

マルチモーダル対話環境における指示照応

中川 裕康 田村 直良

横浜国立大学 工学部 電子情報工学科
{yass,tam}@tamlab.dnj.ynu.ac.jp

瀬川 修

中部電力株式会社 電力技術研究所
sega@rd.chuden.co.jp

1 はじめに

本研究は、マルチモーダル対話環境下での指示照応について、対話実験を通して調査を行ない検討するものである。人間の本来のコミュニケーション形態であるマルチモーダル対話をインタフェース上でも実現することが望まれているが、そのためにはモダリティ個々の特徴や相互の影響を調査・考察する必要がある。そこで本研究は、マルチモーダル化の影響が顕著に反映すると考えられる指示照応について、コ系、ソ系、ア系指示詞・指示代名詞およびゼスチャを対象とし、対話収集実験を行ない、照応表現の産出・理解過程をモデル化した。

2 マルチモーダル対話収集実験

2.1 環境と実現

対話実験の目的は、マルチモーダル対話環境での指示照応の機構を調べるために、マルチモーダル対話文を収集することである。

本研究で対象とするマルチモーダル対話とは、二人の対話者が、あるものや風景を共有し、その環境のなかへ指示照応を行なうような場面を想定する。例としては、机上に地図を広げて道案内をすることがある。対話には自然言語のみではなく、地図への直接のポインティングゼスチャなども併用され、マルチモーダル対話がなされることになる。本実験ではこのようなマルチモーダル対話の環境を、二つの対話インタフェースをネットワークを通して互いに接続し、この対話インタフェースを通して二者が対話することで実現する(図1)。一つの対

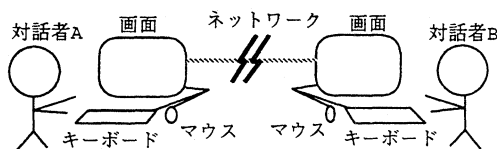


図1: マルチモーダル対話実験環境の構築

話インタフェースは、キーボード・マウス・ビットマップディスプレイから構成され、キーボードでは言語入力を、マウスでゼスチャの入力を受け付ける。ディスプレイにはキーボードおよびマウスからの入力を反映した結果つまり、静止画像および互いのキーボード入力の履歴と、マウス入力による静止画像上への軌跡描画の結果が表示される。このディスプレイの画像は双方の対話インタフェース上で共有される。マウスによる軌跡描画はひと筆書きで、書いた動作がそのまま即時にディスプレイに反映される。一方、キーボード入力の会話文は、送信ボタンを押さない限り通信相手のディスプレイには表示されない(図2)。このインタフェースによってマルチ

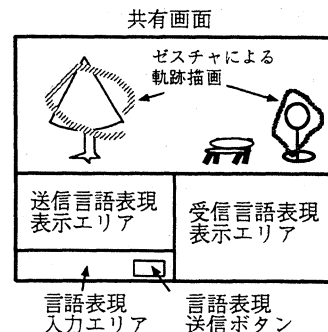


図2: マルチモーダル対話実験環境の画面構成

モーダル対話環境を実現し、同時に対話データを収集する。対話タスクはつぎのように設定した。二人の対話者をそれぞれAおよびBと記す。

● 人物紹介タスク

Aにとって既知、Bにとって未知の人々の集合写真を共有画面に表示、AがBの質問に答えながら人物を紹介してゆく。

● 地図完成タスク

AがBそれぞれ完成した地図と不完全な地図、または、相互に補間するような地図を共有画面に表示し、互いに質問を繰り返して地図を完成する。

マウスによるゼスチャ入力の効果調べるために、マウスの使用を制限した場合についても実験を行なった。また、キーボード入力による会話文入力では、使用単語の制限は設けずに対話者の自由な発話に任せた。対話の交替は、キーボード入力の送信によって行なわれる。

3 実験結果

実験によって、マルチモーダルな表現による対話文を収集することができた。以降は指示表現のうち自然言語による部分を言語表現、マウスのゼスチャによる部分を単にゼスチャと呼ぶ。

ポインティングゼスチャを併用する場合としない場合の言語表現の全体量を比べると、ゼスチャを併用した場合は言語表現の量が非常に少なくなる。

- ゼスチャ無し

「立っている人の中で一番左にいる髪の長い人はどうですか」

- ゼスチャ有り

「この人は誰ですか」（指示する人物の顔の外周を囲う）

ゼスチャを用いない場合の指示表現は指示対象の特徴の記述となり、一方でゼスチャを併用する場合は、ゼスチャが指示対象そのものを指定し、言語表現はそのゼスチャに対する参照を用いた表現となり、その参照は主に指示代名詞によってなされる。

指示表現はその中に相対位置の表現を伴うことが多い。その方向はある基準に基づいて決定され、基準には状況によってさまざまに変化するデフォルトが存在する。対話者の間でこの基準が異なっていれば、対話によるネゴシエーションが行なわれ、基準が統一される。

対話で最初に行なわれる指示の指示対象には、他と比べて顕著な特徴を持っているものが選ばれる。そして次の指示対象は、最初の指示対象を基準にして以降、連続的に選ばれてゆく。実験環境の制約から言語表現とそれに伴うゼスチャは同時ではなく前後するが、それらのどちらが先であってもそれぞれの表現に変化はなく、対話者は対話を進めることができた。

3.1 言語表現の特徴

対話文の言語表現では、積極的に省略を行なっただけ表現を短くする傾向が見られた。

用いられる指示詞はほとんどがコ系・ソ系で、ア系は出現しなかった。また、対話相手が指示したものをコ系で指示することも観測された。

ゼスチャの使用を制限した場合には自然言語表現のみで指示対象を表現しなければならず、その表現は長いものとなった。それでもやはりできるだけ短く表現しようと努力する傾向があり、指示対象を指示するのに必要最小限と思われる表現がなされる。

3.2 ゼスチャの特徴

マウスゼスチャを伴う指示表現においては、ゼスチャの形状はその指し示す対象の形に依存する。ゼスチャの形状による分類を表1に、ゼスチャの指し示す対象別の分類を表2にそれぞれ示す。

ゼスチャ形状名	形状と意味のある部分
ライン	対象の形状に沿って直線または曲線を引く。始点と終点は離れている。意味のある部分は、始点・終点・始点と終点の間、始点から終点に向かう方向、直線または曲線によって分けられる二つの部分。
サークリング	曲線または複数の直線を用いてある対象を囲む。線は必ずしも閉じていないが、その場合に始点と終点は接近している。意味のある部分は線によって囲まれている内部。
矢印	ひと筆書きによる矢印。矢印の指す先の点または方向。
連続アーク	複数の点を連続的に結ぶ。始点と途中の点および終点。
併用	ゼスチャと、すでに画面内にある直線または曲線を併用したもの。それらによって、サークリングと同じ効果を持つ曲線を作る。

表 1: ゼスチャ形状分類

モダリティの点からみると、ゼスチャはいつも指示表現全体の最終的な指示対象を指示するためだけに用いられているわけではない。検索対象の検索空間を限定するための範囲指定としても用いられ、図3のように分類できる。ラインによる範囲指定には、「この線上のどこかに」「この線の上に」「この線の右側に」などの表現がある。ゼスチャによって限定するための範囲が表現されている場合、ゼスチャのみでは指示表現の同定は不可能で、それに伴う対話文も考慮した推論が必要になる。

指示対象	ゼスチャの形状
人物(一人)	サークリングによって人物の顔または体全体を曲線で囲む
人物(複数)	連続アークによって人物の顔を曲線で順に結ぶ
道路	ラインによって道路に沿って道路の中心に線を引く
範囲の指定	サークリングまたはラインによる。範囲を示すのに十分な曲線または直線

表 2: ゼスチャ形状分類 (指示対象別)

ゼスチャの分類		形の分類		
		サークリング	ラインの終点 連続アークの点	ライン
モダリティ の分類	指示			
	範囲指定			

図 3: ゼスチャのモダリティと形状

「ゼスチャが為された」ということ自体にも大きな意味があり、例えば同様な質問が繰り返されている場合などはゼスチャのみの表現によって対話表現が完結し、対話の交替が起こることが観測された。

4 指示照応産出・理解過程のモデル化

実験結果を踏まえ、焦点情報に基づいた指示照応産出・理解過程のモデル化を行なう。産出・理解機構の全体的なモデルを、図4に示す。図4のメンタルシーンとは、指示表現によって対話者の表象空間に喚起される指示焦点についての情報を記録してゆく機構である。メンタルシーン単独で指示照応の産出・理解をするのではなく、意味処理機構や表現産出機構と連係して照応表現の理解・産出を行なう。処理は対話のターンごとに進行する。このメンタルシーンは、次の構成物からなる。

● 焦点

指示表現によって指示されるもの。物理空間に存在するものである場合と、メンタルシーンのイメージの場合とがある。

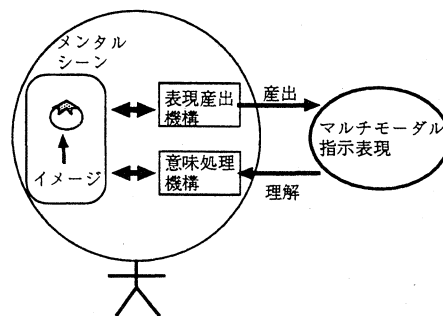


図 4: マルチモーダル指示照応の産出・理解機構

● イメージ

指示表現の意味処理によって表象空間に喚起される指示焦点の表象。指示された焦点についての状態。その時点の五感の感覚、心理の状態。

● シーン

意味処理の進行によって形成されるイメージの情景、場面。対話のターンごとに進行する。

そして、このメンタルシーンの動作は次のようになる。

● 【シーンの形成】

対話のターンごとに、対話内容の意味処理の結果によって現在の焦点についてのイメージが喚起され、そのイメージによって新たなシーンが形成される。シーンはその時点での対話の焦点についての最終状態のイメージを映す。

● 【シーンの進行】

シーンは対話のターンを単位時間として進行する。

● 【シーンの切替え】

これまでのシーンに存在したイメージとは異なるイメージが形成される場合、古いシーンを捨てて、新しくシーンを形成する。

対象とする対話環境をメンタルシーンを用いてモデル化すると、図5のようになる。

図5のマルチモーダル対話環境においては、ポインティングゼスチャによる表現と言語による表現の二つから指示表現が構成されている(図6)。

ところで指示表現は、メンタルシーンに対する働きによって導入表現と参照表現とに分けられ、それぞれの効果は次のようになる。

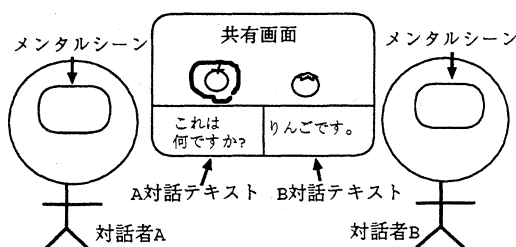


図 5: マルチモーダル対話環境のモデル

● 導入表現

対話環境の中でメンタルシーン以外の場所に存在しているものから、ある特定のものを指定してそれに焦点をあてる。その新しい焦点をイメージとして両方の対話者のメンタルシーンのシーンに導入する。

● 参照表現

すでにメンタルシーンのシーンに存在するイメージを指示する。

導入がおこると、以降は導入されたものを焦点として、対話内容の理解や指示詞の照応などがなされる。新たに導入が起こらない限り、焦点は移動しない。

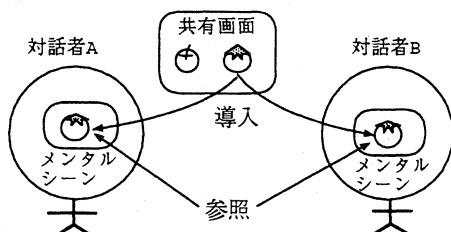


図 6: 焦点の導入と参照

4.1 参照表現の産出

参照表現は、すでにメンタルシーンに導入された焦点に対して行なわれ、ポインティングゼスチャによる参照はない。焦点の導入や移動が無ければ、意味処理の結果が反映されシーンが進行しても、シーンの切替えは起こらない。とくに必要がある場合や誤解を招く場合でない限り、指示詞はコ系ソ系どちらも用いられ、その選択は任意である。ただし、表現者自身がポインティングゼス

チャによって導入するものは、その導入の直前直後(つまり対話のターンが起きる前)で表現者自身が参照する場合に限り、必ずコ系指示詞を用いる。

4.2 導入表現の産出

導入表現は、表現者の知識空間にあるものも導入の対象となるが、本研究では共有画面内にあるものに対象を限る。新しい焦点を指示するには2つの方法があり、一つは対象を直接指示する方法で、もう一つは相対的な位置を示し間接的に指示する方法で、このための基準点をオリゴと呼ぶことにする。現在の焦点をオリゴとすることができれば、このオリゴが積極的に用られ、新しく導入される焦点は間接的に指示表現される。現在の焦点をオリゴとすることができなければ、新しくオリゴを導入してそれを参照しながら新しい焦点の位置を表現するか、直接対象を指示するかのどちらかである。

4.3 参照表現の理解

指示表現の理解過程は、受信した表現の意味処理と、現在のメンタルシーンでの焦点のシーンとの照合処理の相互な処理によって行なわれる。受信した表現がメンタルシーンの現在の焦点についての表現ではないと判断されれば、その表現に基づいて共有画面の中が検索され、該当する対象を新たな焦点としてメンタルシーンに導入し、シーンの切替を行う。

5 おわりに

本研究では、マルチモーダル対話収集実験をとおり、指示照応の産出・理解過程のモデル化を行なった。本研究で対象としたマルチモーダル対話環境においては、一つの焦点情報についての機構によってその産出・理解過程がモデル化できることを示した。今後の課題として、オリゴのデフォルトの決定方法やオリゴからの方向基軸の決定方法についての調査があげられる。

参考文献

- [1] Annette Herskovits, 空間認知と言語理解, 堂下修司, 西田豊明, 山田篤 (共訳), オーム社, 1991
- [2] 谷戸ほか, 道案内タスクにおけるマルチモーダル対話の会話文の特徴分析, 電子情報通信学会論文誌 Vol.J77-D-II No.8, pp1475-1483, 1994