京都観光案内対話コーパスにおける対話行為の分析

大竹 清敬　堀 智織　柏岡 秀紀　中村 哲
情報通信研究機構 / ATR 音声言語コミュニケーション研究所
{kiyonori.ohtake, chiori.hori, hideki.kashioka, satoshi.nakamura}
@ {nict.go.jp, atr.jp}

1　はじめに

人間が自然に対話できる頑健な音声対話システムを構築するためには，ユーザの発話がシステムに対して何を要求しているのかという対話行為の理解と，その要求の具体的な内容の理解，さらにその対話行為の背後にある人間の意図も推定する必要がある [Kaw06]．我々は，頑健な音声対話システムの構築を目指して，人間対人間による音声対話の大規模なデータを収集した．そこでの対話の諸相を解析し，対話データから，人間のふるまいを統計的にモデル化することで，より人間に近い音声対話システムを設計し，データを直接的に用いて対話制御することを目指している．

本報告では，現在我々が収集・解析を進めている京都観光案内音声対話コーパスの概要と，コーパスに対する対話行為タグ付与，ならびにこのコーパスとそれに付与した対話行為タグに基づいて動作する対話制御方式について述べる．

本研究では再利用性，拡張性の高い対話システムの設計を可能にする方法として，WFST に基づく対話制御方式を提案する．この手法は，コーパスから得られるガイドモデルやユーザモデルを WFST に合成する事で，適応的な対話制御の実現が期待できる．

本稿では，特に，WFST に基づく対話制御がなされることを前提にした場合に，どのような事を考慮して対話行為タグを設計すべきかについて議論する．

2　京都観光案内対話コーパス

本研究において我々が収集したのは，京都観光案内のエキスパートガイド (男性 1 名，女性 2 名) が模擬旅行者 (ユーザ) に対して京都市内一日観光の計画立案を行う 2 者による対面対話である．観光ガイドは自身が持つ観光に関する知識，ガイドブック，地図，WEB 上の情報を用いてユーザに情報を提供し，一日の旅程を共同作成した．1 対話は，30 分間である．20 歳代から 50 歳代の 114 名（男性 57 名，女性 57 名）をユーザとして収録した．コーパス全体では，50 時間あまりのデータである．収録の様子を図 1 に示す．

図 1: 収録の様子

ガイドは，ヘッドセットマイクを使用し音声を収録した．ユーザは，スタンドマイクまたは，ヘッドセットマイクのいずれかを用いて音声を収録した．書き起こしや，その他の参考資料として，ビデオも収録した．

114 対話のうち，チェック中のものも含めて 108 対話が書き起こされている．書き起こしは，収録で得られた音声データに対し，1 秒以上の無音区間を発話セグメントの認定単位として，発話セグメントを切り出し，それに対してなされた．つまり，音声データ上の任意の音声波形に対して，その開始，終了時刻と音声波形の書き起しを組とする形式で書き起こした．書き起こしからわかるコーパスのデータを表 1 にまとめる．

本研究で想定しているユーザは，京都の一日観光をしてみたいが，京都市内の観光地に関する充分な知識はなく，京都の観光に対して，強い動機を持ってない者である．現実的には，観光地に関する充分な知識を持っている場合は，他者の知識を利用することは考えにくい．また，京都観光に対して強い動機を持っている者は，現在では，WWW をはじめ各種情報を容易に入手できるので，自分自身で調べ，観光案内システムを利用することは考えにくい．今回のデータ収録では，ユーザ役を担当する者には，事前に充分な説明を与え

ガイド発話セグメント数	40,336 個
ユーザ発話セグメント数	32,182 個
ガイドのみの発話時間	24.76 時間
ユーザのみの発話時間	7.56 時間
両者ともの発話時間	6.27 時間
無音時間	18.4 時間
ガイドがユーザへバージイン	10,731 回
ユーザがガイドへバージイン	20,350 回

表 1: 京都観光案内対話コーパス概要

ず，当日，簡単な説明を行いデータを収録した．したがって，本コーパスにおけるユーザは，我々が想定しているユーザに比較的近いと考える．

3 対話行為

音声対話システムを利用するユーザの対話行為を知ることは，音声対話システムの挙動を決定する上で，非常に重要である．そのため，これまでにも数多くの対話行為タグが設計され，コーパスにタグ付けがなされてきた (たとえば，[Ara99, Jur97] など)．ここでは，本研究において，対話システムを構築することを前提として，対話行為タグ付けを行った結果，現時点で明らかになっている点などを紹介する．

3.1 タグセットとタグ付け

これまでにコーパスの書き起こし作業を行い，ユーザおよびガイドの対話行為に対するタグ付け作業を行った．完全なタグセットをあらかじめ用意することは困難だったので，タグ付けと同時にタグセットを修正するアプローチをとった．

対話行為のタグを付与するために，まず一連の対話が，目的地，移動手段，観光スポットでの行動や各種イベントなどいずれかについて話をすすめているかによって対話を分割し，その分割された単位をエピソードとした．次に，各エピソードにおいて順次対話行為タグを，各対話行為の範囲を認定しながら，その対話行為の種類 (以下 DA タグと呼ぶ) ならびに，それより前のいずれの対話行為に対応するかなどを付与する．各対話行為は，識別コードによって区別される．また，DA タグは，コーパスを解析する過程で新たに観測された現象に対して適切となる対話行為を考慮し，適宜追加，併合ならびに修正した．さらに，対話行為の他に，対話を理解する上で重要と考えられる表現に対してもタグを付与する．本稿では，これを NE タグと呼ぶ．対話行為は，<DA タグ 識別コード 対応識別コード>1 つ以上の発話セグメント</識別コード>の形式でタグ付けされる．識別コードは，対話中の各対話行為を識別するために用いる．また，対応識別コードは，その対話行為に先行し，その対話行為に直接的に対応する識別コードを表現するために用いる．一方 NE タグは<NE タグ>該当表現</NE タグ>の形式でタグが付与される．

```
<Episode 1 Determination DST(general)>
10688-12762:g <Ask OQ DST, 1 0> で　　は　　、<location>京都</location>の<activity>一日観光</activity>。/
11389-11601:u <Nod, 2 1>  はい。</2>/
13138-13310:u <Nod, 3 1>  はい。</3>/
13341-17214:g [えーと]、どちらかご興味あるところありましたらご案内さしていただきますが。</1>/
18743-18916:u <Ack, 4 1>  はい。</4>/
19165-19945:u <List PREFERENCE(DST(general)), 5 1>  [えっと][ね] /
20074-20427:g <Nod, 6 5>  はい。</6>/
22828-24715:u [こう]<spot(type)>景色がいいとこ</spot(type)>とか。</5>/
25080-26019:g <Repeat, 7 5> <spot(type)>景 色 が い い と ころ</spot(type)>。</7>/
25958-26312:u <Ack, 8 7>  はい。</8>/
26506-27666:u <List PREFERENCE(DST(general)), 9 1>
<spot(type)><activity>散歩</activity>コース</spot(type)>みたいな/
27058-27588:g <Nod, 10 9>  はい。</10>/
28127-29960:g <Ack, 11 (5 9)> [あー] なるほど。</11>/
28199-28637:u 所が。</9>/
30399-30569:u <Ack, 12 11>  はい。</12>/
30650-32790:g <Recommend DST(general), 13 (7 9) >[そ う で すね] 時期が、/
33378-36192:g ちょう ど も う す ぐ<season><view>桜</view>の 時期<season>なので/
36225-36494:u <Nod, 14 13>  はい。</14>/
37530-40533:g <spot(type)><view>桜</view>が 綺 麗 に 見 える<activity>散策</activity>コース</spot(type)>なんかが。/
40730-41856:u <nod(positive), 15 13> [あ] 素敵ですね。</15>/
40984-42082:g よろしいですかね。 </13>/
42220-42514:g <Rhythm, 16 13>  はい。 </16>/
42236-42525:u <Accept, 17 13>  はい。 </17>/
42818-43357:g <Ack, 18 17> わかりました。 </18>/
</Episode 1>
```

図 2: コーパスに付与された対話行為タグの例

図 2 は，コーパスに対する対話行為のタグ付け結果の一例である．図はユーザとガイドの一つの対話エピソードに対話行為タグを付与したものである．左の数値は各発話セグメントの開始時刻と終了時刻を表している (ミリ秒単位)．コロンのあとの，g，u はそれぞれその発話セグメントが，ガイドおよびユーザのいずれかによってなされたものかを示す．この例では，識別コードに整数を用いている．0 は対応識別コードに用いられ，直接的に対応する識別コードがないことを意味する．

これまでにも，さまざまな対話行為タグセットが提案され，多くのデータに対してタグ付けがなされた．代表的なタグセットとして DAMSL[All97] がある．我々が作成するコーパスに対しても DAMSL を適用するこ

とが考えられる．しかしながら，対象とする対話のタスクやドメイン，またタグ付与の目的が異なると，特定のタグセットをそのまま使い続けることは困難である (たとえば [Jur97])．

そこで，我々は，対話システムの特に対話制御に重点を置いて，これまでに提案されているタグセットを参考にしつつ新規にタグセットを定義することにした．これまで，2 対話に対して図 2 に示した結果と同様のタグ付けを行い，使用する DA タグをほぼ決定した．現在は，この 2 対話に対するタグ付け結果を参照データとして，さらにタグを付与するための準備をすすめている．

我々が新規に定義する DA タグセットは，これまで提案されてきた対話行為タグセットと基本的な部分は大きく変わらないと考える．たとえば，Ack, Confirm, Decide, Explain, Filler, Greeting, Ask(Open-ended Question), Ask(Yes/No Question) などタスクなどに依存せずよく用いられるものを備えている．一方で，観光対話タスクにおいては，たとえば，Ack としてその対話行為が分類される発話に対して，Ack が肯定的な了承なのか，否定的な了承なのかによって，システムの挙動を変更しなければならない場面が多くある．そのため，現在は，わかる範囲で，肯定的か否定的かを DA タグとともに明示している．そのため，DA タグセットは一部のシンボルについては，階層的になっていると言える．また，我々の対話行為タグは，先行するいずれの対話行為に対応するのかその識別コードを記述しており，ユーザの発話に対するシステムの挙動などを抽出しやすくしている．

3.2 対話システムのための対話行為タグ

本研究で用いている対話行為タグは，対話制御に使用されることを前提として設計された．そのため多目的に用いることを考えて設計された対話行為タグとは若干異なる部分がある．たとえば，ガイドの対話行為が非常に広範にわたってある一つのタグにまとめられる場合がある．具体的には，ある観光地を説明する場合に，1 発話の一般的な対話行為としては，説明が連続するような場合がある．このような場合に，システムのもう一つの側面，つまり，説明の対象が変更されない限り，それは同一の対話行為としてひとつにまとめた．これは，対話システム内部での状態が，対話行為も，内部変数にも変化がないと考えることができるためである．別の見方をすると，エピソードの内の一連の発話をサブトピックに分割していると見ることもできる．現在，ガイドの対話行為タグの一部には，この機能ももたせて運用されているものがあり，他のタグセットとの対応をとるような場合には注意が必要である．

これに対し，ユーザの対話行為は比較的短い範囲 (ほぼ 1 発話ごと) に付与されている．対話システムの運用の際には，ユーザの一発話に対してシステムは適切な動作をしなければならない．そのため，ユーザの発話に対しては，1 発話毎に対話行為タグを付与する．

したがって，対話行為タグの範囲は，対話システムの動作を前提として付与されている．また，ガイドに使用する DA タグと，ユーザに使用する DA タグには同一のシンボルを用いているが，対話行為タグに含まれる DA タグのみによってシステムの挙動が一意に決まるわけではない．システムの挙動は，その発話者（役割），先行する対話行為に対する応答かなどに基づいて決定される．そのため本研究で用いている対話行為タグでは，対応する対話行為も含めている．

4 音声対話システム

ここでは，現在，我々が開発をすすめている音声対話システムについてその概要を説明する．特に，これまで述べてきた対話行為タグによってタグ付けされたコーパスを用いて対話制御をどのように行うかに重点を置いて説明する．

4.1 概要

収録した京都観光案内音声対話コーパスに基づき音声対話システムを構築する．図 3 にシステムの構成を示す．観光案内ならびに、観光計画立案は、複合タスクである．知識処理部で扱う特定の対話については，別々のネットワークとして記述する．システムの拡張性を高めるため，タスク依存の対話とタスク非依存の対話を分割する．たとえば，選択肢を順次提示して決定を促す対話はタスクに依存せず用いることができる．

4.2 WFST による対話制御

本研究では，重み付き有限状態トランスデューサ (Weighted Finite-State Transducer: WFST) による対話シナリオ記述と対話制御を採用する．WFST は有限状態オートマトンの一種であるため，WFST による対話制御は基本的に状態遷移モデルによる対話制御と等価であるが，入力記号列をそれが対応する出力記号列に変換する記号列変換モデルとして記述でき，重みによる遷移コストも付けられる．

本システムでは，ユーザの発話に含まれるキーワー

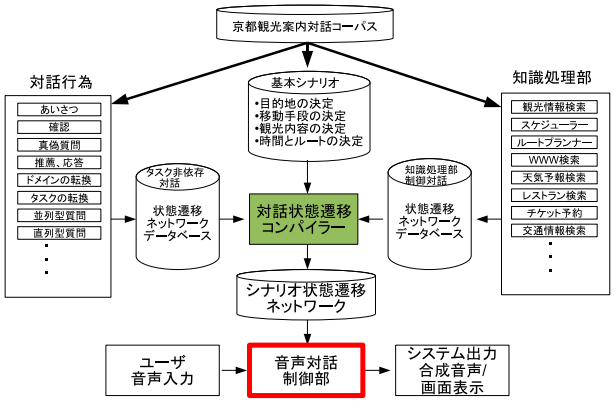

図 3: 京都観光案内音声対話システムの構成

ドとその対話行為を入力記号列，ガイドの対話行為および必要なシステムの手続きを出力記号列とする対話制御 WFST を記述し，これを対話制御に用いる．ここで，システムの手続きとは，データベース検索や情報提示，合成音声による案内再生などである．

開発した対話制御部は，図 4 のように構成される．対話制御部は WFST で記述された対話戦略と手続きのセット，タスク定数・内部変数を読み込んで動作する．

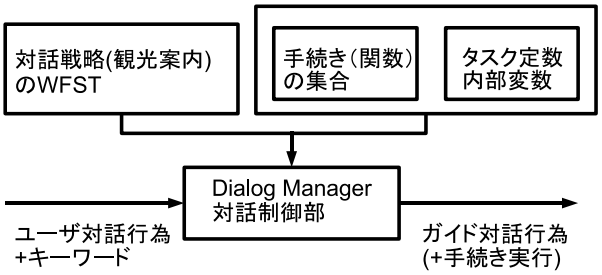

図 4: WFST による対話制御部

WFST に基づく対話制御には，ある WFST に対して任意の WFST で記述された制約（およびコスト）を合成できるという利点がある．たとえば，ガイドの対話行為の N-gram モデルを WFST に変換し，対話制御 WFST と合成 (composition) することで状態遷移に優先度を与えられる．この枠組みは対話制御 WFST が非決定性である時に，適当な状態遷移を選択できる枠組（マルコフ決定過程 [You06]）を提供する．更に，合成された WFST は対話制御部でそのまま駆動させることができる． また，通常のオートマトンと同様に，複数の WFST を連接 (concatenate)，結合 (union) することで，事前に作成された WFST を自在に組み合わせることができる．また，合成や組み合わせによって WFST が肥大化しても，各種最適化演算により対話制御部の負荷軽減が期待できる．このように，WFST による対話制御は，高い拡張性と再利用性を提供する有効なアプローチである．

4.3 ガイドモデル・ユーザモデルの導入

観光案内対話において，ガイドが何かを推薦する仕方は，ガイドによって異なる．たとえば，全ての可能性を述べてからユーザの選択を促す戦略や，ユーザの反応を見ながら一つの選択肢について詳細を紹介し最後に意思決定を迫る戦略などがある．また，ガイドはユーザの状態に応じて案内方法を適応的に変化させている．ガイドの個性やユーザの内部状態に基づき柔軟な対話制御を行わせるため，様々な対話戦略を取り得る非決定性 WFST とコーパスから獲得されるガイドの対話戦略モデルとを合成する事で，ガイド依存の対話戦略を実現する．更に，ユーザの希望が有る場合と無い場合の対話からそれぞれ別々に重みを学習しユーザの状態モデルとする．

5 まとめ

本報告では，我々が構築した京都観光案内対話コーパスの概要と，それに対して付与している対話行為タグについて概観した．また，現在開発を進めている音声対話システムの構成およびその対話制御方式について述べた．今後は，さらにコーパスの解析を進めると共に，音声認識部，音声理解部，音声合成部を対話制御部と結合して，実際にシステムを動作させる予定である．

参考文献

[All97] Allen, J. and Core, M.: Draft of DAMSL: Dialog Act Markup in Several Layers, Technical report, Discourse Research Initiative (1997).

[Ara99] 荒木雅弘, 伊藤敏彦, 熊谷智子, 石崎雅人：発話単位タグ標準化案の作成, 人工知能学会誌, Vol. 14, No. 2, pp. 251–260 (1999).

[Jur97] Jurafsky, D., Shriberg, E., and Biasca, D.: Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual, Draft 13, Technical report, University of Colorado at Boulder & SRI International (1997).

[Kaw06] 河原達也, 荒木雅弘：音声対話システム, オーム社 (2006).

[You06] Young, S.: Using POMDPS for Dialog Management, In *IEEE/ACL Workshop on Spoken Language Technology (SLT)*, pp. 8–13 (2006).