

音声対話コーパスにおける発話意図タグの設計と評価  
入江 友紀<sup>†a)</sup> 松原 茂樹<sup>††(正員)</sup>

河口 信夫<sup>††(正員)</sup> 山口由紀子<sup>††(正員)</sup>

稲垣 康善<sup>†††(正員:フェロー)</sup>

Design and Evaluation of Speech Act Tags in Spoken Dialogue Corpus

Yuki IRIE<sup>†a)</sup>, *Nonmember*, Shigeki MATSUBARA<sup>††</sup>,  
Nobuo KAWAGUCHI<sup>††</sup>, Yukiko YAMAGUCHI<sup>††</sup>, *Members*,  
and Yasuyoshi INAGAKI<sup>†††</sup>, *Fellow*

<sup>†</sup> 名古屋大学大学院情報科学研究科, 名古屋市

Graduate School of Information Science, Nagoya University,  
Nagoya-shi, 464-8601 Japan

<sup>††</sup> 名古屋大学情報連携基盤センター, 名古屋市

Information Technology Center, Nagoya University, Nagoya-  
shi, 464-8601 Japan

<sup>†††</sup> 愛知県立大学情報科学部, 愛知県

Faculty of Information Science and Technology, Aichi Pref-  
ectural University, Aichi-ken, 480-1198 Japan

a) E-mail: irie@el.itc.nagoya-u.ac.jp

あらまし 発話意図タグの設計と評価について述べる。設計した意図タグは、詳細度に応じて階層化している。名古屋大学 CIAIR 車内音声対話コーパスの約 35,000 発話に意図タグを付与した。信頼性を評価するため、コーパスの一部を用いてタグ付与実験を実施した。

キーワード 音声対話コーパス, 談話行為, 発話意図理解, 音声対話システム

## 1. ま え が き

近年、音声対話システムの研究が盛んに行われており、対話研究のための基礎データとして、大規模コーパスの蓄積が進んでいる。このようなコーパスを音声対話システムの開発に効果的に利用するためには、音声の書き起こしにとどまらず、様々な情報が付与されている必要がある。中でも音声対話システムへの入力となるユーザ発話の意図を正確に理解することがシステムが適切に振る舞うために必須であり、コーパスを構成する各発話に意図を表すタグが付与されていることが望ましい。

本論文では、音声対話コーパスにおける発話意図タグの設計について述べる。設計した発話意図タグは、これまでに使用されてきた談話タグ [1]~[5] に比べ、音声対話システムで利用するタグ付けコーパスを想定して体系化している。タスクに依存したレベル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。

名古屋大学 CIAIR 車内音声対話コーパスを用いて、

発話意図タグ付き音声対話コーパスを構築した。コーパスに収録されている対話から、レストラン検索をタスクとする約 35,000 発話に発話意図タグを手付け付与した。

本設計に基づくコーパスの信頼性を評価するために、タグ付け実験を実施した。実験により、本論文で提案した意図タグ体系を用いることによって、高い信頼性を有したコーパスを構築できることを確認した。

## 2. 発話意図タグ

これまでに、発話意図を表すタグとして、発話内行為レベルの情報を表す様々なタグが提案されている [1]~[5]。これらの多くは、Austin [6] や Searle [7] らによる発話行為理論が基礎となっており、例えば、「真偽情報要求」、「未知情報要求」、「依頼」、「情報伝達」等、数個から 20 個程度の談話タグが用いられている。

音声対話システムは、ユーザの発話意図を正しく理解することにより適切な振舞いが可能となる。しかしながら、従来の談話タグに基づく意図理解では、システムの具体的な応答や動作を決定するには必ずしも十分ではない。例えば、「そこは何時までやっているの」というユーザ発話に対して、「未知情報要求」という意図を推定しても、具体的に何を要求されているのかは明らかでなく、更に推論等の処理が必要になる。

そこで本研究では、システム動作を決定できるほどに詳細化された発話意図タグを設計し、コーパスに付与することとした。先の例の場合、「店の営業時間の提示を要求」を表す発話意図タグを与える。

しかしながら、このようなタグ内容の詳細化を行うと、タグの種類が増え、揺れが生じやすくなり、その結果、以下に述べる問題が生じる。

- タグ内容を詳細化すると、そこには、発話内行為などの抽象的な情報から、行為の対象といった具体的な情報までが含まれることになる。例えば、「店の営業時間の提示を要求」は「要求」という発話意図を具体化した意図であり、より詳細な情報を含んでいる。しかし、開発する対話システムによっては、タグ情報の詳細度のレベルを選択的に利用することが考えられる(汎用性の問題)。

- 発話の意図は、話者の表情やしぐさなどに現れることもあり、発話に対して必ずしも一意に決まるわけではない。例えば「今あいているかな」という発話は、「今、店の営業時間内かどうかを尋ねている」とも「空席があるかどうかを尋ねている」とも受け取れる。発話意図タグの付与では、対話に参加していない作業

者が正確に発話意図タグを付与することは難しく、作業によってタグが一致しない、といった問題が生じる（信頼性の問題）。

そこで我々は、タグ情報の階層化及びタグ付与の文脈依存性に注目し、タグの設計を行った。

[タグ情報の階層化] 汎用性の問題に対して、本研究では、意図タグ情報を抽象度によってレベル化することにより階層化する。詳細な発話意図は、各レベルの意図タグを組み合わせることにより表現できる。コーパスの構築では、階層間の関係を考慮することにより、揺れの少ないタグ付けを実現することができる。

[タグ付与の文脈依存性] 信頼性の問題に対して、本研究では、対話参加者による判断を尊重するとともに、対話参加者が十分に協調的であるとし、聞き手が話し手の発話意図をどのように理解したのかに基づいて、付与すべき意図を決める。具体的には、その発話に対する対話相手の応答発話を参照して定める。例えば、「今あいているかな」という発話への応答が「営業時間は9時から20時までです」であれば、もとの発話の意図は「営業時間を尋ねている」とみなし、「ただいま満席となっています」であれば、「空席状況を尋ねている」とみなす。

3. 意図タグの体系化とタグ付け

名古屋大学 CIAIR 車内音声対話コーパスの書き起こしテキスト [14] を使用し、意図タグ付き音声対話コーパスを構築した。

3.1 CIAIR 車内音声対話コーパス

名古屋大学統合音響情報研究拠点（CIAIR）では、実環境下でのロバストな音声対話システムの実現を目指し、道路案内や店情報検索をタスクとする、実走行車内音声対話コーパスの構築を行ってきた [8] ~ [12]。これは、被験者数約 800 名、言語データ量にして約 103 万語に及ぶ世界最大規模のマルチモーダルコーパスであり、貴重な音声言語資源として、今後、多くの研究者によって利用されていくものと予想される。

収録した音声データの書き起こし作業は人手によって行われている。書き起こし作業は日本語話し言葉コーパス（CSJ）の音声書き起こし基準 [13] に準拠している。図 1 に、書き起こしテキストの例を示す。

3.2 意図タグの体系化

2. で示した方針に従い、意図タグ体系を設計した。意図タグの例を図 2 に示す。また、各階層におけるタグの例を表 1 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」の四つの

```

0003 - 00:04:955 - 00:06:560 M:D:N:O:
    じゃあ          &ジャー
    マック          &マック
    教えてください<SB> &オシテクダサイ<SB>
0004 - 00:08:101 - 00:09:952 F:O:N:I:
    はい           &ハイ
    マクナルドです<SB> &マクナルドデス<SB>
0005 - 00:10:665 - 00:14:111 F:O:N:O:
    この先         &コノサキ
    二百メートル先に &ニヒャクメートルサキニ
    マクナルドが   &マクナルドガ
    あります<SB>   &アリマス<SB>
    
```

図 1 書き起こしテキストの例

Fig. 1 A sample of transcription of dialogue speech.

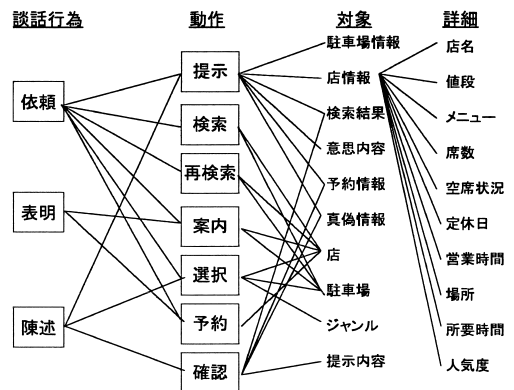


図 2 階層化された発話意図タグ（一部）

Fig. 2 A part of organization of intention tags.

表 1 各レイヤの発話意図タグ（一部）

Table 1 Examples of layered intention tag.

談話行為レイヤ	動作レイヤ	対象レイヤ	詳細レイヤ
依頼	提示	店情報	店名
提案	確認	駐車場情報	ジャンル
表明	検索	予約情報	値段
示唆	案内	検索結果	場所
陳述	選択	意思内容	空席状況

階層を有する。談話行為レイヤは話者の発話内行為を表している。動作レイヤは話者の行為を、対象レイヤは動作レイヤで定まる動作の対象を、詳細レイヤは対象に関する詳細情報をそれぞれ表す。

談話行為レイヤに含まれるタグは、すべてタスクに依存しないタグである。残りのレイヤは、より詳細化された意図を表現しており、タスクに依存したタグも含まれる。上の階層のタグと下の階層のタグに対し

表 2 意図タグ付き音声対話コーパスの例  
Table 2 An example of an intention-tagged corpus.

書き起こし		意図タグ			
話者	発話例	談話行為	動作	対象	詳細
ドライバー	この近くに中華の店ありますか	依頼	検索	店	
オペレータ	近くに大黒天があります	陳述	提示	検索結果	店名
ドライバー	その店にラーメンはあるのかな	依頼	提示	店情報	メニュー
オペレータ	はいございます	陳述	提示	店情報	メニュー
ドライバー	そこに案内して	依頼	案内	店	
オペレータ	大黒天までご案内します	表明	案内	店	

て、図 2 に示すような結び付き制約を定めた。例えば、談話行為レイヤが「表明」である発話の動作レイヤは「案内」、「予約」のいずれかであるとした。

### 3.3 音声対話コーパスへのタグ付け

名古屋大学 CIAIR 車内音声対話コーパスの書き起こしテキスト [14] を使用し、レストラン検索をタスクとする 3,641 対話に含まれる 35,411 発話に発話意図タグを人手で付与した。表 2 に意図タグ付き音声対話コーパスの例を示す。我々は「被験者（運転手）とオペレータ役（人間）との模擬対話」と「WOZ システムとの対話」の 2 種類の対話を対象に意図タグを付与している。これにより、対話相手の違いによる対話の遂行への影響の分析などが可能となる [14]。

発話意図タグの付与作業にあたって、タグ付与マニュアルを作成した<sup>(注1)</sup>。マニュアルでは、意図タグの体系と階層間の結び付き制約、並びに、作業の手順、タグ付与の単位などについて、発話例を用いて説明している。意図タグ付与の単位は、原則として、発話単位（200 ms 以上のポーズで分割された発話区間）とした。一般には、発話単位が、文など、タグ付与の対象となる単位と必ずしも一致するわけではないが、レストラン検索をタスクとする本コーパスではそれらの多くに一致が見られるという観察のもと、作業の効率を優先し、そのような原則を設けた。ただし、以下の例外を認めた。

- 発話意図が複数の発話単位にまたがる場合、それらを統合し、一つの発話意図タグを付与する。例えば、二つの接続する発話単位「人気のあるうどん屋ですと」、「兼松がこの先にございます」を統合し、「人気のあるうどん屋ですと兼松がこの先にございます」に対してタグを付与する。

- 一つの発話単位が、異なる発話意図をもつ複数の発話区間から構成される場合、それぞれに発話意図タグを付与する。例えば、発話単位「ダカル口には駐車場はありませんがよろしいですか」を発話区間「ダ

カル口には駐車場はありませんが」と「よろしいですか」に分割し、それぞれにタグを付与する。

なお、構築したコーパスに現れた発話意図タグは、談話行為、動作、対象、詳細レイヤの組合せで計 95 種類であった。

### 4. 意図タグ体系の評価

提案するタグの付与結果が、複数の作業間で一致しないようであれば、そのタグが付与されたデータから導かれる結論は信頼できるものとはいえない。設計した意図タグ体系の信頼性を検証するために評価実験を行った。

タグ体系の信頼性を議論する先行研究として、文献 [5], [15] が挙げられる。これらの研究では、複数作業者がタグ付けを行い、タグ体系の信頼性を評価している。また、主観が入る判定が複数作業者の間でどの程度一致するかを定量的に評価する指標としては、Cohen の kappa 値がよく使われる [5], [15], [16]。Cohen の kappa 値は、観測された一致率を  $P(O)$ 、期待される一致率を  $P(E)$  とすると、

$$\kappa = \frac{P(O) - P(E)}{1 - P(E)}$$

と表される。 $\kappa = 1$  は完全な一致を示し、 $\kappa = 0$  は偶然の一致程度の一致であることを示す。本研究では、この指標を用いてタグ体系の信頼性を評価した。

タグ付与を行う作業者として、ある程度の知識をもった専門家と特別な知識をもたない一般の作業者が考えられる。本研究では、コーパスの一部に対して、以下の意図タグ付与実験を行った。

[実験 1] 2 名の作業者（タグの設計者）による、28 対話、合計 296 発話に対する意図タグの付与

[実験 2] 4 名の作業者（特別な知識をもっていない）による、51 対話、合計 528 発話に対する意図タグ付与

(注1): <http://www.el.itc.nagoya-u.ac.jp/in-car/LIT-man.pdf>

表 3 意図タグの評価結果

Table 3 The results of evaluation experiments.

	実験 1 (設計者)	実験 2 (一般)
$P(O)$	0.853	0.705
$P(E)$	0.071	0.052
$\kappa$	0.842	0.689

表 4 意図タグの評価結果 (レイヤごと: 設計者)

Table 4 The result for each layer. (designers)

	談話行為	動作	対象	詳細
$P(O)$	0.939	0.911	0.904	0.881
$P(E)$	0.341	0.252	0.184	0.302
$\kappa$	0.907	0.881	0.883	0.829

ただし、各対話 2 名で意図タグを付与した。

作業者は、タグ付与マニュアルを参照して作業を遂行し、3.2 で述べたタグ体系、及び、階層間の結び付き制約に基づき、意図タグを付与した。

一致率について評価した結果を表 3~表 5 に示す。表 3 はすべてのレイヤのタグが一致したときにタグが一致したとみなしたときの値であり、表 4 の「談話行為」「動作」「対象」「詳細」は、実験 1 において、表 5 は実験 2 において、それぞれのレイヤのタグのみを評価対象として一致率を計算したときの値である。

音声対話研究において、kappa 値がどのくらいの範囲であれば、分析結果が信頼できるかの絶対的な基準はない。2 人の作業者の一致率について、文献 [15], [16] では、 $.80 < \kappa$  を good reliability,  $.67 < \kappa < .80$  を usable quality としている。この基準に従うと、設計者間の kappa 値は good reliability の範囲にあるので、このようなタグが付けられたデータから導かれる結論は信頼できるものであるといえる。95 種類と従来のタグよりも種類数が多いにもかかわらず、このような高い値が得られた理由として、タグ情報を階層化したことにより、発話に応じて決定しやすいレイヤに着目してタグを選択できたこと、タグ付与の文脈依存性を利用したことにより、タグの選択基準が明確になったことが挙げられる。

また表 4 から、すべてのレイヤにおいて kappa 値が good reliability の範囲にあり、どのレイヤにおいても信頼できるデータを構築できることが分かった。このことから、レイヤを選択的に利用したとしても、それらから導かれる結論は信頼できるものであるといえる。

今後、より大規模なコーパスを構築していく際には、必ずしも専門家がタグを付与するとは限らない。表 3

表 5 意図タグの評価結果 (レイヤごと: 一般)

Table 5 The result for each layer. (untrained annotators)

	談話行為	動作	対象	詳細
$P(O)$	0.821	0.795	0.833	0.821
$P(E)$	0.356	0.230	0.168	0.302
$\kappa$	0.722	0.733	0.799	0.744

及び表 5 が示すように、一般の作業者に対してトレーニングを事前に行わなくても、usable quality の範囲の一致率が得られている。このことから、本論文で提案する意図タグ体系は、特別な知識を有していなくても信頼性を備えたデータを構築できるといえる。

## 5. むすび

本論文では、名古屋大学 CIAIR 車内音声対話コーパスを用いた発話意図タグの設計と評価について述べた。設計した発話意図タグは、従来のコーパスアノテーションで使用されてきたタグに比べ、対話システムの開発に特化して体系化されている。名古屋大学 CIAIR 車内音声対話データベースに収録されている対話のうち、レストラン検索をタスクとする約 35,000 発話に意図タグを手で付与した。意図タグ付与と実験の結果、信頼性の得られるデータを構築することができることを確認した。

このようにして構築した意図タグ付き音声対話コーパスは、対話に関する基礎から応用まで様々な用途に利用可能である。特に、対話コーパスの実践的利用 [17] といった観点からは極めて重要な役割を担うと予想される。これまでに、談話分析 [14], [18], 発話意図推定 [19], [20], 音声対話システムの開発 [21] などで利用されている。また、これらの研究においては、目的によって注目するレイヤを使い分けており、設計した意図タグの汎用性が示されている。

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致します。また、タグの設計に対して貴重なコメントを頂いた三洋電機 (株) の村尾浩也氏に感謝致します。本研究の一部は、日本学術振興会科研費基盤研究 (B) (2) (No.15300045) によります。

## 文 献

- [1] J. Alexandersson, B. Buschbeck-Wolf, T. Fujinami, E. Maier, N. Reithinger, B. Schmitz, and M. Siegel, "Dialogue acts in verbmobil-2," *Verbmobil Report* 204, 1997.
- [2] J. Allen and M. Core, "Draft of DAMSL: Dialog act markup in several layers," <http://www.cs.rochester>.

- edu/research/cisd/resources/damsl/RevisedManual/RevisedManual.html, 1996.
- [3] M. Walker and R. Passonneau, "DATE: A dialogue act tagging scheme for evaluation of spoken dialogue systems," Proc. First Int. Conf. on Human Language Technology Research, pp.66–73, 2001.
- [4] T. Fukada, D. Koll, A. Waibel, and K. Tanigaki, "Probabilistic dialogue act extraction for concept based multilingual translation systems," Proc. 5th Int. Conf. on Spoken Language Processing, vol.6, pp.2771–2774, 1998.
- [5] 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," 人工知能誌, vol.14, no.2, pp.251–260, 1999.
- [6] J.L. Austin, How to do things with words, Harvard Univ. Press, 1962.
- [7] J.R. Searle, Speech acts: an essay in the philosophy of language, Cambridge Univ. Press, 1969.
- [8] 河口信夫, 牛藤誠一, 松原茂樹, 岩 博之, 梶田将司, 武田一哉, 板倉文忠, "走行車内室内音声対話収録システムの開発," 信学論 (D-II), vol.J84-D-II, no.6, pp.909–917, June 2001.
- [9] N. Kawaguchi, S. Matsubara, K. Takeda, and F. Itakura, "Construction of speech corpus in moving car environment," Proc. 7th European Conf. on Speech Communication and Technology, pp.2027–2030, 2001.
- [10] N. Kawaguchi, S. Matsubara, K. Takeda, and F. Itakura, "Multi-dimensional data acquisition for integrated acoustic information research," Proc. 3rd Int. Language Resources and Evaluation Conf., pp.2043–2046, 2002.
- [11] N. Kawaguchi, S. Matsubara, K. Takeda, and F. Itakura, "CIAIR in-car speech corpus -Influence of driving status," IEICE Trans. Inf. & Syst., vol.E88-D, no.3, pp.578–582, March 2004.
- [12] K. Takeda, H. Fujimura, K. Itou, N. Kawaguchi, S. Matsubara, and F. Itakura, "Construction and evaluation of a large in-car speech corpus," IEICE Trans. Inf. & Syst., vol.E88-D, no.3, pp.553–561, March 2004.
- [13] K. Maekawa, H. Koiso, S. Furui, and H. Isahara, "Spontaneous speech corpus of Japanese," Proc. LREC-2000, pp.947–952, 2000.
- [14] I. Kishida, Y. Irie, Y. Yamaguchi, S. Matsubara, N. Kawaguchi, and Y. Inagaki, "An advanced Japanese speech corpus for in-car spoken dialogue research," Proc. 8th European Conf. on Speech Communication and Technology, pp.1581–1584, 2003.
- [15] M.G. Core and J.F. Allen, "Coding dialogs with the DAMSL annotation scheme," Proc. American Association for Artificial Intelligence Fall Symposium on Communicative Action in Humans and Machines, pp.28–35, 1997.
- [16] J. Carletta, "Assessing agreement on classification tasks," Computational Linguistics, vol.22, no.2, pp.249–254, 1996.
- [17] 松原茂樹, 河口信夫, 外山勝彦, 武田一哉, "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," 人工知能誌, vol.17, no.3, pp.279–284, 2002.
- [18] Y. Irie, N. Kawaguchi, S. Matsubara, I. Kishida, Y. Yamaguchi, K. Takeda, F. Itakura, and Y. Inagaki, "An advanced Japanese speech corpus for in-car spoken dialogue research," Proc. Oriental COCOSDA-2003, pp.209–216, 2003.
- [19] 入江友紀, 松原茂樹, 河口信夫, 山口由紀子, 稲垣康善, "意図タグつきコーパスを用いた発話意図推定手法," 人工知能研資, SIG-SLUD-A301-03, pp.7–12, 2003.
- [20] Y. Irie, S. Matsubara, N. Kawaguchi, Y. Yamaguchi, and Y. Inagaki, "Speech intention understanding based on decision tree learning," Proc. ICSLP-2004, ThB2202p-16, 2004.
- [21] K. Hayashi, Y. Irie, Y. Yamaguchi, S. Matsubara, and N. Kawaguchi, "Speech understanding, dialogue management and response generation in corpus-based spoken dialogue system," Proc. ICSLP-2004, TuA1302p-22, 2004.

(平成 17 年 4 月 4 日受付, 5 月 30 日再受付)