NICT 京都観光案内音声対話コーパスの構築とその利用

大竹 清敬　翠 輝久　堀 智織　柏岡 秀紀　中村 哲
情報通信研究機構 MASTAR プロジェクト
{kiyonori.ohtake, teruhisa.misu, chiori.hori, hideki.kashioka, satoshi.nakamura}
@ nict.go.jp

1　はじめに

本稿では，音声対話システムの構築を目的に収集した京都観光案内対話コーパスの概要と，コーパスに付与している対話行為タグならびにそれらの利用について述べる．近年，音声対話システムを構成する個々の要素技術，音声認識や音声合成技術などはコーパスを前提とした統計的手法によって実現されている．また，対話システムにおける対話管理を，規則に基づいて実現するのではなく，統計的手法によってコーパスから学習する研究がなされている (たとえば [Lev97] など)．したがって，このような統計的手法を用いる場合には，コーパスは必要不可欠であり，その役割はきわめて重要である．コーパスの質が対話システムの性能を決定すると言っても過言ではない．

我々は，統計的手法によって対話管理をする同調的な音声対話システムの構築を目指している．そこでの目標は，タスク達成のための機能を提供することはもちろん，人とシステムとの対話が人間同士の対話と同程度に自然に行われることである．コーパスに基づく統計的対話管理手法を用いたとしても，画一的な応答をする対話システムでは不十分であると考える．学習に利用するコーパスを変更することによって，システムのふるまいを制御できることはもちろん，状況によって柔軟に応答できる枠組みが望まれる．

我々は，ユーザの発話を入力とし，それを対話システムの動作記述へ書き換える WFST を用いた統計的な対話制御手法を提案している [Hor08]．このようなアプローチを用いて対話制御のためのモデルを学習するためには，入力となるユーザの発話を汎化して，記述し，それに対するシステムの動作を詳細に記述する必要がある．本研究では，発話を汎化し，記述するために対話行為を用いる．対話行為タグを設計し，コーパスへタグ付けを行なうことで，統計的な対話制御手法を実現する．

2　京都観光案内対話コーパス

京都観光案内対話コーパスは，京都観光案内のエキスパートガイドが模擬旅行者に対して京都市内一日観光の計画立案を行なう 2 者による対話からなる [Oht08]．現在までに，日本語の対面形式での 114 対話，非対面形式での 100 対話，Wizard of OZ(WOZ) 形式での 80 対話ならびに英語での対面形式での 48 対話を収録している．1 対話は約 30 分である．現在，前年度まで収録した日本語対話のほぼ全て（対面 114 対話，非対面 60 対話，WOZ 形式 80 対話）が書き起こされている．書き起こしは，発話の開始時刻，終了時刻をともなった形式で書き起こされ，500 ミリ秒以上のポーズによってセグメントを認定している．

観光計画立案のような相談対話は，ホテル予約などの単純な情報伝達・確認対話とは異なり，一つの決定を行なうまでに多様な発話がなされるため，従来の音声対話研究で扱われてきた対話と比較して難しい対話であると言える．

3　対話行為タグセットの概要

我々の目的は，ユーザがシステムとの自然なインタラクションを通じて対話目的を達成できるシステムを構築することである．そのためには，対話を計算機で扱うための記述手段として，我々は対話行為に着目し，対話行為を表現するタグを設計した．Bunt[Bun00] は，対話行為の 2 つの側面を指摘している．一つは，情報伝達機能 (communicative function) であり，発話が対話をどう制御しようとしているかという側面である．もう一つは意味内容 (semantic content) であり，発話がどのような内容について言及しているかという側面である．

これまでに対話の様々な諸相の分析や，対話システムの構築を目的として対話コーパスへの対話行為タグの付与が行われてきた．しかし，これらの研究の多くは前述の対話行為のうちのどちらか片方の側面のみを記述したものであり，両者を扱ったものは少ない．今回我々が扱う観光案内のような相談型の対話を詳細に記述するためには，どちらの側面も重要であると考え，2 種類のタグを設計した．発話の情報伝達機能を記述するものとして発話行為タグを，そして発話中の意味記述として意味内容タグを導入する．

4　発話行為タグ

この節では，発話の情報伝達機能を記述するために設計した発話行為タグについて簡単に説明する．まず，対話コーパスに対して発話行為タグを付与した例を表 1 に示す．詳細は，文献 [Oht09b] を参照されたい．

我々は，タグ付け単位に対して該当する全ての機能を

表 1: 発話行為と意味内容タグ付け例

UID	発話時間	話者	書き起こし	発話行為タグ**	意味内容タグ
54	76669–78819	User	あと， 大原が どの 辺に なりますか	WH–Question_Where	null (activity),location (activity),(dmst),interr (activity),(dmst),noun around=”yes” (activity),predictae
55	78889–80069	User	地図が あんまり 言うほど	State==	media adv-p adv-p
56	80069–80926	User	頭に 入ってない.	State_Reason→54	object predicate
57	80788–81358	Guide	この 辺ですね.	State_Answer→54	(dmst),kosoa (dmst),noun around=”yes” pred=”copula”
58	81358–81841	Guide	大原は.	State_Inversion	location
59	81386–82736	User	ちょっと 離れすぎてますね.	State_Evaluation→57	(trsp),(cost),(distance),adv-p (trsp),(cost),(distance),predicate eval=”yes”
60	83116–83316	Guide	あ,	Pause_Grabber	null
61	83136–85023	User	これでも, 一日では どうでしょう？	Y/N–Question	null (activity),(planning),duration (activity),(planning),(dmst),interr

**発話行為タグは，“_” 記号をレイヤー区切り文字として記述し，“null” は省略している．
“→” 記号の後ろの数字は，タグの対象となる発話 ID を示している．

記述しようとする DAMSL (Dialog Act Markup in Several Layers) [Jur97] や，MRDA (Meeting Recorder Dialog Act) [Shr04] と同一のアプローチをとったタグを設計した．

MRDA タグセットを拡張し，*General*，*Response*，*Check*，*Constrain*，*ActionDiscussion*，*Others* の 6 レイヤーを設計した．また，ターン管理および WH-質問の発話に，詳細な情報を付与するために *Pause*，*WH* の 2 つのサブレイヤーを *General* レイヤーの下に設けた．なお，*General* レイヤーのタグを “必須タグ”，*Response*, *Check*, *Constrain*, *ActionDiscussion* レイヤーのタグを “任意タグ” と呼ぶ．各レイヤーの発話行為タグは，基本的に他のレイヤーのタグ情報に影響されることなく独立に付与されるが，各レイヤーではタグを複数付与することはできない．また，*General* を除く各レイヤーでは，該当するタグが存在しない場合に ‘null’ をとることができる．

以下に各レイヤーを説明する．

General このレイヤーのいずれかのタグが各発話に対して必ず一つ付与される．大きく「質問」「断片」「発言」の 3 種類に分類される．

Response 特定の発話を対象とした応答発話に対応するレイヤー．肯定・否定など相手の発話にどのように応答したかを記述するタグを含む．

Check 話し手が自分の発話に対する聞き手の応答に対して，一定の予測を持った上で発話されている確認を扱うレイヤー．

Constrain 条件・理由や評価など対象となる発話の意味を限定・付与する発話を扱うレイヤー．

ActionDiscussion 希望や要求など発話者・聞き手の将来の行動に対して，何らかの拘束を与える発話を扱うレイヤー．

Others 挨拶など対話の細かい諸相を捉えるためのレイヤーである．倒置表現など，必ずしも発話行為とはいえない機能も記述するタグも含まれている．たとえば，**Greeting**, **SelfTalk**, **Welcome**, **Apology** などである．

Pause ポーズが，ターン管理の観点からどのように機能しているかを記述するタグを含むサブレイヤー．General レイヤーにて **Pause** タグが付与されたすべての発話に対して，**Hold**, **Grabber**, **Holder**, **Releaser** のいずれかのタグが付与される．

WH WH-質問のタイプを記述するサブレイヤー．**When**, **Where**, **Who**, **How_much** タグなどが含まれる．

これらの発話行為タグをこれまで対面対話 36 対話，非対面対話 20 対話に対して付与しており，さらに現在もタグを付与している．より詳細な分析などについては，文献 [Oht09a] を参照されたい．ラベラー間の一致率を Kappa 値により調べたところ，General レイヤーのみの場合で 0.74，全レイヤーで 0.68 であった．

5 意味内容タグ

この節では，意味内容タグの設計について説明する．意味内容を表現するために，文節を単位とする依存構造を用いて，各文節に意味クラスを割り当てる．また，意味クラスの他にも，付加的な情報を記述する．

意味内容タグは，発話の意味を直接的に記述しようとしたものではなく，発話に含まれる単語に対してその属性を与えようとするものである．京都観光案内の計画立案において，重要な属性は，固有名ならびに，数値や日付，時刻などをあらわす固有表現である．しかしながら，一般的な名詞であっても，対象とするドメインによってその重要度は異なるため，固有表現に限定せず重要な表現を網羅的に収集する必要があると考える．また，観光計画立案という相談対話において，ユーザの好みや，観光地の候補に対する印象などの表現も，

対話システムを駆動していく上で重要である．

本研究では，観光案内における相談対話に対して意味内容タグを付与する場合に，(1) 決定事項，(2) 決定要因，(3) 印象評定，(4) コンサルティングの 4 要素が重要であると考える．決定事項は，観光計画を立てる上で最終的に決定する事項，たとえば，観光するお寺などのスポットや，交通手段などである．決定要因は，決定事項を確定する上で，その要因になり得るもの，たとえば，桜が有名であるとか，混雑していないなどを指す．また，印象評定は主に決定要因に対する利用者，あるいはガイドの印象を述べた表現を指す．その他，一般的な情報伝達をコンサルティングとしてとらえる．意味内容タグのための意味クラスの設計にあたっては，これら 4 つの事象を意識して意味クラス階層を定義した．

5.1 意味内容タグセット

意味内容タグは，具体的には，依存構造上の単位への意味クラスラベルとして記述される．意味クラスはネットワーク形状の階層構造を持っている．階層構造の最上位には，33 のクラスが存在する．たとえば，activity, event, meal, spot, cost などである．それぞれのクラスは子を持ち，子には，さらに子を持つノードと子を持たない葉の 2 種類がある．具体的なタグ付与の例を表 1 に発話行為タグとともに示す．表中では，1 文節を 1 行ずつ記載しているが，依存関係が省略されており，各文節に付与された意味クラスのパスを明記している．また，付加情報も併記している．

現在，タグセットの設計をほぼ終えているが，実際にタグ付けをしながら，最終調整を行っている．また，階層構造を全展開すると，まったく使用されないであろうパスも含めて 26,800 ほどのパスが存在する．これまでににタグ付けした対面対話 36 対話のデータでは，2,000 ほどのパスのタグが使用されている．

5.2 意味内容タグ付与

ここでは，意味内容タグがどのように付与されているかその概要を説明する．

発話行為タグは，節を単位としてタグが与えられたが，意味内容タグは依存構造に基づいてタグを付与するため，節を単位とすると，同一処理単位内に係り先が存在しないことが多くなり，タグ付けがしにくくなる．そこで，書き起こしの際に付与された句点を用いて文を認定し，これを発話行為タグの単位とする．各文に対して形態素解析，依存構造解析を行なう．依存構造解析結果まで含めたデータを 1 文 1 ファイルとして，格納する．意味内容タグを付与した後に，発話行為タグで用いた単位とのアライメントをとる．

形態素解析器には chasen ，依存構造解析器には cabocha を用いている．形態素解析辞書は，ipadic-2.6.3 を元に話し言葉の解析に向けてチューニング (主に接続表の修正) されている．辞書項目は，固有名詞などが拡張され，80 万形態素ほどのサイズである．また，依存構造解析器は，京都テキストコーパス[1] ならびに，ATR 音声対話コーパス，IPAL 辞書の例文など計約 8 万文で学習したモデルを使用している．

依存構造解析結果を読み込み，文節単位で，意味内容タグを付与するための専用のツールを開発し，利用している．意味内容タグの付与は，文節単位で行なうため，タグ付与の際に文節認定が過っているとわかった場合，これを修正する．また，依存構造解析結果，形態素解析結果も同じように修正可能になっている．したがって，意味内容タグの付与を行ないつつ，形態素解析，文節まとめあげ，依存構造解析のすべての結果について，検証し，修正を行なうことになる．

現在までに 40 対話ほどにたいして意味内容タグを付与した．さらに現在もタグを付与している．

6 京都観光案内対話コーパスの利用

この節では，NICT で現在構築をすすめている京都観光案内対話コーパスを実際に音声対話システムのために利用することについて述べる．

音声対話システムは，非常に複雑なシステムである．ここでは，音声対話システムは大きく分けて音声認識，対話管理，音声合成の 3 つのモジュールと，タスク遂行のための各種データベースから構成されると考える．そして，これらの全てを音声対話コーパスが支える．図 1 に関係を示す．

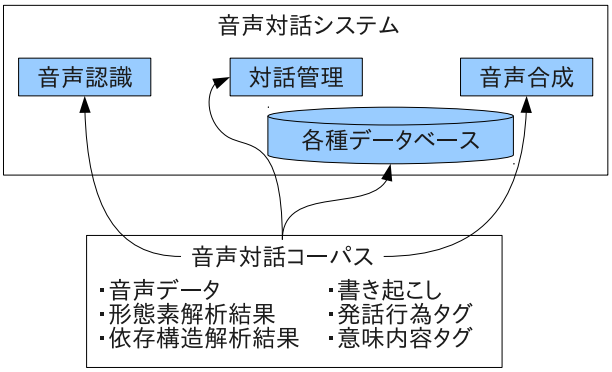

図 1: 音声対話システムと音声対話コーパス

6.1 音声認識

現在，我々はコーパスの形態素解析結果を用いて言語モデルを構築し音声認識に直接利用している．固有表現などは，辞書に含めておかないとその認識が難しいが，それ以外の一般的に思える表現であっても，n-gram

[1] http://nlp.kuee.kyoto-u.ac.jp/nl-resource/corpus.html

として網羅しなければ，ユーザが思うように認識できない．また，音声データを用いて音響モデルを構築することも考えられるが，収録の際に，対話の自然性を重視したために，音響環境がその他のモデル構築データと離れていることもあり，音響モデル構築に音声対話コーパスを現在は用いていない．

6.2 対話管理

対話管理において一番の問題となるのは，ユーザ発話の理解である．ユーザの多様な発話をいかに理解し，システムの次の動作をいかに決定するかは大きな問題である．我々は，コーパスからユーザ発話の理解，そしてガイドの動作と発話の系列を統計的に扱い，対話管理を行なおうと考えている．そのために，対話行為タグと呼ばれる発話行為タグと，意味内容タグという2種類のタグからなるタグセットを設計し，コーパスに付与している．この対話行為タグを用いて対話管理を行なうが，実際には，システムの動作に対応するアノテーションが必要であり，現状では，そのままコーパスから直接的に対話管理を行えるだけの情報が揃っているとは言えない．これについては，現在検討中である．

また，実際にタスクが明確な対話をすすめるためには，サービスを提供するための各種データベースが必要になる．特に，観光案内タスクにおいては，収録した人間同士の対話に，ガイドが提供した経験に裏打ちされた知識が豊富に含まれており，これをマイニングし，データベースとして整備することができる．

6.3 音声合成

一方的な，情報伝達のみを目的とした音声合成ではなく，より自然な対話のための音声合成を考える．既存の音声合成ソフトウェアは，テキストを読み上げるという点では，十分に高性能と言えるかもしれないが，音声対話システムにおいては，同じテキストであっても状況に応じて読み分けられることが求められる．たとえば,「はい.」を音声合成する場合として，相槌，ターン取得，承認（[はい，わかりました」の「はい」）などいくつかのパターンが考えられ，異なる音声を合成することで不自然さを回避できる．

我々は，実際の人間同士の収録においてガイドの音声をそのまま音声合成に適用できる程度の品質で録音し，音声合成装置を構成することを考えている．コーパスには，発話行為タグなどの発話機能を示すタグがあるので，これによって，同一表現でも異なる機能を持つ発話を区別することができる．音声合成時にこれらの機能をあわせて指示することで，同一表現であっても，異なる調子で読み上げさせることが可能になる．NICTでは，2009年度にこのような目的で，非対面対話を40対話収録しており，同様にコーパス整備をすすめている．ただし，観光ガイドは，発声という点では，あくまでも素人であるので，発声のぶれが大きいことが考えられる．そのため，書き起こしから台本を作成し，ガイドとユーザを声優に演じてもらい，それを収録し，対話システムのための音声合成装置で使用するデータを構築することも進めている．

7 むすび

本稿では，NICT 京都観光案内対話コーパスの概要を紹介し，統計的手法による対話管理を前提としたタグ付けについて概観した．対話を記述するために対話行為に着目し，それを表現するために発話行為タグと意味内容タグを設計し，コーパスに付与している．また，タグ付与結果を含む音声対話コーパスをいかに対話システムに利用するかについて言及した．今後は，タグ付けを大規模に行ない，大量のタグ付きコーパスを整備するとともに，対話システムとして実際に動作する際に必要となる自動識別器をコーパスから統計的手法によって構築する予定である．

参考文献

[Bun00] BUNT, H.: Dialogue pragmatics and context specification, In BUNT, H. and BLACK, W., editors, *Abduction, Belief and Context in Dialogue*, pp. 81–150, John Benjamins (2000).

[Hor08] HORI, C., OHTAKE, K., MISU, T., KASHIOKA, H., and NAKAMURA, S.: Dialog Management using Weighted Finite-state Transducers, In *Proc. Interspeech*, pp. 211–214 (2008).

[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).

[Lev97] LEVIN, E., PIERACCINI, R., and ECKERT, W.: Learning dialogue strategies within the Markov decision process framework, In *Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp. 72–79 (1997).

[Oht08] 大竹清敬, 堀智織, 柏岡秀紀, 中村哲：京都観光案内対話コーパスにおける対話行為の分析, 言語処理学会第 14 回年次大会 発表論文集 (2008).

[Oht09a] OHTAKE, K., MISU, T., HORI, C., KASHIOKA, H., and NAKAMURA, S.: Dialogue Acts Annotation to Construct Dialogue Systems for Consulting, In *Proceedings of FIRST INTERNATIONAL WORKSHOP ON SPOKEN DIALOGUE SYSTEMS TECHNOLOGY* (2009).

[Oht09b] 大竹清敬, 翠輝久, 堀智織, 柏岡秀紀, 中村哲：統計的手法による対話管理のための発話行為と意味内容タグ, 言語処理学会 第 15 回年次大会 (2009).

[Shr04] SHRIBERG, E., DHILLON, R., BHAGAT, S., ANG, J., and CARVEY, H.: The ICSI Meeting Recorder Dialog Act (MRDA) Corpus, In *Proc. 5th SIGdial Workshop on Discourse and Dialogue*, pp. 97–100 (2004).