

## 3つの異なる種類の音声認識器を利用した照応解析

嶋田 和孝<sup>†</sup> 棚町 範子<sup>†</sup> 遠藤 勉<sup>†</sup>

<sup>†</sup>九州工業大学 情報工学部 知能情報工学科  
〒 820-8502 福岡県飯塚市川津 680-4  
E-mail: †{shimada,endo}@pluto.ai.kyutech.ac.jp

あらまし 本論文では、複数の異なる認識器を組み合わせた音声理解手法に基づく照応解析について述べる。提案手法は(1)1つの大語彙認識器と(2)複数の小語彙なドメイン依存認識器(3)中規模なドメイン依存認識器の3種類で構成される。提案手法では、介護支援ロボットをタスクとし、小語彙認識器で、ロボットへの命令発話についてロバストな認識を実現し、一方で、大語彙認識器により、命令発話以外の雑談中に存在する単語やトピックを獲得する。ユーザからの命令発話中に、照応詞が存在する場合は、それ以前のそれぞれの音声認識器の結果から適切な先行詞を推定する。先行詞は、大語彙認識器が扱う発話中に存在することが多いが、大語彙認識器の認識精度は十分ではない。そこで、提案手法中の小語彙認識器の語彙を組み合わせた中規模な認識器を用意し、大語彙認識器の認識精度の問題を解消する。実験により、中規模認識器を含む手法の有効性が確認された。

キーワード 照応解析, 複合音声認識器, 統合

## Combination of 3 types of speech recognizers for anaphora resolution

Kazutaka SHIMADA<sup>†</sup>, Noriko TANAMACHI<sup>†</sup>, and Tsutomu ENDO<sup>†</sup>

<sup>†</sup> Department of Artificial Intelligence, Kyushu Institute of Technology  
680-4 Kawazu Fukuoka, 820-8502 Japan  
E-mail: †{shimada,endo}@pluto.ai.kyutech.ac.jp

**Abstract** In this paper, we propose a combination method of 3 different types of speech recognizers. The method consists of a large vocabulary continuous speech recognizer (LVCSR) and 2 types of domain-specific speech recognizers (DSSR). By integrating them, we can realize a simple and flexible speech understanding method with an anaphora resolution process. The DSSRs consists of some DSSRs with a small vocabulary for particular topics in a target task and a DSSR with a medium-scale vocabulary for the task. The task of this system is speech understanding for a livelihood support robot. First we describe the basic idea of our multiple speech recognizer. The selection process in the recognizer is based on the similarity between the LVCSR and each DSSR. Then we explain an anaphora resolution method in the recognizer. By using the outputs from the LVCSR and the medium-scale DSSR, we resolve anaphoric expressions in the current outputs from small-scale DSSRs. The experimental results show the effectiveness of the proposed method.

**Key words** Anaphora resolution, Multiple speech recognizer, Combination

### 1. はじめに

近年の音声認識技術の向上により、音声入力を用いた実用的な対話システムの実現を目指した研究が進められている。しかしながら、実用的な音声対話システムの構築には、音声認識誤りへの対応や音声認識そのものの精度向上が不可欠な状況である。

音声認識の精度向上のための一つのアプローチは、キーワー

ドやキーフレーズの抽出に基づく発話理解である[1]~[3]。もう一つのアプローチとしては、対象となるタスクやドメインにかなり限定した文法や言語モデルを利用することである。しかしながら、このようなアプローチは音声認識・理解の精度向上には繋がっても、対象外の発話や予期しない形の発話には対応できず、十分とはいえない。

音声認識の精度を上げるための効果的な手法の一つは、複数の音声認識器を統合的に、もしくは選択的に用いることである。

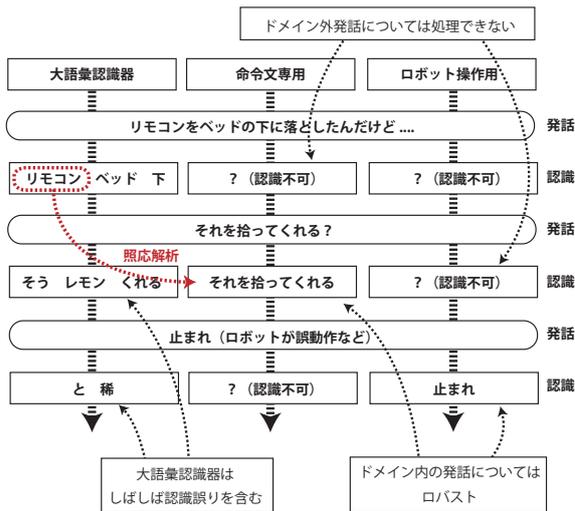


図 1 提案システムのイメージ。

磯部ら [4] は、話題に依存した複数の音声認識器を利用し、その結果を選択的に用いることで、複数の話題に対応可能な音声認識システムについて報告している。この手法により、単語正解率は若干向上している。さらに、別の話題を追加する場合、その追加する話題に関するのみ音声認識器を構築するだけで良いため、システムの拡張が容易という利点もある。しかし、この手法は、それぞれのタスク依存の認識器が基本的に排他的な状態であり、相互の情報を統合的には扱えない。さらに、タスクに依存しないような自由発話部分についても柔軟に処理できないという問題点がある。一方で、統合的に扱う手法では、複数の音声認識器を併用し、その一致部分を検出することで、信頼度の高い単語の特定や、音声認識誤り箇所を推定している [5] ~ [7]。これらの研究は単語レベルでは精度が向上しているが、発話単位では処理をしていない。また、汎用的な認識器を用いた場合、ドメインに依存する言語モデルを利用した場合と比べて、認識精度が低くなる傾向もある。

本論文では、複数の音声認識器を併用することで、特定のタスクやドメインにロバストでかつ自由発話にも柔軟に対応可能な音声理解手法について提案する。我々は現在、介護施設や病院などで暮らしている施設入居者の日常生活や施設スタッフの作業を支援する施設内生活支援ロボットの構築を進めている。本研究はその入力部の 1 つである音声認識部にあたる。提案手法は、大きな分類では、タスク依存の認識器と一般的な言語モデルに基づく認識器から構成される。タスク依存の認識器は、主にユーザからの指示を理解する役割を持ち、正確な認識が可能なが望まれる。一方で、タスク依存の認識器では十分に認識できないような発話に対しては、大語彙認識器を用いて、認識を行う。それぞれの認識結果を選択的にもしくは統合的に扱うことができれば、柔軟でかつロバストな認識器を実現できる。図 1 に提案システムのイメージを示す。

提案手法を用いて照応解析を実現する場合、大語彙認識器の認識精度が大きな問題となる。これは一般に、先行詞となる語は、大語彙認識器が扱う範囲の発話（非命令発話：雑談）に存在することが多いためである。しかし、大語彙認識の精度は十

分ではないことが多い。そこで、本論文では、非命令発話のための認識器として、ドメイン依存認識器を組み合わせた新たな認識器（以降、中規模認識器と呼ぶ）を導入する。中規模認識器の目的は、雑談などに含まれる先行詞となる可能性の高い単語群（すなわち命令発話の扱う単語群）の認識精度を高めることである。

以降、2. 節では、提案する複合認識器の概要について述べ、3. 節でその複合認識器における照応解析を提案する。4. 節で中規模認識器導入の有効性に評価し、5. 節でまとめる。

## 2. 音声認識

### 2.1 基本的な考え方

複数の特性の異なる認識器を組み合わせて処理を行う場合、どの認識器の結果を最終的に利用するか、という問題が発生する。今回のシステムでは、入力された発話がロボットへの命令発話なのか、それとも雑談やそれ以外の発話なのかを分別する必要がある。これは、大語彙認識器とタスク依存認識器のどちらの認識結果を採用するか、という問題になる。例えば、図 1 で「それを拾ってくれる?」という入力について、どの認識器の結果を採用するかは大きな問題である。これは発話検証というタスクの範疇に入る。

本論文では、それぞれの認識器の出力の類似性に着目する [8]。人間でさえも似たような音を持つ単語は誤認識することがあり [9]、音声認識器にも同様の傾向があることは自明である。タスク依存認識器は、もし入力が想定された発話であれば、高い認識精度を得られるはずである。一方、大語彙認識器は、自身の音響モデル・言語モデルに基づいて、最適と思われる単語列を出力する。この場合、2 つの認識器の結果は、少なくとも音素レベルでは比較的類似していると考えられる。逆に、想定外の発話の場合、大語彙認識器は命令発話を認識したときと同様に、最適と思われる単語列を出力するが、タスク依存認識器は持ち合わせている語彙と文法を基にあまり適切でない単語列を出力すると考えられる。すなわち、この場合は、2 つの認識結果は、音素レベルでも必ずしも一致しない可能性がある。

そこで、提案手法では、2 つの認識結果の編集距離を算出し、それを分別のための特徴とする。編集距離のような一致度は、1. 節でも示したように、複数の認識器の出力の高信頼度部分の推定 [6] や誤認識部分の抽出 [5] などによく使用される指標である。Komatani ら [9] も音声対話システムにおいて、人間が聞き取りにくいと思われる音的に類似した単語ペアの抽出に編集距離を基にした距離値を計算している。

本論文では、それぞれの認識器の出力全体（すなわち発話単位）での編集距離と、各単語ごとの編集距離の 2 つの距離値（ともに音素レベル）を DP マッチングにより算出し、以下のように分別処理を実行する。

- (1) 発話単位の編集距離が閾値未満の場合、編集距離が最小のドメイン依存認識器の結果を採用する。
- (2) 単語単位の編集距離の平均値で比較する。
  - 閾値以上：大語彙認識器の結果を採用する。
  - 閾値未満：最小のドメイン依存認識器の結果を採用する。

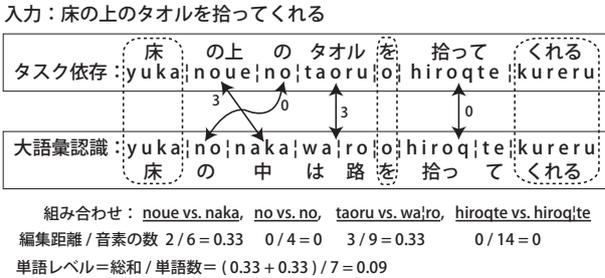
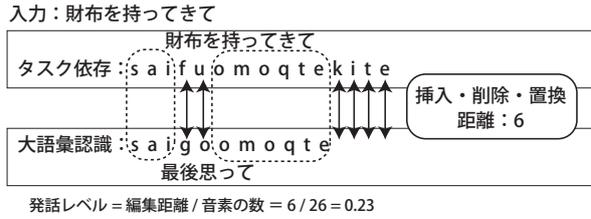


図 2 編集距離の計算例 .

単語単位で編集距離を求める場合、2つの認識結果から編集距離を計算するペアを決定する必要がある。本論文では、まず完全一致する単語を検索し、それを編集距離計算から除外した後、残りの単語群について総当たりで編集距離を求める。その計算の結果、最も編集距離の低い組み合わせの距離値をそれぞれ採用している<sup>(注1)</sup>。具体的な例を図2に示す。図において、破線部は完全一致した箇所を表し、矢印に付随する数値はそのペアの編集距離に相当する。

## 2.2 認識器

提案手法は、大まかな分類では、認識対象を制限しない大語彙認識器と特定の語彙や発話に特化したドメイン依存認識器の2つで構成される。照応解析で対象となる先行詞は、一般に大語彙認識が処理すべき範囲にある。しかしながら、大語彙認識器から必ずしも十分な認識結果が得られるとは限らない。もし、大語彙認識器が認識に失敗し、先行詞がそもそも過去の音声認識結果の履歴に含まれていなければ、照応解析は必ず失敗する。

本論文では、ドメイン依存認識器を拡張することで、この問題に対応する。先行詞となる語は、基本的にそれぞれのドメイン依存認識器の語彙（本論文では、介護支援ロボットが行動する範囲内にある名詞群など）に含まれる。そこで、すべてのドメイン依存認識器に含まれる語を扱う、中規模なドメイン依存認識器を新たに導入する。中規模認識器の導入により、雑談や非命令発話中に存在する、先行詞となりやすい語の認識精度が向上する。

本論文では、大語彙認識器として Julius を、中規模および小語彙ドメイン依存認識器として、その派生形である Julian を利用する。Julius は数万語の語彙を対象とした音声認識ソフトウェアである [10]。Julian は Julius を基にした記述文法による音声認識器である。言語モデルとして有限状態文法 (DFA) を用い、ユーザがBNF風の記述で、認識用の構文規則を作成可

(注1): ただし、組み合わせを求める際には隣接する3つの単語までという制約を設けている。

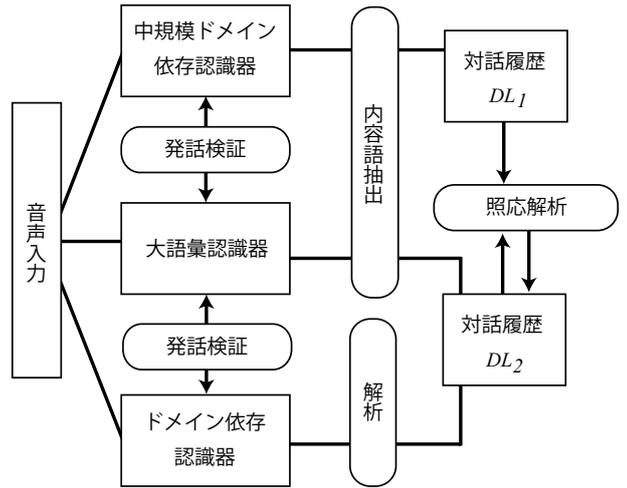
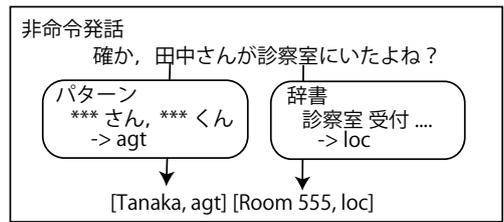


図 3 照応解析の処理の流れ .



(a)



(b)

図 4 音声認識結果の解析 .

能である。語彙辞書は、文法ファイル中で使用される単語の意味的なカテゴリとそれに属する単語で構成される。

## 3. 照応解析

本節では、2.節で述べた複合音声認識器における照応解析処理について述べる。提案手法の流れを図3に示す。提案手法では2種類の対話履歴を利用する。図中の  $DL_1$  は、中規模ドメイン依存認識器から得られた内容語リストを保持する。発話検証において、中規模認識器と大語彙認識器の出力が類似していれば、その認識結果の内容語が  $DL_1$  に格納される。同様に、小語彙なドメイン依存認識器と大語彙認識器についても発話検証を行い、もし、編集距離が閾値以上であれば(すなわち雑談と判定)、大語彙認識器の出力結果から内容語を抽出し、 $DL_2$  に格納する。発話検証の結果、ドメイン依存認識器の出力が採用されれば(すなわち命令発話と判定)、その内容を解析する(解析結果は  $DL_2$  に格納される)。

まず、命令発話の解析と対話履歴  $DL_1$  および  $DL_2$  への処理について述べる。命令発話は格構造へと変換される。ここで、我々は、ドメイン依存認識器の持つ、構文規則と意味カテゴリを利用する。図4の(a)に解析例を示す。ドメイン依存認識器

の文法と語彙辞書には、その音声認識精度を上げるために、意味的な制約が設けられている。例えば、この図では「～を飲みたい」のヲ格には飲み物 (Drink\_N) に関する単語のみを受理すること、「～の { 上 or 中 or 下 } にある」では場所情報を表す単語が受理されることとなる。これらの情報を基に命令発話の解析を行う。また、事前に照応解析の対象となる格を設定することで、ゼロ代名詞化されている要素の特定も行う。

大語彙および中規模認識器の結果については、いくつかの表層的なパターンやルール、辞書を利用して、照応解析に必要な内容語を抽出する。その例を図4の(b)に示す。図の例では、表層パターンとして「～さん」を利用している。辞書的なルールとしては、単語「診察室」は場所情報に属することを利用している。この辞書的なルールは、前述の音声認識に利用した文法における意味カテゴリを利用する。

次に、解析・抽出された結果をもとに照応解析を行う。本論文での照応解析は、対話履歴  $DL_1$  および  $DL_2$  中の単語へのスコアリングに基づき実行される。

まずはじめに、単語への重み付与について説明する。 $DL_1$  および  $DL_2$  における、ある単語  $k_i$  の重みは以下のように計算される。ここで、 $i$  は単語  $k$  の  $DL_1$  および  $DL_2$  中での出現位置を表す。

$DL_1$  の単語に対しては、以下のように重みを設定する。

$$Conf_{k_i}^1 = CN_{k_i}^{DL_1} \quad (1)$$

ここで  $CN_{k_i}$  は単語  $k_i$  の信頼度である。この信頼度は認識器の出力する信頼度に基づくものであり、 $[0,1]$  の範囲の値を取る。

$DL_2$  については (1) 大語彙認識器の認識結果 (2) 小規模ドメイン依存認識器の認識結果および (3) 照応解析によって得られた結果の3種類の出力が格納されている (1) および (2) の場合については、 $DL_1$  と同様に、単語の信頼度をそのまま用いる。

$$Conf_{k_i}^2 = CN_{k_i}^{DL_2} \quad (2)$$

一方で、 $k_i$  が照応解析によって得られた単語である場合は、以下のように一定値を設定する。

$$Conf_{k_i}^2 = 0.7 \quad (3)$$

これは、照応解析の正解率が十分でないことを考慮し、設定されている。

次に、各々の  $k_i$  について、スコアを計算する。ここで、我々は、距離と状況変化による、次のような減衰係数をスコアリングに適用する。

$$DF_{k_i} = \frac{1}{dist_k^2} \times sc^n \quad (4)$$

ここで、 $dist_k$  は、照応詞を含む文と単語  $k$  を含む文との距離を表す。 $sc$  は状況変化に関するパラメータである。ここで、「状況変化」とは、「ロボットの所在地が変わること」および「ロボットとの話者が変わること」を表している。 $n$  は状況が変化した数であり、状況に変化がない場合を0とする。 $sc$  は0.1とした。この減衰係数  $DF_{k_i}$  を  $Conf_{k_i}^1$  と  $Conf_{k_i}^2$  に掛け合わせる。

$$dConf_{k_i}^1 = Conf_{k_i}^1 \times DF_{k_i} \quad (5)$$

$$dConf_{k_i}^2 = Conf_{k_i}^2 \times DF_{k_i} \quad (6)$$

最終的な、ある単語  $k_i$  に関するスコアは、

$$Score_{k_i} = \alpha \times dConf_{k_i}^1 + \beta \times dConf_{k_i}^2 \quad (7)$$

で計算される。ここで、 $\alpha$  と  $\beta$  はそれぞれの  $dConf$  への重みである。このスコアを  $N$  文前までに含まれるすべての単語に計算し、その中で最も高い値を持つ単語  $k$  を先行詞とする。本論文では、 $N = 10$ 、 $\alpha = 0.5$ 、 $\beta = 1.5$  とした。ここで、 $\alpha$  の値が  $\beta$  の値と比較して小さいのは、中規模認識器の結果が挿入誤り(注2)を含む傾向があるためである。これは、中規模認識器はタスクに関連する語しか辞書中に含んでいないため、雑談のタスク外の語を無理に認識しようとした結果、辞書中に含まれる関係のない語を認識結果に含むことがあるためである。

#### 4. 実験・考察

本節では、提案手法の評価実験について述べ、考察したのち、関連研究について言及する。

##### 4.1 実験

###### 4.1.1 実験設定

実験では、前述のように、音声認識器として Julius/Julian を用いた。Julius については、音響モデル、言語モデルとも同ソフトウェアに添付されているオリジナルのモデルを利用した。タスク依存認識器としては、以下のような4種類の認識器に関する文法・語彙辞書を用意した。

- 患者からの命令発話  
例) 机の上のリモコンを取ってくれる? など
- 看護師からの命令発話  
例) この食事を501号室に運んで、など
- ロボット制御命令  
例) 止まれ、50cm 右に移動、など
- 質問発話  
例) 田中さんはどこにいる? など

各々の認識器は200単語弱の語彙とそれに基づく100パターンほどの文法で構成されている。

照応解析のための中規模認識器も Julian で実現された。語彙は、前述の4つのドメイン依存認識器のもつ語彙を統合したものとした。文法については、その単語の任意の長さの組み合わせとした。具体的には「名詞 助詞 名詞 助詞 動詞」のようなものである。このようなナイーブな文法を用いたのは、中規模認識器の目的が、照応解析の際の先行詞の欠落を防ぐ、いわばキーワード認識のようなものであり、先行詞が認識されれば、文としての正しさは必要がないためである。

###### 4.1.2 実験結果

発話検証については、各ドメイン(患者命令、看護師命令など)に20発話、雑談として20発話(「おはよう」や「今日は天気が良いね」など)の合計100発話を用意し、被験者を10

(注2): 本来は含むべきではない単語を誤って含んでいる間違い。

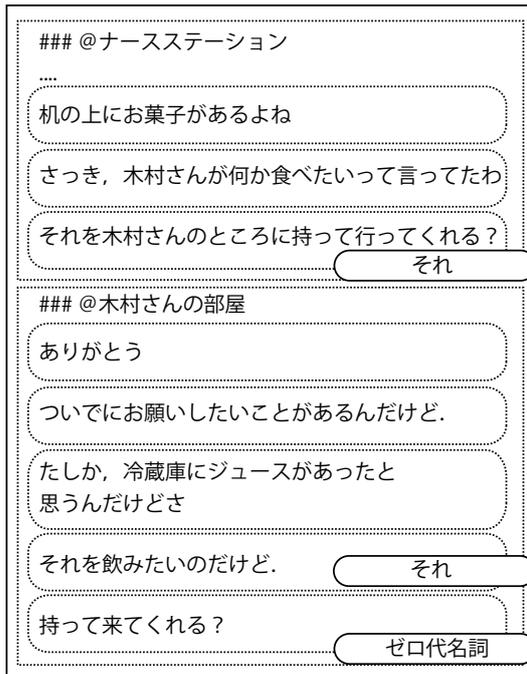


図 5 対話スクリプト.

表 1 照応解析の精度.

手法	ベースライン	先行研究	提案手法
精度	64.2%	68.9%	71.7%

名として実験した。実験の結果、発話単位の閾値を 0.20、単語単位の閾値を 0.08 とした場合の分別精度が最も高く、F 値で 0.924 となった。閾値については、発話単位の閾値を 0.20-0.26 の範囲で変更し、F 値の変化を調査したが、閾値を変化させることで F 値に大きな揺れは生じなかった。この結果は、提案手法のロバストさを示している。

次に、照応解析の実験結果について述べる。実験には 206 発話からなる仮想対話を用意し、2 名の被験者で実験した。206 発話中には 53 個の照応詞が存在する。図 5 にその対話の一部を示す。図中の「###」は状況変化（場所および話者変化）を表す。

実験結果を表 1 に示す。表中の「ベースライン」とは、本論文でのスコアリング手法において、中規模認識器を利用しない場合を指す。すなわち、 $DL_1$  は用いず、 $DL_2$  のみで実行した結果である。「先行研究」とは、[11] で使用された手法であり、提案手法が候補中で最大のスコアを持つ単語を選ぶのに対し、この手法は、対象となる  $N$  発話中における各単語の累積値に基づき先行詞を特定する手法である。提案手法と公平に比較するために、先行研究に基づく手法にも、提案手法と同様に中規模認識器を適用している。

中規模認識器を利用した提案手法は、中規模認識器を利用しないベースラインおよび先行研究よりも高い精度を得た。中規模認識器を適用することによって、音声認識結果中の先行詞の欠落が減少し、その結果、精度が 64.2% から 71.7% まで向上した。この結果より、中規模認識器を導入することの有効性が確

められた。提案手法と先行研究を比較すると、精度が 3% 程度向上している。これは、先行研究が、履歴中の累積値を用いていることに起因している。中規模認識器を導入することで、音声認識における削除誤りは低減するが、今回の手法では、挿入誤りについては十分に対応できない。これは、累積値を用いた場合、挿入誤りによって対話履歴中に存在する多くのノイズに過剰反応してしまい、精度が低下したものと考えられる。

3 種類の認識器を用いることで、照応解析の精度は向上したものの、その精度は 70% 程度にとどまっており、十分とはいえない。今回の実験は実際の音声認識結果を用いて行ったが、もし、実際の音声認識結果ではなく、書き起こしのデータ（オラクルデータ）を用いた場合、照応解析の精度は 95% を越えた。これは、いかに音声認識の精度を向上させるかが、照応解析の精度を向上させるのに重要であることを表している。一つの大きな問題は、中規模認識器の言語モデルである。今回の実験では、簡易的に、シンプルな単語の組み合わせに基づく決定的な文法で中規模認識器を実装している。より高精度な音声認識器を実現するために、ある程度トピックを限定した統計的な言語モデルに基づく中規模認識器を適用する必要があると考えられる。

さらに、提案手法では、「話者の変化」と「場所の変化」については扱ったが、それだけでは十分とはいえない。例えば、現在の発話のトピックを推定することで、より詳細な状況的な変化に対応する手法について考察する必要がある。

#### 4.2 関連研究

複数の認識結果から最適なものを選ぶことは、発話検証と呼ばれる研究に属する。発話検証については、数多くの研究がなされている。佐古ら [12] は、AdaBoost を用いて、入力をシステムへの問い合わせかと雑談に分別する手法を提案している。磯部ら [4] は、モデルの尤度に基づくマルチドメインへ対応した音声認識手法を提案している。これらの手法では、新たな認識器を加えたり、タスクを変更する場合に、学習器作成のための十分な訓練データやモデルの再構築が必要になる。提案手法において分別に必要なのは、編集距離に対する閾値のみである。一方で、Komatani ら [13] は、音声認識器が出力する音響尤度に着目した発話検証の枠組みについて提案している。発話検証に音響尤度の差を利用することは、理論的に適しており、この手法を我々の手法を統合することは、有効であると考えられる。

照応解析については、我々は信頼度を利用したスコアリングに基づく手法を提案している。テキストを対象とした照応解析の研究では、様々な機械学習に基づく手法が提案されており、その有効性が示されている [14], [15]。しかし、発話検証の場合と同様に、機械学習には十分な学習データが必要であるという問題点がある。照応解析に関する研究では、センタリング理論 [16] やそれを拡張した Salience Reference List (SRL) [17] などがある。峯脇ら [18] は関連性理論に基づく照応解析について提案している。これらの言語学的知見を組み込むことは、興味深い今後の課題である。

提案手法では、先行詞に関する音声認識誤りの問題を解消するために、中規模認識器を導入した。この手法は、ROVER 法 [19] の範疇にある。さらなる音声認識誤りの低減のために、

異なる認識や言語モデルを適用する必要がある。

また、そもそも音声認識誤りを含むことを前提に、照応解析を行うことも興味深い課題である。例えば、音声文書検索においては、音節同士のマッチングを取ることで、未知語（すなわち音声認識の結果には単語として含まれないもの）の検索を行う手法が提案されている [20]。あるクエリにおける音声文書検索の問題を、照応解析における先行詞の探索だと考えることで、音声認識結果に正しい先行詞が単語として含まれていない場合でも、音節的に類似箇所を検出することが可能になり、その結果、照応解析の精度が向上すると考えられる。

本研究が対象としているのは、ロボットとの対話である。そのため、音声認識結果に十分な信頼が持てない場合は、対話により、音声認識の誤りによる問題を低減するということが可能である。また、音声認識結果を可視化し、ユーザが修正できるような枠組み [21] を導入することも、実用的なシステムを構築するためには必要になってくるであろう。

## 5. おわりに

本論文では、3つの異なる音声認識を組み合わせた音声理解手法における照応解析について述べた。複数の性質の異なる認識器を選択的もしくは統合的に扱うことで、柔軟でかつロバストな音声理解が可能となった。

認識器の選択（発話検証）においては、出力の類似性に着目した。その分別処理の F 値は 0.92 と高い値であった。照応解析については、先行詞抽出のために、中規模認識器を導入した。中規模認識器の導入により、音声認識の先行詞に対する削除誤りが減り、その結果、照応解析の精度が向上した。先行研究と比較して、提案手法の方が高精度であることも確認された。

しかしながら、照応解析の精度は 70% 強であり、十分に高いとはいえない。一方で、先行詞に関する音声認識精度が十分であれば、さらに精度が向上することが確認された。先行詞に関する音声認識精度向上のために、中規模認識器に、より適切な言語モデルを適用することや、さらに別の特性の異なる認識器を加えることが今後の課題となる。

謝辞 本研究は、次世代ロボット知能化技術開発プロジェクト（独立行政法人新エネルギー・産業技術総合開発機構）における「施設内生活支援ロボット知能の研究開発」の成果の一部である。

## 文 献

- [1] K. Komatani and T. Kawahara: “Flexible mixed-initiative dialogue management using concept-level confidence measures of speech recognizer output”, Proceedings of COLING 2000, Vol. 1, pp. 467–473 (2000).
- [2] 宮崎, 中野, 相川: “逐次発話理解法による対話音声理解”, 電子情報通信学会論文誌 D-II, **J87-D-II**, 2, pp. 456–463 (2004).
- [3] K. Shimada, T. Endo and S. Minewaki: “Speech understanding based on keyword extraction and relations between words”, Computational Intelligence, **23**, 1, pp. 45–60 (2007).
- [4] 磯部, 伊藤, 武田: “複数の認識器を選択的に用いる音声認識システムのためのスコア補正法”, 電子情報通信学会論文誌 D, **J90-D**, 7, pp. 1773–1780 (2007).
- [5] 伊藤, 西崎, 関口, 中川: “音声文書インデキシングのための web 文書を利用した自動誤り訂正”, 第 3 回情報科学技術フォーラム講演論文集, 第 2 巻, pp. 343–344 (2004).
- [6] 宇津呂, 西崎, 小玉, 中川: “複数の大語彙連続音声認識モデルの出力の共通部分を用いた高信頼度部分の推定”, 電子情報通信学会論文誌, **7**, pp. 974–987 (2003).
- [7] 山口, 酒向, 山本, 菊井: “信頼度尺度に基づく音声認識誤りの検出および誤り訂正”, 電子情報通信学会技術研究報告 音声 SP2003-65, pp. 7–12 (2003).
- [8] K. Shimada, S. Horiguchi and T. Endo: “An effective speech understanding method with a multiple speech recognizer based on output selection using edit distance”, Proceedings of the 22nd Pacific Asia Conference on Language, Information and Computation (PACLIC22), pp. 350–357 (2008).
- [9] K. Komatani, R. Hamabe, T. Ogata and H. G. Okuno: “Generating confirmation to distinguish phonologically confusing word pairs in spoken dialogue systems”, Proceedings of 4th IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems, pp. 40–45 (2005).
- [10] 河原, 李: “連続音声認識ソフトウェア julius”, 人工知能学会誌, **20**, 1, pp. 41–49 (2005).
- [11] K. Shimada, A. Uzumaki, M. Kitajima and T. Endo: “Speech understanding in a multiple recognizer with an anaphora resolution process”, Proceedings of the 11th Conference of the Pacific Association for Computational Linguistics (PACLING2009), pp. 262–267 (2009).
- [12] 佐古, 滝口, 有木: “Adaboost を用いたシステムへの問い合わせと雑談の判別”, 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, pp. 19–24 (2006).
- [13] K. Komatani, Y. Fukubayashi, T. Ogata and H. G. Okuno: “Introducing utterance verification in spoken dialogue system to improve dynamic help generation for novice users”, Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue, pp. 202–205 (2007).
- [14] R. Iida, K. Inui and Y. Matsumoto: “The issue of combining anaphoricity determination and antecedent identification in anaphora resolution”, International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE), pp. 244–249 (2005).
- [15] V. Ng and C. Cardie: “Improving machine learning approaches to coreference resolution”, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 104–111 (2002).
- [16] M. Kameyama: “A property-sharing constraint in centering”, Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics, pp. 200–206 (1986).
- [17] S. Nariyama: “Grammar for ellipsis resolution in Japanese”, In Proceedings of the 9th International conference on Theoretical and Methodological Issues in Machine Translation, pp. 135–145 (2002).
- [18] S. Minewaki, K. Shimada and T. Endo: “Interpretation of utterances based on relevance theory: Toward the formalization of implicature with the maximum relevance”, Proceedings of the 9th Conference of the Pacific Association for Computational Linguistics (PACLING2005), pp. 211–216 (2005).
- [19] J. G. Fiscus: “A post-processing system to yield reduced word error rates: Recognizer output voting error reduction (rover)”, Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), pp. 347–352 (1997).
- [20] 西崎, 中川: “音声認識誤りと未知語に頑健な音声文書検索手法”, 電子情報通信学会論文誌 D-II, **J86-D2**, 10, pp. 1369–1381 (2003).
- [21] J. Ogata and M. Goto: “Speech repair: Quick error correction just by using selection operation for speech input interfaces”, Proceedings of Interspeech 2005, pp. 133–136 (2005).