

## マルチモーダル対話システムにおける対話管理

5E-6

高橋 和子 竹澤 寿幸  
ATR 音声翻訳通信研究所

## 概要

音声および画像入出力を持つユーザ主導型の質疑応答システムの対話管理について述べる。特に省略・照応に焦点をあて、入力発話で直接/間接に言及されたタスクの情報とデータベースの内容を使ってこれらを解決する手法を提案する。これによって、より相互作用性の高い自然な対話が可能になる。

## 1 はじめに

音声および画像インタフェースを持つマルチモーダル対話システム [Tak94][KMMN94][NDI98][TM98]における対話では一度言った語句を省略した短い発話が多く、インタラクションが頻繁に起こるという特徴がある。そのため、データベースを検索するためには省略されたものの補完や指示代名詞によって指示された対象を同定する照応の問題を解決しなければならない。

一般に、省略・照応が何をさすかを判定するためには過去の発話履歴を木構造やスタックとして記憶し、それを参照して決定するという方法がとられている。そのため対話の連続性が高いと履歴の探索に時間を要する。本稿では、発話履歴を記憶するのに話題の遷移を反映した構造を導入する。これを発話ごとに書き換えていくことでメモリを節約できる。また、探索時間もそれほどかからない。

本稿は以下のように構成される。まず、第2章で、対話システムの一例である MMGS の概要を示す。次に、第3章で対話の管理機構について述べ、最後に第4章で結論を述べる。

## 2 マルチモーダル対話システム

MMGS[TM98]は音声および3次元画像上のポイントングによる入出力を持つユーザ主導型の質疑応答システムである。データベースは階層型オブジェクトで記述されている。

ユーザからの入力は認識部を通った後、意味解析が行われ適当な意味表現の形になる。問題解決部ではこの意味表現をもとに照応や省略の解決を行ってデータ

ベースの検索を行い、画面の更新情報や応答文を生成する。出力部ではそれを音声合成して出力したり3次元グラフィックスや写真によって表示する。

問題解決部への入力は (*class, action, object*) の3項組で表現される。*class* は質問の種類を、*action* は主動作の種類をそれぞれ表す。*object* はその文で言及されている対象物を表し、名前 (*label*) および素性 (*ftr*) の  $n$  個の ( $n = 0$  を含む) 有限列であり、その構造は *class, action, label* によって決まる。各素性は素性名 (*fn*) と素性値 (*fv*) から成る。素性名は意味的な種類を表す。たとえば、喫茶店 (*coffee-shop*) という *label* を持つ *object* に対しては場所 (*loct*) という素性名を持つ構造がつけられる。主語や目的語が省略された発話では素性名ができて素性値が決まらない (すなわち変数のままである) 構造ができる。以後、値の決まっていない部分は '?' で始める文字列で表す。

## 3 対話管理

## 3.1 対話例

簡単のため各発話は単文1つで話者交代が生じるものと仮定する。

以下に対話例を示す。

例1. けいはんな地区の案内

- U1: これはなんですか [画面上を○で囲む].  
S1: 「けいはんなプラザ」です。  
U2: 喫茶店はありますか。  
S2: 「ライン」があります [写真を見せる].  
U3: ATR にはありますか。  
S3: カフェテリアがあります [写真を見せる].

[ ] 内は画面上の操作を示す。U はユーザの発話、S はシステムの発話をそれぞれ表す。U2 では「けいはんなプラザに」が、U3 では「喫茶店は」がそれぞれ省略されている。

一般に質疑応答型システムでは、ユーザが質問しそれに続いてシステムが答えるという隣接ペアが存在す

るものとしてよい。各隣接ペアごとに Q 話題および A 話題を対応させる。これらは現発話で直接言及された対象だけでなくそれが参照する過去の発話で言及されたものも含む。

### 3.2 Q 話題の生成

入力である意味表現 (*Rep*) が変数を含む場合、直前の隣接ペアで生成された A 話題 (*Ainfo*) を使って補完を行い、現在のユーザ質問に対応する Q 話題 (*Qinfo*) を生成する。object のある素性は判明しているが label が不明な場合、*Ainfo* に素性名の一致するものがあるれば、その素性値といれかえる。object 自体が変数の場合、*Ainfo* そのものを代入する。いずれでもない場合は、データベースを検索して同じ構造を持つオブジェクトを探し出し、補完する。(詳細は [TT99] を参照。)

例 2. 例 1 の U3 「ATR にはありますか」

```
Rep = (yn-q, exist,
      (?LABEL, (loct, ATR), ?FEATURE))
```

直前の発話で「けいはんなプラザの喫茶店」が言及されているので

```
Ainfo = (coffee-shop, (loct, keihanna-plaza))
```

が成り立っており、ここから *rhein*(ライン) へポインタがはられている。

この時、object の変数部分 ?LABEL および ?FEATURE を補完して *Qinfo* を生成する。ここで、*Ainfo* には *loct* という素性名があるので、この素性値を置き換えて、

```
Qinfo = (coffee-shop, (loct, ATR))
```

を得る。同時に ?LABEL には *coffee-shop* が代入され、?FEATURE には *NIL*(情報なし) が代入される。

例 3. 例 1 の U3 の代わりに U3' 「広いですか」という質問がきた場合

```
Rep = (yn-q, wide, ?OBJECT)
```

この場合、?OBJECT の構造には制約がないので、*Ainfo* そのものを代入してよいが、*Ainfo* からポインタがはられているので、その先の *rhein* を代入し、*Qinfo* = *rhein* を得る。

これらの例で、省略されていた「喫茶店」も「ライン」もともに名詞であり、いずれも喫茶店という範疇でとらえられるため、文法や意味的な制約からだけではどちらが適切か判断できない。

### 3.3 A 話題の生成

ユーザの質問に対応した Q 話題が生成されると、システムはこれに追加、訂正などの修正を加えた A 話題を生成し、これを適当な形にして回答とする。ユーザの質問の種類によって A 話題の生成方法も変わるが、ここでは *yn-question* によるオブジェクトの存在

の質問の場合を例として *Qinfo* から *Ainfo* を生成する方法を述べる。

対応するオブジェクトはデータベースに存在する場合は、*Qinfo* に対する具体値である *instance* を返すので、*Qinfo* から *instance* へのポインタをはったものを *Ainfo* とする。ない場合は、*NIL* とする。

例 4. 例 1 の S3 「カフェテリアがあります」

```
Qinfo = (coffee-shop, (loct, ATR))
```

データベースを検索すると対応するオブジェクトとして *cafeteria* が見つかるのでここにポインタをはる。

### 3.4 特徴

*Qinfo/Ainfo* の生成の際、唯一履歴として参照するのが *Ainfo/Qinfo* だが、この参照では語句の長さや構造にかかわらず一度のアクセスによって補完すべき語句を見つけることができる。また、これまでの発話すべての履歴を持つ必要がないので探索も速い。さらに、音声認識がうまく行われなかった場合でも意味表現として、*class, action* さえ認識できていれば、不明な部分の補完が可能である。なお、以上では省略について述べたが、照応に関しても同様に扱える。

## 4 おわりに

本稿では、発話ごとに直接/間接的に言及された対象を Q 話題および A 話題として定義し、これを使って省略の補完を行う手法を述べた。この手法の特徴として、ドメインに依存しない、実時間応答が可能である、データベースの語彙を増加させるのが比較的簡単であるなどがあげられる。これによってより使い勝手のよい相互作用性の高いシステムが実現できる。

### 参考文献

- [KMMN94] 神尾広幸, 松浦博, 正井康之, 新田恒雄. マルチモーダル対話システム MultiksDial. 電子情報通信学会論文誌, vol.J77-D-II, no.8, pp.1429-1437, August, 1994.
- [NDI98] 中川 聖一, 傳田 明弘, 伊藤 敏彦. マルチモーダル観光案内対話システム. 人工知能学会誌, vol.13, no.2, pp.241-251, March, 1998.
- [Tak94] 竹林 洋一. 音声自由対話システム TOSBURG II - ユーザ中心のマルチモーダルインタフェースの実現に向けて -. 電子情報通信学会論文誌, vol.J77-D-II, no.8, pp.1417-1428, August, 1994.
- [TM98] T. Takezawa and T. Morimoto. A Multimodal-Input Multimedia-Output Guidance System: MMGS. In *Fifth International Conference on Spoken Language Processing* vol.2, pp.285-288, 1998.
- [TT99] 高橋 和子, 竹澤 寿幸. マルチモーダル対話システムにおけるインタラクション機構. In *インタラクション'99, to appear*, 情報処理学会, 1999.