マルチモーダル対話における参照表現パタンの分析

加藤 恒昭　　中野 有紀子
NTT 情報通信研究所

1. はじめに

参照行為もしくは参照物同定要求とは、名詞句等で言語的に表現したり指差し等を用いて直示することを通じて、意図した対象物の心的表現を聞き手に持たせようとする話し手の行為であり、そのための言語表現は参照表現と呼ばれる。人間どうしの対話における参照表現がどのような特徴を有しているか [Levelt89, Cohen84]、参照行為を生成する機構はどのようなものであるか [Appelt87, Dale95] 等、参照行為に関する様々な研究が続けられている。我々は、マルチモーダル説明システム[Kato96] への実装を目指し、音声言語と指差し等の動作を組み合わせた効果的な参照行為を生成する枠組みについて検討を続けている。その第一歩として、音声だけが伝達できる状況（以下、音声状況）と音声と指差し等の動作に関する情報が伝達できるマルチモーダル状況（以下、MM状況）とで、人間どうしの教示対話を収録し、その差に注目することで、参照行為に対する対話状況の影響について、経験論的な研究を行ってきた[Kato97]。本稿では、このふたつの状況で観察される参照表現を、参照される対象物の特徴と関係づけることで整理し、参照表現のパタンの抽出を試み、その決定要因について考察する。まず、第2章で、対話コーパスの分析を通じて得られた参照表現パタンとその出現頻度について報告する。第3章では、従来研究を参考に、それらのパタンを分析する。第4章で今後の課題を述べてまとめる。

2. 参照表現のパタンとその出現頻度

2.1 コーパスと対象物の分類

解析の対象とした対話コーパスは、文献[Kato97]で用いたものと同じで、図1に示す留守番電話機能付きの電話機の初期設置の教示を対象としている。電話機の組み立てと各種初期設定の過程で、ボタン等の対象物が参照される。その際に用いられた表現を分析した。分析においては、同じ5人の専門家がそれぞれ異なる初心者に対して、音声状況とMM状況とで行った説明各3回ずつを解析した。つまり、各状況で各対象物について15回ずつの参照表現を解析している。ここで解析した参照表現は、ある対象物を同定させて、それに対する何らかの行為を行わせるという文脈でなされたもので、その対象物が初めて説明対話に導入されるという初回参照の際のものだけを選んでいる。

対象物は、その場所、色、形状、大きさ、マーク、種類等、様々な特徴で分類されるが、予備的な分析の結果、次の2つの特徴が、参照表現に大きく影響を与えていることが明らかとなった。以下では、これらに沿って対象物を分類し、そこで使われる参照表現のパタンをまとめる。

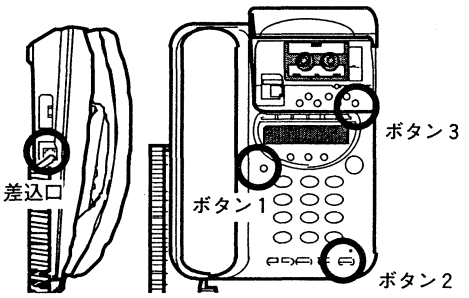

図1　教示の材料となった電話機

可視であるか：ある種の対象物は電話機の側面や背面にあって、電話機本体を回転する、覗き込む等しないとそれを見ることができない。また、カバーの内側にあって、そのカバーを開かない限り見ることができない対象物もある。これらを不可視な対象物と分類し、電話機正面にある可視であるものと区別する*。

グループのメンバであるか：同じ形状をしたボタンが一列に並ぶ等して、グループを構成しているように認知される場合がある。ある対象物が、このようなグループのメンバであるか、それともそれひとつで孤立しているかを区別する。

例えば、図1において、ボタン1は、可視で孤立した対象物である。側面の差込口は不可視で孤立した対象物である。ボタン2は、可視のグループメンバである。ボタン3は、図では開いているがカバーの内側なので、普段は不可視なグループメンバである。

2.2 可視で孤立した対象物の参照

可視で孤立した対象物を参照する表現として、以下の2つのパタン（ [] は、［意味内容 統語カテゴリ］で示される要素を表現する。()内の要素は省略可能である）が頻繁に観察された。

NP：［対象物の記述 NP］（、［代名詞］）を［行為 V］してください。
EX：［場所 NP］に［対象物の記述 NP］があります。［代名詞］を［行為 V］してください。

* この特徴は対象物に固有ではなくその対象物が参照される時点での電話機の状態に依存する。

ここで、［対象物の記述NP］は、対象物の場所や大きさや色等の特徴の表現によって修飾され、その主辞が対象物が属すクラスを表現する一般名詞であるような名詞句である。

第一のパタンNPは、行為の依頼を行う命令文の中の対象の表現に、対象物を同定させるための表現が含まれるものである。第二のパタンEXは２つの発話からなり、最初の発話は、ある場所にその対象物が存在することを述べるもので、場所の情報を運ぶ後置詞句が最初に来ている。それに行為の依頼を行う命令文が続く。パタンによって［対象物の記述］は若干異なり、NPの場合、対象物の場所が最初に述べられ、その後にその他の特徴が続けられることが殆どである。EXの場合、場所の情報は含まれない。それぞれの例を示す*。

NP：ダイヤルボタンの//左上にある受話器のマークがついたボタン、//これを押してみてください。
EX：こちら数字ボタン１の、//隣に、受話器のマークのついたボタンありますよね。//これを押してみてください。

もうひとつのパタンとして、参照表現中に現われるランドマーク（上例のダイヤルボタンや数字ボタン１）の参照が発話として分節されるものがある。このパタンをLM+EXと呼ぶ。以下に例を示す。

ダイヤルボタンの１ありますね。//その左に丸いスイッチあると思うんですが、//それを押してみてください。

以上のパタンの出現頻度を表１に示す。コーパス中で２つの対象物が可視で孤立した対象物であったので、各状況での合計は３０である。MM状況ではパタンNPが、音声状況ではパタンEXが多く見られる。

2.3 不可視な対象物とグループメンバの参照

不可視で孤立した対象物は、コーパス中に５つ現われたが、これらは全て正面以外の面にあるスイッチや差込口であった。前節で述べた２つのパタンNP, EXにおいて、場所の情報としてどの面に存在するかを述べることで、これらを参照することが可能である。例えば、以下のような表現になる。

表１　可視で孤立した対象物の参照パタン
出現数（括弧内は総数に対する%）

	MM状況	音声状況
NP	16 (53)	5 (17)
EX	9 (30)	18 (60)
LM+EX	0	5 (17)
Others	5 (17)	2 (6)

NP：～を、本体の左側の差し込み口に入れてください。
EX：～を、本体の左の部分に、//差し込み口があると思いますので、//そちらの方に爪を合わせて入れてみてください。

また、先のパタンLM+EXのように、ランドマークや対象物が存在する面の参照だけでひとつの発話とするものも観察された。面等の参照は、その面を見ることやその面を向けることを依頼する命令文や、その面を含む copula によって行われる。例えば、以下のような発話である。

本体の左の側面を見てください。//その真中へんに～本体の左側面ですね。//そこに～

これらの後にはパタンEXだけではなく、パタンNPが続くこともある。表２にこれらの出現頻度をまとめる。各状況での合計は７５である。MM状況ではパタンNPが、音声状況ではパタンLM+EXが多い。

可視のグループメンバについては、それを参照する場合、まずグループ全体を参照し、その後にそのメンバである対象物を参照するというパタンが観察される。グループ全体の参照は多くの場合、パタンEXの第一発話と同様に行われる。例えば、

ダイヤルボタンの下の方に、本体と同じ色をしたボタンが５つ並んでいると思いますが。

その後、パタンNPやパタンEXで対象物が参照される。これらのパタンをそれぞれGR+NP, GR+EXと呼ぶ。例えば、GR+EXでは以下のような発話が続く。

その一番右にスピーカのマークがついたボタンがありますね。//…

一方、グループの参照を分節しないNP, EXのパタンも観察される。例えば、

一番下のボタンの列の//右端、スピーカのマークのついているボタンがありますね。//…

表２　不可視で孤立した対象物の参照パタン

	MM状況	音声状況
NP	22 (29)	7 (9)
EX	13 (17)	7 (9)
LM+NP	6 (8)	3 (4)
LM+EX	7 (9)	37 (49)
Others	27 (36)	21 (28)

* 今回の調査では、多様な表現をパタンとしてまとめるために、言い直しや言い淀み、助詞の有無等を捨象して分類を試みている。実際の発話はここに示すものよりはるかに多様で、あるパタンに属するかが不明確なものも少なくない。また、以後の議論で「ひとつの発話」「発話として分節する」等の表現を使うが、これについても、何をもって一発話もしくは発話の区切りとするかという問題が残る。例えば、「～に～がありますので、それを～」をひとつの発話としてNPに分類するか、前半が統語的に存在文の形式であることを重視してEXとするかも、議論の対象となる。ちなみに今回は後者の立場をとった。これらの問題と関連するが、本稿で示した例には若干の編集を施してある。なお、句読点はポーズやイントネーションの情報を基に付与した。//はそこに初心者の相槌が入ったことを示している。

この例では、グループの参照が分節されていないとはいえ、対象物がグループのメンバであることが伝達され、位置の情報としてそのグループ内の相対位置もしくはグループメンバであることが用いられているが、グループと関連したこれらの言及が全くない以下のような参照のバタンも観察される。

本体の右下、// スピーカボタンを押してみてください。

これらを区別して、グループに関する言及があるものをNPG, EXG、ないものをNPD, EXDと呼ぶ。

不可視のグループメンバについては、コーパス中に現われたのはふたつで、共に本体右上のカセットカバーの内側に並んだボタンであった。その参照においては、全てのデータで、最初にカセットカバーを開けさせる依頼があった。その後の参照表現では、可視のグループメンバの参照と同じバタンが観察された。

可視のグループメンバの参照表現バタンと、依頼の部分に続く不可視のグループメンバの参照表現バタンの出現頻度を表3に示す。前者が4対象物、後者が2対象物であったので、各状況での合計は９０である。MM状況ではグループに関する言及のないNPD, EXDが、音声状況ではグループの参照を分節するGR+*が多く見られる。

3．参照表現バタンを決定する要因

3.1　全体的傾向との関連

我々は参照表現が対象物のどのような特徴に言及しているか、参照がどのような形式で行われているかに着目し、それらへの対話状況の影響が以下のようにまとめられることを示した[Kato97]。

P1：MM状況では参照行為において指差しが頻繁に用いられ、言語的に伝達される情報の量は音声状況より少なくなる。初回参照において見ると、場所に関する情報には有意な差がないが、形状、マーク、グループ等に関する情報の量はMM状況で有意に減少している。

P2：音声状況では、参照行為の達成を独立したゴールとすることが多く、更にそれを細かい副ゴールの列とし、それらを順次、確認しながら達成していく形で参照行為が進めてられていく。

今回の結果はこれらの知見を説明するものである。まず、MM状況と較べた場合、音声状況では参照を存在文で行い、行為の依頼を発話として分節するバタンEXが、両者をひとつの発話で行うバタンNPより多く、加えてランドマークやグループの参照を分節して行うLM+*, GR+*のバタンも多い。これらがP2を導いている。また、バタンEXでは常に場所情報の伝達が行われ、バタンNPでも［対象物の記述］の先頭に場所に関する情報が来ることが多い。これについてはほぼ固定的で、参照のために必要な情報の量（これは利用可能なモードと関連して対話状況によって異なるはずである）はそれ以外の［対象物の記述］の部分で調節されており、これがP1を導いていると考えられる。更に、MM状況ではグループメンバを参照する際にそれが属すグループを介さないことも多く、これによりグループ等の情報が減っている。この点もP1に寄与することになる。

表3　グループメンバの参照バタン

	MM状況	音声状況
NPD	29 (32)	8 (9)
EXD	22 (24)	14 (16)
NPG	3 (3)	11 (12)
EXG	6 (7)	11 (12)
GR+NP	11 (12)	22 (24)
GR+EX	9 (10)	16 (18)
Others	10 (11)	8 (9)

3.2　参照表現を構成する情報

文献[Dale95]では、参照行為において、以下の２種類の情報を分離して考えることが重要であると述べられている。

誘導情報：同定させたい対象物を含む集合（文脈集合）に聞き手の注視の焦点を向けるための情報
識別情報：同定させたい対象物を文脈集合の他のすべての対象物から区別させるための情報

更に、少なくとも物理的な対象については場所情報は誘導情報の典型的な形式であるとされている。前章で示した参照表現のバタンは、日常的な場面における参照表現においては、誘導情報と識別情報の組み合わせと分節の仕方に様々なバラエティがあり、対話状況の影響がそれらの決定に一定の役割を果たしていることを示唆している。例えば、可視の孤立対象物に関する次の参照表現を考えてみる。

液晶画面の右下の方に留守と書いている大きなボタ
¦--- (1) ---¦------ (2) ------¦---------- (3) ----------
ンありますね。それを押してください。
---¦ ¦---------- (4) ----------¦

(1)+(2) が広義の誘導情報、(3) が識別情報を与えていて、両者で参照表現を構成している。(4)は行為を依頼している部分である。更に (1)+(2) は、識別情報のみからなるランドマーク（この場合は液晶画面）の参照(1) と聞き手の焦点をそこから対象物の場所へと誘導する（狭義の）誘導情報(2) からなっている。誘導情報は常にこのような構成を持つわけでなく、「一番下」「カバーを開けてください」等も誘導情報であると考えられる。ここでは、これら単独で誘導情報なるものと、(2) のようなランドマークの参照と組み合わせて誘導情報となるものとを合わせて誘導情報と分類する。

このように捉えると、可視の孤立対象物に関するそれぞれの参照表現バタンは分節化の違いとして以下のように特徴づけられる。

NP：(1) - (4) 全部がひとつの発話に含まれる（「留守と書いている大きなボタンを押してください」のように (3) (4)のみで参照と行為の依頼が行われることもある）。

EX：(1) - (3) がひとつの発話で提示され、(4)はそれとは別の発話で行われる。

LM+EX：(4) がひとつの発話となることに加えて、(1) - (3)においても、(1)もしくは (1) - (2) が (3)とは異なる発話で提示される。「液晶画面があります。その右下の方に…」と「液晶画面の右下を見てください。そこに…」がそれぞれの例である。

前章の分類を基に、その他の対象物の場合も含めて、参照表現のパタンを遷移ネットワークの形でまとめたものを図 2 に示す。nav, discはそれぞれ誘導情報、識別情報の提示を示している。REFはこのネットワークの参照表現全体を示し、これにより再帰的な定義となっている。actは行為の依頼である。上述の参照表現パタンのバラエティは、図の最上縁を通るパスでの分節の違いに相当する。グループメンバにおけるパタン間の違いのうち、NPD, EXDとそれ以外のものとの違いは、前者が図の上半分のパスを通り、後者が下半分を通ることにある。

このネットワークにおいて、提示される情報（どのパスを通るか）と分節化（どの部分をひとつの発話とするか）における対話状況間での違いを図 3 に示す。情報の提示や分節化に関する幾つか選択点（これらの選択の結果、前章で示したいずれかパタンとして参照表現が決定される）において、情報の少ない方、分節化を行わない方が選ばれた割合を対象物ごとに算出し、状況間で比較するために散布図としている。殆どの点が左上にあることから明らかなように、さまざまな点で、MM状況ではより少ない情報、より少ない分節化が好まれている。

4．おわりに

本稿では、MM状況と音声状況で行われた教示対話に現われる参照表現を分析し、対象物を可視であるか、グループのメンバであるかという観点で分類することでそれらが幾つかのパタンに分類できることを示した。このパタンとその対話状況毎の出現頻度は、従来研究で得られた知見を裏付けるものである。更に参照表現を、そこで提示される誘導情報と識別情報の組み合わせとその分節という観点で捉え、パタンの一般

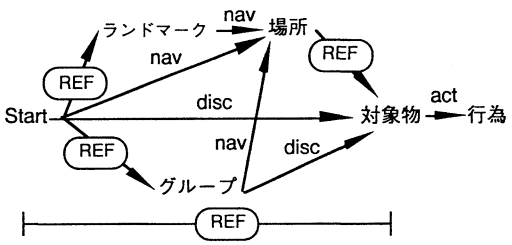

図 2　参照表現生成の遷移ネットワーク

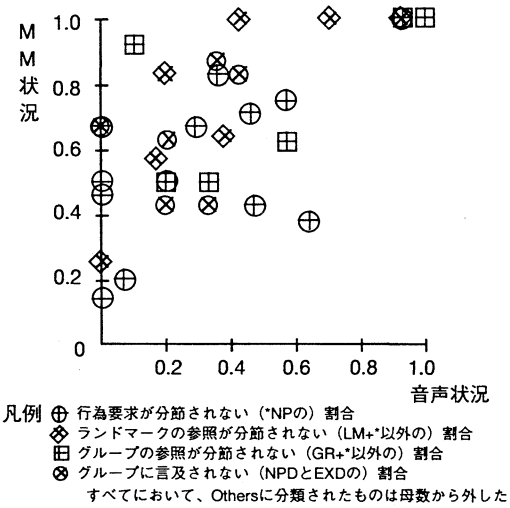

図 3　提示情報と分節化の観点での対話状況間比較

化を行い、その観点での対話状況の影響を示した。

同じ対話状況、ほぼ同じ文脈で発せられる同じ対象物への参照表現が一意に定まらないことから、観察に基づく経験論的研究だけから参照表現生成アルゴリズムを抽出することが困難であることは明らかである。しかし、これらの分析はアルゴリズム化において極めて有益な指針を与えると考えている。今後は、評価実験等、他の方法論と組み合わせてアルゴリズム化を進めていきたい。また、従来は識別情報についての研究が主で、誘導情報として何が適切かについての議論が少なかったが、この点については、文献[Herskovits97]で示されたFigureとGroundの考えを基に整理していくことを考えている。

参考文献

[Appelt87] Appelt, D. and Kronfeld, A. "A Computational Model of Referring" in IJCAI 87, pp. 640 - 647 (1987)

[Cohen84] Cohen, P.R. "The Pragmatics of Referring and the Modality of Communication", Computational Linguistics, Vol. 10, No. 2, pp. 97 - 146 (1984)

[Dale95] Dale, R. and Reiter, E. "Computational Interpretations for the Gricean Maxims in the Generation of Referring Expressions" in Cognitive Science, 19(2), pp. 233 - 263 (1995)

[Herskovits97] Herskovits, A. "Language, Spatial Cognition, and Vision" in Spatial and Temporal Reasoning eds. Stock, O. ch. 6 pp. 155 - 202 (1997) Kluwer Academic

[Kato96] Kato, T., Nakano, Y.I., Nakajima, H., and Hasegawa, T. "Interactive Multimodal Explanations and their Temporal Coordination" in ECAI-96, pp. 261 - 265 (1996)

[Kato97] 加藤 and 中野 "マルチモーダル対話における参照行為" 人工知能学会 Vol.12, No.4, pp. 627 - 634 (1997)

[Levelt89] Levelt, W.J.M. "Speaking: From Intention to Articulation" Ch 4.3.3 pp. 129 - 134 (1989) ACL-MIT Press