

大規模音声対話コーパスを用いた対話理解

松原 茂樹[†] , 加藤 真吾^{††} , 山口 由紀子[†] , 河口 信夫[†]

[†] 名古屋大学情報連携基盤センター ^{††} 名古屋大学大学院情報科学研究科
{matubara,gotyan,yamaguchi,kawaguti}@el.itc.nagoya-u.ac.jp

自然な音声対話システムの実現のために、これまでに多くの音声対話コーパスが作成されてきたものの、対話理解においてそれらを効果的に利用する方法については、必ずしも明らかではない。本論文では、対話システムにおけるコーパスの実践的な利用について述べる。著者らが開発を進めているコーパスベース対話システムについて概説し、対話理解で使用するコーパスとその利用法について説明する。また、システムの高度化のために進めているタグアノテーションとして、発話文構造タグ、及び、対話構造タグの設計と付与について述べるとともに、その活用法について論じる。

Corpus-based Spoken Dialogue Understanding

Shigeki Matsubara[†] Shingo Kato^{††} Yukiko Yamaguchi[†] Nobuo Kawaguchi[†]

[†] Information Technology Center, Nagoya University

^{††} Graduate School of Information Science, Nagoya University

{matubara,gotyan,yamaguchi,kawaguti}@el.itc.nagoya-u.ac.jp

Several spoken dialogue corpora have been made for the achievement of spontaneous speech dialogue systems so far. However, the method of effectively using them for the dialogue understanding is not necessarily clear. This paper describes practicing use of the corpus in the dialogue processing. It outlines a corpus-based dialogue system to which we are advancing development, and explains the corpus for the dialogue understanding and its utilization. Moreover, we describe the design of the sentence structural tags and the dialogue structure tag as advanced tag annotation for the upgrade of the system.

1 はじめに

自然な音声対話システムの実現を目指し、これまでに多くの大規模音声対話コーパスが構築されてきた。多様な対話音声を大量に収集することにより、音声認識や音声合成におけるロバスト性や自然性を向上のための統計データとして利用するなど、音声処理技術の開発に必要な不可欠なリソースとなっている。

一方、言語処理や対話処理では、人手で作成したルールやテンプレートを用いることが多い。コーパスの利用といっても、対話データを参照してルールを作成したり、他にも、言語的現象の分析や対話システムの評価など、システム開発の観点からは間接的な利用がほとんどである¹²⁾。最近では、コーパスへの豊富なタグ付けを背景に、対話処理におけるコーパスの利用法がいくつか提案されているものの、その効果的な方法はまだ明らかになっていない。対話システムにおけるコーパスの利用には、以下の二つのアプローチがある。

- 統計データとしての利用^{1, 3, 4)}
コーパスから統計的に獲得した対話モデル(単

語生起、言語文法、発話行為等の確率モデル)を利用して対話処理を実行する。タスクが一定の対話データを大量に使用することにより、精緻な対話モデルを学習することができる。

- 事例データとしての利用^{12, 13)}
コーパスから入力発話や対話状態に合致した発話等のデータを取り出し、それを利用して対話理解・生成を実行する。あらゆる発話や状態に対応するデータが存在すれば、高い精度での対話処理が可能になる。

どのようなアプローチが適しているかは、利用可能な対話データ、適用する処理モジュール等に依存するため、二つのアプローチの混合も含め、適切な利用法を選択することが重要となる。

本論文では、対話システムにおけるコーパスの実践的な利用について述べる。これまでに我々が開発しているコーパスベース対話システムについて述べる。システムは、発話理解、対話制御、発話生成から構成され、モジュールのすべてにおいて対話コーパスを用いて処理を実行する。コーパスとして、名

古屋大学 CIAIR 車内音声対話データベースを使用している。書き起こしデータの各発話には階層的な話者意図タグが付与されており (35,411 発話), 対話の流れを発話意図の系列として制御することができる。

また本論文では, 我々がシステムの高度化のために進めているタグアノテーションについて述べる。対話における構造的情報の重要性に着目し, 文構造タグ (10,995 ターン), 及び, 対話構造タグ (8,150 発話) を付与している。それぞれ, 発話意図推定, 次発話予測に用いた実験を行っており, 現在, システムにおいて実践的に利用するための検討を進めている。

本論文の構成は以下の通りである。次節では, コーパスベース対話システムについて説明する。3 節では, コーパスに付与した新たなタグについて述べる。4 節では, 対話システムにおけるコーパスの利用について論じる。

2 コーパスに基づく音声対話処理

実走行車内での自然な音声対話の実現を目指し, 著者らはこれまでに車内音声対話コーパスを利用した対話システムの開発を進めている。本節では, 使用した対話データ, 及び, システムの実装について概説する。

2.1 車内音声対話コーパス

コーパスベース対話システムを実現するためのデータとして, 名古屋大学 CIAIR 車内音声対話コーパス⁹⁾を使用した。このコーパスは, 実走行環境下でのロバストな車内音声対話処理の高度化を目的として構築されており, 道路案内や情報検索をタスクとするドライバとオペレータとの会話を収録している。ロバストな音声認識・理解技術を開発するために, ドライバ役として 800 名を超える被験者を使用しており, 運転作業中の多様な音声・言語データを利用することができる。収録した音声データの書き起こし作業はすべて人手によって行われた。書き起こしデータの例を Fig. 1 に示す。日本語話し言葉コーパス (CSJ)¹⁰⁾ の書き起こし基準に準拠しており, 読みデータのほか, フィラーや言い淀み等の言語タグ, 発話時間タグ等が付与されている。

2.2 発話意図タグ

高い性能を備えた音声対話システムを実現するために, ユーザの発話意図を正確に理解し, 適切なシステム発話意図を生成することが不可欠である。コーパスを用いた対話システムを構築するにあたり, 話者の発話意図に関する情報が有用であると考え, 上述の車内音声対話コーパスに対して発話意図タグを付与した⁵⁾。

まず我々は, システム開発での利用を意識して意図タグ付きコーパスの設計を行った。設計方針として,

0002 - 00:05:772-00:08:452 F:D:I:C:	ドライブスルーの	&	ドライブスルーノ
	ある	&	アル
	ファーストフード	&	ファーストフード
	お願いします<SB>	&	オネガイシマス<SB>
0003 - 00:08:803-00:10:520 F:O:I:C:	お店の	&	オミセノ
	指定は	&	シテーワ
	ございますか<H><SB>	&	ゴザイマスカ<H><SB>
0004 - 00:11:008-00:12:729 F:D:I:C:	(F え)	&	(F エ)
	マクドナルドで	&	マクドナルドデ
	お願いします<SB>	&	オネガイシマス<SB>
0005 - 00:12:795-00:15:917 F:O:I:I:	はい	&	ハイ
	この	&	コノ
	近くですと	&	チカクデスト
	マクドナルドー社店が	&	マクドナルドイッシャテンガ
	ありますが<SB>	&	アリマスカ<SB>
0006 - 00:16:765-00:18:958 F:D:N:C:	じゃ	&	ジャ
	そちら	&	ソチラ
	そこへ	&	ソコエ
	案内してください<SB>	&	アンナイシテクダサイ<SB>
0007 - 00:19:135-00:22:483 F:O:N:O:	はい	&	ハイ
	マクドナルドー社店まで	&	マクドナルドイッシャテンマデ
	ご案内いたします<SB>	&	ゴアンナイタシマス<SB>

Fig. 1 車内音声対話コーパスの書き起こしデータ

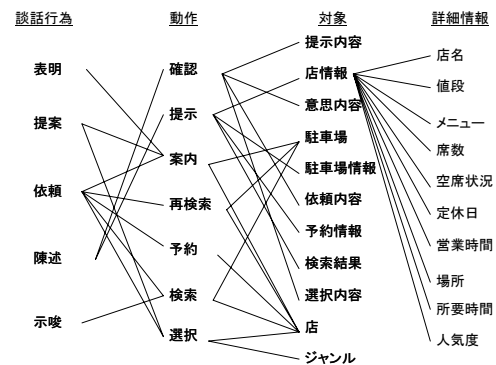


Fig. 2 意図タグ体系の例

- 従来の発話行為理論を背景とした談話タグに対して, より詳細化したタグを設定する。システム動作を決定できるほどに具体化することにより, 意図タグ情報を直接的に利用できる。
- タグ利用の汎用性を高めるために, タグ情報をその抽象度に応じてレベル化する。階層化により意図タグを選択的に利用することができる。
- タグ付与の信頼性を高めるために, 対話進行における発話の効果という観点からタグ名を定める。対話相手の応答発話をもとにタグを付与することができ, 話者意図のあいまい性を軽減することができる。

を定めた。設計した意図タグ体系を Fig. 2 に示す。意図タグは、「談話行為レイヤ」「動作レイヤ」「対象レイヤ」「詳細レイヤ」の 4 つの階層を有し, それぞれ「話者の発話内行為」「行為」「動作対象」「対象に関する詳細情報」を表す。談話行為レイヤは, 従来の発話行為に相当し, あらゆる対話タスクに出現

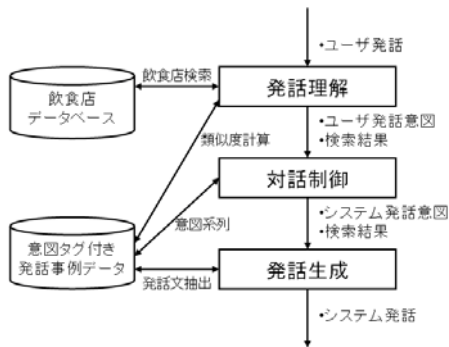


Fig. 3 コーパスベース対話システムの構成

するタグである．それ以外のレイヤは，より詳細化された意図を表現しており，対話タスクに依存したタグを含んでいる．

名古屋大学 CIAIR 車内音声対話コーパスに対して意図タグ付与を実施した．これまでに飲食店案内をタスクとする 3,641 対話に対して作業を実施し，35,411 発話に対して人手でタグ付けを実施した．Table 1 に意図タグ付き対話データの例を示す．構築したコーパスに出現した発話意図タグは，4 つのレイヤの組み合わせで 95 種類存在する．

2.3 コーパスベース対話システム

車内音声対話データ，及び，発話意図タグを用いた音声対話のプロトタイプシステムを開発している²⁾．システムの構成を Fig. 3 に示す．システムは，発話理解，対話制御，発話生成から構成されており，すべてのモジュールをコーパスを利用して実現している．対話タスクとして，車内での飲食店案内を選定し，システムの運用では車内対話コーパスのうち，同一タスクの対話データのみを利用している．

発話理解では，コーパスを事例として用いた意図推定手法¹²⁾を採用している．すなわち，コーパス内の発話事例のうち，入力発話に最も類似している発話を抽出し，それに付与された発話意図タグを入力発話の意図とする．類似度は，発話を構成する形態素の一致度に基づいて計算する．すなわち，入力発話 j の意図を I_j ，形態素数を M_j ，コーパス中の発話 k の意図を I_k ，形態素数を M_k ，また，発話 j と k で一致する形態素数を M_{jk} とするとき， I_j を

$$I_j = \operatorname{argmax}_{I_k} \frac{2M_{jk}}{M_j + M_k} \quad (1)$$

で計算する．ただし，抽出する発話は，直前の発話の意図が入力発話の直前の発話意図と一致する場合に限るとし，これにより文脈を考慮した意図推定を実現している．例として，ユーザ発話「パスタをお願い」が“依頼+選択+ジャンル”を意図とするシステム発話に続いて行われたときには，発話事例デー

タから同一意図の発話に続いて現れるユーザ発話のうち，最も類似度が高い発話，例えば「え マクドナルドをお願いします」を抽出し，その発話意図“陳述+選択+ジャンル”を入力発話の意図として推定する．

対話制御では，コーパスを統計情報として利用してシステム応答発話の意図を決定する．すなわち，対話コーパスから作成した意図タグの N-gram 確率に基づいて最尤の意図タグを選ぶ．意図タグ系列 $I_{j-N+1} \cdots I_{j-1}$ のもとでの次発話の意図 \hat{I}_j は，

$$\hat{I}_j = \operatorname{argmax}_{I_j} P(I_j | I_{j-N+1} \cdots I_{j-1}) \quad (2)$$

で計算する．本システムでは基本的に 3-gram 確率を採用している．すなわち，直前の二つの発話の意図タグと一致する発話系列に対して，最も出現頻度の高い発話意図をシステム発話の意図とする．

最後に，発話生成では，コーパス内のシステム発話を利用して発話を作り出す．具体的には，発話意図が同一で，かつ，必要な内容語 (content word) が含まれる発話をコーパスから抽出し，それをを用いて適切な発話文を作り出す．内容語とは，直前のユーザ発話に含まれる内容語，及び，飲食店データベース検索により該当した店名である．コーパスから抽出した発話文の内容語を置き換え，不要な語句を削除することにより，文を生成する．

3 タグの高度化

コーパスに基づく対話システムの高度化のために，使用するコーパスへのアノテーションを高度化する必要がある．現状のシステムでは，発話を形態素列，また，対話を発話列として利用しており，よりきめ細かく参照するために，発話及び対話を構造的にとらえることは有用である．本研究では，意図タグ付き車内音声対話コーパスに対して，発話文構造タグ，及び，対話構造タグを付与することにより，アノテーションの高度化を試みた．Fig. 1 に対する構造タグを付与したデータを Fig. 4 に示す．

3.1 文構造タグ

車内音声対話コーパスの書き起こしデータでは，発話はポーズで単位分割されているだけでなく，文節区切りデータも付与されている．コーパスベース対話システムにおける発話理解では，類似した発話を抽出するにあたり形態素情報のみを使用している．しかしながら，文間の類似度の計算では，文の構造を考慮することの有用性が示されており，この場合，コーパスに文構造情報が付与されている必要がある．

我々は，文構造データを発話理解で利用することを目的に，ユーザ発話に対して係り受け構造を付与している¹⁴⁾．データには，形態素情報及び構文情報を付与し，係り受け文法は，京大コーパス⁸⁾の

Table 1 意図タグつき対話の例

ID	話者	発話内容	意図タグ			
			第一 (談話行為)	第二 (動作)	第三 (対象)	第四 (詳細)
#1	D	ドライブスルーのあるファーストフードをお願いします	依頼	検索	店	
#2	O	お店の指定はございますか	依頼	選択	ジャンル	
#3	D	え マクドナルドをお願いします	陳述	選択	ジャンル	
#4	O	はいこの近くですとマクドナルド一社店がありますが	陳述	提示	検索結果	店名
#5	D	じゃそちらそこへ案内してください	依頼	案内	店	
#6	O	はいマクドナルド一社店までご案内いたします	表明	案内	店	

作成基準に準拠した。なお、対話音声における文区切りのあいまい性を考慮し、対話ターンを単位として係り受け構造を与えた「きょう朝パン食べてお昼はおそばをたべたんですよ」に対する係り受け構造データの例を Fig. 5 に示す。この例では、係り受け構造を係り受け関係の列で表し、係り受け関係を係り文節と受け文節の対で記している。

係り受け構造の付与にあたっては、効率的な付与作業のために、統計的係り受け解析¹⁴⁾とデータ修正作業を逐次的に実行する増殖的な手法を採用した。すなわち、係り受け解析の結果を手で修正したデータを、係り受け構造データに追加し、別のデータを解析するときの統計情報として利用する。統計情報は修正したデータから自動的に獲得できるので、解析結果の修正以外に人手による作業は必要としない。

Fig. 1 の各発話文に対する文構造タグを Fig. 4 の第 2 段に示している。現在までに、飲食店案内をタスクとする対話のユーザ発話 10,995 ターンに対して係り受け構造を付与している。

3.2 対話構造タグ

開発したコーパス対話システムでは、対話データを意図タグが付与された発話の系列として利用している。しかしながら、対話データの正確な参照という観点から、以下が問題となる。

- 発話理解では、類似する発話を抽出するうえで、直前の発話意図を考慮しているが、発話が行われた文脈情報としては必ずしも十分ではない。一般に、タスク指向対話は、いくつかの部分対話から構成される。発話がどの部分対話内で行われたかなど、対話の構造を考慮した意図推定が望まれる。
- 対話制御では、直前のいくつかの発話系列を考慮して、次発話の意図を決定しているが、より洗練された対話を遂行するには、対話の進行状況や遂行を要する部分対話などを考慮した制御が望まれる。

そこで我々は、発話理解、及び、対話制御で利用することを目的に、車内音声対話コーパスに対話構造

- (1) ((きょうキョウ きょう名詞 副詞可能なし))
-> (4 ((食ベタベ 食べる 動詞 自立 一段 連用形)
(ててて 助詞 接続助詞 なしなし)))
- (2) ((朝 アサ 朝名詞 副詞可能なし))
-> (4 ((食ベタベ 食べる 動詞 自立 一段 連用形)
(ててて 助詞 接続助詞 なしなし)))
- (3) ((パン パン パン 名詞 一般なし))
-> (4 ((食ベタベ 食べる 動詞 自立 一段 連用形)
(ててて 助詞 接続助詞 なしなし)))
- (4) ((食ベタベ 食べる 動詞 自立 一段 連用形)
(ててて 助詞 接続助詞 なしなし))
-> (7 ((食ベタベ 食べる 動詞 自立 一段 連用形)
(た た た 助動詞 なし 特殊・タ 基本形)
(ん ン ン 名詞 非自立なしなし)
(です デス です 助動詞 なし 特殊・デス 基本形)
(よ よ よ 助詞 終助詞 なしなし)))
- (5) ((お屋 オヒル お屋 名詞 副詞可能なしなし)
(は ハ は 助詞 係助詞 なしなし))
-> (7 ((食ベタベ 食べる 動詞 自立 一段 連用形)
(た た た 助動詞 なし 特殊・タ 基本形)
(ん ン ン 名詞 非自立なしなし)
(です デス です 助動詞 なし 特殊・デス 基本形)
(よ よ よ 助詞 終助詞 なしなし)))
- (6) ((お オ お 接頭詞 名詞接続なしなし)
(そば ソバ そば 名詞 一般なしなし)
(を ヲ を 助詞 格助詞 なしなし))
-> (7 ((食ベタベ 食べる 動詞 自立 一段 連用形)
(た た た 助動詞 なし 特殊・タ 基本形)
(ん ン ン 名詞 非自立なしなし)
(です デス です 助動詞 なし 特殊・デス 基本形)
(よ よ よ 助詞 終助詞 なしなし)))
- (7) ((食ベタベ 食べる 動詞 自立 一段 連用形)
(た た た 助動詞 なし 特殊・タ 基本形)
(ん ン ン 名詞 非自立なしなし)
(です デス です 助動詞 なし 特殊・デス 基本形)
(よ よ よ 助詞 終助詞 なしなし))
-> (NO (なし))

Fig. 5 発話文構造データ

データを付与した⁶⁾。発話意図タグをもとに、発話間の関係や部分対話間の関係を記すことにより対話構造を作成する。

本研究では、対話を木構造で表すこととし、部分対話を部分木として表現する。部分対話の内容を記すために、POD と呼ばれる部分対話に関するカテゴリを 11 種類定め、部分木のルートノードに与えた。POD の種類と内容を Table 2 に示す。

対話構造の付与作業では、発話意図の系列に対して以下の操作を繰り返し実行する。

- 発話の融合：意味的に対応関係にある隣接する 2 つの発話を発話対とする。その対に最も合致

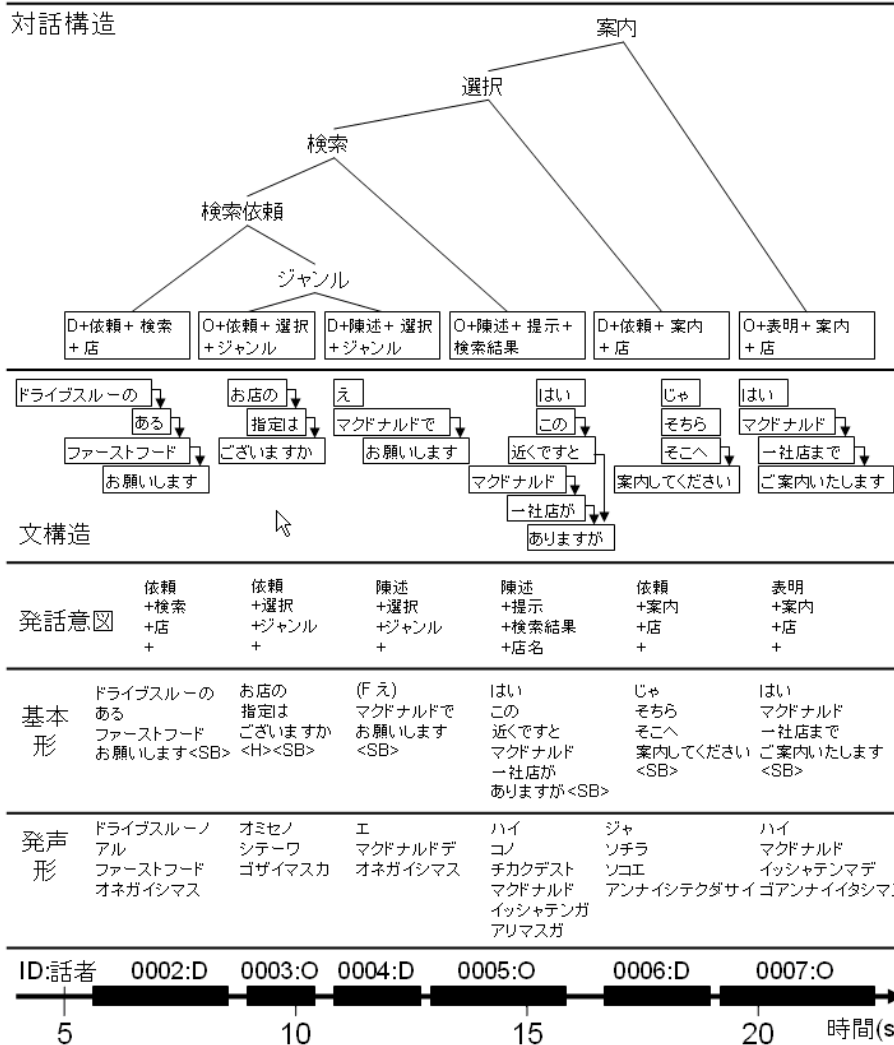


Fig. 4 高度アノテーションが与えられた対話コーパス

案内→ 選択 O+表明+案内+店
 選択→ 検索 D+依頼+案内+店
 検索→ 検索依頼 O+陳述+提示+検索結果
 検索依頼→ D+依頼+検索+店 ジャンル
 ジャンル→ O+依頼+選択+ジャンル D+陳述+選択+ジャンル

Fig. 6 対話構造規則の例

するをラベルとして付与する。

- 部分対話の融合：意味的に対応関係にある隣接する部分対話を対話対とし、その対話対に最も合致する POD をラベルとして付与する。

現在までに、飲食店案内をタスクとする 789 対話 (8,150 発話) に対する対話構造の付与が完了している。作成した木構造から、各ノードに対する親ノードと子ノードとの関係から、対話構造規則を取り出すことができる。構造規則の例を Fig. 6 に示す。これまでに、297 種類の構造規則を獲得しており、これらは対話構造解析に使用することができる。

4 対話理解におけるコーパスの利用

コーパスベース対話システムでは、コーパスを統計データとして、あるいは、事例データとして利用することができる。例えば、2.3 節で概説したシステムでは、発話理解及び発話生成では事例データとして、また、対話制御では統計データとしてコーパスを直接的に用いている。

より洗練されたシステムへと拡張するために、前節で示したような構造化された対話データを利用することが望まれるが、そのための効果的な方法についてはまだ明らかになっていない。我々は、構造化データを利用したシステム開発に向けた準備として、以下を試みている。

- 文構造タグを発話事例データとして発話意図推定に利用している¹²⁾。係り受け構造の一致度に基いて文間類似度を計算することにより、

Table 2 PODの種類と内容

POD	内容
ジャンル	店のジャンルを決定する
案内	店や駐車場に案内する
駐車場情報	駐車場情報, 駐車場の有無, 近辺の駐車場等に関する情報のやり取りを行う
駐車場検索	駐車場の検索を行う
店情報	店情報値段, メニュー, 場所, 定休日等に関する情報のやり取りを行う
選択	店や駐車場を選択する
検索	店を検索する
検索依頼	検索依頼をする
予約	予約を行う
予約情報	予約依頼の詳細情報のやり取りを行う
予約依頼	予約依頼をする

入力発話に類似した発話をコーパスから抽出する。形態素情報のみを利用した場合に比べ、係り受け情報を併用すると推定精度が高いことを確認している。

- 対話構造タグを統計データとして次発話予測に利用している⁷⁾。対話の進行に応じてその構造を解析し、解析途中の対話構造データを作成する。作成した(部分的な)対話構造に対して、コーパスから獲得した統計モデルにより次の発話の意図を予測する。対話構造解析が正しく実行されれば90%を超える精度で次発話を予測できることを確認している。

今後は、コーパスベース対話システムにおいて、構造化されたデータを利用するためのより汎用的かつ実践的な方法を検討する予定である。

5 まとめ

自然な音声対話システムの実現のために、これまでいくつかの音声対話コーパスが作成されてきたものの、対話理解においてそれらを効果的に利用する方法については、必ずしも明らかではない。そのような現状のもと、本論文では、対話システムにおけるコーパスの利用について述べた。開発を進めているコーパスベース音声対話システム、及び、システムで使用している発話意図タグつき車内音声対話コーパスについて概説した。また、システムの高度化のために進めているタグアノテーションとして、発話文構造タグ、及び、対話構造タグの設計と付与について述べるとともに、その利用法について論じた。

謝辞

車内音声対話コーパスの開発に貢献された名古屋大学 CIAIR のメンバー諸氏に感謝いたします。また、タグの設計及びシステムの実装にご協力いただいた名古屋大学の入江友紀氏、大野誠寛氏、林啓太氏、ならびに、三洋電機の村尾浩也氏に感謝いたします。本研究は、一部、日本

学術振興会科研費基盤研究(B)(2)(No.15300045)によります。

参考文献

- 1) Blaylock, H. and Allen, J.: Corpus-based, Statistical Goal Recognition, *Proc. of 8th Int. Joint Conf. on Artificial Intelligence (IJCAI-2003)*, pp. 1303-1308 (2003).
- 2) Hayashi, K., et al.: Speech Understanding, Dialogue Management and Response Generation in Corpus-based Spoken Dialogue System, *Proc. of Int. Conf. on Spoken Language Processing (ICSLP-2004)*, (2004).
- 3) He, Y. and Young, S.: A Data-Driven Spoken Language Understanding System, *Proc. of IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU-2003)* (2003).
- 4) Irie, Y., et al.: Speech Intention Understanding based on Decision Tree Learning, *Proc. of 8th Int. Conf. on Spoken Language Processing (ICSLP-2004)* (2004).
- 5) 入江ほか: 音声対話コーパスにおける発話意図タグの設計と評価, 電子情報通信学会論文誌, J88-D-II(10), pp. 2169-2173 (2005).
- 6) Kato, S., et al.: Dialogue Structure Annotation of In-car Speech Corpus based on Speech-Act Tag, *Proc. of Int. Conf. on Speech Databases and Assessment (Oriental COCOSDA-2005)*, pp. 159-163 (2005).
- 7) 加藤ほか: 対話構造木コーパスを用いた発話意図予測, 情処研報 (SIG-SLP-59), pp. 73-78 (2005).
- 8) 黒橋, 長尾: 京都大学テキストコーパス・プロジェクト言語処理学会第3回論文集, pp. 115-118 (1997).
- 9) Kawaguchi, N., et al.: CIAIR in-car Speech Corpus -Influence of Driving Status-, *IEICE Trans. on Information and Systems*, E88-D(3), pp. 578-582 (2005).
- 10) 前川ほか: 日本語話し言葉コーパスの設計, 音声研究, 4(2), pp. 51-61 (2000).
- 11) 松原ほか: 音声対話コーパスの収集と利用 -より豊かな車内音声対話システムを目指して-, 人工知能学会誌, 17(3), pp. 279-284 (2002).
- 12) Matsubara, S., et al.: Example-based Speech Intention Understanding and Its Application to In-Car Spoken Dialogue System *Proc. of 17th Int. Conf. on Computational Linguistics (COLING-2002)*, Vol. 2, pp. 633-639 (2002).
- 13) Murao, H., et al.: Example-based Query Generation for Spontaneous Speech, *IEICE Trans. on Information and Systems*, E88-D(2), pp. 324-329 (2005).
- 14) Ohno, T., et al.: Robust Dependency Parsing of Spontaneous Japanese Spoken Language, *IEICE Trans. on Information and Systems*, E88-D(3), pp. 545-552 (2005).