

6T-5

# 方言対訳コーパスを用いた日本語方言変換システム

平山 直樹<sup>†</sup>

森 信介<sup>†‡</sup>

奥乃 博<sup>†</sup>

<sup>†</sup> 京都大学 大学院情報学研究科 知能情報学専攻

<sup>‡</sup> 京都大学 学術情報メディアセンター

## 1. はじめに

日常生活において、人々はそれぞれの方言で会話している。異なる地域の人々の会話では、互いの方言がすぐに理解できず、コミュニケーションがたどたどしくなる場合がある。本稿の目的は、話し手の方言から聞き手の方言への変換によるコミュニケーションの円滑化である。単に方言といっても、変化する特徴は様々である。ここでは方言の特徴を以下の3つに分類する。

- 発音 同じ単語の発音の部分的変化
- 語彙 同じ概念を表す別の単語を使用
- 語順 文における単語の順序交替

日本語の場合には発音および語彙の変化が多いので、本稿ではこの2つを対象とする手法を開発する。

方言音声認識の実現には、2点の課題の解決が必要である。以下にその課題と、本稿における解決策を述べる。

- 1) 方言言語資源の不足 方言は元来話し言葉であるため、方言言語資源が不足しており、信頼性の高い言語モデルの構築が困難である。本稿では、大量に存在する共通語言語資源を用いて方言言語資源を模倣することで、この課題を解決する。
- 2) 方言の混合 地域間の人々の移動により、方言は互いに影響し合い、「純粋な」方言は仮定できない。方言の混合による多様性を考えるため、本稿では方言間の確率的変換ルールを導入する。

## 2. 関連研究

方言音声変換では、方言音声認識を行い、その認識結果を別方言に変換するという処理の流れとなる(図2)。方言音声認識および方言間機械翻訳は従来から取り組まれてきた課題ではあるが、その多くは音声認識に用いる単語発音辞書を人手で構築したり[1]、翻訳ルールを人手で列挙したり[2]する必要があった。本稿で述べる手法では、後述する方言対訳コーパスを用いてこれらの作業を自動化する。対訳コーパスを用いる手法の利点は、自動化による省力化、各ルールに対する重みの統計的推定による確率的ルールの構築である。

## 3. 手法

方言音声変換を3つの部分問題: 1) 方言音声認識, 2) 方言テキスト変換, 3) 方言音声合成に分ける(図2)。本稿では、1)および2)、すなわち方言音声別方言テキストとするまでを対象とする。方言音声認識については、言語モデル・音響モデルそれぞれに対する手法を述べる。

### 3.1 方言音声認識: 言語モデル

本稿では、大規模方言言語コーパスの模倣により、統計的に信頼できる方言言語モデルを構築する。コーパスの模倣には、小規模な共通語-方言対訳コーパス[3]から学習した、重み付き有限状態トランスデューサ(WFST)を用いる。学習は、大きく次の3要素から成る(図1)。

An automatic speech recognition system for Japanese dialects using parallel corpora between dialects: Naoki Hirayama, Shinsuke Mori, and Hiroshi G. Okuno (Kyoto Univ.)

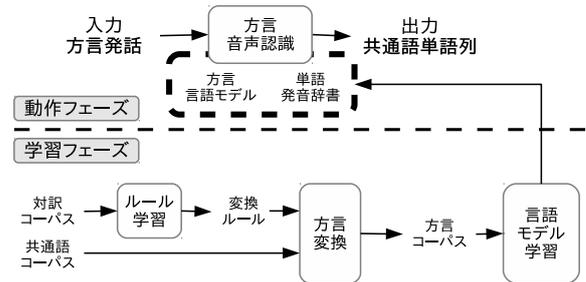


図1: 方言音声認識の流れ。

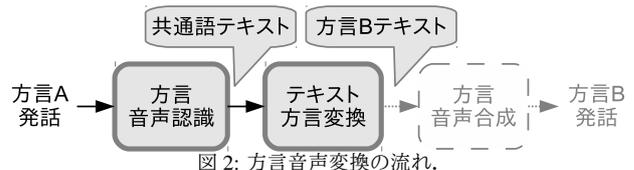


図2: 方言音声変換の流れ。

1. 方言変換ルール学習
2. コーパスの方言テキスト変換
3. 言語モデル学習

以下では、各要素について具体的な手法を述べる。さらなる詳細については[4]を参照されたい。

### 3.1.1 方言変換ルール学習

方言発話の書き起こしと共通語訳が対になった方言対訳コーパスから、各共通語単語に方言発音を自動で求める。処理の流れを以下に示す。

- (a) 共通語と方言を共に発音(音素列)で表現
- (b) 動的計画法(DP)により音素単位のマッチングを推定
- (c) マッチング結果を用いて、共通語単語ごとに対応する方言発音を推定

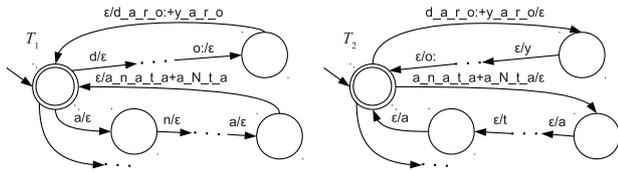
得られた単語と発音の対応を用いて3つのWFST  $T_1, L, T_2$ を構築する。 $T_1$ は、各単語に対応する共通語音素列を共通語と方言の発音組に変換する(図3(a))。Lは、発音対応の文脈依存性を発音組3-gramモデルで表現したもので、マッチングで得られた発音対応の列から学習される。 $T_2$ は、共通語と方言の発音組を方言音素列に変換する(図3(b))。以上により、WFST  $T = T_1 \circ L \circ T_2$ が、発音の変換と同時に、各変換候補の尤度を計算する。

### 3.1.2 コーパスの方言テキスト変換

大規模な共通語言語コーパスの各文を音素列に変換し、前述のWFST  $T$ に入力する。方言の多様性を表現するため、尤度に比例する確率で候補をランダムに選択する。

### 3.1.3 言語モデル学習

本稿では共通語単語をクラスとし、各クラスが複数種類の方言発音を含むクラス  $n$ -gramモデルを構築する。これは、共通語単語と方言発音の組では種類が膨大になりうるためである。発音のクラス内確率は、コーパスの模倣により出現した方言発音の回数を共通語単語ごとに集計し、総和が1になるよう正規化して求める。ここで、



(a) 共通語音素列から音素列対応. (b) 音素列対応から方言音素列.  
図 3: 方言変換 WFST の構造.  $\epsilon$  は入出力記号なしを示す.

表 1: 実験に用いたデータセット. 語数は共通語ベースでカウント.

用途	データセット	規模
言語モデル学習	Yahoo! 知恵袋 <sup>1</sup>	300 万文, 6,030 万語
音響モデル学習	CSJ	500 講演, 70.2 時間
	JNAS	308 話者, 23.3 時間
対訳コーパス	[3]	619 発話, 2.5 万語

クラス内確率を共通語発音のみ (確率 1) の場合と重み付き平均すると, 認識精度が向上すると報告されている [4]. 後の実験では重みを 1:1 とした場合の結果を示す.

### 3.2 方言音声認識: 音響モデル

話し言葉認識に適した音響モデルを学習する場合には, 日本語話し言葉コーパス (CSJ) [5] の利用が効果的であることが報告されている [6]. 本稿では, これに加えて最尤線形回帰 (MLLR) 法を用いた方言発話への音響モデル適応を試みる.

### 3.3 方言テキスト変換

方言テキスト変換を行うには, 共通語で出力された音声認識結果を, 3.1 節で学習した WFST に入力すればよい. WFST での変換は発音ベースであるため, コーパス変換と同様に, 事前に認識結果をすべて発音に変換する.

## 4. 評価実験

本手法の有効性を評価するため, 関西弁を対象とし, 音声認識および方言変換のそれぞれで実験を行った. 共通語話者 (東京都, 埼玉県出身者) および方言話者 (大阪府, 兵庫県出身者) 各 5 名に共通語 100 文 (すべての話者に共通) を提示し, 共通語話者はそのまま, 方言話者は自分の方言に翻訳して読み上げを行った. 実験に用いたデータセットを表 1 にまとめる. 音響モデル学習には, CSJ に加え JNAS [7] に含まれる音素バランス文も用いた. 音響特徴量は MFCC (25 次元) とした.

### 4.1 音声認識

音声認識実験 (Julius<sup>2</sup>による) では, 共通語単語認識精度を評価尺度とし, 以下の 2 つの実験を行った.

最初に, 音響モデルは共通とし, 言語モデル学習に共通語言語コーパスを用いた場合と方言変換したコーパスを用いた場合とで認識精度を比較した. 方言話者, 共通語話者に対する結果をそれぞれ表 2, 3 に示す. 方言話者に対しては, 方言変換後コーパスで学習した言語モデルの方が平均 5.7 ポイント高くなり, 本手法が方言話者により適応した言語モデルを構築することが示された.

続いて, 音響モデルを方言発話に適応させた場合の認識精度の変化を調べた. 対象話者はすべて方言話者とし, 言語モデルは話者の方言に合わせたものを用いた. 適応に用いた音声は, 対象話者以外の 4 名の方言発話 (計 400 文) である. 結果を表 4 に示す. 適応させた場合には, 適応させない場合と比較して平均 5.4 ポイントの認識精度向上がみられた.

<sup>1</sup>国立情報学研究所 (NII) が提供.

<sup>2</sup><http://julius.sourceforge.jp/>

表 2: 方言話者に対する単語音声認識精度 [%]. LM は言語モデル (language model) の略.

話者	#1	#2	#3	#4	#5	平均
共通語 LM	52.5	50.8	60.4	53.3	53.0	54.0
方言 LM	<b>60.3</b>	<b>56.7</b>	<b>65.9</b>	<b>59.0</b>	<b>56.8</b>	<b>59.7</b>

表 3: 共通語話者に対する単語音声認識精度 [%].

話者	#1	#2	#3	#4	#5	平均
共通語 LM	84.5	<b>78.2</b>	<b>83.8</b>	79.3	<b>81.6</b>	<b>81.5</b>
方言 LM	<b>85.5</b>	77.7	81.7	<b>79.9</b>	79.2	80.8

表 4: 音響モデルの適応による単語音声認識精度 [%] の変化.

適応	#1	#2	#3	#4	#5	平均
なし	60.3	56.7	65.9	59.0	56.8	59.7
あり	<b>63.8</b>	<b>61.6</b>	<b>73.4</b>	<b>64.6</b>	<b>62.1</b>	<b>65.1</b>

表 5: 共通語 100 文の方言変換に対する BLEU.

	変換前	変換後
BLEU	0.819	<b>0.834</b>

## 4.2 方言テキスト変換

方言テキスト変換については, BLEU [8] を評価尺度とした. BLEU は, 正解文に現れる  $n$ -gram を多く含む翻訳が優れているという考え方に基づく. ここでは話者に提示した共通語 100 文を方言変換し, 5 話者の方言翻訳文を正解文としてカナ列表記で比較した. その際, 句読点および表記ゆれの起こりやすい長音・促音を除去し, カナ 4-gram までで BLEU を計算した. 結果を表 5 に示す. 方言変換により, 方言話者が翻訳した文により近い発音列が得られたといえる.

## 5. おわりに

本稿では, 方言音声変換システムの実現のために, 方言音声認識および方言テキスト変換の手法を検討し, 実験によりその有効性を確認した. しかし, 音声認識に関していえば, 音響モデルの適応を行った場合でも, 共通語の場合と比較して単語認識精度で 17.8 ポイントの差がある. 方言特有の誤りで多かったパターンは, 方言特有の表現が無理やり共通語単語の組合せで認識される<sup>3</sup>ものであった. これは, その表現が対訳コーパスに現れなかったり, 言語モデル制約が弱かったためと考えられる. この点については, 今後詳しく検討する必要がある.

## 謝辞

本研究の一部は, 科研費 (S) (No. 24220006) の援助を受けた.

## 参考文献

- [1] Lyu, D. et al.: Speech recognition on code-switching among the Chinese Dialects, *Proc. of ICASSP 2006*, Vol. 1, pp. 1105–1108 (2006).
- [2] Zhang, X.: Dialect MT: a case study between Cantonese and Mandarin, *Proc. of ACL and COLING 1998*, Vol. 2, pp. 1460–1464 (1998).
- [3] 国立国語研究所 (編): 全国方言談話データベース 日本のふるさとことは集成 (全 20 巻), 国書刊行会 (2001–2008).
- [4] Hirayama, N. et al.: Statistical Method of Building Dialect Language Models for ASR Systems, *Proc. of COLING 2012*, pp. 1179–1194 (2012).
- [5] Maekawa, K.: Corpus of Spontaneous Japanese: Its design and evaluation, *Proc. of ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition* (2003).
- [6] 篠崎隆宏ほか: 話し言葉コーパスを用いた音声認識の検討, 日本音響学会 2001 年春季研究発表会講演論文集, pp. 31–32 (2001).
- [7] Itou, K. et al.: JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research, *Acoustical Society of Japan (English Edition)*, Vol. 20, pp. 199–206 (1999).
- [8] Papineni, K. et al.: BLEU: a method for automatic evaluation of machine translation, *Proc. of ACL 2002*, Association for Computational Linguistics, pp. 311–318 (2002).

<sup>3</sup>「ようさん (たくさん)」が「予算」と認識されるなど.