中には多くの省略や照応が使われており，人間でもその場の状況を参照しないと，これらの発話を正確に理解することは難しい．ここで扱っている世界は単純で，エージェントの行動も限定されているので，センタリング理論のような表層的な手がかりを利用した手法でも照応が解消できることも多い．しかし，以下の例でわかるとおりユーザがキーボードを介したシステム対話によって表層的な手法には明らかな限界がある．

(2-1) 雪だるまは赤い球を押して

(2-2) 青い球の前に移動して

(2-3) (青い球を) 押して

発話 (2-3) で省略されている目的語は「青い球」であり，これはセンタリング理論でも扱える．しかし，最初と最後の発話の動詞だけを変えた以下の例では，3番目の発話で省略されている要素は変ってくる．

(3-1) 雪だるまは赤い球を取って

(3-2) 青い球の前に移動して

(3-3) (赤い球を) 置いて

文法的な構造や各要素の役割は例 (2-1)–(2-3) も (3-1)–(3-3) も同じなので，この違いは表層的な手がかりだけでは説明できない．

これらの発話の背後にあるユーザのプランを考慮するとこれらの例は以下のように説明できる．あるオブジェクトを「置く」ためには，前提としてそのオブジェクトを持っていなければならない．例 (3-1)–(3-3) の場合，発話 (3-1) によってエージェントは赤い球を持つことになり，(3-3) で置く対象となるのは現在持っている赤い球であると考えるのが自然である．一方，例 (2-1)–(2-3) の場合，一般にあるオブジェクトに対して操作をするためにはその近くに行く必要があるという知識があれば，青い球の前に移動するのはそれを押すための動作であると理解できる．

我々が対象としているシステムでは，ユーザは自分のゴールを達成するためにシステムと対話する．特に Cohen が指摘しているように，音声対話では，発話が断片化しやすく，いくつかの発話によってあるゴールを達成する傾向が観察される [3]．したがって，ユーザの発話はゴールを達成するプランの一部に対応していることが多く，これらの関連したプランの一部の中では照応や省略が起りやすくなると考えられる．このような観察に基づき，次節では，ユーザの発話の背後にある部分プラン (プラン演算子) を同定し，その部分プラン同士の前提条件と効果の関係に基づいて照応を解消する手法について述べる．

3 プランに基づく照応解消

システムに入力された音声発話は音声認識器によって認識され，統語・意味解析され，解析結果は格フレームの形式で照応・省略解析モジュールに渡される．照応・省略解析では，この格フレームに対応するプラン演算子をプラン・ライブラリから検索する．現在の対象領域は単純なので，この検索は格フレームの動詞をキーとしておこなう．このシステムで使っているプラン演算子は STRIPS [4] と同様に，前提条件，効果，行為記述からなっている．

検索したプラン演算子中の変数は格フレームの要素によって埋められるが，照応・省略がある場合には，すべての要素が埋まらない．欠けた要素がある場合はこの図 3 に示すアルゴリズムによって照応・省略を解消する．

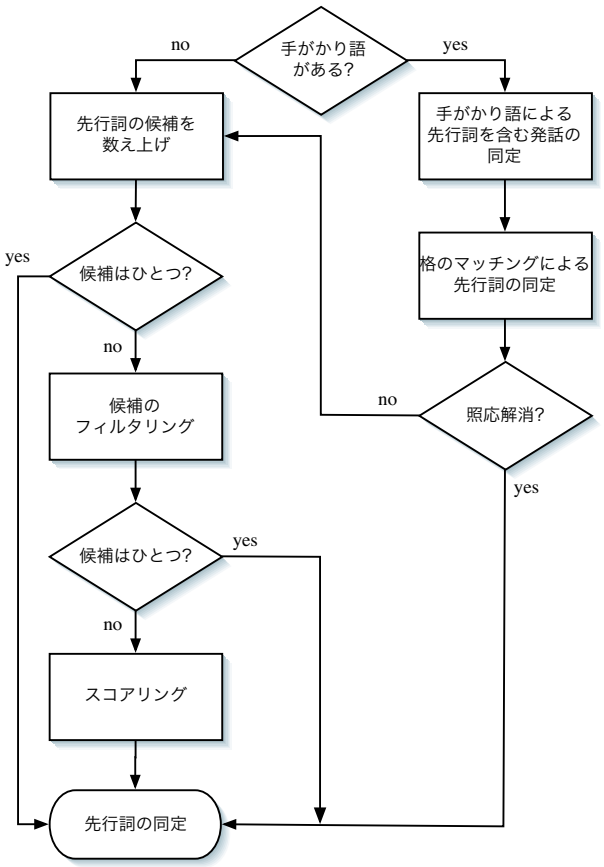


図 1: 照応・省略解消アルゴリズム

発話中に「もっと」や「〜も」などの手がかり語がある場合は，発話履歴を参照し，格フレームのマッチングによって同じゴールを持っている発話を検索し，その要素を使って先行詞を同定する．手がかり語があっても，発話履歴から先行詞を同定できなければアルゴリズムは左の枝分かれに移る．

左の枝分かれでは，まず，先行詞の候補を表層情報を手がかりにして数え上げ，それを言語的な制約とプラン・ライブラリの情報を用いてフィルタリングする．たとえば,「これ」「それ」などの指示詞で指される対象は，ユーザの視界に入っていることが多いので，このような場合はユーザの視界を考慮する．

ある発話に対応するプラン演算子の効果が別の発話のプラン演算子の前提条件になっている場合は，これらは共通のプランの部分プランである可能性が高い．プラン演算子の効果と前提条件の連鎖の関係を考慮して，プラン演算子のグループを作り，それに対応する発話の中に現われる照応・省略はそのグループ内で先行詞を探す．

一般にはプラン演算子の前提条件は複数の命題の連言で記述されることがあるが，前提条件中のすべての命題が別のひとつのプラン演算子の効果によって成立することはまれなので，プラン演算子の連鎖を探すときには，あるプラン演算子の効果が他のプラン演算子の前提条件中の命題の最低ひとつを満足するという条件で連鎖を作る．

フィルタリング後も複数の先行詞の候補が残る可能性があるが，その場合は以下の要因を考慮してスコアを付ける．

- 対象の顕現性
- エージェントの視界
- ユーザの視界

顕現性は対象が最近どれくらいの頻度で言及されたか，エージェントとユーザの視界は，視線の方向と距離を考慮する．現在の実装では，これらの要因の組み合わせは経験的におこなっている．

4 評価実験

本稿で述べた手法を評価するために，同じ対象領域について Wizard of Oz 方式で収集した日本語の音声対話コーパスを用いて実験をおこなった．コーパスは 15 対話を含み，合計で 532 発話からなる．各発話は統語解析し，人手によって格フレームに変換されている．ここで用いた格は動作主，対象，始点，終点，方向，程度，様体の 7 つである．表 1 にそれぞれの格の照応・省略の数を示す．

表 1: 照応・省略の分布

格	動作主	対象	始点	終点	方向	程度	様体
照応	0	30	0	5	1	1	0
省略	287	183	6	65	30	0	3

次にこのコーパスに 2 種類のセンタリング理論を人手によって適用した．ひとつは BFP 版と呼ばれるもの [1, 11]，もうひとつは機能センタリング理論 (FC) [9] である．誤りの連鎖を防ぐため，直前の発話は常に正しく処理されていると仮定してアルゴリズムを適用した．

提案手法については以下のプランを仮定してアルゴリズムを人手によって適用した．

- ある位置を見る
- オブジェクトを見る
- ある位置に移動する
- ある位置に近くに移動する
- オブジェクトの近くに移動する
- オブジェクトを押す
- オブジェクトを写す (カメラのみ)

表 1 からわかるとおり動作主と対象における照応・省略が顕著なので，この 2 つの格について調査した．

表 2: 照応解消の結果

	格	BFP	FC	PB	合計
照応	動作主	0	0	0	0
	対象	21	22	29	30
省略	動作主	277	275	284	287
	対象	141	149	162	183

表 2 に各アルゴリズムの適用結果を示す．ここで PB は提案手法である．いくつかの例についてはセンタリング理論 (BFP，FC) がうまく適用できなかった．主な原因として以下のようなものがあった．

- 先行詞が直前の発話に現われない
- 先行詞の候補に優先順位がつかない
- 優先順位の予測が正しくない

これらの中でも最初の場合が一番多かった．たとえば，以下のような例を考えよう．

(4-1) 馬は赤い球のところへ行け.

(4-2) (馬は) もう少し前へ (行け).

(4-3) (馬は) (赤い球を) 前へ押せ.

この例では発話 (4-3) で省略された「赤い球」は直前の発話 (4-2) には出現しないので，センタリング理論ではうまく扱えない．我々の手法では，この例は図 3 の左の枝分れによって処理される．つまり，システムはプラン・ライブラリを検索し,「オブジェクトを押すためには，エージェントその十分近くにいる必要がある」という関係を見つけだし，省略の対象を「赤い球」であると同定する．

コーパス中には提案手法でも解決できない例がいくつかあった．その多くは,「青い球と赤い球が両方見えるように写して」などのように，ユーザが自分のゴールを目的とする状態を述べることによって命令する場合であった．この場合,「写す」の対象は上記のような条件を見たすシーンであるが，このような省略を補完するためにはより深い推論が必要となる．

5 結論

本稿では音声対話におけるプランに基づく照応解消の手法について述べた．基本的な考え方は，ユーザの発話の背後にある部分プランを同定し，プランの効果と前提条件の関係を利用し，これらの部分プランを同一のゴールを共有するグループにまとめ，照応・省略の先行詞の検索対象範囲を同一の発話グループに限定するというものである．Wizard of Oz 方式によって収集した音声対話コーパスを用いて，提案手法とセンタリング理論を評価した結果，センタリング理論ではうまく扱えない現象も提案手法で扱える場合があることがわかった．

参考文献

[1] S. E. Brenann, M. Walker, W. Friedman, and C. J. Pollard. A centering appoach to pronouns. In *Proceedings of ACL'87*, pp. 155–162, 1987.

[2] D. Byron and A. Stent. A preliminary model of centering in dialog. In *Proceedings of ACL'98*, pp. 1475–1477, 1998.

[3] P. R. Cohen. The pragmatics of referring and the modality of communication. *Computational Linguistics*, Vol. 10, No. 2, pp. 97–146, 1984.

[4] R. E. Fikes. STRIPS: A new approach to the application of theorem problem solving. *Artificial Intelligence*, Vol. 2, pp. 189–208, 1971.

[5] B. J. Grosz, A. K. Joshi Joshi, and S. Weinstein. Providing a unified account of definite noun phrases in discourse. In *Proceedings of ACL'83*, pp. 44–49, 1983.

[6] B. J. Grosz, A. K. Joshi, and P. Weinstein. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, Vol. 21, No. 2, pp. 203–226, 1995.

[7] B. J. Grosz and C. L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, Vol. 12, No. 3, pp. 175–204, 1986.

[8] Y. Shinyama, T. Tokunaga, and H. Tanaka. Kairai – Software robots understanding natural language. In *Third International Workshop on Human-Computer Conversation*, pp. 158–163, 2000.

[9] M. Strube and U. Hahn. Functional centering – Grounding referential coherence in information structure. *Computational Linguistics*, Vol. 25, No. 3, pp. 309–344, 1999.

[10] M. A. Walker, A. K. Joshi, and E. F. Prince, editors. *Centering Theory in Discourse*. Clarendon Press Oxford, 1998.

[11] M. Walker, M. Iida, and S. Cote. Japanese discourse and the processing of centering. *Computational Linguistics*, Vol. 20, No. 2, pp. 193–232, 1994.

[12] T. Winograd. *Understanding Natural Language*. Academic Press, 1972.

[13] 志賀聡子, 徳永健伸, 田中穂積. 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.