

## 京都観光案内対話コーパスにおける 対話行為タグの設計と分析

翠 輝久 大竹 清敬 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 / 国際電気通信基礎技術研究所

〒 619-0288 京都府「けいはんな学研都市」光台 2-2-2

e-mail: teruhisa.misu@nict.go.jp

**あらまし** 我々は、人間同士が行っているような自然な対話を行う対話制御モデルを、対話コーパスから学習することを目指している。そのためには、コーパスには対話の流れに影響を与える諸相が詳細に記述されている必要がある。本研究では、コーパス中の発話を汎化し記述するために、対話行為に着目し、これを記述するために2種類のタグの設計を行った。一つは発話行為(発語内行為)タグであり、発話の情報伝達機能を記述する。もう一つが、意味内容タグであり、発話の中に含まれる語句の意味を記述する。これらのタグを京都観光案内対話コーパスに対して付与し、対話の諸相の分析を行った。

## Dialog Act Annotation Scheme for the Kyoto Tour Guide Dialog Corpus

Teruhisa Misu Kiyonori Ohtake Chiori Hori  
Hideki Kashioka Satoshi Nakamura

NICT, MASTAR Project/ATR,

2-2-2 Hikaridai, "Keihanna Science City," Kyoto, 619-0288, Japan

e-mail: teruhisa.misu@nict.go.jp

**Abstract** Our goal is to train a dialog manager that can handle complex dialog flows and spontaneous user utterances from a tagged dialog corpus. The corpus must contain rich information enough to describe factors that affect the flow. In this work, we adopt dialog acts (DAs) for this information and provide our taxonomy that can describe two aspects of DA. One is a communicative function (speech act), and the other one is a semantic content of an utterance. We have annotated the speech act tags on the part of our corpus by several annotators.

## 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、コーパスに付与する対話行為タグについて述べる。近年、音声対話システムを構成する個々の要素技術（たとえば、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そのものを、ルールなどで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。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、同調的な音声対話システムの構築を目指している。それは、既存の音声対話システムが提供するタスク達成のための機能だけではなく、人同士が行っているように自然に対話を実現するものである。このような対話システムを実現するためのアプローチとして、人同士による複雑な対話を人手により詳細に記述する方法は、非現実的である。また、画一的なルールによる対話システムを構築するのではなく、学習に利用するコーパスを変更することによって、システムそのもののふるまいを制御できる枠組みが望まれる。したがって、対話システムの動作はコーパスからボトムアップに学習することが望ましい。コーパスが大量に利用できる場合には、対話をMDPやPOMDP[1]によりモデル化して強化学習を用いて学習することができる。

我々は、入力シンボルをユーザの発話とし、出力シンボルをシステムの動作とするWFSTを用いた統計的な対話制御手法を提案している[2]。このようなアプローチを用いてモデルを学習するためには、入力シンボルとなるユーザの発話がある程度汎化した状態で、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

この節では、我々が現在構築している京都観光案内対話コーパスについてその概要を述べる。

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う2者による対話からなる。

現在までに、対面での対話(114対話)、非対面での対話(20対話)、Wizard of OZ(WOZ)形式での対話(40対話)を収録している。1対話は約30分である。現在、対面対話の全対話ならびにWOZ形式と非対面対話の4分の3ほどが書きおこされている。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、1秒以上のポーズによってセグメントを認定している。

対面対話の例を図1に示す。この例は、ユーザが訪れる観光スポットを決める対話の一部である。コーパスには、スポットを決める対話の他に、レストランやスポット間の交通手段を決める対話が含まれている。このような、相談を行う対話は、ホテルの予約などの単純なフォームフィリングとは異なり、一つの決定を行うまでに多様な発話行為がなされるため、従来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いえる。

## 3 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、人間の自然な発話を柔軟に解釈できる必要がある。また、ガイド(システム)はユーザに対してプロアクティブに働きかけ(推薦など)を行い、対話を進める必要がある。

コーパス中のこれらの現象を記述するための手段として、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する。対話行為は、発話者が聞き手の解釈を通じてどのように対話を制御しようとしているかを記述するものであり、従来から盛んに研究が行われてきた。Bunt[3]は、この対話行為の2つの側面を指摘している。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。もう一つは、情報伝達機能(communicative function)であり、どのような意図で発話されているかという側面である。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。一つは発話行為(発話内行為)タグであり、発話の情報伝達機能を記述する。もう一つが、意味内容タグであり、発話の中に含まれる語句の意味を記述する。

## 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説明する。

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	地蔵が 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),ksoa (dmst),noun around="yes" pred="copula"
58	81358-81841	Guide	大原は、 ちょっと	State.Inversion	location
59	81386-82736	User	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。

図 1: 対話コーパスに対するタグ付けの例

表 1: タグの一覧と出現頻度

タグ名	頻度 (%)		タグ名	頻度 (%)		タグ名	頻度 (%)	
	User	Guide		User	Guide		User	Guide
<b>(General)</b>			<b>(Response)</b>			<b>(ActionDiscussion)</b>		
Statement	45.25	44.53	Acknowledgment	19.13	5.45	Opinion	0.52	2.12
Pause	12.99	15.05	Accept	4.68	6.25	Wish	1.23	0.05
Backchannel	26.05	9.09	PartialAccept	0.02	0.10	Request	0.22	0.19
Y/N-Question	3.61	2.19	AffirmativeAnswer	0.08	0.20	Suggestion	0.16	1.12
WH-Question	1.13	0.40	Reject	0.25	0.11	Commitment	1.15	0.29
Open-Question	0.32	0.32	PartialReject	0.04	0.03	<b>(Constrain)</b>		
OR-after-Y/N	0.05	0.02	NegativeAnswer	0.10	0.10	Reason	0.64	2.52
OR-Question	0.05	0.03	Answer	1.16	2.57	Condition	0.61	3.09
Statement==	9.91	27.79	<b>(Check)</b>			Elaboration	0.28	4.00
			RepetitionRequest	0.07	0.03	Evaluation	1.35	2.01
			UnderstandingCheck	0.19	0.20			
			DoubleCheck	0.36	0.15			
			ApprovalRequest	2.01	1.07			

#### 4.1 タグ付けの単位

対話行為タグの基本単位については、これまでに様々な議論がなされてきた。最も単純な単位として、文や発話といった単位が考えられる。しかしながら、人同士の対話には、必ずしも明示的な文末表現が現われるとは限らず、境界が明確でない。また、単位を長くすると複数の対話行為が含まれやすいといった問題がある。そのため、対話中の事象を正確に捉えるためには、意味のまとまりを持つできるだけ短い単位に対してタグを付与することが望ましい。そこで、話し言葉の節単位認定を目的に開発されたCBAPを用いて発話の節単位を認定し、タグ付けの単位とする。

#### 4.2 タグセットの設計

タグセットの設計には、二つの方針がある。一つは、タグ付け単位に対して最も適切なタグを一つだけ付与するアプローチであり、談話タグワーキング

グループの発話内行為タグ [4] や AMI コーパスの DA タグなどがこれに該当する。もう一つは、タグ付け単位に対して該当する全てのタグを付与するアプローチであり、DAMSLや、MRDA[5]などがこれに該当する。本研究では、対話の同調性の一つの側面は発話の多機能性にあると考え、発話の情報をできるだけ多く記述するために後者のアプローチを採用する。

MRDA タグセットを拡張し、*General*, *Response*, *Check*, *Constrain*, *ActionDiscussion*, *SocialObligation*, *Others* の 7 レイヤーと、*Pause*, *WH* の 2 つのサブレイヤーからなるタグセットを作成した。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”と呼ぶものとする。2 つのサブレイヤーは、*Pause* および *WH-Question* の発話に、詳細な情報を付与するためのものである。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*以外のレイヤーは、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。

**General** 各発話に対して必ず一つ付与される必須タグのレイヤーであり、タグは大きく「質問」「断片」「発言」の3種類に分類される。また、発話が次の節に継続している場合には、“Statement==”タグを付与する。

**Response** 聞き手の特定の発話を対象とした応答発話に付与されるタグのレイヤーであり、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。

**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記述するレイヤーである。

**Constrain** 条件・理由や評価など対象となる発話の意味を限定・付与する発話を記述する。

**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記述する。

**SocialObligation** 挨拶や謝罪など円滑にコミュニケーションを行うための発話行為を記述する。

**Others** 相手発話の模倣や、補完など対話の細かい諸相を捉えるためのタグのレイヤーである。倒置表現など、必ずしも発話行為とはいえないタグも含まれている。

**Pause** *Pause*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。*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-Question*の質問タイプを記述する。(When, Where, Who, How\_much など)

本タグセットは、相談型の対話を行うシステムの対話管理を統計的手法によって実現することを前提として設計している。ユーザがシステムと相談し、意思決定などを行う過程を、可能な限り詳細に記述することができる。たとえば、“確認”という発話行為は、*Check*レイヤーにおいて、*UnderstandingCheck*と*RepetitionRequest*の2種類の発話行為に分類される。*UnderstandingCheck*が、多くの場合、発話意図の明確化のための質問として用いられるため、対話の流れを記述する観点から重要な機能である。

それに対し、*RepetitionRequest*は相手の発話の聞き取れなかった箇所の再発話の要求であり、対話の流れに直接関係なく任意の場所に出現する可能性がある。したがって、同じ“確認”でも区別すべきと考えた。

*SocialObligation*, *Others*レイヤーのタグ以外のタグ一覧を表1に、コーパスに対するタグ付けの例を図1に示す。

### 4.3 タグを用いたコーパスの分析

設計したタグセットを用いて、対面対話30対話に対して発話行為タグの付与を行った。タグ付けを行う際には、テキスト情報に加えて発話時間と音声情報を利用して判断している。また、ラベラー間でクロスチェックを行っているため、前回の報告時[6](ラベラー間の一致率が約70%)よりも高い精度で付与できていると期待される。

#### 4.3.1 使用頻度の分析

付与されたタグの出現頻度を表1に併記する。*General*レイヤーにおいて、発話全体の5割近くが*Statement*に分類されているが、これらの発話の過半数に任意タグが付与されていたため、頻度の偏りは特に問題にならないと考える。

#### 4.3.2 対話の過程におけるタグの出現傾向の分析

コーパス中で行われている相談型対話の特徴を調べるために、対話のフェーズごとに出現するタグの傾向を調べた。本コーパスにおいて、ユーザは観光の内容を決定し、旅程表に記入している。決定事項には訪問するスポット、スポットで行う観光行為、スポット間の交通手段などがある。このうちスポットを決定する対話の区間(エピソード)を12対話に対して認定した。認定された62エピソードのうち、発話数が50以下のものを除いた37エピソードを対象に対話の進行状況とタグの出現傾向の分析を行った。これは、発話数が少ないエピソードでは、ユーザが希望するスポットを最初から決めている場合が多く、決定に際してガイドとの相談がほとんど行われていないためである。発話数が均一になるように各エピソードを5つの区間に区切り、各区間におけるタグの出現比率を求めた結果を図2に示す。図中のタグの出現比率は、各タグの該当区間における出現頻度の、全区間で出現した総数に対する割合である。

分析の結果、タグの出現傾向は、大きく3つのパターンに分類できることがわかった。一つ目は、エピソードの前半に集中して出現するものであり、*Open*



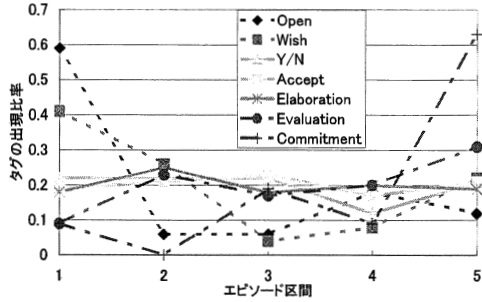


図 2: エピソードの進行とタグの出現傾向の関係

Question, WH-Question, Wishなどが該当する。二つ目は、エピソードの段階によらず出現頻度が一定のものであり、Y/N-Question, Accept, Elaborationなどが該当する。三つ目は、タグがエピソードの後半部分に多く出現するパターンであり Evaluation, Commitment, Opinion タグなどが該当する。これらの出現傾向から、エピソード前半でガイドが Open, WH-Question によりユーザに希望の伝達 (Wish) するように働きかけ、ユーザがスポットに対する評価 (Evaluation, Opinion) を行い、決定に至る (Commitment) 相談型対話の流れが読み取れる。

#### 4.3.3 タグ推定に関する検討

実際にコーパスを利用して音声対話システムの対話制御を行うためには、ユーザの発話内容から、発話意図 (=対話行為タグ) を推定する必要がある。自動付与を行う際の予測の難しさを調べるために、初期分析として 10-fold クロスバリデーションにより、タグのテストセットパープレキシティを求めた。タグの種類は、すべてのレイヤーの組み合わせになるため、定義上は約 5,000 通りが考えられるが、実際にタグ付けた 30 対話では 125 種類タグが使用されていた。なお、パープレキシティの計算には Backchannel と Pause タグを除外している。この結果を表 2 に示す。

必須タグのみのパープレキシティは、4.25 であった。任意タグの中では、Response, Constrain レイヤーのタグを計算に使用した場合にパープレキシティの上昇率が大きかった。

## 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こととする。

表 2: 発話行為タグのテストセットパープレキシティ

評価対象	Perplexity
必須タグのみ	4.25
必須タグ + Response レイヤーのタグ	6.84
必須タグ + Check レイヤーのタグ	4.63
必須タグ + Constrain レイヤーのタグ	6.19
必須タグ + ActionDiscussion レイヤーのタグ	5.20
必須タグ + 任意タグすべて	14.75

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、そのような属性のうちの重要なものの多くは固有名ならびに、数値や日付、時刻などをあらかず固有表現である。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、固有表現のみならず、幅広く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は、また同様に重要である。

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 つ事象が重要であると考えた。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。

一方で、このような意味を記述しようとするために、Lexical Functional Grammar (LFG)[7] や HPSG[8] の枠組みによるアプローチも考えられる。しかしながら、発話が本来断片的であることを鑑みると、より頑健な処理を実現するためには、断片的な表現からその意味クラスを推定できるような枠組みが望ましいと考えた。そのため、特定の意味処理の枠組みを前提にせず、文節単位の依存構造(データ)へ直接意味クラスを付与するアプローチを取る。

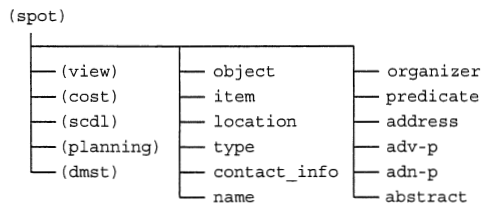


図 3: spot クラスの構造

## 5.1 意味内容タグセットの設計

意味内容タグのための意味クラスは階層構造を持っている。階層構造の最上位には、32のクラスが存在する。たとえば、activity, event, meal, spot, transportation, 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。図3に spot クラスの構造を例示する。ここで、丸括弧 ( ) で囲まれたものはノードを意味し、囲まれていないものは葉を表現している。具体的なタグ付与の例を図1に示す。

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 21,900 ほどのパスが存在する。現在、試験的にタグ付けしている 10 対話のデータでは、1,380 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

意味内容タグの処理単位は発話行為タグと同様に、節を単位として分割されたものとする。これをセグメントと呼ぶ。各セグメントに対して形態素解析、依存構造解析を行う。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。

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、文節単位で行うため、タグ付与の際に文節認定が過っているとかかった場合、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っている、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、修正を行うことになる。

## 6 むすび

本稿では、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。発話を近似するために対話行為に着目し、対話行為を表現するために2種類のタグ、発話行為タグと意味内容タグを設計した。発話行為タグを30対話に対して付与し、タグの出現傾向について分析を行い、コーパスの特徴を調べた。また、意味内容タグに関しては、設計をほぼ終えつつあり、大規模なタグ付けを開始する準備が整いつつある状況である。今後は、タグ付けを大規模に行い、大量のタグ付きコーパスを整備する。さらに、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。

## 謝辞

タグ設計に際して、多くの有益な助言を頂いた情報通信研究機構の新美康永先生、及び京都大学の河原達也先生に深く感謝します。

## 参考文献

- [1] J. Williams and S. Young. Scaling POMDPs for Spoken Dialog Management. *IEEE Trans. on Speech and Audio Processing*, Vol. 15, No. 7, pp. 2116–2129, 2007.
- [2] C. Hori, K. Ohtake, T. Misu, H. Kashioka, and S. Nakamura. Dialog Management using Weighted Finite-state Transducers. In *Proc. Interspeech*, pp. 211–214, 2008.
- [3] H. Bunt. Dialogue pragmatics and context specification. In Harry Bunt and William Black, editors, *Abduction, Belief and Context in Dialogue*, pp. 81–150. John Benjamins, 2000.
- [4] 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.
- [5] E. Shriberg, R. Dhillon, S. Bhagat, J. Ang, and H. Carvey. The ICSI Meeting Recorder Dialog Act (MRDA) Corpus. In *Proc. 5th SIGdial Workshop on Discourse and Dialogue*, pp. 97–100, 2004.
- [6] 翠輝久, 大竹清敬, 堀智織, 柏岡秀紀, 中村哲. 音声対話システム構築のための発話行為・意味内容タグの設計. *人工知能学会研究会資料*, SIG-SLUD-A802-6, 2008.
- [7] M. Dalrymple, R. Kaplan, J. Maxwell III, and Annie Zaenen, editors. *Formal Issues in Lexical-Functional Grammar*. CSLI Publications, 1994.
- [8] C. Pollard and I. Sag. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press, 1994.