

音声対話システムにおける 音韻的類似表現の混同を防ぐための確認の生成

浜辺 良二[†]

駒谷 和範[‡]

尾形 哲也[‡]

奥乃 博[‡]

[†] 京都大学 工学部 情報学科

[‡] 京都大学大学院 情報学研究科 知能情報学専攻

1. はじめに

音声による対話システムでは、音声の聞き誤りを防ぐことが重要である。しかしシステムの語彙の中に、音韻的に類似した識別の難しい語（類似表現）が含まれる場合には、これらはシステムの音声認識誤りの原因となる。さらには人間がシステムの音声合成を聞き間違える可能性も高める。

本研究ではこのような類似表現に対して、識別の容易な別の単語を付加した確認表現を生成する。この際の処理はタスクメインに依存していないため、様々なシステムにおいてこのような確認を自動的に生成し、対話中で用いることができる。

本手法では、複数の知識源を用いて確認表現を生成する。これは単一の知識源から必ずしも適切な確認表現が得られるとは限らないためである。我々は複数の候補に対して適用可能な確認の良さを測る尺度を定義することにより、複数の知識源より得られた候補から最適な確認表現を自動的に選択可能とする。

2. 確認の生成手順

確認の生成は次の手順で行う（図1）。

1. 単語間の音韻的距離を定義することで音韻的に類似した単語の組をシステムの語彙リストから抽出する。
2. 得られた単語の組に対し一般に用いられている辞書を利用して、付加する音韻的に遠い単語の候補を取得する。
3. 得られた候補の中から最適なものを選択して確認を生成する。

以降は京都市バス運行情報案内システム [1] に本手法を適用した場合を例にとり説明する。

本研究で想定する対話例を図2に示す。類似表現間で音声認識結果が曖昧だった場合、まず信頼度の高い方の

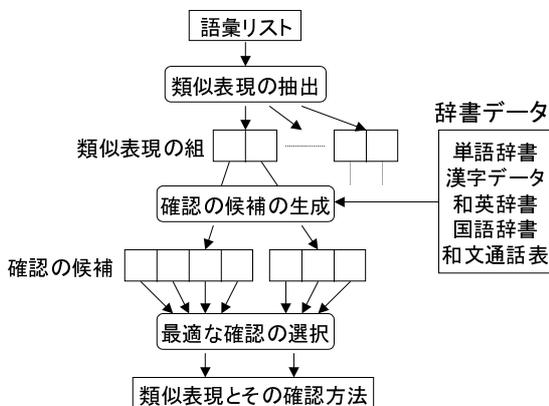


図1: 確認の生成の流れ

Generating confirmation to prevent confusion between phonologically similar expressions for a spoken dialogue systems: Ryoji Hamabe, Kazumori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)

S: 乗車駅, 降車駅, 系統番号をおっしゃってください
 U: 金閣寺道から
 S: 金閣寺道か銀閣寺道が分からないのですが, 先頭に英語でシルバーの銀の字が付きますか?
 U: いいえ
 S: では, 先頭に英語でゴールドの金の字が付きますか?
 U: はい

図2: 確認を用いた対話例

表1: 得られた類似表現の例

京都府立盲学校	京都府立聾学校
地下鉄・四(し)条駅	地下鉄・二(に)条駅
金閣寺道	銀閣寺道
洛西(らくさい)高校前	洛水(らくすい)高校前

単語をユーザが発話したかどうかを、生成した確認表現を用いてシステムが尋ねる。ユーザの返答が No であった場合は、組の他方の単語について確認を行う。

2.1 類似表現の自動抽出

単語同士が音韻的に類似している（すなわち音声による識別が難しい）のは、共通する音素が多く、また逆に異なる音素が少なく、さらにモーラの数が近い場合である。よって、単語 w, w' 間の音韻的距離 $p.d.(w, w')$ を次式で定義する。

$$p.d.(w, w') = \frac{e.d.(w, w')}{\log(\min(|w|, |w'|) + 1)}$$

ただし、 $|w|, |w'|$ はそれぞれの単語の音素数を表す。 $e.d.(w, w')$ は w, w' の音素列の編集距離 (edit distance) であるが、編集の際にモーラの数が変化する場合のコストを1、変化しない場合のコストを2として、DP マッチングにより求める [2]。

この値が閾値より小さいものを類似表現として抽出する。バスシステムでは、閾値を 0.6 としたところ、単語辞書 1574 語の中から 29 組の類似表現が得られた。表1にその一部を示す。

2.2 確認の候補の生成

上記の手法で得られた類似表現の識別を容易にする単語（付加表現）を取得し、確認の候補を生成する。タスクに依存せず扱える情報として、類似表現間の文字の違いおよび音の違いに着目し確認を行う。

そのためにあらかじめ、類似表現間の非共通文字および非共通音を抜き出しておく。「洛西高校前」と「洛水高校前」の例では、非共通文字は「西」と「水」、非共通音は「サ」と「ス」である。この違いを含む単語を以下に述べる4つの方法で取得する。それぞれの方法で得られた確認を表2に示す。

単語辞書から熟語を取得する

非共通文字を含む、日常で頻繁に用いられる熟語を単語辞書から取得し付加表現とする。単語辞書として Juman の JC1[3] を用いる。この辞書は単語の難易度に応じて5つのファイルに分かれており、難易

表 2: 「洛西高校前」と「洛水高校前」を識別するための確認の候補

手法	「洛西高校前」の確認	「洛水高校前」の確認
熟語	洛の後に西洋の西という字が付きますか？	洛の後に水曜日の水という字が付きますか？
訓読	洛の後に西(にし)という字が付きますか？	洛の後に水(みず)という字が付きますか？
英訳	洛の後に英語でイーストの西の字が付きますか？	洛の後に英語でウォーターの水の字が付きますか？
音	ラクの後の音は「さくら」のサですか？	ラクの後の音は「すずめ」のスですか？

表 3: ホテル検索タスクの語彙内の類似表現と生成された確認の例

類似表現 1	類似表現 2	類似表現 1 の確認	類似表現 2 の確認
有田温泉 柏原(かいばら)町 美浜(みはま)町	有馬温泉 米原(まいばら)町 海山(みやま)町	アリの後の音は「煙草」のタですか？ 最初の音は「為替」の力ですか？ 先頭に美術館の美の字が付きますか？	アリの後の音は「マッチ」のマですか？ 先頭に米(こめ)という字が付きますか？ 先頭に英語でシーの海の字が付きますか？

度の低いものから順に 2337 語, 2618 語, 4892 語, 8989 語, 16239 語が含まれている。

非共通文字を含む熟語は複数存在するが, もっとも難易度が低くさらにその中で発音の一意性が最大のものを取得する。発音の一意性は, 辞書に含まれる単語の中で音韻的距離が小さいもの上位 10 個の平均値とする。この値が大きいくほど, その熟語を聞いたときに他の熟語と混同せずに漢字を想像することが容易である。非共通文字が複数ある場合は, 全ての文字の中で難易度最小, 一意性最大となる熟語が得られる文字に着目する。

漢字の訓読みを用いる

全ての漢字に対して, 音読み, 訓読みが与えられている既存のテキストデータを用いて, 非共通文字の訓読みを調べる。それが前述の単語辞書に含まれていれば, その訓読みを付加表現として取得する。

漢字の英訳を用いる

和英辞典のテキストデータから漢字一文字から成る見出し語を取り出し, その英訳を得る。さらに, 国語辞典のテキストデータ(カタカナ英語として一般的に用いられる英単語が見出し語に含まれている)からそのカナ発音を得て, それを付加表現とする。これにより 275 種類の英訳を得た。非共通文字が複数ある場合は熟語を取得する場合と同様である。その際の難易度は辞書中の学習レベルを用いる。

音の違いを強調する

和文通話表を用いて最も先頭にある非共通音を強調する。和文通話表とは日本語の 1 音を表現するために, その音で始まる単語を集めたものである。使われる単語は全て単語辞書における難易度が最小のものである。

(例) 朝日の「あ」, いろはの「い」, ...

特殊な場合として「金閣寺道」と「銀閣寺道」のように, 類似表現間の非共通音が濁点の有無のみの違いである場合は「最初の音は濁点が付かないですか(付きますか)?」といった質問をする。

2.3 最適な確認の選択

得られた確認の候補から最適となるを選択する。本研究では, 最適な確認は以下の 2 条件を満たすものとする。

- 他の単語と聞き分けやすい
- 内容の理解がしやすい

聞き分けやすさを測る尺度として, 付加表現 w_i の発音の一意性 u_i , および元の類似表現 w と付加表現との音韻的距離 $p.d.(w, w_i)$ を用いる。また, 内容の理解しやすさは, 付加表現の難易度によって測る。よって, 確

認の候補それぞれのスコアをこれらの重み付き和で次のように定義する。

$$score_1(i) = W_1 d_i + W_2 u_i + W_3 p.d.(w, w_i)$$

d_i は付加表現の難易度に基づく値で, 次式で与える。

$$d_i = \frac{(w_i \text{より難易度が大きい単語の数})}{(\text{辞書の単語の総数})}$$

この値の重み付き和が最大となるものを最適な付加表現とみなし, 実際の対話中に使用する。重みは, $W_1 = -10, W_2 = 2, W_3 = 1$ とした。

さらに, 2 回目の確認(1 回目の確認でユーザの返答が No だった場合に行う確認)を行う際には, 1 回目の確認に用いた付加表現との音韻的距離 $p.d.(w_i, w_j)$ を確認の良さを測る指標に加える。つまり, 2 回目の確認を選択する際のスコアは次式で与える。重みは $W_4 = 1$ とした。

$$score_2(j) = score_1(j) + W_4 p.d.(w_i, w_j)$$

これらの重みを用いたところ, 表 2 の「洛西高校前」と「洛水高校前」の例では, 訓読みを用いた手法が 1 回目, 2 回目とも最適な確認として選ばれた。

3. 他ドメインへの適用

バスシステムにおいて設定した閾値と重みを用いて, ホテル検索タスクの語彙リストに本手法を適用した。その結果, 単語辞書 865 語の中から 25 組の類似表現が得られ, 最適な確認として表 3 のような確認が選択された。このように本手法は, バスシステムに限らず様々なタスクドメインに対して有効であり, 自動的に適用することが出来る。

4. おわりに

類似表現の混同を防ぐための確認を自動的に生成する手法について述べた。今後, 実際に本手法を適用したシステムをユーザに利用してもらい, 類似表現の混同が削減できるかを評価する予定である。

本研究の一部は, 科研費, 21 世紀 COE プログラムの支援を受けた。また, JC1 の使用を許可して下さい。京都大学の佐藤理史先生に感謝します。

参考文献

- [1] 安達史博, 河原達也, 岡本隆志, 中嶋宏. Voice XML の動的生成に基づく協調的な電話自動応答システム. 日本音響学会研究発表会講演論文集, 1-5-18, 春季 2002.
- [2] G. Navarro. A guided tour to approximate string matching. ACM Computing Surveys, 33(1):31-88, 2001.
- [3] 佐藤理史. 異表記同語認定のための辞書編纂. 情報処理学会研究報告 2004-NL-161, pp. 97-104.