

マルチモーダル対話における対象物の同定について

4 R-5

加藤 恒昭 中野 有紀子

NTT情報通信研究所

1 はじめに

言語表現や指差し等によって対象物を指示することでそれを聞き手に同定させるという行為が、参照物同定要求である。物理的対象の操作等を題材とし、言語情報と視覚情報の両方が伝達されるマルチモーダル対話においては、参照物同定要求が特に重要な役割を果たすと共に、様々なバラエティを示すこととなる。本稿では、マルチモーダル対話での参照物同定要求における、適切な情報の種類と量の選択に関する実験結果を報告し、その結果に基づき、マルチモーダル対話システムにおける参照物同定要求生成の枠組みについて考察する。

2 対話データ

今回利用した対話データはマルチモーダル対話システム設計のための基礎データ取得を目的に収録したものである。タスクは留守番電話機の設置の教示で、初心者から専門家からの教示を受けながら組み立てや各種設定を実際に行ってゆく。専門家も初心者同様、自分の前の電話機を操作することができる。対話は音声対話状況SDとマルチモーダル対話状況MDの2状況で収録した（全体の収録は独白も含めて4状況）。音声状況SDでは、ヘッドセットを用いて音声チャンネルのみで説明が行われる。マルチモーダル状況MDでは、音声に加えて専門家側の電話機への指差しという視覚情報の伝達が可能である。この状況ではビデオカメラとディスプレイを組み合わせることによって、専門家側電話機とそこへの両者の指差しが完全に共有されている。これは、画面出力と音声言語出力を持ち、ユーザからのポインティングと音声入力を受け付けるマルチモーダル対話システム操作の模擬となっている。

3 参照物同定要求について

今回の調査では専門家5人の各状況3説明、計30説明を対象とし、音量設定、時計合わせ、応答メッセージ録音等の各種初期設定を行なっている部分に現れる参照物同定要求を取り出した。この部分での参照物同定要求は、登録ボタンを同定しそれを押す等、対象物を同定し、それに対してある行為を行うという文脈[2]で行われる。以下に、抽出された

部分の例を示す。記号//は初心者のあいづち等を、[]の部分はそれに合わせて指差し等の行為があったことを示す。

状況SD（音声対話状況）

ではまず、あの、カセットカバーをあけてください。
// そうすると、ボタンが並んでると思いますけれども//その左から2番目、登録ボタンを押してください。

状況MD（マルチモーダル対話状況）

[まずカセットカバーを開けてください。(指-本体、指-カセットカバー、開ける)]登録ボタンを押します。

抽出・注釈付けされた参照物同定要求について、参照物についてのどのような情報がどの程度の量、専門家から提供されたかを調査した。ここでは情報を極めて広い意味で捉え、以下の様に分類した。
代名詞 「これ」「それ」等の代名詞、なお、状況MDでのコ系の代名詞はほぼ常に指差しを伴う、deicticである。

文脈参照「以前使った」「先程の」等、談話中で対象物が既に同定されたことがあるという記述

一般名称「ボタン」「スイッチ」等、対象物の種類を示す概念で、電話機内でその対象を一意化するのには不十分なもの

固有名称「登録ボタン」等、電話機内でその対象を一意化するのに十分な概念

形状「丸い」「赤い」「大きい」等、対象物の形状、大きさ、色についての記述

位置「本体右上の」等、対象物がどこにあるかという位置に関する記述

紋様「留守と書かれた」等、対象物やその近傍に記入されたマークや文字列等についての記述

機能「音量を調節する」等、対象物が果たす役割、機能についての記述

関連物 対象物が属す集団やその近傍の事物についての言及

その他 上記のいずれにも分類できない記述

指差し 対象物への指差し

3.1 初回同定

談話的要因の影響を受けない参照物同定要求のモード依存性を明らかにするために、ある対象物が最初に談話中に導入される際の同定要求について調査した。12の対象物が導入された。結果を表1に

Referent Identification Requests in Multi-Modal Dialogs
Tsuneaki Kato and Yukiko I. Nakano
NTT Information and Communication Systems Labs.
1-2356 Take, Yokosuka-shi, Kanagawa 238-03, Japan

表1 初回同定における利用情報数の分布

状況	代名詞	一般名称	固有名	形状	位置	文字模様	機能	関連物	その他	指差し
SD	1.01	1.45	0.68	0.34	1.08	0.70	0.11	0.27	0.18	-
MD	0.99	1.24	0.78	0.12	1.06	0.44	0.12	0.11	0.10	0.98

表2 同一談話セグメント内第2回、新談話セグメント内第3回同定における利用情報数の分布

(1) 第2回同定						(2) 第3回同定							
状況	代名詞	一般名称	固有名称	位置	指差し	状況	代名詞	文脈参照	一般名称	固有名称	位置	文字模様	指差し
SD	0.06	0.17	0.89	0.02	-	SD	0.05	0.10	0.17	1.03	0.23	0.05	-
MD	0.25	0.15	0.81	0.08	0.81	MD	0.27	0.12	0.10	1.03	0.07	0.03	0.82

示す。各セルの値は一回の同定に用いられた情報の量の平均である。以下の傾向が観察された。

- ・状況MDではdeicticの代名詞を伴った指差しが頻繁に用いられる。
- ・状況MDでは一部の言語的な情報量が減少する。形状、紋様、関連物においてその差は顕著である(順に、 $F(1, 55) = 15.91$, $P < 0.001$, $F(1, 55) = 28.96$, $P < 0.001$, $F(1, 55) = 11.48$, $P = 0.001$)。位置についても差があるがさほど顕著ではない($F(1, 55) = 5.193$, $P = 0.027$)。

3.2 同定と談話的要因

談話セグメントとの関連等、談話的要因によって参照物同定要求がどのように変化してゆくかを調査した。対話中で4つの対象物が、初回同定の後、同じ項目設定の中でもう一度、更に異なる項目の設定の中で同定されている。タスクの構造と談話構造との関連から、第1回、第2回と第3回とは異なる談話セグメントに属していると考えられる。表2に第2回、第3回の同定で利用された情報を示す。全ての状況で0になった項目は除いてある。以下の傾向が観察された。

- ・同定要求に利用される言語的情報量は2回目以降、劇的に減少する。deicticな代名詞の利用は減少するが、指差し自体の減少は他に較べると顕著でない。
- ・談話セグメントの境界を挟むと、同定要求に利用される言語的情報量はやや増加する。状況MDにおいては文脈指示が、状況SDにおいては位置情報が増加するようである($F(1, 15) = 4.29$, $P = 0.06$, $F(1, 15) = 6.82$, $P = 0.02$)。

4 議論

Appelt[1]によれば、参照物同定要求とは聞き手の心的状態においてある概念を活性化させることであり、そのために、その概念を他と区別して一意化す

るために必要十分な情報を与えなければならない。今回の実験から、この枠組みを伝達モードの観点で拡張した次の様なモデルが示唆される。つまり、各種の情報、形状や色についての言語的な記述、指差し、文脈的想起等を含んだ広い意味での情報が、それを伝達するモードと結び付けられており、それぞれに、それが対象を活性化する効果とそれを利用するためのコストが与えられている。発話者は、これらの情報を適当に組み合わせ、対象物を活性化するのに必要十分な効果を得る。

効果とコストという観点で比較できる情報の選択を考えることでモード間の依存性が導ける。つまり、視覚的な情報が利用できれば、それに応じて言語的な情報は減少する。興味深いのは、指差しが利用できるということが、一定の言語的情報の減少に繋がっていないことである。先に伝達すべき情報が決定され、その後にそれをどのモードで伝達するかを決定するというモデル[3]では、同じ指差しが状況によって様々な役割を持つことになってしまい、この結果をうまく説明できない。モードと結び付いた情報を選択する過程を考える方がより簡単に指差し等の役割を扱うことができる。更に談話構造中での対象物の位置とそれを活性化するために必要となる効果との関係を考えれば、既に導入された対象物を再び同定する際に必要な情報が変化することも説明できる。

参考文献

- [1] Appelt, D. E. "Planning English Referring Expressions" Artificial Intelligence 26, 1985 pp. 1 - 33
- [2] Cohen, P. R. "The Pragmatics of Referring and the Modality of Communication" Computational Linguistics Vol. 10 No. 2 pp. 97 - 146 1984
- [3] Feiner, S. K. and McKeown, K. R. "Coordinating Text and Graphics in Explanation Generation" in Procs. AAAI -90 pp. 442 - 449 1990