

# ポスター発表における発表者と聞き手の間の対話シーンの意味的構造化

Detecting Microstructures of Conversations Using Contextual Data such as Physical References of Posters

熊谷賢\*1\*2  
Ken Kumagai

角康之\*1\*2  
Yasuyuki Sumi

間瀬健二\*2\*3  
Kenji Mase

西田豊明\*1\*2  
Toyoaki Nishida

\*1京都大学情報学研究科

Graduate School of Informatics, Kyoto University

\*2ATR メディア情報科学研究所

ATR Media Information Science Laboratories

\*3名古屋大学情報連携基盤センター

Information Technology Center, Nagoya University

The purpose of this paper is reusing conversational archives. So far we have given structure to conversations by analyzing gaze and utterance data from an interaction corpus, which is a semi-structured set of a large amount of interaction data collected by various sensors.

In this paper, we give more detailed structure to conversations by analyzing poster touch data. We devise a poster touch capturing method and investigate co-relation between poster touch and conversation structure. As a result, we have found a tree is suitable structure for representing conversation with a poster. By this tree model, we can interpret conversation scene in detail, and give detailed structures to conversation.

## 1. はじめに

近年, Web 環境の普及により, 知識流通への関心が高まってきた。従来はテキスト文書が知識流通の主なメディアであったが, ネットワークの高速化, ハードディスクの大容量化に伴い, 映像や音声といったマルチメディアを用いた知識流通の研究が盛んになってきた [1] [2]。

本研究では, 知識流通のためのメディアとして会話に注目して研究を進めている。日常生活において, 人は他者との会話を通じて, さまざまな知識のやり取りを行なっているため, 会話はそうした知識流通メディアの一つとして有効と考えられる [3]。我々は, 誰でも手軽に会話を用いて, 知識を伝えることができるので, 生の思考が反映されやすいという反面, テキストのように整理というプロセスを経ない場合が多いため, そのままでは扱うことが困難である。知識流通においては, 知識の生成, 編集, 管理が重要となるが, メディアとしての会話により, これらを実現するためには, 会話に適切な構造を与える必要がある。会話の構造には「誰が話したか」「いつ話したか」「どこで話したか」「何を話したか」といった属性が考えられるが, この中でも特に「何を話したか」を会話に与えることが非常に困難となる。従来より, 音声認識でなされているように, 発話の音波に注目してこの問題を扱おうという音声そのものへのアプローチがある一方で, 会話の内容を特定しようとする場合, 会話内容と密接な関連のある対象物が外界に存在するという状況は, 実世界に数多く存在し, それに注目するというアプローチの仕方もある。例えば, ミーティングでは, 資料ドキュメントが会話の対象物となるし, 博物館ではそれぞれの場所における展示物がその対象物となる。こうした環境においては, 会話内容を特定する際に, 外界に存在する対象物に関する情報を有効利用することができる。

本研究では, 学会発表の場となるポスター展示会場において, 展示員と見学者がポスターブースで行なうポスターを介した会話(以下, これをポスター会話と呼ぶ)を対象としている。ポスター会話においては, 展示員が, 見学者に対して, 現在の話題を明示的に示すためにポスター指差し行為を行なう。

本稿では, このポスター指差し行為を有効利用することで, 会話内容に詳細な構造を与えることを目的としている。

以降, 2章で会話の構造化における本研究の位置づけについて述べ, 3章でポスター指差し行為データ取得のための環境とポスターの構造について述べる。4章で指差し行為データを用いて構造を与える手法及び, 援用例について述べた後, 5章でまとめと将来の展望について述べる。

## 2. 本研究の位置づけ

ポスター会話を扱うのに先立ち, 会話を持つ構造について考察する。一般的に会話は次のような属性を持つと考えられる。

1. 誰が話したか
2. いつ話したか
3. どこで話したか
4. 何を話したか

ATR メディア情報科学研究所で開発されている体験キャブチャシテム [4] では, 実世界における人のインタラクションを, 装着するセンサや環境側に設置されたセンサといった複数センサ群により強制的に記録し, それらを体系的に解釈する。体験キャブチャシテムを用いることで, 1~3の属性を与えることが可能となるが, 4の属性に関しては, これまで漠然とポスター全体に関してという以上に踏み込めなかった。しかし, ポスターには内容に応じたより詳細な構造が存在すると考えられ, ポスターの展示員はその中の一箇所を指差すことで, 現在, 会話の対象となっている箇所を見学者に示す。本研究では, ポスターに存在する構造及び, ポスター指差し行為に注目して, ポスター会話に詳細な構造を与える手法を提案する。

我々が扱うポスター会話は, 次のような構造を持つと考えられる(図1)。すなわち, 一つ一つのポスター会話は Poster-Conversation クラスのインスタンスであると考えられる。それぞれ Object id は, 会話の生成者, Start, End は 会話の開始時間, 終了時間を表し, Booth id は, 展示会場におけるブースを表す。Role, Type は, ポスター会話特有の属性

連絡先: 熊谷賢, 京都大学情報学研究科, 京都市左京区吉田本町(工学部10号館), kumagai@lab1.kuis.kyoto-u.ac.jp

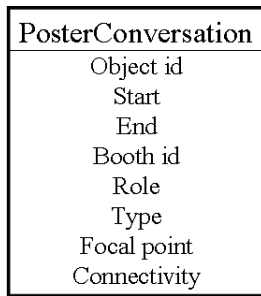


図 1: ポスター会話のクラス図

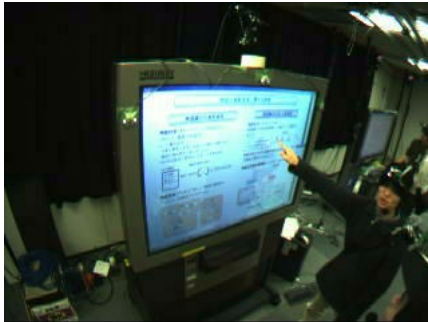


図 2: ポスター指差し行為データ取得環境

で、Role は、展示員が見学者のいずれかをとり、Type は、展示員の発話の連続性や、展示員と見学者間のターンテイキングに着目することにより、説明が質疑応答のいずれかをとり [5]。そして、Focal Point は、会話の対象を表し、Connectivity は、会話と Focal Point との結びつきの度合いを表している。なお本稿で扱うのは、主に Focal Point と Connectivity である。

### 3. ポスター指差し行為データ

ポスター指差し行為データを用いて、会話に構造を与えるのに先立ち、ポスター指差し行為データの取得環境、及び、ポスター構造に関する分析・モデル化について述べる。

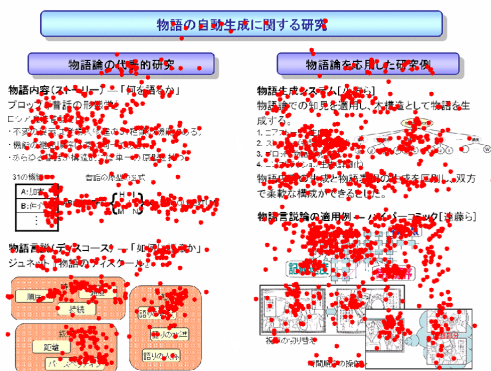


図 3: 蓄積されるポスター指差し行為データ



図 4: 交互指差し行為

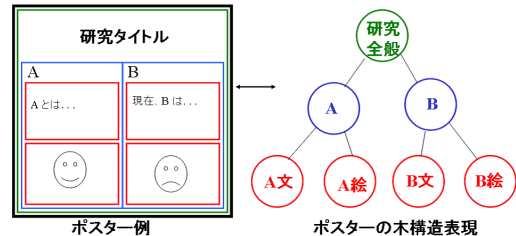


図 5: ポスター構造の木表現

#### 3.1 ポスター指差し行為データ取得環境

本研究では、ポスター指差し行為データを取得するために、大型のタッチパネルディスプレイを利用した。そして、タッチパネルディスプレイに電子化されたポスターを表示し、その上に透明ウィンドウを重ねて表示させるという手法を用いた。この手法により、ポスターを視覚的にさえぎることなく、透明ウィンドウを介して、タッチパネルディスプレイ上の接触イベントを取得する。展示員は、タッチパネル上に写されたポスターに対して、ポスター指差し行為を行うことで、ポスター指差し行為データを蓄積していくことが可能となる (図 2)。実際に蓄積されたデータを赤い点で表示すると図 3 のようになる。

#### 3.2 ポスター構造に関する分析

このような環境の下で実験を行い、蓄積されるデータ及び、映像の分析を進めていたところ、展示員の指差し行為は大きく二つの種類に分かれていた。

- 一箇所を指差ししながら、会話を行なう場合 (通常の指差し行為)
- 複数の箇所を交互に指差ししながら、会話を行なう場合 (図 4: 交互指差し行為)

当初、同じ箇所への指差し行為は、全て同じものとして扱っていたが、上記の分析により、次の仮説を立てる。

仮説 比較的短い時間で複数の領域を交互に指差す場合、展示員は、それらを内包する内容を説明しようとしている。

通常の指差し行為に加え、このような指差し行為を同様に扱うためには、ポスター構造の木表現が適切であると考えられる (図 5)。ポスター構造の木表現により、通常の指差し行為は、木の葉への指差しとして解釈することができ、交互指差し行為の場合は、交互に指差された葉の共通親となるノードへの指差しと解釈することができる。この過程は、図 6 で説明すると、通常の指差し行為がマイクロ指差し行為であり、そして、マイクロ指差し行為を比較的短い時間内に交互に行なった場合、そ

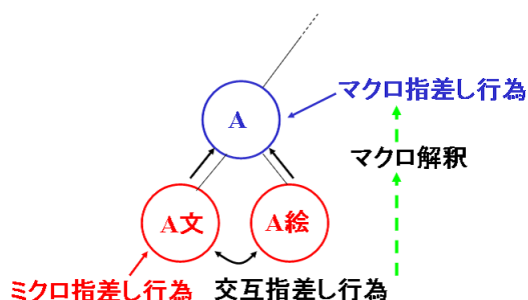


図 6: マクロ解釈

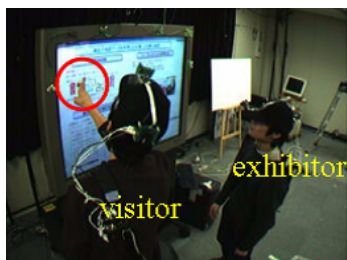


図 7: 見学者 (Visitor) による指差し行為シーン

これらの共通親を求める処理 (マクロ解釈) により、マクロ指差し行為と解釈する。このモデル化を用いて、PosterConversation (図 1) の Focal Point を、木表現におけるノードのうちの一つに対応づける。

また、実験を繰り返す過程で、見学者がポスターの一部を指差して、展示員に説明を求めるといったシーンも確認された (図 7)。本研究では、展示員が指差し行為を行なうという暗黙の仮定の下に研究を進めていたが、こうしたシーンは開放的な空間である展示会場においては十分に想定されるものである。しかしながら、この場合でも、指差し行為により指差された箇所が、会話の対象となっているという点では、展示員の指差し行為と共通しているため、同様のモデルで扱うことができる。このようなシーンは、外界に存在する対象物が会話を促進していると考えられ、興味深い。

本研究では、ポスター構造を木表現でモデル化した。他の表現方法としてはネットワーク表現が考えられる。しかし、次のような 2 つの理由から、本研究では木表現の方が適切であるとしている。

- ポスター構造は静的な構造である
- 端的に情報を整理する必要がある場合、より簡易な構造を用いるほうがよい

ポスターは一度生成されると、そこに新たな情報が追加されることはないため、構造は静的である。したがって、生成後、ノード間に新たな関係が生じることは無く、ネットワークのように動的に構造を変化させる際に有利となるモデルを用いる必要はない。また、ポスターのように非常に大きいとは言えない媒体を用いて、研究内容を説明する場合、情報を整理して載せる必要があり、端的に情報を整理して提示する場合、木表現で概念を整理して記述するという手法は、さまざまな場面で用いられる。本研究に近い状況としては、例えば研究発表のプレゼンテーションが考えられる。当初、頭の中で、概念が複雑な

ネットワーク状に結び付いていたとしても、資料を作成する段階で、木表現に整理して、相手に伝えるというプロセスはよく行なわれる。このように、端的に説明するという動機を持つ人は、より簡易な表現を用いて情報を整理する、と考えられる。

なお、ポスターから木表現を構築する部分に関しては、手動で設定するようにした。画像処理を行なうことで、領域から各ノードに対応する部分を見つけることは可能だが、意味的な関係である木表現を再現することはできない。ノード間の意味的な関係を正確に理解しているのは、ポスター作成者自身であるため、ポスター作成者が手動で設定するという指針が適切であると考えられる。

#### 4. ポスター会話の構造化

この章では、ポスター指差し行為データを用いてポスター会話に構造を与える方法について述べる。その際、ポスター会話の評価尺度として、Connectivity という指標を導入した。この指標は、ポスター会話とポスターの木表現におけるノードとの結びつきを定量的に表す。なお、Connectivity の算出においては、次の仮説を用いている。

仮説 ポスター会話は対応する指差し行為データとの時間間隔が離れるほど、その会話内容と指差し行為データで示された内容との結びつきが弱くなる

以下、ポスター会話に構造を与える過程を説明していく。始めに、発話データに対し、クラスタリング、ノイズ除去を行い、発話クラスタを生成する。発話クラスタには、開始時間、終了時間の属性が存在し、各発話クラスタを以下の 3 つの場合に分けて処理する。

1. 発話クラスタの時間内に、指差し行為データが一つ、もしくは一種類ならば、発話クラスタをその指差し行為データと対応づけ、Connectivity を 1 とする。
2. 発話クラスタの時間内に、指差し行為データが二種類以上存在する場合、指差し行為データごとに発話クラスタを分割する。ただし、時間的に先頭の指差し行為データでは分割を行わない。分割後、先頭の発話クラスタは、時間内の指差し行為データに対応付け、それ以外の発話クラスタは、時間的に直前の指差し行為データに対応付ける。なお、Connectivity は、それぞれ 1 とする。
3. 発話クラスタの時間内に、指差し行為データが存在しない場合、発話クラスタの開始時間前、終了時間後で、最も時間的に近傍にある指差し行為データに対応付ける。次に、発話クラスタの開始時間、終了時間、指差し行為データの時間をそれぞれ、start, end, time とし、定数  $m, n$  を用いて、Connectivity を以下の式で求める。

$$Connectivity = \frac{1}{1 + K}$$

$$K = \min\{m \times |start - time|, n \times |time - end|\}$$

2 は、展示員が説明を行なっている会話に多い場合で、この場合、比較的時間幅の大きい発話クラスタが形成されるが、指差し行為データに注目することで、1 つの指差し行為データに対応付けられるように発話クラスタを分割する (図 8: 場合 2)。なお、そうした説明シーンにおいては、多くの場合、指差し行為を起点として会話の対象が切り替わるため、時間的に



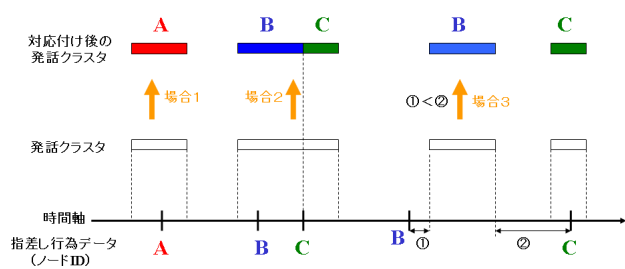


図 8: 発話クラスタ処理例

直前の指差し行為データに各クラスタを結びつけている。3は、会話と同時にポスター指差し行為が行なわれていない時に生じる現象で、この場合、時間的近傍の指差し行為データを、発話クラスタに対応付ける(図8:場合3,  $m=n=1$ )。しかし、前方と後方それぞれの時間幅に同じ重みを与えることが適切かどうかは、自明ではなく、 $m, n$ 間の関係に対して、今後さらに考察が必要である。

このように構造化された会話は、会話コンテンツとして、Webを介した知識流通に用いることができる。展示会を訪れた人に対しては、本人と興味の近い他の見学者の会話を提供することができ、訪れなかった人は、自分の興味のある項目についての会話を聞くことができる。その際、会話内容を保障する指標の一つとして、Connectivityを利用することができる。また、このプロセスにエージェントを用いることで、知識流通を促進させるという手法も考えられる [5]。

## 5. まとめと将来の展望

本研究では、会話に注目し、会話による知識流通を促進するために、会話に詳細な構造を与える試みを行なった。その際、困難となる点は、2章でも述べたとおり、「何について」の情報を、いかにして会話に与えるかという点にある。そこで、本研究では、会話と密接に関連する対象物が存在する状況を対象とし、このような環境における会話として、ポスター展示会場のポスター会話を対象に、構造を与える手法を提案した。

本研究では、ポスター指差し行為に注目したが、見学者の視線情報を用いる方法も考えられる。すなわち、各時間における見学者の視線情報から、ポスターのどの領域を注視しているかを判定し、それを会話に結びつけるという手法である。しかしながら、見学者の視線は常に話題となっている領域に集中しているというわけではなく、展示員が見学者を見たとき、見学者はそれに応えるように展示員の方を向くといった状況や、話題となっている領域と展示員の両方を視界に納めるために、顔がその中間方向を向くといった状況もあり、さらなる分析が必要となる。今後の方向性の一つとして、このような視線情報による手法と、我々が提案する手法それぞれでポスター会話に構造を与え、その結果を比較検討していくといったものが挙げられる。

ポスター会話においては、展示員が説明した後、展示員と見学者が議論を行なうという推移をみせることが多いが、議論シーンにおいて、見学者は常に質問を行なっているわけではなく、会話の流れに応じて、提案、同意、否定、賛同といった発言を行なう。今後、展示員や見学者といった属性情報を併用しつつ、そのような会話の流れに注目した構造化を進めるという方向性が考えられる。例えば、見学者の質問に注目して、議論シーンを段落分けするといったもので、見学者の質問から一

つの段落が始まり、次の見学者の質問で、次の段落が始まるといったものである。

ポスター展示会場以外のドメインで、研究を進めるという方向性も考えられる。具体例としては、美術館や、博物館といったものがあるが、こうした環境においては、実験的な環境と異なり、あまり詳細に環境を作りこむことができないという制約が生じる。こうしたドメインにおいては、会話の際に、そうした制約を満たすという条件の下で、対象を指し示すための簡易な手法を提案することが課題となる。

今後はこうした方向性を考察しつつ、ミーティングルームにおける資料を交えた会話やホワイトボード上で動的に生成される書き込みに対する会話なども対象として、研究を進めていきたい。

## 謝辞

本研究を進めるにあたり、多分のご意見を賜りました久保田秀和氏、多分の協力を賜った川口洋平氏を始め、関係者の方々に感謝します。なお、本研究は情報通信研究機構の委託研究「超高速知能ネットワーク社会に向けた新しいインタラクション・メディアの研究開発」により実施した。

## 参考文献

- [1] P. Chiu, A. Kapuskar, S. Reitmeier, L. Wilcox: Meeting capture in a media enriched conference room, In Proceedings of CoBuild'99 (Springer LNCS1670), pages 79-88, 1999.
- [2] 村上正行, 角所考, 美濃導彦: マルチメディア一斉講義における内容に基づく受講生の注視行動の分析, 人工知能学会論文誌, Vol.17 No.4 PP.473-480, 2002.
- [3] T. Nishida: Conversation quantization for conversational knowledge process, Special invited talk, S. Bhalla (Ed.): DNIS 2005, LNCS 3433, Springer, pp. 15-33, 2005.
- [4] Y. Sumi, S. Ito, T. Matsuguchi, S. Fels, K. Mase: Collaborative capturing and interpretation of interactions, Pervasive 2004 Workshop on Memory and Sharing of Experiences, pp. 1-7, 2004.
- [5] 川口洋平, 角康之, 西田豊明, 間瀬健二: 展示会場における過去の対話データを利用した分身プレゼンテーション, IPSJ SIG Technical Reports, Vol. 2005, No. 28, pp. 225-232.