

CIAIR 車内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^{†1}

松原 茂樹^{†2}

河口 信夫^{†2}

山口 由紀子^{†2}

稲垣 康善^{†3}

^{†1}名古屋大学大学院情報科学研究科 ^{†2}名古屋大学情報連携基盤センター

^{†3}愛知県立大学情報科学部

irie@el.itc.nagoya-u.ac.jp

1 はじめに

近年、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利用するためには、音声の書き起こしに留まらず、様々な情報が付与されている必要がある。なか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を付与されていることが望ましい。

本稿では、CIAIR 車内音声対話コーパス [1, 2, 3, 4] 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行為タグ [5, 6, 7, 8, 9] に比べ、タスクに依存したレベルまで意図を詳細化している。このことにより、システムの動作に直結した意図記述が可能となる。

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 3,641 対話を抽出し、それに含まれる約 35,000 発話に発話意図タグを手で付与することによって、意図タグつきコーパスを構築した。

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、本稿で提案した意図タグ体系を用いることにより、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 発話意図タグの設計

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 [5, 6, 7, 8, 9]。これらの多くは、Austin[10] や Searle[11] 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る舞いが可能となる。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、さら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、ゆれが生じやすくなり、その結果、

構築されたコーパスの信頼性が損なわれるといった問題が起こる。そこで我々は、タグ情報の階層化、タグ付与の文脈依存性に着目し、タグの設計を行った。

1. タグ情報の階層化

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、タグ情報のレベルを選択的に利用することが考えられる。

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、ゆれの少ないタグ付けを実現することができる。

2. タグ付与の文脈依存性

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。このため、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い、また、発話が意味的に曖昧であり作業者によってタグが一致しない、といった問題が生じる。

本研究では、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、元の発話の意図は「営業時間を探している」とみなし、「ただいま満席となっています」であれば「空席状況を探している」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 [4] を使用し、意図タグつき音声対話コーパスを構築した。まず、3.1 節で、CIAIR 車内音声対話コーパスについて説明し、3.2 節では、発話意図タグの設計



図 1: 音声データの収録車両

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> &オシエテクダサイ<SB>
 0004 - 00:08:101 - 00:09:952 F:O:N:I:
 はい &ハイ
 マクドナルドですね<SB> &マクドナルドデスネ<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 &ニヒヤクメートルサキニ
 マクドナルドが &マクドナルドガ
 あります<SB> &アリマス<SB>

図 2: 書き起こしテキストの例

について述べ、3.3 節では、コーパスへのタグ付けについて述べる。

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収集を行ってきた [3]。本拠点では、図 1 に示すような専用の音声データベース収録車を構築し、音声に加え、画像、車両操作情報、車両位置といったマルチモーダルな情報を 800 名に上る被験者に対し収録した。これは、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 [12] に準拠した。図 2 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 節で示した方針に従い、意図タグ体系を設計した。設計した意図タグ体系を図 3 に示す。意図タグは、「談話

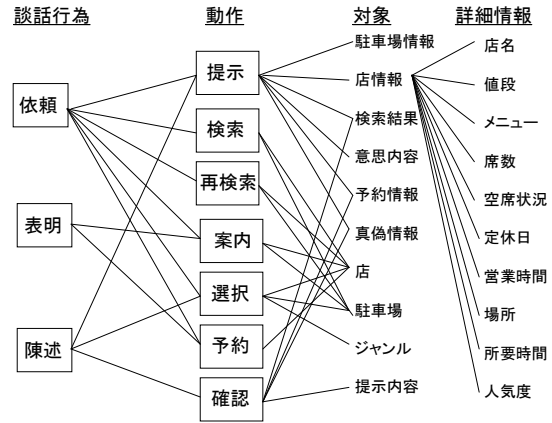


図 3: 階層的な発話意図タグ (一部)

表 2: 意図タグつきコーパスの規模

被験者数	1,256
対話数	3,641
ドライバー発話数	16,224
オペレータ発話数	19,187

行為レイヤ」「動作レイヤ」「対象レイヤ」「詳細情報レイヤ」の 4 つの階層を有する。談話行為レイヤは話者の発語内行為を表しており、動作レイヤは話者の行為を表している。対象レイヤは動作レイヤで定まる動作の対象を表し、詳細情報レイヤは対象に関する詳細情報を表す。談話行為レイヤに含まれるタグは、全てタスクに依存しないタグである。動作レイヤ、対象レイヤ、詳細情報レイヤには、タスクに依存したタグも含まれる。

また、上位の階層のタグと下位の階層のタグには図 3 に示すような、結びつき制約がある。

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 [4] 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 1 に意図タグつき音声対話コーパスの例を示す。

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ている¹。マニュアルには、それぞれのタグの説明を発話例を用いて説明してある。書き起こしコーパスの作成では、200msec 以上のポーズで発話を分割しているため、原則として 1 発話単位に 1 つの発話意図タグを付与した。ただし、次の例外を認めた。

- 複数の発話単位にまたがる意図
複数の発話単位を統合し、1 つの発話意図タグを付与する。
- 複数の発話意図を持つ発話単位
節を目安に発話を分割し、分割したそれぞれの発話に発話意図タグを付与する。

作成した意図タグつき音声対話コーパスの規模を表 2 に示す。

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

表 1: 意図タグつき音声対話コーパスの例

書き起こし		意図タグ			
話者	発話	談話行為	動作	対象	詳細情報
ドライバー	この近くに中華の店あるかな	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	んーこっからどのくらい	依頼	提示	店情報	所要時間
オペレータ	10分ほどになります	陳述	提示	店情報	所要時間
ドライバー	じゃあそこに案内して	依頼	案内	店	
オペレータ	かしこまりました	陳述	提示	意思内容	了解
オペレータ	大黒天までご案内します	表明	案内	店	

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、詳細情報レイヤの組み合わせで計 95 種類であった。

4 意図タグ体系の評価

本稿で提案する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。

提案するタグの付与結果が、複数の作業員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言えない。このような、タグ体系の信頼性を議論する先行研究としては、文献 [9, 13] があげられる。これらの研究では、複数作業員がタグ付けを行い、タグ体系の信頼性を評価している。

主観が入る判定が複数作業員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 値がよく使われている [9, 13, 14]。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。

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。

タグ付与を行う作業員として、ある程度の知識を持った専門家と特別な知識をもたない一般の作業員が考えられる。そこでコーパスの一部について、以下の意図タグ付与実験を行った。

実験 1 作業員数は 2 名 (タグの設計者) で、28 対話、合計 296 発話に対して意図タグを付与した。

実験 2 作業員数は 4 名 (うち、3 名は音声・言語処理に関する知識をほとんど持っていない) で、51 対話、合計 528 発話に対して、各対話 2 名で意図タグを付与した。

一致率の評価の結果を表 3、表 4 に示す。表中の「全レイヤ」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、「談話行為」「動作」「対象」「詳細情報」は、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

表 3: 意図タグの評価結果 (設計者)

	全レイヤ	レイヤ別			
		談話行為	動作	対象	詳細情報
$P(O)$	0.853	0.939	0.911	0.904	0.881
$P(E)$	0.071	0.341	0.252	0.184	0.302
κ	0.842	0.907	0.881	0.883	0.829

表 4: 意図タグの評価結果 (一般)

	全レイヤ	レイヤ別			
		談話行為	動作	対象	詳細情報
$P(O)$	0.705	0.821	0.795	0.833	0.821
$P(E)$	0.052	0.356	0.230	0.168	0.302
κ	0.689	0.722	0.733	0.799	0.744

い。2 人の作業員の一致率について、文献 [13, 14] では、 $0.80 < \kappa$ を good reliability、 $0.67 < \kappa < 0.80$ を usable quality としている。

設計者間の kappa 値について、文献 [13, 14] の基準では 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関わらず、このような高い値を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あげられる。

また表 4 から、全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、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 4 より、一般の作業員に対してトレーニングを事前に行わなくても、usable quality の範囲の一致率を得られたことから、本稿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の得られるデータを構築することができるといえる。

5 おわりに

本稿では、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から、レストラン検索をタスクとする 3,641 対話を抽出し、それに含まれる約 35,000 発話に意図タグを人手で付与した。意図タグ付与実験の結果、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つ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。我々はこれまでに、談話分析 [4, 15, 16, 17]、発話意図推定 [18, 19]、音声対話システムの開発 [20] などで成果を得ている。

今後は、頑健性を備えた音声対話システムの実現はもちろん、話し言葉文法と発話意図の関係付け、コーパスからの対話文法の知識獲得、音声対話からの知識獲得等に利用できると考えられる。

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いたします。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いたします。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。

参考文献

- [1] 河口信夫, 牛窪誠一, 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, no.6, pp.909-917, 2001.
- [2] 河口信夫, 松原茂樹, 岩博之, 梶田将司, 武田一哉, 板倉文忠, “実走行車内における音声データベース”, 情報処理学会研究報告, SLP-30-12, pp.57-62, 2000.
- [3] N. Kawaguchi, S. Matsubara, Y. Yamaguchi, K. Takeda, and F. Itakura, “CIAIR in-car speech database”, Proc. of ICSLP-2004, 2004.
- [4] 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. of EuroSpeech-2003, 2003.
- [5] J. Alexandersson, B. Buschbeck-Wolf, T. Fujinami, E. Maier, N. Reithinger, B. Schmitz, and M. Siegel, “Dialogue acts in verbmobil-2”, Verbmobile Report 204, 1997.
- [6] 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.
- [7] M. Walker and R. Passonneau, “DATE: A Dialogue Act Tagging Scheme for Evaluation of Spoken Dialogue Systems”, Proc. of the First Int. Conf. on Human Language Technology Research, pp.66-73, 2001.
- [8] T. Fukada, D. Koll, A. Waibel and K. Tanigaki, “Probabilistic Dialogue Act Extraction for Concept Based Multilingual Translation Systems”, Proc. of ICSLP-1998, Vol.6, pp.2771-2774, 1998.
- [9] 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.2519-260, 1999.
- [10] J. L. Austin, How to do things with words, Harvard Univ. Press, 1962.
- [11] J. R. Searle, Speech acts: an essay in the philosophy of language, Cambridge Univ. Press, 1969.
- [12] K. Maekawa, H. Koiso, S. Furui, and H. Isahara, “Spontaneous speech corpus of Japanese,” Proc. of LREC-2000, pp.947-952, 2000.
- [13] M. G. Core and J. F. Allen, “Coding dialogs with the DAMSL annotation scheme”, Proc. of the American Association for Artificial Intelligence Fall Symposium on Communicative Action in Humans and Machines, pp.28-35, 1997.
- [14] J. Carletta, “Assessing agreement on classification tasks”, Computational Linguistics, vol.22, no.2, 1996.
- [15] 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. of Oriental COCOSDA-2003, pp.209-216, 2003.
- [16] 加藤真吾, 入江友紀, 山口由紀子, 松原茂樹, 河口信夫, “CIAIR 車内音声対話コーパスを用いた対話フロー解析”, 情報処理学会全国大会講演論文集, vol.2, pp.157-158, 2004.
- [17] 加藤真吾, 松原茂樹, 山口由紀子, 河口信夫, “発話意図タグに基づく対話構造の記述”, 言語処理学会第 11 回年次大会発表論文集, 2005.
- [18] 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”, 人工知能学会研究会資料, SIG-SLUD-A301-03, pp.7-12, 2003.
- [19] Y. Irie, S. Matsubara, N. Kawaguchi, Y. Yamaguchi, and Y. Inagaki, “Speech Intention Understanding based on Decision Tree Learning”, Proc. of ICSLP-2004, 2004.
- [20] 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. of ICSLP-2004, 2004.