

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

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/ ATR 音声言語コミュニケーション研究所

{kiyonori.ohtake, teruhisa.misu, chiori.hori, hideki.kashioka,
satoshi.nakamura} @ {nict.go.jp, atr.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、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のある程度汎化して、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う 2 者による対話からなる [大竹 08]。現在までに、対面形式での 114 対話、非対面形式での 20 対話、Wizard of OZ(WOZ) 形式での 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、1 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、我々是对話行為に着目し、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。

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。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。

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

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 頭に	State==	media adv-p adv-p object
56	80069-80926	User	入ってない。	State_Reason→54	predicate
57	80788-81358	Guide	この 辺ですね。	State_Answer→54	(dmst),koso
58	81358-81841	Guide	大原は。	State_Inversion	location
59	81386-82736	User	ちょうど 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。

表 2: タグの一覧と予備実験における出現頻度

タグ名	頻度 (%)	タグ名	頻度 (%)	タグ名	頻度 (%)	タグ名	頻度 (%)
(General)		(Response)		(Check)		(ActionDiscussion)	
Statement	58.55	Acknowledgment	12.43	RepetitionRequest	0.26	Opinion	1.50
Pause	18.52	Accept	7.56	UnderstandingCheck	1.64	Wish	0.76
Backchannel	15.55	PartialAccept	0.24	DoubleCheck	0.55	Request	0.12
Y/N-Question	4.68	AffirmativeAnswer	0.24	ApprovalRequest	1.66	Suggestion	0.38
WH-Question	1.24	Reject	0.40	(Constrain)		Commitment	1.45
Open-Question	1.16	PartialReject	0.05	Reason	1.88		
OR-after-Y/N	0.24	NegativeAnswer	0.19	Condition	1.47		
OR-Question	0.07	Answer	2.21	Elaboration	3.73		
				Evaluation	4.90		

表 3: 一致率によるタグセットの評価

評価対象	必須 タグのみ	必須タグ、任意タグ サブレイヤーのタグ全て
2 者一致率	86.3%	67.9%
3 者一致率	79.9%	55.9%
κ 統計量	0.74	0.62

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、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。また、タグを音声対話システムの入力として使用するためには、機械的に認定可能な単位である必要がある。そこで、話し言葉の節単位認定を目的に開発された CBAP[Mar04]を用いて発話の節を認定し、タグ付けの単位とする。

本研究では、書き起こしテキストを形態素解析し、CBAP を適用するとともに、音声にフォースアライメントを適用する。CBAP によって認定された節と、フォースアライメントの結果を照合することによって、各節の開始ならびに終端時刻を推定し、これらの節をタグ付けの単位とする。

4.2 発話行為タグの設計

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

るもの。談話タグワーキンググループの発話内行為タグ [荒木 99] や AMI コーパスの対話行為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ものである。DAMSL や、MRDA[Shr04]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方針を採用する。

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、ポーズおよび 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”と呼ぶものとする。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* 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．

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．

4.3 タグセットの評価

タグセットの評価を行うために3名のラベラー(1名は筆者である)により対話コーパス中の同一データに対してタグを付与した．評価セットとして，対面対話3セッションの前半部分(計45分，節数1,709)を用いた．タグ付けを行う際には，テキスト情報に加えて音声情報を利用して判断した．

4.4 使用頻度の分析

付与されたタグの出現頻度を表2に併記する．General レイヤーにおいて，発話全体の6割近くがStatementに分類されているが，これらの発話の66%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．また，Constrain レイヤーのタグが付与された発話の割合が比較的多いことは，ガイドが決定事項を提示し，ユーザが評価しながら決定を下すという流れが多い本コーパスの特徴を表している．

4.5 一致率による評価

次に，タグの安定性，再現可能性を評価するためにラベラー間でのタグの一致率を調べた．評価尺度には2者一致率，3者一致率と κ 統計量を用いた．この結果を表3に示す．なお，図中の2者間一致率は，3者間で可能な3通りの組み合わせの平均値である．

必須タグは，表層表現や音声から比較的容易に判別可能であるため，高い一致率を得られている．任意タグならびにサブレイヤーのタグを考慮に入れた場合に一致率が低下しているが， κ 統計量は0.62であり文献[荒木99]などよりも難しい対話を扱っていることを考慮すると，比較的高い一致率を得られているといえる．

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．

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)[Dal94] や HPSG[Pol94] 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．

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木構造の階層構造を持っている．階層構造の最上位、つまりルートの直下には、32 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 に示す．表中では、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意味内容タグとして明記している．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 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

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．

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．依存構造解析結果まで含めたデータを 1 セグメント 1 ファイルとして、格納する．また、話者役割、発話開始時刻、発話終了時刻などは別途メタファイルにて管理する．

形態素解析器には chasen¹、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．また、発話行為タグのタグ付与結果について、複数人による一致率などを概観し、概ね良好な結果が得られていることを確認した．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．

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).
- [Dal94] DALRYMPLE, M., KAPLAN, R. M., III, J. T. M., and ZAENEN, E. A., editors: *Formal Issues in Lexical-Functional Grammar*, CSLI Publications (1994).
- [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).
- [Lev97] LEVIN, E., PIERACCINI, R., and ECKERT, W.: Learning dialogue strategies within the Markov decision process framework, In *Proc. ASRU*, pp. 72–79 (1997).
- [Mar04] 丸山岳彦, 柏岡秀紀, 熊野正, 田中英輝: 日本語節境界検出プログラム CBAP の開発と評価, *自然言語処理*, Vol. 11, No. 3, pp. 39–68 (2004).
- [Pol94] POLLARD, C. and SAG, I. A.: *Head-Driven Phrase Structure Grammar*, The University of Chicago Press (1994).
- [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).
- [荒木 99] 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).
- [大竹 08] 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 会年次大会発表論文集* (2008).

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