

# 音声対話システムにおける ユーザの固有名詞の簡略化に対処する語彙拡張

勝丸 真樹<sup>†</sup>      駒谷 和範<sup>‡</sup>      尾形 哲也<sup>‡</sup>      奥乃 博<sup>‡</sup>

<sup>†</sup> 京都大学 工学部情報学科      <sup>‡</sup> 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

音声対話システムにおいてシステムに慣れていない初心者には、しばしば正式名称を簡略化した表現で発話する [1]。この問題は、データベースから得た音声認識辞書内の語と、ユーザが想定する表現とのギャップに起因する。このような発話による音声認識誤りを防ぐには、システムの音声認識辞書に簡略表現を追加する必要がある。我々は、データベースから得た音声認識辞書 (既存辞書) の語の任意の部分列を省略することで簡略表現を生成し、辞書へ追加する。この際に以下の課題がある。

1. 簡略表現生成のための固有名詞の分割  
音声対話システムでは固有名詞が重要な役割を果たすため、それらの簡略表現への対処は必須である。簡略表現生成には、既存辞書の単語を分割する必要があるが、形態素解析器による分割では、固有名詞を正しく分割できない。
2. 生成した簡略表現に起因する認識率低下の抑制  
生成した簡略表現を単純に音声認識辞書に加えると、語彙の増加により認識率低下を招く。特に、既存辞書の語と音素列が似ている簡略表現は音声認識上混同されやすい。

本研究では、まず既存辞書内の単語を文字の接続確率を併用して分割し、その組み合わせにより簡略表現を生成する (2 章)。それらの生起確率を、既存辞書の語との類似度に応じて操作する (3 章)。この概要を図 1 に示す。簡略表現を辞書に追加する従来研究 [2] では、固有名詞の分割に対して形態素解析器しか用いておらず、[3] では、語彙の追加による認識率低下を考慮していない。本研究では文字間の接続確率を定義して固有名詞を分割し、さらに音声認識時に混同される度合いを考慮することで、簡略表現の再現率を高めたうえで音声認識率低下を抑制する。

## 2. 固有名詞の分割と簡略表現生成

まず、辞書内の単語  $w$  を要素列  $e_1, e_2 \dots e_n$  に分割する。形態素解析器では分割できない固有名詞に対処するために、形態素解析器 MeCab [4] に加え、辞書内での文字間の接続確率を考慮した。辞書固有の文字間の接続確率による分割は次のように行なう。ある単語  $w$  を文字列  $c_1 c_2 \dots c_{i-1} c_i \dots c_n$  で表す。単語の分割の尺度として、 $c_{i-1}$  と  $c_i$  との接続確率を音声認識辞書内の文字 N-gram 確率を用いて

$$\min\{P(c_i|c_{i-1}c_{i-2}\dots c_1), P(c_{i-1}|c_i c_{i+1}\dots c_n)\} \quad (1)$$

と定義する。つまり  $c_i$  への前からの N-gram 確率と、 $c_{i-1}$  への後からの N-gram 確率との小さい方の値を、文字  $c_i$  と  $c_{i-1}$  との接続確率とした。 $c_{i-1}$  と  $c_i$  との接続確率が閾値  $\theta$  以下の場合、 $w$  を  $c_{i-1}$  と  $c_i$  の間で分割する。形態

Expanding Vocabulary to Cope with User's Abbreviations for proper nouns in Spoken Dialogue Systems: Masaki Katsumaru, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)

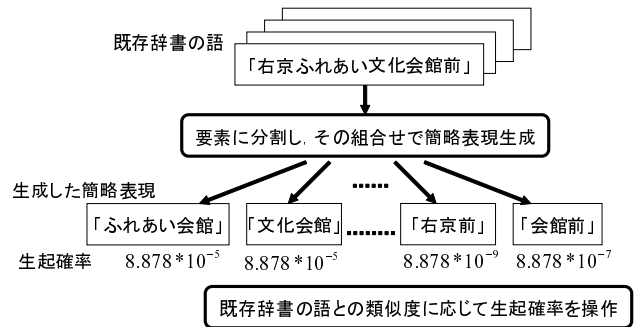


図 1: 簡略表現追加までの流れ

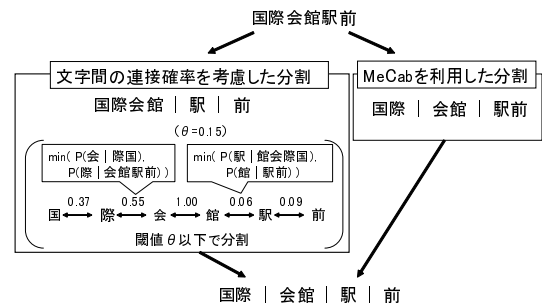


図 2: 「国際会議駅前」の分割例

素解析結果と接続確率のいずれかで区切られていれば、その位置で単語を分割する。本手法により「国際会議駅前」を分割した例を図 2 に示す。MeCab による分割では「駅前」を分割できないが、文字間の接続確率をみることで「駅」と「前」に分割できる。これは、例えば簡略表現として「国際会議駅」を生成する上で必要な分割である。

次に、要素列  $e_1, e_2 \dots e_n$  から任意個の要素の省略によって、 $(2^n - 1)$  個の簡略表現を生成した。生成した簡略表現の読みは、各要素ごとの読みを、MeCab による読み付与と簡略前の読みとのマッチングにより同定し付与した。

## 3. 追加した簡略語による音声認識低下の抑制

既存辞書の語の部分列と音韻的に類似した簡略表現は、音声認識時に悪影響を与えやすい。そこで音韻的距離を定義し、既存辞書の語に類似した簡略表現の生起確率を減少させる。

### 3.1 音韻的距離の定義

生成した簡略表現  $w$  と既存辞書  $D_{org}$  との音韻的距離を以下の式で定義する。

$$dist(w, D_{org}) = \min(e.d.(w, part(D'_{org}))) \quad (2)$$

ただし、 $D'_{org}$  は  $D_{org}$  の辞書の語彙から  $w$  の簡略前の語を除いた語彙、 $part(D'_{org})$  は  $D'_{org}$  内の語の音素部分列である。 $e.d.(x, y)$  は単語  $x, y$  の音素列間の編集距離 (edit distance) であり、DP マッチングにより求める。編集距

表 1: 京都市バス運行情報案内システムの語彙の部分列と音韻的に近い簡略表現の例

生成した簡略表現	既存辞書の語	min(e.d.)
大門(大門町より) [daimoN]	大文字寺 [daimoNjidera]	0
野田(野田町より) [noda]	花園大学 [hanazonodaigaku]	0
京都私学(京都私学会館より) [kyo:toshigaku]	京都役所前 [kyo:toshiyakushomae]	1
東側(東側町より) [higashigawa]	東山仁王門 [higashiyamano:moN]	2

離のコストは、母音・促音・撥音の置換・挿入・削除は2、子音の置換・挿入・削除は1とする。つまり、音韻的距離  $dist(w, D_{org})$  は、 $w$  と  $D'_{org}$  の語内の全ての音素部分列との編集距離の最小値である。生成した簡略表現と既存辞書の語との  $\min(e.d.(w, part(D'_{org})))$  が小さい例を表1に示す。

### 3.2 音韻的距離に基づく生起確率の操作

既存辞書内の語と生成した簡略表現には、 $\sum P_{org} = 1$  になるように一様に生起確率  $P_{org}$  を割り当てる。さらに、 $dist(w, D_{org}) \leq d$  となる生成簡略語  $w$  に対しては

$$P_{new}(w) = P_{org} * f(dist(w, D_{org})) \quad (3)$$

となるよう新たな生起確率を  $P_{new}$  を割り当てる。予備実験より  $d = 4$ ,  $f(dist(w, D_{org})) = 10^{dist-5}$  とした。  $P_{new}$  計算後、生起確率の和が1になるように正規化を行なった。

## 4. 評価実験

### 4.1 評価用データ

京都市バス運行情報案内サービス [5] において、2002年5月から2005年3月に収集した実際のユーザの発話データを用いた。本研究では、システムの持つ語彙の知識がないユーザを対象とするため、使用回数が1回だけの初心者ユーザの発話を集めた。ここで無音やタスクの進行に関係のない発話は除いた。その結果、183名1494発話を得た。そのうち150発話に既存辞書の語を簡略した表現が70種類162単語含まれていた。

### 4.2 生成簡略表現の評価

京都市バス運行情報案内サービスの初期の音声認識辞書(語彙サイズ1668)のうち、固有名詞であるバス停名・施設名(1481単語)に対して簡略表現を自動生成した。予備実験の結果、分割時の閾値を  $\theta = 0.15$  とした。結果、15012語の簡略表現を得た。これによりユーザが発話した70種類の簡略表現のうち、66種類(94.3%)を再現することができた。なお、MeCabのみの単語の分割による簡略表現生成では、60種類(85.7%)である。つまり、MeCabによる分割だけでなく、文字間の接続確率も考慮することで、簡略表現の再現率が8.6ポイント向上した。

### 4.3 単語正解精度の評価

生成した簡略表現の生起確率を操作して音声認識辞書に追加し、単語正解精度の評価をおこなった。音声認識にはJulius [6]を用いた。言語モデルは、京都市バス運行情報案内サービスの音声認識用文法から内容語のクラス化した後、全ての文パターンを生成することで単語N-gramモデルを学習した。音響モデルは、電話用2000状態16混合トライフォンモデルを用いた。以下の各認識辞書を用いて認識実験を行い、単語正解精度を比較した。

既存辞書 初期の音声認識辞書(語彙サイズ1668)。

表 2: 評価用データ1494発話に対する単語正解精度

辞書の種類(語彙サイズ)	単語正解精度%
既存辞書(1668)	44.4
+生成簡略表現(16680)	35.3
+生成簡略表現+確率操作(16680)	47.7
+実際に発話された簡略表現(1738)	49.1

既存辞書+生成簡略表現 生成簡略表現を単純に既存辞書に加えた辞書(語彙サイズ16680)。

既存辞書+生成簡略表現+確率操作 生成簡略表現を既存辞書に加え、生起確率を操作した辞書(本手法)(語彙サイズ16680)。

既存辞書+実際に発話された簡略表現 ユーザが実際に発話した簡略表現(70種類)のみを既存辞書に加えた辞書。本研究における上限値となる(語彙サイズ1738)。

各辞書に対する単語正解精度を表2に示す。音声認識率が全体的に低いのは、対象ユーザに初心者が多く、文法外の発話が多いためである。例えば「京都駅からじゃなくて北白川の駅からなんですけど」(下線部は文法外)という発話では、文法外部分が別の単語に置き換わり誤認識されていた。単純に簡略表現を追加しただけでは、既存辞書の語の発話に対し誤認識が増加し、単語正解精度が9.1ポイント下がった。そこで生成簡略表現の生起確率を操作することで、操作前から12.4ポイント精度が向上し、既存辞書よりも3.3ポイント高い精度を得た。また、実際に発話された簡略表現70種類のみを加えた辞書で認識させた場合との差は1.4ポイントである。これは確率の操作により、語彙の追加による誤認識増加が抑制された結果を示している。

## 5. おわりに

本研究では、音声対話システムにおける固有名詞の簡略表現の認識を目的として、音声認識辞書の語から簡略表現の候補を自動生成した。さらに、語彙サイズ増加による音声認識率の低下を抑えるために、音韻的距離に基づき生起確率を決定し操作した。評価実験では、初心者ユーザの発話の簡略表現の再現率は94.3%であり、元の音声認識辞書と比べ3.3ポイント高い単語正解精度を得た。さらに、生起確率を操作することで、語彙の拡張による認識率低下を抑えることができた。

謝辞 本研究の一部は、科研費、グローバルCOE、SCATの支援を受けた。

## 参考文献

- [1] Geoffrey Zweig et al. The Voice-Rate Dialog System for Consumer Ratings. In *Interspeech-2007*, pp. 2713-2716.
- [2] 櫻井功, 他. 日本語における略語自動生成法の検討とその音声インタフェースへの応用. *情処学研報*, 2007-SLP-69-54.
- [3] E.E. Jan et al. Automatic Construction of Unique Signatures and Confusable Sets for Natural Language Directory Assistance Applications. In *Eurospeech-2003*, pp. 1249-1252.
- [4] 工藤拓. MeCab: Yet Another Part-of-Speech and Morphological Analyzer. <http://mecab.sourceforge.net/>.
- [5] 駒谷和範, 他. 音声対話システムにおける適応的な応答生成を行うためのユーザモデル. *電子情報通信学会論文誌*, Vol. J87-D-II, No. 10, 2004.
- [6] 河原達也, 李晃伸. 連続音声認識ソフトウェア Julius. *人工知能学会誌*, Vol. 20, No. 1, 2005.