

〔論 文〕

高齢者のための音声対話システム

—方言への対応に向けて—

高橋 雅仁^{*1}・原 圭孝^{*1}・佐野 弘明^{*1}
服部 雄紀^{*2}・小田まり子^{*3}

A Spoken Dialog System for the Elderly:
Toward the Realization of a Dialect-enabled Dialog System

Masahito TAKAHASHI^{*1}, Yoshitaka HARA^{*1}, Hiroaki SANO^{*1}
Yuki HATTORI^{*2}, Mariko ODA^{*3}

Abstract

The Intelligent Mobility Laboratory at the Kurume Institute of Technology is developing an interactive electric wheelchair called “Partner Mobility.” In this study, we propose an improvement to the spoken dialog system to enable the use of dialects. This will allow elderly people, who are the main users of Partner Mobility, to interact with it without stress. Specifically, correction of speech recognition errors in dialectal utterances is added as a speech recognition post-processing process for the user utterances; this is followed by the extraction of keywords such as nouns and verbs from the utterances. Then, by matching the extracted keywords with pre-prepared case frame patterns centered on verbs, the system can grasp the user's intention and advance the user's dialog with Partner Mobility. In addition, we report the results of an experiment comparing the speech recognition accuracy of the system for standard Japanese versus a dialect (the Chikugo dialect).

Key Words: Spoken Dialog System, Dialect, Speech Recognition, Elderly, Partner Mobility, Dialog Agent

1. はじめに

久留米工業大学では、2015年より、インテリジェント・モビリティ研究所（IML）を中心に、介助者なしで移動を楽しめる人工知能搭載の対話型電動車いす「パートナー・モビリティ」（図1）の開発を進めてきた⁽¹⁾。2019年度から、空港、病院、シニアマンションなどで実証実験が開始され、2020年11月には、福祉施設への導入も行われている。本稿では、このパートナー・モビリティの主たる利用者である高齢者がストレスなしにパートナー・モビリティとの対話を行えるように、音声対話システムにおいて方言を使用可能とするための改善案に



図1 パートナー・モビリティ

^{*1} 情報ネットワーク工学科

^{*2} インテリジェント・モビリティ研究所

^{*3} AI 応用研究所

令和2年12月26日受理

について検討を行う。

人工知能分野における深層学習の研究開発の進展により、近年、音声認識や機械翻訳の性能が飛躍的に向上し、Web上の翻訳サービスやAIスピーカ、携帯型音声翻訳機等の商品を誰でも安価に利用できるようになった。一方、高齢者のための音声対話システムでは、高齢者がストレスなく対話を行えるように、高齢者が日常使用している方言への対応を図っておくことが望ましい。しかしながら、深層学習によってこの課題に対応するためには、方言毎に、機械学習用の大量の音声発話データを収集する必要がある、その実現のためのコスト負担が大きく、国立国語研究所による諸方言データベース構築や弘前大学による津軽弁の文例データベース構築のプロジェクトが進行中であるが、現状では方言への対応は十分に行われているとは言い難い⁽²⁾⁽³⁾⁽⁴⁾⁽⁵⁾。

本研究では、パートナー・モビリティに搭載されている音声対話システムにおける方言への対応方法について検討を行う。具体的には、汎用の音声認識システムを使用し、音声認識処理の後処理として、方言による発話文の音声認識誤りの訂正処理を加え、さらに、音声認識誤り訂正後の発話文から名詞や動詞などのキーワードを抽出し、それらのキーワードと予め準備した動詞の格フレームパターンとのマッチングを行い、その結果を基に、ユーザ発話文に対する意味表現を生成することによって、ユーザの意図を把握し、対話を進める方式を提案する。また、パートナー・モビリティで用いられている音声認識システムにおける標準語による発話と方言（筑後方言）による発話とでの音声認識精度の比較実験結果についても報告する。なお、本研究では、シニアマンション内でのパートナー・モビリティの利用を想定して、シーン別の対話シナリオを作成し、この対話シナリオに基づいて音声対話システムの方言への対応方法について検討を行った。

2章では現状のパートナー・モビリティの音声対話システムの概要について説明を行い、3章では方言に対応可能なパートナー・モビリティ用の音声対話システムの構想について述べる。4章では標準語と方言による音声認識精度の比較実験の結果について報告を行う。5章で今後の課題等について述べる。

2. パートナー・モビリティの音声対話システムの概要

2・1 パートナー・モビリティの音声対話処理

パートナー・モビリティの音声対話システムにおける音声対話処理の概要を以下に記す。

- ・対話処理は、(株)NTTドコモ社のドコモAIエージェントを用いて、Web上で行われる⁽⁶⁾。
- ・利用者の発話は、AIエージェントの音声認識機能を利用して、テキストデータに変換される。
- ・AIエージェントには、対話シナリオ作成機能があり、これを用いて対話処理を行う。あらかじめユーザ発話文に現れるキーワードとパートナー・モビリティの応答文の組を登録しておき、ユーザ発話の音声認識結果として得られたテキストデータとキーワードとのマッチングによって、パートナー・モビリティの応答文が選択され、ユーザとパートナー・モビリティとの対話が行われる。
- ・ユーザ発話文に、場所のキーワードが含まれていれば、パートナー・モビリティ内の自動運転機能にクラウド上のモビリティサービス統合サーバーから目的地の情報が渡され、移動の制御が行われる。

2・2 パートナー・モビリティの音声対話システムにおける課題

パートナー・モビリティの音声対話システムの課題を以下に記す。

- ・現在は、施設内での移動指示や限られたサービスへの対応が可能であるが、今後、病院、介護施設、美術館、空港など様々な場所における広範なサービス内容に応じたより柔軟な音声対話機能が必要となる。
- ・主な利用者である高齢者の発話に対する音声認識精度が不十分である。特に、方言による発話が行われた場合の対応が必要である。

3. 方言に対応した音声対話システムの構想

3・1 シーン別の対話シナリオの作成

シニアマンション内でのパートナー・モビリティの利用を想定し、シーン別の対話シナリオを作成した。各利用エリアに応じた対話シナリオは、パートナー・モビリティの現在位置の情報を取得することにより、自動的に切り替えられ、選択された対話シナリオを用いて適切かつ柔軟な対話処理を行うことを目指している。本研究では、パートナー・モビリティの利用エリアとして、ユーザの居室前、レストラン、ロビーの3か所での対話シーンを想定し、まず、標準語に

よる対話シナリオの作成を行った。対話シナリオの一部を図2に示す。なお、現時点では、パートナー・モビリティの対話機能は、場所の移動やインターネットを用いた情報収集に関するものに限られているが、将来的には、各エリアでのサービスを代行するエージェント機能をパートナー・モビリティがもつものと考えて、対話シナリオを作成した。

対話シナリオ	ユーザの発話	パートナー・モビリティの発話
居室前から レストランへの移動	レストランに行きたいんだけど。	
		了解しました。レストランですね。
	はい。	
		それでは、今から、レストランに移動します。
	レストランに着いたね。	
		レストランに入ってよいですか？
	入っていいよ。	
レストラン での料理の注文		それでは、レストランに入ります。
	メニューを見せて。	
		かしこまりました。こちらがメニューです。（ディスプレイに表示）
	今日のランチは、何？	
		ビーフハンバーグとアサリのパスタの2種類がございます。
	オーダーを取ってくれる？	
		少々お待ちください。
	ビーフハンバーグをお願いします。	
		パンとライスがありますが、どちらになさいますか？
	パンをお願いします。	
		パンですね。飲み物はどうなさいますか？
	食後にコーヒーをお願いします。	
		食後にコーヒーですね。
		ご注文は、以上でよろしいでしょうか。
	はい。	
ロビーでの タクシー予約		かしこまりました。しばらくお待ちください。
		いらっしゃいませ。何か御用ですか。
	13時に車いすで乗れるタクシーを呼んでほしいのですが。	
		承知しました。13時ですね。どちらにお出かけですか。
	美術館に行きたいんです。	
		美術館までですね。予約をしますので少々お待ち下さい。
		13時に予約がとれました。
	ありがとう。	
		どういたしまして。

図2 対話シナリオの例

3・2 解析処理の難易度による対話シナリオにおけるユーザ発話文の分類

文献（7）にしたがって、対話シナリオに含まれるユーザ発話文計78文に対して、それらを解析処理の難易度によって以下の3つのレベルに分類した。

- ・レベル1：名詞のみの受け答えや修飾句のない口語表現の単文である。

- ・レベル2：修飾句を含む、あるいは、助詞や文末表現の省略などがある口語表現の単文である。
- ・レベル3：重文や複文、並列句、倒置表現などやや複雑な構造を含む口語表現の文である。

分類結果を表1に示す。表1より、レベル1とレベル2の文で全体の91%を占めていることがわかり、文のパターンの特徴を利用することで、かなりの程度、従来の解析手法に沿った解析処理が可能であると推察される。

各レベルの文例を以下に記す。

- ・レベル1：「レストランに行きたいんだけど。」、「メニューを見せて。」、「お勘定をお願いします。」
- ・レベル2：「こんにちは、予約をしていたんだけど。」、「今日のランチは何？」、「ここは持ち帰り可能ですか？」
- ・レベル3：「予約はしていないのですが、はいれますか？」、「13時に車いすで乗れるタクシーを呼んでほしいのですが。」

表1 ユーザ発話文の解析処理の難易度による分類

レベル	文数 (%)
レベル1 (解析処理の難易度が低い)	42文 (53.8%)
レベル2 (解析処理の難易度がやや高い)	29文 (37.2%)
レベル3 (解析処理の難易度が高い)	7文 (9.0%)
計	78文 (100.0%)

3・3 ユーザ発話文に対する意味表現

3・3・1 各対話シーンにおけるサービス依頼のための設定項目

ユーザ発話文に対する意味表現の検討に先立ち、シニアマンション内のユーザの居室前、レストラン、ロビーの3か所の各エリアの対話シナリオに現れたサービスの種別とサービス提供に必要な設定項目を以下に示す。

(1) ユーザの居室前

- ・移動指示：①目的地（例：レストラン、ロビー、…）

(2) レストラン

(i) レストラン受付

- ・予約の確認：①人数（例：1人、大人2人と子供1人、…）②時間＜任意項＞（例：13時、…）
- ・座席の決定：①テーブルの場所（例：窓際、個室、…）②禁煙／喫煙＜任意項＞（例：禁煙席、…）

(ii) レストランのテーブル

- ・メニューの確認：①メニュー情報（例：メニュー、本日のおすすめ、…）
- ・料理の注文：①注文内容（例：カレーライス、コーヒー、…）②アレルギー＜任意項＞（例：エビ、…）
- ・追加注文：①注文内容（例：コーヒー、お茶、水、…）
- ・その他の要求：①要求物（例：ナイフ、スプーン、箸、…）

(iii) レストランの会計

- ・料金の確認：①料金（例：勘定、会計、…）
- ・支払い処理：①支払い方法（例：現金、クレジットカード、PayPay、…）

(3) ロビー

- ・タクシーの予約：①出発時間 ②行先 ③車いすでの移動の有無＜任意項＞

3・3・2 深層格の種類

本研究では、文の意味を表現するために Fillmore が提唱した格文法⁽⁸⁾⁽⁹⁾を用いている。格文法では、文の意味構造を「動詞 - 深層格 - 名詞」という関係の集合として捉え、これを格構造表現と呼ぶ。ここで、深層格とは、名詞が動詞に対して果たしている意味的な役割のことであり、動作主格、対象格、場所格、目的格などがある¹。日本語では、格助詞が深層格を示すマーカとして機能しており、格助詞を表層格と呼ぶ。たとえば、「太郎が本を買う。」という文の格構造表現は、「買う - 深層格：動作主体（表層格：が） - 太郎」と「買う - 深層格：対象格（表層格：を） - 本」の2つの動詞と名詞間の関係表現により記述される。意味解析処理を行うためには、各動詞に対して、それらの動詞が取り得

¹ Fillmore は、以下の8種類の深層格を提案している。動作主格 (Agent)、経験者格 (Experiencer)、道具格 (Instrument)、対象格 (Object)、源泉格 (Source)、目標格 (Goal)、場所格 (Location)、時間格 (Time)。

る格構造パターンを記した格フレーム辞書を用意しておく必要がある。国立国語研究所における研究では、日本語の深層格として35種を提案している⁽¹⁰⁾。

3・1で説明したシーン別の対話シナリオに現れたユーザ発話文の意味を上記の格構造表現を用いて記述するために、以下の15種類の深層格を設けた。パートナー・モビリティの現在位置の情報に基づいて対話シーン別に作成した対話シナリオを切り替えることにより、使用する格フレームパターンが限定され、一般の文で発生する動詞の多義の問題も大幅に軽減でき、適格かつ柔軟な対話処理を実現できる可能性が高いと予想される。なお、動詞には、通常、時制（現在、過去、未来）やアスペクト（事象の発生、継続、終了など）、法（平叙文、命令文、疑問文）などのモダリティに関する情報が付随するが、これらの情報は、ユーザ発話の内容が各エリアでのサービスの要求が中心である場合は、必ずしも必要でないため、記述しないこととした。また、取り扱うサービス内容を追加する場合は、適宜深層格の増補が必要になる。

＜格フレームで用いる深層格＞

- ・主体（行為を起こす主体となる人やもの）＜ユーザ、あるいは、音声対話システム＞
- ・対象（行為の対象となるもの）＜メニュー、タクシーなど＞
- ・相手（行為の受け手となる人やもの）＜ユーザ、あるいは、音声対話システム＞
- ・場所（行為を行う場所）＜パートナー・モビリティの現在位置、天気予報の予報対象地など＞
- ・場所起点（出発地）＜居室前、ロビーなど＞
- ・場所終点（目的地）＜レストラン、美術館など＞
- ・日付（行為を行う年月日）
- ・日付起点（行為を開始する年月日）
- ・日付終点（行為を終了する年月日）
- ・時間（行為を行う時間）
- ・時間起点（行為を開始する時間）
- ・時間終点（行為を終了する時間）
- ・原因・理由（行為をもたらす原因・理由）
- ・属性（動詞が表す対象を限定する属性）＜例：“料理がうまい”の“料理”など＞
- ・天気状態（晴れ、曇り、雨、雪など） 注：状態、程度などを示す特殊な深層格

3・3・3 ユーザ発話文に対する意味表現の例

ユーザ発話文に対する格フレームを用いた意味表現の例を以下に示す。1つの発話文は、1個以上の格フレームの組として表現される。ここで、格フレームは、LISPのS式を用いて表記し、1つの動詞における複数の動詞と名詞間の関係を1つのS式で示している。なお、S式の空白を‘ ’に置き換えている。

(i) 発話文：“メニューを見せて。”

(見る, ((深層格, 主体), (見出し, ユーザ)),
((深層格, 対象), (見出し, メニュー)))

(ii) 発話文：“13時に車いすで乗れるタクシーを呼んで。”

(呼ぶ, ((深層格, 主体), (見出し, 音声対話システム)),
((深層格, 対象), (見出し, タクシー)),
((深層格, 時間起点), (見出し, 13時)))
(乗る, ((深層格, 主体), (見出し, ユーザ)),
((深層格, 対象), (見出し, タクシー)),
((深層格, 道具), (見出し, 車いす)))

(iii) 発話文：“今日の久留米の天気を教えて。”

(天気である, ((深層格, 日付), (見出し, 今日)),
((深層格, 場所), (見出し, 久留米)))
(%天気, ((深層格, 天気状態), (見出し, #X1)))²

² 疑問文の場合は、問い合わせ対象となっている深層格に「#Xi」の見出しを設定する。また、“～を教えて。”の部分は、格フレームの作成を省略している。

3・4 方言によるユーザ発話に対応した対話処理の流れ

ユーザ発話文に対する汎用の音声認識システムを用いた音声認識処理に続く、方言による発話に対応した対話処理の流れを以下に示す。なお、パートナー・モビリティの現在位置の情報を用いて、対話シーン別に用意した対話シナリオおよび格フレーム辞書が自動的に切り替わるものとする。

(i) 方言によるユーザ発話文の音声認識誤り訂正

方言による発話文の音声認識結果に対して、音声認識誤り訂正表を用いて音声認識誤りの訂正を行う。

(ii) 方言によるユーザ発話文からのキーワードの抽出

音声認識誤り訂正済みの方言による発話文から名詞、動詞などのキーワードを抽出する。

(iii) キーワードの方言から標準語への変換

キーワードに対して、方言から標準語への変換表を用いて、方言から標準語への変換を行う。

(iv) ユーザ発話文に対する意味表現の生成

標準語に変換された発話文中の名詞、動詞などのキーワードと、選択された発話シーンに対応する格フレーム辞書に登録された各エントリーとのマッチング処理を行い、その結果を基に、ユーザ発話文に対する意味表現を生成する⁽⁷⁾。

(v) 応答文の音声出力

ユーザ発話文に対する意味表現に対応して予め用意しておいた応答文を選択し、その音声を入力する。

(vi) エージェントへのサービス依頼

ユーザ発話文に対する意味表現を基に、ユーザが要求したサービスを実行するエージェントにサービス依頼を行う。各エリアでのエージェント機能を実現するためには、レストランでの料理の注文の依頼など、エージェントと外界とのメッセージ交換についても検討を行う必要がある。また、過去のユーザの行動履歴を蓄積することで、個々のユーザの意向に沿ったより適切なサポートを行うことが可能となる。なお、ドコモの先読みエンジンは、このような考えに基づく能動的ユーザサポート機能を提供している⁽¹¹⁾。図3にエージェント機能を含む音声対話システムのイメージを示す。将来的には、ユーザの訴えや感情を読み取ることができる「寄り添い型」の対話エージェントシステムへの発展が必要になると考えられる⁽¹²⁾。

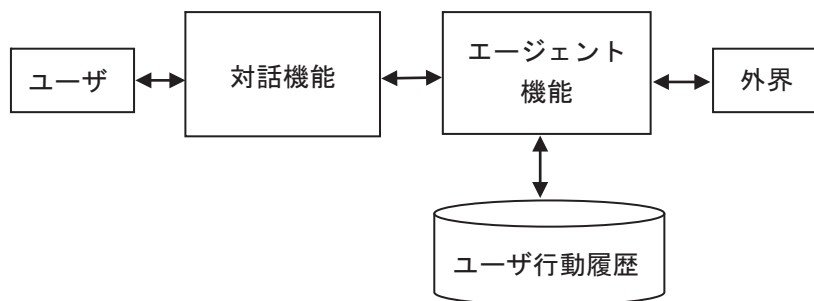


図3 エージェント機能を含む音声対話システムのイメージ

4. 標準語と筑後方言による音声認識精度の比較実験

4・1 筑後方言の概要

本研究では、大学所在地の久留米市近辺で用いられている筑後方言を研究対象としている。そこで、標準語と方言との音声認識精度の比較実験の説明を行う前に、筑後方言の概要を以下に記す⁽¹³⁾⁽¹⁴⁾。筑後方言は、筑前方言（博多弁、福岡弁など）や佐賀弁、熊本弁などとともに、肥筑方言に位置付けられる。下位方言として、久留米弁、柳川弁、大牟田弁などがある。

(1) 発音の特徴

筑後方言の発音の特徴の一部を記す。なお、筑後方言は、無アクセント方言に区分される。

- ・連母音が融合する。（例：「大事」→「でーじ」）
- ・ナ行・マ行音節が撥音化する。（例：「鬼」→「おん」）
- ・促音化がおこる。（例：「するけん」→「すっけん」）

(2) 文法

筑後方言の文法の特徴の一部を以下に記す.

- ・一段活用が, ラ行五段活用化する. (例:「見らん」(否定))
- ・形容詞の終止形・連体形は, 語尾が「か」になる. (例:「しょんなか」=しょうがない)
- ・動詞の否定は, 未然形に「ん」を付けて表す. (例:「帰らんじゃった」=帰らなかった)
- ・様態の助動詞「ようだ」に, 「ごとある」, 「ごたる」を用いる.
- ・対格「を」に, 「ば」を用いる.
- ・順接の接続助詞「から」に「けん」を, 逆接の接続助詞「けれども」に「ばってん」, 「ばって」を用いる.

(3) 語彙

筑後方言の語彙の一部を以下に記す.

(i) 名詞

「いん」=犬, 「すったく」=手抜き, 「たんなか」=田んぼ, 「よめご」=嫁

(ii) 動詞

「うてあう」=相手にする, 「かたる」=参加する, 「がまだす」=精を出す, 「からう」=背負う, 「せれ」=しろ(命令), 「つんのていく」=ついて行く, 「のうなる」=無くなる, 「ひく」=敷く, 「ほがす」=穴をあける

(iii) 形容詞

「えーらしか」=可愛らしい, 「こすか」=ずるい, 「しかとんなか」=くだらない, 「しょんなか」=しょうがない, 「ばさらか」=たくさん, 「ぬっか」=暖かい, 「ふとか」=大きい, 背が高い

(iv) 副詞など

「いっちゃん」=まったく, 「かば」=とても, 「ごつ」=みたいな

(v) 助詞・助動詞など

「～げな」=～だそうだ(伝聞), 「～けん」=～から(原因・理由), 「～さん」「～さね」=～(場所)へ, 「～しきらん」=～できない, 「～しこ」=～だけ, 「～しとと?」=～しているの?, 「～しよらす」=～している, 「～せやん」=～をしなければならない, 「～たい」=～ね, ～よ(例:やるたい=あげるよ), 「～ちゃなか?」=～ではないだろうか?(推定), 「～ちゃる」=～てあげる(例:しとっちゃる=してあげる), 「～でん」=～でも, 「～ばい」=～よ(寒かばい=寒いよ), 「～やけん」=～だから

4・2 実験方法

(1) 音声認識用対話例文

対話シナリオ内のユーザ発話を音声収録の対象文とした. 音声収録は, 標準語と筑後方言(久留米市およびその近辺を対象地域とする)で行うこととした. なお, 当初は, 対話シナリオの標準語の文例に加え, 筑後方言の文例を作成して被験者が方言で発話を行う際の参考にしてもらうことにしていたが, 筑後方言といっても被験者の育った地域や性別, 年齢等により発話内容がかなり異なることが分かり, 方言による発話文例を予め用意するのではなく, 方言による発話は標準語の文例を参考にして自由に行ってもらう方法に改めた.

(2) 被験者

音声収録の被験者として, 久留米市およびその近辺に永らく居住する本学の事務職員5人を選んだ³.

- ・被験者1 性別:男性, 年齢:40歳台, 出身地:久留米市
- ・被験者2 性別:男性, 年齢:40歳台, 出身地:鳥栖市
- ・被験者3 性別:女性, 年齢:50歳台, 出身地:小郡市
- ・被験者4 性別:女性, 年齢:50歳台, 出身地:田主丸町
- ・被験者5 性別:女性, 年齢:40歳台, 出身地:大川市

(3) 音声収録方法

Sony製のPCMレコーダを使用し, 被験者の音声をwave形式のファイルとして保存した.

(4) 音声認識方法

NTTドコモAIエージェントの音声認識エンジンを用いて音声認識を行った. 1つの発話文に対する音声認識候補は複数個出力されるが, 確信度がもっとも高い候補を音声認識結果として採用した.

³ 当初は, 高齢者施設の居住者を被験者として予定していたが, プライバシーの問題等で実現困難となったため, 本学職員を被験者として実験を行った.

(5) 音声認識精度の評価方法

標準語と方言とによる音声認識精度の違いを文字レベルと単語レベル（キーワードとなる名詞および動詞）の2種で調べた。なお、方言による発話については、各被験者の発話を書き起こし、その結果を基に音声認識精度を調べた。

4・3 標準語と筑後方言との音声認識精度の比較実験の結果

(1) 標準語による発話の音声認識の結果

標準語による発話の音声認識の結果を表2に記す。表2において、標準語による発話の音声認識精度は文字レベルで95%、単語レベルで93%となっており、ある程度の誤りが発生するものの実用水準にあると思われる。

表2 標準語による発話の音声認識の結果

被験者	文の文字数 A	正解した 文字数B	正解した文字数 (B/A)*100(%)	文中の名詞・ 動詞の数C	正解した名詞・ 動詞の数D	正解した名詞・動詞の数 (D/C)*100(%)
1	750	708	94%	149	136	91%
2	751	702	93%	154	141	91%
3	750	745	99%	156	154	98%
4	750	740	98%	157	153	97%
5	818	754	92%	153	133	86%
合計	3819	3649	95%	769	717	93%

(2) 筑後方言による発話の音声認識の結果

表3に各被験者の筑後方言による発話内容の一例を示す。表3に示したように、被験者により発話内容に違いがあることがわかる。被験者には、普段話しているように発話を行うようお願いしたが、極端に方言の影響が強い被験者はいなかった。発話内容は、標準語が基本となっており、動詞に続く文末表現などに方言が混じる傾向があるようである。

表3 各被験者の筑後方言による発話内容の例

標準語	被験者1	被験者2	被験者3	被験者4	被験者5
もう1回言って.	もう1回言ってくれん?	もう1回言って.	もう1回言って.	もう1回言ってもらってよか?	もう1回言ってくれん.
予約はしていないのですが入れる?	予約はしとらんばってん入れるね?	予約はしとらんですけど入れますか?	予約はしてないけど入れる?	予約はしてないんですけど入れますか?	予約はしてないけど入れる?

筑後方言による発話の音声認識の結果を表4に記す。表4において、方言による発話の音声認識精度は文字レベルで92%、単語レベルで91%となっており、標準語による場合と比較し、それぞれ3ポイント、2ポイント低いことわかった。特に、単語レベルでは、10%弱の音声認識誤りが発生しており、格フレーム辞書に登録された格フレームパターンとのマッチング処理の精度に影響を与えるため、改善が必要と考えられる。なお、今回の実験では、被験者が高齢ではなく、方言を多用しない傾向があったため、標準語による発話との比較で、音声認識精度にあまり大きな差が生じなかったが、方言の影響がより強いと思われる高齢者が発話した場合には、さらに大きな差が生じる可能性がある。

筑後方言による発話の音声認識誤りの例を以下に示す。左側に書き起こし文、右側に音声認識結果を記す。音声認識の誤り部分は、太字で記している。“～ばい.”が“～場合.”に、“パンば～”が“馬場～”に認識される例などが目立つ。

- ・“レストランに行きたいっちゃけど.” → “レストランに行きたいって言うけど.”
- ・“レストランに着いたばい.” → “レストランに着いた場合.”
- ・“予約はしとらんばってん入れるね?” → “予約はしとらんばってん晴れるよね.”
- ・“パンばお願いします.” → “馬場お願いします.”

表4 筑後方言による発話の音声認識の結果

被験者	文の文字数 A	正解した 文字数B	正解した文字数 (B/A)*100(%)	文中の名詞・ 動詞の数C	正解した名詞・ 動詞の数D	正解した名詞・動詞の数 (D/C)*100(%)
1	812	737	90%	149	137	92%
2	769	681	86%	154	135	88%
3	757	715	94%	156	143	91%
4	851	802	94%	158	146	92%
5	809	749	92%	155	145	93%
合計	3998	3684	92%	772	706	91%

・“水ば持ってきて.” → “水場持って来て.”

・“小麦アレルギーがあるばってん、この料理には使われとると?” → “小麦アレルギーがあるばってん、この料理にお疲れとると.”

・“現金で支払うばい.” → “現金で支払う場合.”

5. おわりに

5・1 まとめ

本研究では、パートナー・モビリティに搭載されている音声対話システムにおける方言への対応方法について検討を行った。具体的には、汎用の音声認識システムを使用し、音声認識処理の後処理として、方言によるユーザ発話文の音声認識誤りの訂正処理を加え、さらに、音声認識誤り訂正後のユーザ発話文から名詞や動詞などのキーワードを抽出し、それらのキーワードと予め準備した動詞の格フレームパターンとのマッチングを行い、その結果を基に、ユーザ発話文に対する意味表現を生成することによって、ユーザの意図を把握し、対話を進める方法を提案した。シニアマンションでのパートナー・モビリティの利用を想定し、音声対話システムの検討を行ったところ、パートナー・モビリティの現在位置の情報を用いてシニアマンション内の各エリアに対応したシーン別対話シナリオを切り替えることにより、ユーザ発話文の意味を表現するために使用する格フレームパターンが限定され、一般の文で発生する動詞の多義の問題も大幅に軽減でき、的確かつ柔軟な対話処理を実現できる可能性が高いと予想されることがわかった。また、パートナー・モビリティで採用された汎用の音声認識システムを用いた標準語による発話と方言（筑後方言）による発話とでの音声認識精度の比較実験を行ったところ、方言による発話の音声認識精度は文字レベルで92%、単語レベルで91%となり、標準語による場合と比較し、それぞれ3ポイント、2ポイント低いことがわかった。特に、単語レベルでは、10%弱の音声認識誤りが発生しており、格フレーム辞書に登録された格フレームパターンとのマッチング処理の精度に影響を与えるため、改善が必要と考えられる。なお、今回の実験では、被験者が高齢ではなく、方言を多用しない傾向があったため、標準語による発話との比較で、音声認識精度にあまり大きな差が生じなかったが、方言の影響がより強いと思われる高齢者が発話した場合には、さらに大きな差が生じる可能性がある。

5・2 今後の課題

(1) 方言に対応した音声対話処理の検証実験

パートナー・モビリティの音声対話システムに、今回提案した方言への対応処理を組み込み、検証実験を行う。

(2) 各地域での方言への対応方法のマニュアル化

標準語でのシーン別の対話シナリオおよび格フレーム辞書を基本とし、本音声対話システムが筑後方言以外の他の地域の方言にも対応できるようにするため、音声認識誤り訂正表、キーワードの方言から標準語への変換表等の作成手順のマニュアル化が可能かどうかについて検討を行う。

謝 辞

本研究は、平成30年度文部科学省私立大学研究ブランディング事業（事業名：先進モビリティ技術で多様な人々が能力を発揮できる、Society 5.0に基づく「いきいき地域づくり」）の支援を受けており、謝意を表します。また、方言に

よる音声認識の実験を行うために、音声収録に協力していただいた本学事務職員みなさまに謝意を表します。

文 献

- (1) 東大輔, 田中基大, 服部雄紀, 金子寛典, リチャードリー, “AI 搭載対話型自動運転パートナーモビリティを用いた新たな福祉サービスデザイン”, 久留米工業大学インテリジェント・モビリティ研究所研究報告第2号 (2019), pp. 15-20.
- (2) 河原達也, “音声対話システムの進化と淘汰—歴史と最近の技術動向—”, 人工知能学会誌, Vol. 28, No. 1 (2013), pp. 45-51.
- (3) 木部暢子, 佐藤久美子, 中西太郎, 中澤光平, “『日本語諸方言コーパス』の構築について”, 言語資源活用ワークショップ発表論文集, 巻1 (2017), pp. 57-68.
- (4) 弘前大学, “「弘大×AI×津軽弁プロジェクト」の開始について”, <https://www.hirosaki-u.ac.jp/45240.html>, 2019.
- (5) 阿部香央莉, 松林優一郎, 岡崎直観, 乾健太郎, “ニューラルネットを用いた多方言の翻訳と類型分析”, 言語処理学会第24回年次大会発表論文集 (2018), pp. 304-307.
- (6) ドコモ AI エージェント <https://docs.sebastien.ai/>
- (7) 高橋雅仁, “音声対話システムにおけるユーザ発話文の解析処理”, 久留米工業大学研究報告 vol. 40 (2019), pp. 75-84
- (8) 長尾真編, “岩波講座ソフトウェア科学15 自然言語処理”, 1996年, 岩波書店.
- (9) チャールズ・J. フィルモア著, 田中春美訳, “格文法の原理—言語の意味と構造”, 1975年, 三省堂.
- (10) 国立国語研究所編, “日本語における表層格と深層格の対応関係”, 1997年, 国立国語研究所.
- (11) 山田直治他, “ユーザの行動を能動的にサポートする先読みエンジン”, NTT DOCOMO テクニカル・ジャーナル, Vol. 26 No 3, pp.20-26, 2018.
- (12) 総務省, “高度対話エージェント技術の研究開発・実証補足説明資料”, https://www.soumu.go.jp/main_content/000627820.pdf, 2019.
- (13) ウィキペディア, “筑後方言”, https://ja.wikipedia.org/wiki/%E7%AD%91%E5%BE%8C%E6%96%B9%E8%A8%80#cite_note-FOOTNOTE%E5%B2%A1%E9%87%8E198378-79-25, 2020.
- (14) weblio 辞書, “筑後弁の概要”, https://www.weblio.jp/wkpja/content/%E7%AD%91%E5%BE%8C%E5%BC%81_%E7%AD%91%E5%BE%8C%E5%BC%81%E3%81%AE%E6%A6%82%E8%A6%81, 2020.