

# 大規模音声言語コーパスの高度化と 走行車内音声対話の特徴分析

岸田 樹<sup>1</sup> 入江 友紀<sup>1</sup> 山口 由紀子<sup>2</sup> 松原 茂樹<sup>2 4</sup> 河口 信夫<sup>2 4</sup> 稲垣 康善<sup>3</sup>

<sup>1</sup>名古屋大学工学部電気電子情報工学科 <sup>2</sup>名古屋大学情報連携基盤センター

<sup>3</sup>名古屋大学大学院工学研究科情報工学専攻 <sup>4</sup>名古屋大学統合音響情報研究拠点 (CIAIR)  
itsuki@inagaki.nuie.nagoya-u.ac.jp

## 1 はじめに

名古屋大学統合音響情報研究拠点 (CIAIR) では、ロボットな実環境音声対話システムの実現を目指し、実走行車内音声対話コーパスの収集を行ってきた [3]。このコーパスは、被験者数約 800 名に対して、ドライバとナビゲータとの間で遂行される対話について、音声、画像、運転情報、書き起しテキストを収録した世界最大規模のマルチモーダルコーパスである。我々は、このコーパスに基づき、言語現象の分析 [3] や事例ベース対話システムの開発 [9] を推進してきた。大規模コーパスは、多様な研究を推進するための重要な情報源であり、様々な情報が付与されていれば、その応用可能性は大きく広がる [8]。

本稿では、本コーパスの高度化の事例として、言語構造情報、および発話意図情報の付与について報告する。これらの情報の付与により、本コーパスは、様々な観点からの分析・利用が可能な多層的なコーパスとなった。また、本コーパスは、人対人、人対 WOZ システムの対話を収録している点に特徴がある。本稿では、この違いについて、コーパスの多層性を利用した比較を行った。

## 2 実走行車内音声対話コーパス

実走行車内対話音声の収録は、分析、調査、利用が可能なデータの収集を目的としている [4]。本収録においては、対話相手の違いによる対話の遂行への影響を調査するために、対話相手として以下に述べる 3 種類のナビゲータを設定した。

- 人間：事前にナビゲータとして訓練を受けており、タスク達成のための詳細な情報を備えている。ただし、対話が発散することを避けるために、発声スタイルに制限を設けている。
- WOZ システム：人間によるタッチパネル入力と機械による音声合成からなる対話システムである [9]。
- 音声対話システム：レストラン検索を対話ドメインとした、システム主導の対話システムである [7]。

収録の様子を図 1 に示す。被験者 (左下) は運転席に、実験補助者 (右下) は後部座席に位置し、タッチパネルを使って WOZ (右上) 等を操作する [3]。表 1 に、1999 年度から 2001 年度の 3 年間に収録した CIAIR 車内音声対話データベースの概要を、人間との対話 (HUM)、WOZ



図 1: コーパスの収録状況

システムとの対話 (WOZ)、音声対話システムとの対話 (SYS) に分けて示す。なお、1999 年度は人間との対話のみを、また、残りの 2 年間は全種類の対話を収録した。その総収録時間は約 179 時間である。

対話音声の収録は、店検索や道案内など、車内での遂行が想定されるいくつかのタスクを設定し、実走行環境下で模擬的に行っている。収録された対話タスクの分布を図 2 に示す。また、収録音声の書き起こしは、日本語話し言葉コーパス (CSJ) の書き起こし基準 [6] に準拠した。文字化データに加え、時間情報や話者情報、言語現象などのタグ付与を実施している。書き起こしテキストの例を図 3 に示す。

## 3 音声対話コーパスの高度化

一般に、音声対話システムは、音声処理、言語処理、対話処理など、レベルの異なる要素技術の組み合わせによって実現できる。収集した対話データをシステムの性能向上に利用するには、単なる音声の収録や書き起こしだけでなく、より高度化された情報を付与することが求められる。そこで我々は、本コーパスのテキストデータに対し、構文や意味などの多様な言語分析を与えることにより、対話コーパスの高度化を進めてきた。これによ

表 1: 音声対話データベースの概要

項目	99HUM	00HUM	00WOZ	00SYS	01HUM	01WOZ	01SYS	合計
収録時間 (秒)	141822	94692	65746	77922	93465	93862	78169	645678
データ数	209	294	199	288	295	294	287	1866
総発話時間 (秒)	98100	69390	31672	54056	67635	47424	48877	417154
ドライバ	44722	28085	12425	11515	26055	18127	11001	151930
オペレータ	53328	41305	19247	42541	41580	29297	37876	265174
総発話単位数	38760	25251	11992	24944	24178	19993	22904	168022
ドライバ	20493	12555	6099	10567	11985	9245	10722	81666
オペレータ	18267	12696	5893	14377	12193	10748	12182	86356
総発話文数	36691	23892	10767	23088	22582	16172	21270	154462
ドライバ	19007	11675	5628	9515	10983	8475	9722	75005
オペレータ	17684	12217	5139	13573	11599	7697	11548	79457

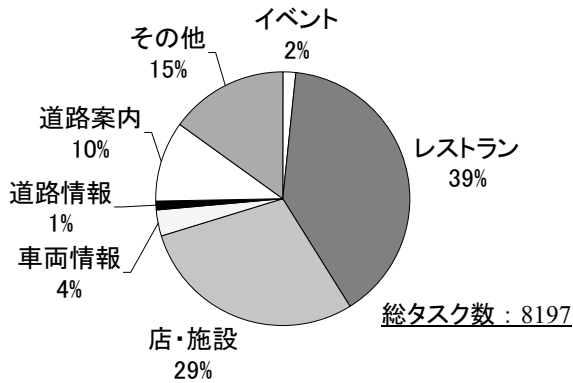


図 2: 収録された対話のタスク分布

0032 - 01:48:502-01:54:821 M:D:I:DI:  
 (F んー) & (F ンー)  
 辛い & カライ  
 台湾ラーメンが & タイワンラーメン  
 食べたいんだけど & タベタインダケド  
 どっか & ドッカ  
 お店 & オミセ  
 ないかなあ<SB> & ナイカナア<SB>  
 0033 - 01:55:975-01:58:093 M:O:I:DB:  
 はい & ハイ  
 ニ軒 & ニケン  
 あります<SB> & アリマス<SB>

図 3: 書き起しテキストの例

り、図 6に示すような多層化された音声対話コーパスを構築することができた。以下では、その事例として、係り受け分析、および、発話意図分析が与えられたコーパスについて述べる。

### 3.1 係り受けタグ付きコーパス

ドライバ発話に対して、係り受け分析を与えている [2]。ポーズにまたがる係り受けや前方に係る係り受けも認めており、自然発話に対応したデータ仕様を採用している。係り受けタグ付きコーパスの例を図 4に示す。文節間の係り受けだけでなく、形態素情報や発話単位情報、ターン情報等も含んでおり、多様なレベルの言語情報を有している。このコーパスは、統計に基づく係り受け解析の係り受け確率の獲得に利用されている [10]。

### 3.2 意図タグ付きコーパス

人間との対話、および、WOZシステムとのレストラン検索対話のドライバ、ナビゲータ双方の発話に対し、意図を付与している [1]。付与するにあたり、意図の抽象度に応じて階層化した意図タグの体系を設計した。意図タグ付きコーパスの例を図 5に示す。話者情報や時間情報、言語現象に関するタグも含んでおり、定量的な談話分析に適した仕様となっている。なお、このデータは、事例に基づく発話意図推定において、事例データベースとし

て活用している [11]。

## 4 音声対話コーパスの特徴分析

高度化された音声対話コーパスを用いて、車内対話の特徴分析を行った。本稿では、発話意図と発話長の関係、および、発話意図と言語現象の関係の分析結果について述べる。特に、人間との対話とWOZシステムとの対話におけるドライバ発話に着目し、その比較を行う。

### 4.1 発話意図と発話長の関係

対話相手に意図を伝えるときに必要な情報の量は、その意図のタイプによって異なると予想される。そこで、発話の意図とその長さとの関係を調べた。調査では、意図タグ付きコーパスを用いて発話を分類し、係り受けタグ付きコーパスを用いて発話の文節数を調べた。

生起頻度が上位 10 位までの意図タグに対する、発話に含まれる平均文節数を図 7に示す。これらの 8 つの発話意図は全体の 91.4 %を占める。このグラフから、「依頼」に関連した意図をもつ発話は長くなる傾向があることがわかる。車内対話では、ドライバがナビゲータに物事を依頼するとき、依頼するに至った理由を述べる傾向があり、それが原因で発話が長くなると考えられる。

また、意図の種類に関わらず、人間との対話における発話の方が、WOZシステムとの対話よりも長い。実際、HUM データにおけるドライバ発話は、平均 3.4 文節から

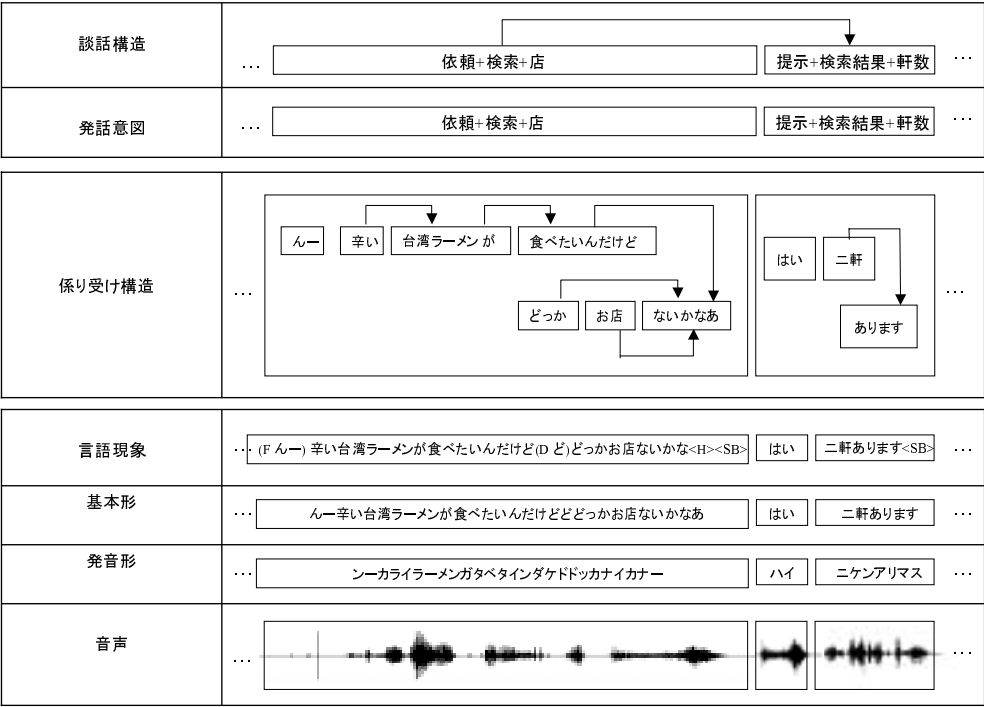


図 6: 多層化された音声対話コーパス

(TIME 01:48:502-01:54:821)

((1 (([F んー] [F んー] [F んー] フィラー なしなし))

-> (NO (なし)))

((2 ((辛い ツライ 辛い 形容詞 自立 形容詞・アウオ段 基本形) ))

-> (3 ((台湾ラーメン タイワンラーメン 台湾ラーメン 名詞 一般 なし) (が が 助詞 格助詞 なし)) ))

((3 ((台湾ラーメン タイワンラーメン 台湾ラーメン 名詞 一般 なし) (が が 助詞 格助詞 なし)) ))

-> (4 ((食べ タベ 食べる 動詞 自立 一段 連用形) (たい タイ たい 助動詞 なし 特殊・タイ 基本形) (ん ん ん 名詞 非自立 なし) (だ ダ だ 助動詞 なし 特殊・ダ 基本形) (けど ケド けど 助詞 接続助詞 なし)) ))

((4 ((食べ タベ 食べる 動詞 自立 一段 連用形) (たい タイ たい 助動詞 なし 特殊・タイ 基本形) (ん ん ん 名詞 非自立 なし) (だ ダ だ 助動詞 なし 特殊・ダ 基本形) (けど ケド けど 助詞 接続助詞 なし)) ))

-> (7 ((ない ナイ ない 形容詞 自立 形容詞・アウオ段 基本形) (か カ か 助詞 副助詞／並立助詞／終助詞 なし) (なあ ナア なあ 助詞 終助詞 なし)) ))

((5 ((どっか ツツカ どっか 名詞 代名詞 なし)) ))

-> (7 ((ない ナイ ない 形容詞 自立 形容詞・アウオ段 基本形) (か カ か 助詞 副助詞／並立助詞／終助詞 なし) (なあ ナア なあ 助詞 終助詞 なし)) ))

((6 ((お お 接頭詞 名詞 接続 なし) (店 ミセ 店 名詞 一般 なし)) ))

-> (7 ((ない ナイ ない 形容詞 自立 形容詞・アウオ段 基本形) (か カ か 助詞 副助詞／並立助詞／終助詞 なし) (なあ ナア なあ 助詞 終助詞 なし)) ))

図 4: 係り受けタグ付きコーパスの例

構成されるのに対して、WOZ システムでは 3.0 文節であった。これは、ドライバがシステムの対話性能を考慮し、簡潔に発話していることによると思われる。

4.2 発話意図と話し言葉に特有の現象

これまでフィラーや言い淀みなどの対話に特有な言語現象の出現頻度や出現位置などが様々なデータで調査されてきた [5]。本研究では、言語現象の出現と発話意図の関係について分析を与える。調査では、意図タグ付きコーパスを用いて発話を分類し、また、書き起こしコーパスに付与された言語現象タグを数え上げるにより言語

0032 D 依頼+検索+店  
んー辛い台湾ラーメンが食べたいんだけどどっかお店ないかなあ  
0033 O 提示+検索結果+軒数  
はい二軒あります

図 5: 意図タグ付きコーパスの例

現象の出現頻度を調べた。

前節と同様、生起頻度が上位 10 位までの意図タグに着目した。文節あたりのフィラーの出現頻度の平均を調査した結果を図 8 に示す。なお、図における値は、コーパス全体の文節あたりの平均出現回数 (HUM では 0.15, WOZ では 0.12) に対する比を示している。「依頼」に関連した発話の場合、その出現密度はいずれも平均的な値を示しているが、それ以外の発話では、意図によって異なる値を示している。また、「提示」に関連した意図をもつ発話において、人間との対話における発話の方が、WOZ システムとの対話よりもフィラーの出現頻度が小さいことが分かった。これは、意思や情報を提示する発話の多くは、ナビゲータから提示を要求されたときの応答であり、WOZ システムと対話するときには合成音声聞くことになるため、ドライバは簡潔に応答することによって考えられる。なお、表 2 に示すように、WOZ システムとの対話と人間との対話では、一般に前者の方でフィラーが少なくなることが知られている [5]。上記のような事象が、その原因の一端を担っている可能性がある。

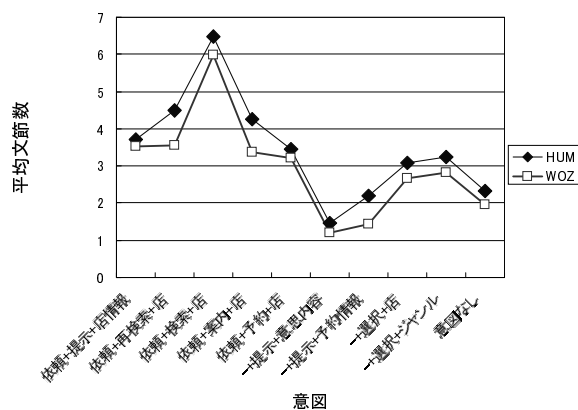


図 7: 発話意図と平均文節数の関係

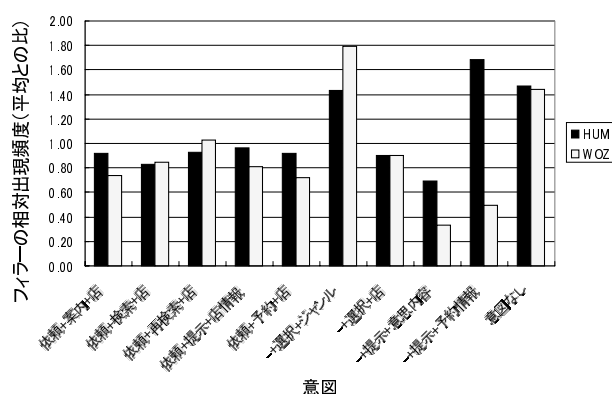


図 8: 発話意図とフィラー出現頻度の関係

## 5 まとめ

本稿では、名古屋大学 CIAIR で収集を行ってきた車内音声対話コーパスの高度化について述べた。音声データおよび、書き起こしデータに対して、言語構造情報や発話意図情報等のデータを新たに作成し、それらを関連付けることにより、多様な観点から分析可能な多層化音声対話コーパスを構築できる。本稿ではまた、係り受けタグ付きコーパスと意図タグ付きコーパスを用いて、発話意図と発話長の関係、および、発話意図と言語現象の関係について分析した結果について述べた。

## 謝辭

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、データ分析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、ならびに、係り受け分析データの使用についてご教示いただいた稲垣研究室の大野誠寛君に感謝いたします。本研究の一部は、文部科学省科学研究費補助金 COE 形成基礎研究費（課題番号 11CE2005）の補助を受けて行われた。

表 2: 発話単位あたりのフィラーの出現率

	HUM	WOZ
項目	出現率 (%)	出現率 (%)
フィラー (D)	46.8	34.6

## 参考文献

- [1] 入江友紀, “意図タグ付き音声対話コーパスの構築とそれを用いた発話意図推定”, 名古屋大学工学部情報コース平成 14 年度卒業研究報告 (2003).
- [2] 大野誠寛, 松原茂樹, 河口信夫, 稲垣康善, “統計的構文解析器を用いた音声言語係り受けコーパスの構築”, 言語処理学会第 9 回年次大会論文集 (2003).
- [3] 河口信夫, 松原茂樹, 武田一哉, 板倉文忠, 稲垣康善, “実走行車内音声対話コーパス”, 情処研報, SLP-39, pp. 141-146 (2001).
- [4] 河口信夫, 牛窪誠一, 松原茂樹, 梶田将司, 武田一哉, 板倉文忠, “走行車室内音声対話収録システムの開発”, 電子情報通信学会論文誌, Vol.J84-D-II, No. 6, pp. 1122-1129 (2001).
- [5] 河口 信夫, 松原 茂樹, 若松 佳宏, 梶田 将司, 武田 一哉, 板倉 文忠, 稲垣 康善, “実走行車内音声対話コーパスの設計と特徴”, 情処研報, SLP-3, pp.179-184 (2000).
- [6] 小磯花絵, 土屋菜穂子, 間淵洋子, 斉藤美紀, 籠宮隆之, 菊池英明, 前川喜久雄, “「日本語話し言葉コーパス」の書き起し基準について”, 信学技報, SP, Vol.100, No.52, pp.55-60 (2000).
- [7] 早川昭二, 磯部俊洋, 河口信夫, 武田一哉, 板倉文忠, “音声対話システムを用いた車内対話の収集”, 音響学会講演論文集, 3-P-25, pp.213-214 (2001).
- [8] 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 -”, 人工知能学会誌, Vol.17, No.3, pp.279-284 (2002).
- [9] 村尾浩也ら, “WOZ システムのログ情報を利用した事例データベース音声対話システムの開発”, 第 4 回音声言語シンポジウム論文集 pp.135-140 (2002).
- [10] S. Matsubara, T. Murase, N. Kawaguchi, and Y. Inagaki, “Stochastic Dependency Parsing of Spontaneous Japanese Spoken Language”, Proc. of COLING-2002, Vol. 1, pp. 640-645 (2002).
- [11] S. Matsubara, S. Kimura, N. Kawaguchi, Y. Yamaguchi and Y. Inagaki, “Example-based Speech Intention Understanding and Its Application to In-Car Spoken Dialogue System”, Proc. of COLING-2002, Vol. 1, pp. 633-639 (2002).