

車内音声対話システムのための 事例に基づく発話意図推定

木村 晋一† 松原 茂樹†* 河口 信夫‡*
山口 由紀子§ 稲垣 康善†

†名古屋大学大学院工学研究科

‡名古屋大学言語文化部

§名古屋大学大型計算機センター

*名古屋大学統合音響情報研究拠点 (CIAIR)
shinichi@inagaki.nuie.nagoya-u.ac.jp

概要

音声対話システムがユーザとの間で自然なインタラクションを遂行するために、ユーザの発話を正しく理解する必要がある。本論文では、事例に基づきユーザ発話の意図を推定する手法を提案する。本手法では、発話意図を表すタグが各発話に付与された意図タグ付きコーパスを使用する。入力発話とコーパス中の発話との類似度を計算し、類似度が最大の発話に付与された意図を入力発話の意図とする。発話間の類似度は各発話の形態素と係り受けの一致度にしたがって決定し、対話履歴を用いてそれを重み付ける。意図タグ付きコーパスを用いて意図推定実験を行った結果、68.9%の正解率を得た。また、本推定手法に基づき音声対話レストラン検索システムを作成し、動作実験を行った。

Example-based Speech Intention Understanding for In-Car Spoken Dialogue System

Shinichi KIMURA†, Shigeki MATSUBARA†*, Nobuo KAWAGUCHI‡*
Yukiko YAMAGUCHI§ and Yasuyoshi INAGAKI†

†Graduate School of Engineering, Nagoya University

‡Faculty of Language and Culture, Nagoya University

§Computation Center, Nagoya University

*Center for Integrated Acoustic Information Research, Nagoya University

shinichi@inagaki.nuie.nagoya-u.ac.jp

ABSTRACT

This paper proposes a method of speech intention understanding based on the dialogue examples. In the method, a spoken dialogue corpus with intention tags is used, and the intention of each input utterance is regarded as that of the utterance to which it is the most similar in the corpus. The degree of similarity is calculated based on the degree of the correspondence in morphemes and Japanese dependency relations between utterances, and it is weighted by the dialogue context information. An experiment using an in-car speech corpus: CIAIR-HCC has shown 68.9% accuracy. And we constructed a restaurant search task-oriented spoken dialogue system based on the method, and we have confirmed the feasibility of the system.

1 はじめに

音声対話システムがユーザとの間で自然なインタラクションを遂行するために、ユーザの発話意図を正しく理解する必要がある。発話意図を推定するための手法として、木村らは、ルールを用いるアプローチを提案している [1]。そこでは、52種類の発話意図を定義し、各発話から意図を抽出するために、直前の発話

意図や発話文の動詞、様相などに基づいてルールを作成している。しかし、ルールの生成に要する膨大な作業は人手に頼らざるを得ず、また、ルールの修正も困難である。ほかにも、談話行為タグを推定する手法 [2, 3] が提案されているものの、音声対話システムの動作を具体的に決定するには、「真偽情報要求」や「未知情報要求」などといった発話内行為のレベルよりも、さらに詳細な意図を推定する必要がある。

そこで本論文では、対話事例に基づくユーザ発話の意図を推定する手法を提案する。本手法では、発話意図を表すタグが各発話に付与された意図タグ付きコーパスを使用する。意図タグは、談話行為タグをタスクに合わせて拡張することにより定義し、現在 78 種類のタグを定めている。タスクに依存したレベルにまで意図を詳細化することにより、システムの動作に直結した意図記述が可能となる。意図推定では、入力発話とコーパス中の発話との類似度を計算する。発話間の類似度は、各発話の形態素と係り受けの一致度に基づいて決定する。対話履歴を用いて類似度の重み付けを行い、スコアが最大の発話に付与された意図を入力発話の意図とする。事例を用いることにより、各発話意図ごとにルールを作成する必要がなく、また発話の多様性に対してもコーパスに対処することができる。

本論文では、また、意図タグ付きコーパスを用いて行った意図推定実験とその結果について報告する。さらに、本推定手法に基づく音声対話レストラン検索のプロトタイプシステムを作成し、動作実験を行ったのでそれについても報告する。

2 事例に基づく発話意図推定

ユーザがある意図に基づいて発話するとき、その影響は音韻、形態素、キーワード、文構造、文脈など、発話に関連する事象に様々な形態となって現れる。人間の複雑かつ多様な発話に対応できるシステムを実現するために、事例を用いるアプローチが有効である。

本手法では、発話意図を表すタグが各発話に付与された意図タグ付き対話コーパスを使用する。以下では、コーパスを用いた推定手法の概要を例で以って説明する。

入力発話「この近くに和食の店ある」の意図を推定する処理の流れを図 1 に示す。まず、入力発話に対して形態素解析、ならびに係り受け解析を行い、その結果をコーパス内の全ての発話との間で比較する。コーパス内の発話には、あらかじめ形態素情報、及び係り受け情報が与えられており、それらの一致率によって各発話との類似度を計算する。ただし、意図推定の精度を高めるために、発話が入力された時点での文脈を考慮する。すなわち、意図タグ付き対話コーパスから意図系列の生起確率をあらかじめ学習し、入力発話の前に生起した意図系列に従って各発話との類似度を重み付けする。その結果、入力発話との類似度が最大である発話が「この近くに洋食の店ありますか」であれば、その発話意図「検索」を入力発話の意図と定める。

3 類似度計算

著者らは、これまでに係り受け関係を用いた発話間類似度計算手法を提案しているが [4]、本論文では、それに加えて形態素に関する情報も使用する。

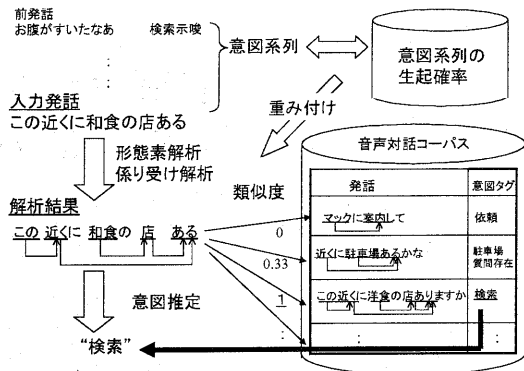


図 1: 事例に基づく意図推定の流れ

3.1 形態素と係り受け関係を用いた類似度計算

発話間の類似度を計算するために、形態素や係り受け情報を用いることが考えられる。形態素のみによる類似度計算では、表層的な単語のみの類似性を考慮することとなり、発話文の構造としては類似していても、類似度が大きくなる場合がある。また、係り受け関係のみに基づく類似度計算では、構文的、あるいは意味的類似性を捉えることはできるものの、話し言葉の場合、係り受け構造だけで文全体の意味関係を表現することは難しいといった問題がある。形態素と係り受け関係の情報を併用することにより、より信頼度の高い発話間類似度計算を行うことが期待できる。

形態素に基づく類似度 α_m と係り受け関係に基づく類似度 α_d の凸結合として、本稿で用いる類似度 β を式 (1) で定義する。

$$\beta = \lambda \alpha_d + (1 - \lambda) \alpha_m \quad (1)$$

α_d : 係り受け関係に基づく類似度

α_m : 形態素に基づく類似度

λ : 重み係数 ($0 \leq \lambda \leq 1$)

α_d , α_m については、本章の以下の 3.2 節、及び 3.3 節で説明する。

3.2 係り受け関係に基づく類似度計算

係り受け関係とは、2つの文節間に成り立つ修飾・被修飾関係を意味する。例えば、「この近くに和食の店ある」という文は、「この」「近くに」「和食の」「店」「ある」という5つの文節から構成されており、例えば、文節「店」は文節「ある」を修飾する。このとき、修飾文節「店」に係り文節、被修飾文節「ある」を受け文節といい、この2つの文節は係り受け関係にあるという。同様に「この」は「近くに」、「近くに」は「ある」、「和食の」は「店」に係る。本稿の以下では、係り受け関係を(店, ある)、(この, 近くに)のように文節の順序対で表現する。

係り受け関係は、発話の構文的、意味的特性の一端を表わしており、その意図内容とも大いに関連する。すなわち、係り受け関係が類似した2つの発話は、その意図も類似する可能性が高いと予想される。

2つの発話 S_A と S_B の間の類似度 α_d を式 (2) で定義する [4].

$$\alpha_d = \frac{2C_D}{D_A + D_B} \quad (2)$$

D_A : S_A の係り受けの数
 D_B : S_B の係り受けの数
 C_D : 一致する係り受けの数

ここでは、2つの係り受けの係り文節と受け文節のそれぞれの自立語の基本形が共に一致するとき、それらの係り受けが一致するとみなす。例えば、(近くに、ある) と (近くに、ありますか) という2つの係り受けの係り文節、及び、受け文節の自立語はそれぞれ、共に「近く」、「ある」であり、この2つの係り受けは一致する。また、対話タスクに特徴的な名詞や固有名詞には単語クラスを付与し、各係り受けを構成する単語が同一クラスに属していればそれらの係り受けも一致しているとみなす。

3.3 形態素に基づく類似度計算

2つの発話 S_A と S_B の間の形態素を用いた類似度 α_m を式 (3) で定義する。

$$\alpha_m = \frac{2C_M}{M_A + M_B} \quad (3)$$

M_A : S_A の形態素数
 M_B : S_B の形態素数
 C_M : 一致する形態素数

本研究では、対話タスクに特徴的な名詞や固有名詞には単語クラスを付与し、形態素が同一クラスに属していればそれらの形態素も一致しているとみなす。また、文全体としてより類似したコーパス中の発話の意図を抽出するために、自立語やキーワードだけではなく、名詞や助詞など、全ての形態素を一致の対象とした。

3.4 類似度計算の例

重み係数 $\lambda = 0.4$ のときの入力発話 S_1 「この近くに和食の店ある」とコーパス中の発話 S_2 「近くに洋食の店ありますか」との間の類似度計算の例を図2に示す。 S_1, S_2 の係り受けの数はそれぞれ4, 3であり、一致する係り受け数は(近く, ある), (店, ある)の2つである。また、和食と洋食は同じ単語クラスに属するため、(和食, ある)と(洋食, ある)の係り受けも一致する。したがって、係り受けに基づく類似度 α_d は式(2)より0.857となる。また、 S_1, S_2 の形態素数はそれぞれ7, 8であり、一致する形態素は「近く、に、の、店、ある」の5つに加え、係り受け同様、和食と洋食は同じクラスに属するのでこれらも一致する形態素となる。したがって、形態素に基づく類似度 α_m は

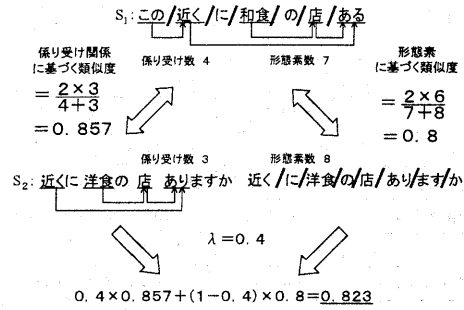


図2: 類似度計算例

式(3)より0.8となる。以上より、形態素と係り受けを用いた類似度 β は式(1)より0.823となる。

4 対話履歴を用いた類似度の重み付け

ユーザの発話の意図は、多くの場合、それ以前の自らの発話や対話相手の発話の意図に依存して生起している。したがってユーザの入力発話もまた、それ以前の対話や応答の内容に影響を受けることとなる。例えば、システムの質問発話に対してはそれに対する応答を返すのが普通であり、また、システムの応答の後には、さらにその応答に関して質問する場合がある。そこで、本手法では、ユーザの発話に至るまでの対話履歴を用いることにより、3章で説明した定義に従って求めた類似度 β に重み付けを行う。対話履歴として、入力発話以前の発話意図の履歴情報を使用する。入力発話以前の発話の意図系列情報を保持しておき、入力発話の類似度を重み付けするときの情報として使用する。すなわち、ある時点 n での発話意図 I_n の生起は直前の $N-1$ 発話の発話意図に依存すると考える。そのときの条件付生起確率 $P(I_n | I_{n-N+1}^{n-1})$ は式(4)となる。

$$P(I_n | I_{n-N+1}^{n-1}) = \frac{C(I_{n-N+1}^n)}{C(I_{n-N+1}^{n-1})} \quad (4)$$

ここで、発話意図の履歴 $I_{n-N+1} \dots I_n$ を I_{n-N+1}^n と書き、それを意図 N-gram と呼び、その対話コーパス中での出現回数を $C(I_{n-N+1}^n)$ と書く。また、式(4)の条件付生起確率を意図 N-gram 確率という。

発話意図の履歴による重み付けは、意図 N-gram 確率がある閾値よりも小さい場合には、類似度の値を減らし、発話意図の系列として生起する可能性の低い発話意図に対してその類似度を軽減することによりそれを実現する。すなわち、意図 N-gram 確率による重み付けを使った類似度 γ を式(5)で定義する。

$$\gamma = \begin{cases} \omega\beta & (P(I_n | I_{n-N+1}^{n-1}) \leq \theta) \\ \beta & (\text{otherwise}) \end{cases} \quad (5)$$

ω : 重み係数 ($0 \leq \omega \leq 1$)

β : 類似度

$P(I_n | I_{n-N+1}^{n-1})$: 意図 N-gram 確率

θ : 閾値

表 1: 意図タグと対応する談話行為タグ、及びその発話例

意図タグ	談話行為タグ	発話例
検索	未知情報要求	この近くに和食の店ありますか
依頼	依頼	マックに案内して
駐車場質問存在	真偽情報要求	駐車場はありますか
店質問距離	未知情報要求	ここからどれくらいの距離ですか
近い質問	未知情報要求	どちらが近いですか
店質問メニュー	真偽情報要求	ラーメンはありますか

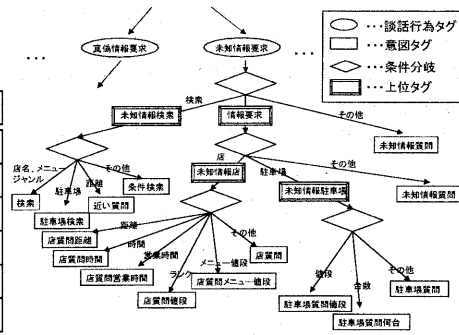


図 3: 決定木の一部

意図 N-gram を用いる効果の例を次に示す。「近くに駐車場はありますか」という入力発話に対しては、「検索したお店に駐車場があるかどうかを質問すること」を意図する「駐車場質問存在」と、「近くの駐車場を検索すること」を意図する「駐車場検索」の発話の類似度が大きくなる。しかし、入力発話が、お店の駐車場がないという応答の後に生じたのであればコーパスから学習した意図 N-gram 確率より、「駐車場質問存在」ではないとことが分かり、結果として、「駐車場検索」であるとして正確な発話意図推定ができる。

5 実験

本手法の有効性を確認するために、意図タグ付きコーパスを用いた発話意図推定実験を行った。また、本研究では、本手法に基づく音声対話システムの実現可能性を調べるために、推定した発話意図の情報に従って動作する音声対話レストラン検索のプロトタイプシステムを作成し、動作実験を行った。5.1, 5.2節で意図推定実験について述べ、5.3節でプロトタイプシステムとその評価結果について述べる。

5.1 実験に使用したデータ

意図タグ付きコーパスを作成するために、名古屋大学 CIAIR 車内音声対話データベース [5] の係り受け分析済みデータ [6] を使用した。すなわち、データベースに収録されている 73 人分の対話から、レストラン検索をタスクとする 412 対話に意図タグを人手で付与した。そのうちの 174 対話の発話系列から意図 2-gram 確率を学習した。作業にあたり、発話に付与すべき意図タグを判断する基準を設ける必要があるが、我々は、発話単位タグ標準化案 [7] の決定木を拡張することによりそれを実現した。その結果、全部で 78 種類（そのうち 38 種類がドライバー発話）の意図タグが設けられた、拡張した決定木をたどることにより、各発話に付与すべき意図タグを定めることができる。意図タグの一部とその発話例を表 1 に、また、ドライバー発話を対象とした決定木の一部を図 3 に示す。ただし、ここでは簡単のため条件分岐における条件の記述を省略している。また、412 対話のドライバー発話

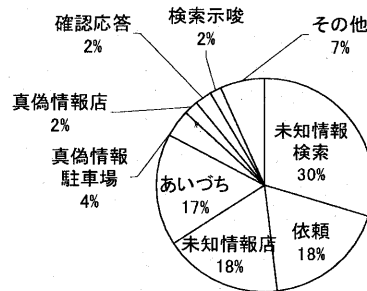


図 4: 意図タグの割合

に出現した意図タグの出現傾向を調べた。意図タグの種類ごとの割合を図 4 に示す。ただし、割合は決定木における上位タグでまとめた値となっている。例えば、「検索」、「駐車場検索」などの上位タグは未知情報検索である。

一方、形態素、係り受けの一致度を計算する上で使用する単語クラスとして、本コーパスをもとに作成した単語クラスデータベース [8] を使用した。また、形態素解析には茶筌 [9] を使用した。

5.2 発話意図推定実験

5.2.1 実験方法

意図 2-gram 確率の学習に用いていない 238 対話のドライバー発話 1609 文に対して、それを 10 個に等分割し、交差検定を行った。すなわち、1609 文すべての意図推定を行い、その再現率と適合率を求めた。実験では、以下の 4 つの類似度計算方法に基づき意図推定を行い、その結果を比較した。

1. 形態素のみによる類似度計算
2. 係り受け関係のみによる類似度計算
3. 形態素と係り受け関係を併用した類似度計算
重み係数の ω を変化させて実験を行った。
4. 3. の条件に加え、意図 2-gram による重み付けを行った類似度計算
重み係数の ω を変化させて実験を行った。ただし、 $\theta = 0$ とした。

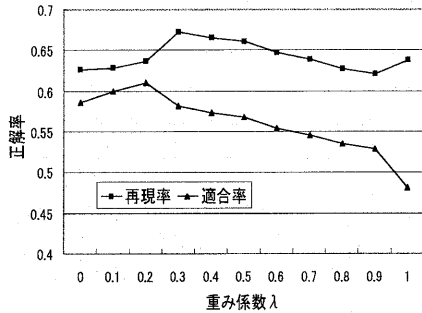


図 5: 重み係数 λ と正解率の関係

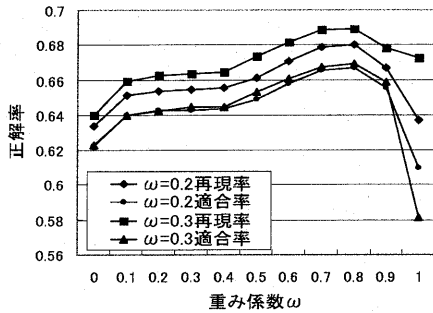


図 6: 重み係数 ω と正解率の関係

5.2.2 実験結果

意図推定実験の結果を図 5 に示す。方法 1. の形態素一致 ($\lambda = 0$) による推定で再現率 62.6%, 適合率 58.6%, 方法 2. の係り受け一致 ($\lambda = 1.0$) による推定で再現率 63.7%, 適合率 48.2% を得た。また, 方法 3. の形態素と係り受け関係の双方を用いた推定では, $\lambda = 0.2$ で適合率が最大の 61.0%, $\lambda = 0.3$ で再現率が最大となり 67.2%, 適合率は 58.1% であった。この結果から形態素と係り受け関係の情報を併用した手法が有効であることが分かった。

$\lambda = 0.3$ 以上で形態素による推定より適合率は, 低くなっている。これは, 車内対話コーパスでは, ユーザが運転作業中であるということもあり, 比較的短い発話が多いため, 1 発話あたりの係り受けの数が少なく [10], 係り受け情報を用いた推定では, 類似度が最大となる発話が数多く存在することになるためである。

次に, $\lambda = 0.2$, 及び 0.3 としたときの類似度に対して, 意図 2-gram による重み付けを用いた意図推定実験の結果を図 6 に示す。 $\omega = 0.8$ のときに, 適合率, 再現率とも最大となった。これは, コーパスの中の対話において, 発話意図の系列で生起しない意図系列を学習しておくことにより, 意図系列としては不自然となる意図を軽視する効果が現れたことによる。以上より, 意図 2-gram による類似度の重み付けが有効であることを確認した。

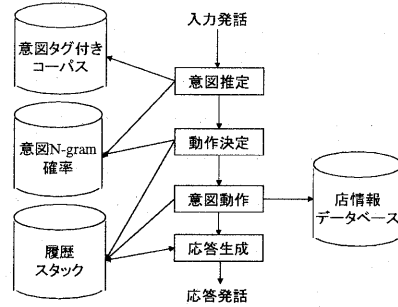


図 7: システム構成

5.3 対話システムによる動作実験

推定された発話意図の情報を利用して動作する音声対話レストラン検索のプロトタイプシステムを作成し, 動作実験を行った。

5.3.1 システムの概要

システム構成を図 7 に示す。各モジュールの説明は以下の通りである。

意図推定: 入力発話に対して 3 章, ならびに 4 章で説明した類似度計算を行い, その意図を推定する

動作決定:

- ユーザ発話が入力された場合には, その発話意図に応じた処理内容を決定する。
- 一定時間ユーザ発話が入力されない場合は, 履歴スタックの意図履歴と意図 N-gram を参照し, システム動作を決定する。

意図動作: 各意図 (現在 78 種類の意図) ごとにシステム動作を定めた。その動作に従って処理を行い履歴スタックを更新する。意図動作の処理は意図とキーワードにより行う。例えば, 「検索」の意図を持つ「この近くに和食の店ありますか」に対する処理は, 意図が「検索」, キーワードは「近く」, 「和食」となり, 動作は「近く」, 「和食」をキーワードとしてデータベース検索を行うことである。

履歴スタック: 発話意図の履歴と店データベース検索結果を対話履歴としてスタックに保持しておく。

応答生成: 店名や軒数などをスロットとしたテンプレートを用いてユーザ発話に対する応答文を生成する。

5.3.2 実験方法

意図推定の学習データとして 5.2.1 節で使用した, 238 対話のドライバー発話 1609 文, 意図 2-gram 確率として 5.1 節で学習したものを使用した。また, テストデータとして学習データに使用していない 10 対話のドライバー発話 60 文を使用した。テストデータに対するプロトタイプシステムの動作結果を表 2 に示した基準で分類した。意図推定の結果に基づく動作実験 (意図推定), 及び, 入力発話に対して正しい意図を人手で付与し, その意図に基づく動作実験 (意図付与) を行った。

表 2: システム動作の評価基準

分類	内容	例	
		発話	動作
正解	正しい動作を行う	和食が食べたい	和食のお店を検索
部分正解	一部情報が欠落しているが、 誤った動作をしていない	醤油味のラーメンが食べたい	ラーメン屋は検索できるが 醤油味の店は検索できていない
誤動作	誤った動作をする	デニーズまでの距離は	距離を答えてほしいが デニーズに案内する
無動作	何も動作しない	禁煙席もあるんだよね	何も答えることができない

表 3: 動作実験の結果

	意図推定 発話数 (割合)	意図付与 発話数 (割合)
正解	31 (51.7%)	42 (70.0%)
部分正解	5 (8.3%)	4 (6.7%)
誤動作	7 (11.7%)	2 (3.3%)
無動作	17 (28.3%)	12 (20.0%)

5.3.3 実験結果

動作実験の結果を表 3 に示す。意図付与の場合の部分正解を含めた正解率は 76.7%，意図推定では正解率が 60.0% となった。意図付与のときに、不正解となった 14 発話に対して誤りの原因を調査した。その内訳は、意図動作を行うときに必要なキーワードを正しく抽出できないなど、キーワード処理で不正解となるものが 6 発話、「そのお店には禁煙席がありますか」などのシステムの想定外のものが 8 発話であった。一方、意図推定の場合には、不正解となる原因として、意図付与と同じ原因の他に、意図推定失敗が原因で不正解となるものが 10 発話あった。キーワード処理で不正解となるものについては、意図動作の処理の語彙や処理を改善することにより、正しく動作できると考えられる。また、システム想定外の発話に対しては、「それについてはお答えできません。〇〇の範囲の質問ならお答えできます」などのような応答を返すことが可能である。本手法に従って正しく意図推定することができれば、適切なシステム動作を行える可能性が高いことを確認した。

6 おわりに

本論文では、事例に基づくユーザの発話意図を推定する手法を提案した。CIAIR 車内音声対話コーパスのドライバー発話 1609 文を用いて意図推定実験を行った結果、形態素、係り受け関係、対話履歴を用いた発話間類似度計算の有効性を確認した。また、プロトタイプシステムを作成し、発話意図を用いたシステムの

実現可能性を確認した。今後、推定精度の向上のために、意図推定に重要な形態素やそれに関する係り受けに重みを付けたり、述部を受け文節とする係り受けの一致などに重みをつけるなど、工夫することが考えられる。

参考文献

- [1] 木村, 徳久, 目良, 甲斐, 岡田: 対話における相手意図の理解と応答のためのプランニング, 信学技報, TL98-15, pp.25-32 (1998).
- [2] 木村, 新美, 荒木, 西本: 機械学習を用いた談話行為タグ推定手法, 人工知能学会研究会資料, SIG-SLUD-A003-8, pp.43-48 (2000).
- [3] 駒谷, 荒木, 堂下: 対話コーパスにおける発話単位タグの一推定手法, 人工知能学会誌, Vol.14, No.2, pp.273-281 (1999).
- [4] 木村, 松原, 河口, 稲垣: 係り受け関係を用いた発話意図推定手法, 情報処理学会第 63 回全国大会講演論文集, vol.2, pp.197-198 (2001)
- [5] Kawaguchi, N. et al.: Multimedia Data Collection of In-Car Speech Communication, Proc. of Eurospeech2001, pp.2027-2030 (2001)
- [6] 松原, 佐藤, 河口, 稲垣: 統計データに基づく話し言葉音声の係り受け解析, 情報処理学会研究報告, SLP36-4, pp.23-28 (2001).
- [7] 荒木, 伊藤, 熊谷, 石崎: 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).
- [8] 村尾, 河口, 松原, 稲垣: 対話事例を利用した音声対話システム, 情報処理学会研究報告, SLP34-34, pp.197-202 (2000)
- [9] 松本, 北内, 山下, 平野: 日本語形態素解析システム「茶筌」 version2.0 使用説明書 第 2 版, Information Science Technical Report, NAIST-IS-TR9908, 奈良先端技術大学院大学 (1999).
- [10] 河口, 松原, 武田, 板倉, 稲垣: 実走行車内音声対話データベース, 情報処理学会研究報告, SLP39-24, pp.141-146 (2001)