

発話中の方言変化に頑健な方言変換システム

平山 直樹[†]

高橋 徹[‡]

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

本稿では、異なる地域の人々のコミュニケーションを円滑にする方言変換手法を開発する。方言とは、1言語の中で地理的要因により異なる特徴を持つ言葉を指すが、特定の人が話す方言は複数の方言種別に渡って変動するので、複数の方言種別を扱える方言変換が要請される。

任意の方言間の変換を実現するため、変換の中間表現に共通語を用いる。なぜなら、従来の方言研究は全地域の基準の言葉である共通語 [1] と比較されることが多く、方言間での比較研究が少ないため、ある方言種別から別の方言種別への変換ルールを見出すのが困難だからである。方言から共通語、共通語から方言への変換器として、対訳コーパスから重みを自動学習した WFST (Weighted Finite-State Transducer) を用いる。方言種別の数だけ方言・共通語間の音素変換器を学習すればよく、文献 [2, 3] など方言・共通語対訳コーパスを利用可能である。

方言の特徴には、アクセントやイントネーション、音素の違いといった音韻的側面、語彙や文法といった言語的側面がある。我々は、言語的側面の違い、つまり語彙の違いを吸収することがコミュニケーションを円滑にするために最も重要と考え、方言語彙の変換について述べる。

2. 方言変換システム

2.1 問題設定とシステムの構成

問題設定を以下に示す。

入力 方言発話音声, 変換先方言種別

出力 発話内容の変換先方言音素列

仮定 入力発話は日本語, 方言種別による語順不変, 音素集合は共通語と同一

前提 入力発話の方言種別は未知, 発話中の方言種別の変化を考慮

まず発話音声の方言種別を既知とし、方言発話音声から共通語単語列への変換を考え、次に発話音声の方言種別が未知かつ発話中に方言種別が変化する状況に対応する。

システムの全体構成を図1に示す。本問題を、音素変換器の統計的学習(2.2節)と、統計的言語モデルを用いた Julius [4] による音声認識器(2.3節)で解決する。音素変換器の統計的学習を行うのは、人手によるルール付与 [5] では方言種別ごとに方言の知見からのルール記述が必要であるが、変換器の自動学習 [6] ならば複数の方言種別に対しても機械的に学習が可能だからである。音響モデルと単語 n -gram (Julius では前向き 2-gram・逆向き 3-gram) は、共通語と同一の方法で学習する。以後、言語モデルとは単語 n -gram と単語辞書の組とする。

2.2 WFST を用いた音素変換器の構成

音素変換器は、方言音素列を共通語音素列に、あるいは共通語音素列を方言音素列に変換するものである。

音素変換ルールの生成には、方言と共通語の音素列対訳コーパスを用いる。対応する音素列の各組に対して動

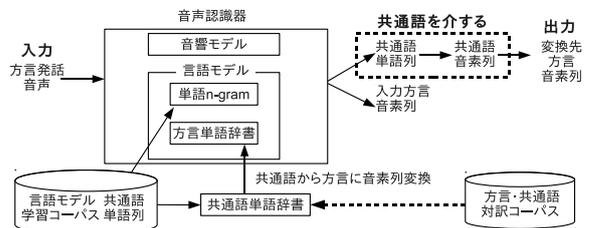


図1: システムの全体構成

方言	koNnichiwa okusaN itsumo	[k+k] [o+o] [N+N] ... [m+m] [o+o]
共通語	koNnichiwa okusaN itsumo	[d_e+o_l_a] [m+m] [a+a] [sh+sh] [i+i] [l+l]
	de mashitekaran i erai	[e+e] [k_a_r_a_n_i+NULL] [e_r_a_i+d_o_m_o]
	ojamashite do:mo	[sh+sh] [i+i] [ts+ts] [u+u] [r+r] [e+e] ...
	shitsure: itashimasu...	
	shitsure: itashimasu...	

図2: 左が音素マッチング結果, 右が音素変換ルールの例。音素変換ルールの各音素対は [方言音素列+共通語音素列] の形式。

的計画法による音素マッチングを行い、音素が一致する箇所は1音素ずつ、異なる箇所は対応する音素列を組にし、各組を1つの記号(以後、音素対と呼ぶ)に変換した記号列により、音素変換ルールを表現する(図2)。

この音素変換ルールから、共通語音素列の入力に対し、尤度付きで方言音素列候補を出力する WFST $T = T_1 \circ L \circ T_2$ を学習する。ここで、 \circ は (W)FST の合成演算である。 T_1 は共通語音素列を音素対列に変換する FST, T_2 は音素対列を方言音素列に変換する FST であり、共通語あるいは方言音素列の入力に対して音素対を出力し、その0回以上の繰り返しを受理する。音素マッチング結果に現れる音素対のみを対象にすればよいので、状態数は有限である。 L は音素対列の 3-gram モデルを表現する WFST である。 L において、3-gram 確率 $P(w_i|w_{i-1}, w_{i-2})$ は、状態 (w_{i-1}, w_{i-2}) から (w_i, w_{i-1}) への遷移に重み $\ln P(w_i|w_{i-1}, w_{i-2})$ を与えることで表現される。3-gram モデルにより、文脈に依存した音素変化特徴を学習できる。共通語と方言の立場を入れ替えれば、方言から共通語への音素変換 WFST も作成できる。 T_1, T_2 の生成に OpenFst [7] を、 L の生成に Kylm [8] を用いた。

2.3 方言音声認識・変換

入力発話の方言種別が既知のとき、対応する音素変換器から方言単語辞書を作成し、方言音声認識を行う。また、共通語発話音声から方言音素列への変換は、通常共通語単語辞書による音声認識結果の単語列を音素列に変換し、それを T に入力して方言音素列を得る。

次に、発話音声の方言種別が未知かつ発話中に変化する場合は、方言種別を N 個定め、発話音声の方言種別はそれらの組み合わせと仮定する。各方言種別の単語辞書を混合して、複数方言種別に対応する方言単語辞書を作成する。単語 w に対する方言音素列 p の生起確率を $P_i(p|w)$ ($i = 1, 2, \dots, N$) とし、新しい方言単語辞書における、単語 w に対する方言音素列 p の生起確率 $P(p|w)$ を $P(p|w) = \sum_{i=1}^N \alpha_i P_i(p|w)$ で与える。 α_i は重みであり $\sum_{i=1}^N \alpha_i = 1$ を満たすよう定める。重みは各方言種別の入力確率で決定するが、入力確率が不明の場合は $\alpha_i = 1/N$ とする。これにより、方言種別が未知かつ発話中に変化

A Robust Dialect Transformation System against Dialect Variation in Each Utterance: Naoki Hirayama, Toru Takahashi, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)

表 1: 音素変換器学習データ

方言	文章数	単語数
京都	1,190	16,334
岐阜	905	16,527

表 2: 方言発話音声・音素列データ

方言	文章数	単語数
京都	111	1,316
岐阜	60	639

表 3: 認識精度 [%]

単語辞書	京都方言		岐阜方言		単語辞書	京都方言		岐阜方言	
	(1)	(2)	(1)	(2)		(A)	(B)	(A)	(B)
共通語	51.7	35.3	52.8	31.3	共通語	2.98	2.71	3.14	2.65
方言	52.5	32.9	52.5	28.7	方言	2.84	2.74	3.09	2.82
混合	52.8	34.8	53.3	30.9	混合	2.79	2.62	3.02	2.60

表 4: 平均エントロピー [bit/音素]

する方言発話音声の認識が実現できる。

3. 評価実験

3.1 実験条件

学習に用いるデータを以下に示す。

音素変換器: 文献 [2] の対訳コーパスから学習する。

言語モデル: 文献 [3] の共通語訳文章のうち、入力方言の地域を除いた 47 地域から学習する。全 48 地域で、合計 11,366 文、277,722 語。語彙数は 5,000。

評価に用いるデータを以下に示す。

方言発話音声: 文献 [3] の方言文章の一部について、方言話者の読み上げ発話を録音して用いる。

音素列: 文献 [3] の方言音素列を基にして、2 種類のデータで評価する: (A) 方言音素列, (B) 共通語と方言の混合音素列 (文献で与えられた分かち書き単位 (全 N 個) で、前半 $\lfloor N/2 \rfloor$ 個の相当部分を共通語音素列、残りを方言音素列としたもの)

データの規模を表 1, 2 に示す。音素変換器と言語モデルの学習には、共通語訳文章に対する KyTea [9] の読み推定結果に基づく音素列、単語列を用いる。方言単語辞書の読みは各単語で尤度順に最大 5 個とし、1 位に対する尤度比が $10^{-1.5}$ 以上のエントリを登録する。以下、音声認識による評価・方言変化頑健性による評価の 2 種類の実験を行う。

3.2 実験 1: 認識精度による評価

以下の評価基準を用いる。

- (1) **音素列認識精度:** 認識した音素列部分の音素単位の認識精度を求める。正解音素列長、置換誤り数、挿入誤り数、削除誤り数をそれぞれ N, S, I, D とすると、認識精度は $Acc = (N - S - I - D) / N$ で表される。
- (2) **音声認識精度:** 方言音声認識結果の共通語単語列と、発話内容の共通語訳の正解文章を比較し、文字単位の認識精度を求める。計算方法は音素に準ずる。

3 種類の単語辞書: (a) 共通語, (b) 方言, (c) 共通語と方言の混合、に対して評価する。(b), (c) において、方言は入力方言種別と同一とする。(c) において、混合時の重みはともに 1/2 とする。

実験結果を表 3 に示す。(1) 音素列認識は、すべての単語辞書で 50% 程度の精度であるのに対し、(2) 音声認識の精度は 30% 程度と低い。音素列認識では、多くの単語を登録すれば、単語の組み合わせで得られる音素列の種類が増え、精度向上が期待できる。一方、音声認識では文脈により同じ音素列が異なる単語 (同音異義語) と認識されるなど、より言語モデルの影響を受けやすい。このことが両者の結果に差を生じていると推測される。

この原因を、言語モデルを構成する単語 n -gram と単語辞書について議論する。単語 n -gram の問題は、学習データの少なさ (文献 [10] では、本実験の約 10 倍の 12 万文としている) や、学習データと発話のドメイン不一致が挙げられる。

単語辞書の問題として、発音できない不適切な音素列が含まれていた。音素変換器の文脈制約の弱さ、KyTea の読み推定誤りの影響、単語単位のみの方音素列付与などが原因であろう。

単語辞書ごとの関係を議論する。(1) 音素列認識精度は、混合単語辞書で高くなった。単語辞書に登録される音素列の種類が増えるからである。(2) 音声認識精度は、方言単語辞書で低下した。単語の読みとして方言音素列しか登録されないことによる誤認識が原因である。

3.3 実験 2: 方言変化頑健性の評価

単語 2-gram と方言単語辞書の下で、音素当たり平均エントロピーにより評価する。未知語モデルには音素 0-gram を用いる。音素列集合を C 、 C 全体の音素数を $|C|$ 、音素列 x の尤度を $P(x)$ と書くと、平均エントロピー H [bit/音素] は $H = -\frac{1}{|C|} \log_2 \prod_{x \in C} P(x)$ で定義される。

実験結果を表 4 に示す。音素当たり平均エントロピーは、音素列が 1 方言か混合かによらず、単語辞書の混合により減少した。定義より、音素列集合が同一のとき、平均エントロピーが小さいほど音素列集合の尤度 $\prod_{x \in C} P(x)$ は大きくなる。これにより、方言音素列の予測精度の向上、すなわち言語モデルにおける方言変化頑健性の向上が示された。

4. おわりに

本稿では、方言変換システムの実現のための音素変換器の構成方法、方言音声認識器の構成方法を提案した。検討課題として、統計的言語モデルにおいては Web 上の文章の活用や言語モデル語彙数の検討、音素変換器においては音素変化の結果の尤度による選択的な読み変換、音節単位での変換器構成が挙げられる。単語 n -gram や音響モデルの改良を行えば、音声認識においても方言変化頑健性の向上が期待できる。

最後に、言語処理に関して助言を頂いた京都大学森信介准教授に感謝する。

本研究の一部は、科研費 (S), GCOE の援助を受けた。

参考文献

- [1] 小林隆, 篠崎晃一 (編): ガイドブック方言研究, ひつじ書房, 2003.
- [2] 国立国語研究所編: 方言録音資料シリーズ (全 15 巻), 国立国語研究所 話しことば研究室, 1965-1973.
- [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ば集成 (全 20 巻), 国書刊行会, 2001-2008.
- [4] 大語彙連続音声認識エンジン Julius: <http://julius.sourceforge.jp/>
- [5] 佐藤守, 横山晶一, 西原典孝: 方言から共通語への翻訳システムに関する基礎的研究, 信学技報. NLC, 言語理解とコミュニケーション, Vol.102, No.199, pp.85-90, 2002.
- [6] David Chiang: A Hierarchical Phrase-Based Model for Statistical Machine Translation, in *Proceedings of the 43rd Annual Meeting of the ACL*, pp.263-270, 2005.
- [7] OpenFst Library: <http://www.openfst.org/>
- [8] Kylm - The Kyoto Language Modeling Toolkit: <http://www.phontron.com/kylm/>
- [9] Kyoto Text Analysis Toolkit: <http://www.phontron.com/kytea/>
- [10] 西崎博光, 中川聖一: 音声認識誤りと未知語に頑健な音声文書検索手法. 信学論. Vol.J86-D-II, No.10, pp.1369-1381, 2003.