

広島弁音声認識のためのコーパスと言語モデルの構築

中本典子 目良和也 黒澤義明 竹澤寿幸
 広島市立大学情報科学部
 [nakamoto, mera, kurosawa, takezawa]@ls.info.hiroshima-cu.ac.jp

1 はじめに

今日、音声認識の技術は様々な分野で使用され、キーボードからの入力に代わる文字入力方法だと注目を集めている。しかし、従来の音声認識は標準語のみを対象としており、方言を用いた認識率は標準語の認識率に加えてかなり低い。この問題を解決するために、我々の研究グループでは方言の一つである広島弁を対象とした、大語彙連続音声認識装置 Julius[1]の改良に取り組んでいる。そのためまず方言独特の単語を 418 語追加した。しかし、音声認識には単語情報だけでなく発音情報や活用情報が必要なため、単語を追加しただけでは広島弁の認識率はあまり向上しなかった。そこで、本研究では Julius の言語モデル作成に使用する日本語形態素解析システム ChaSen[2]を広島弁の活用情報に対応できるように改良する。さらに、改良された ChaSen によって単語辞書と言語モデルを作成し、広島弁認識用に Julius を改良することにより、広島弁の認識率の向上を目指す。

2 関連研究

2.1 音声認識装置 Julius[1]

大語彙音声認識エンジン Julius はフリーでありながら非常に高性能であり、数百万語彙の連続音声認識を一般の PC 上でほぼ実時間で実行することが可能である。また、発音辞書や言語モデル・音響モデルなどの音声認識の各モジュールを組み替えることで、幅広い用途に応用可能である。また、ソースコードを含めてオープンソースなのが特徴である。

Julius は起動時にまず設定ファイルを読み込む。設定ファイルには、単語辞書、2-gram 辞書、逆向き 3-gram 辞書、そして音響モデルのファイル指定と細かい動作設定が記述されている。設定ファイルの読み込みが終わると、音声入力を待つ状態になる。音声入力が行われると、入力の音声波形から特徴抽出が行われ、特徴量ベクトルの時系列が抽出される。この入力特徴量ベクトル系列から、音響モデル、単語辞書、2-gram 辞書、逆向き 3-gram 辞書を元に単語列がテキストで出力される。

2.2 形態素解析システム ChaSen[2]

日本語形態素解析システム ChaSen は、自然言語処理研究に資するためにフリーのソフトウェアとして開発されたものであり、15 万語/秒の速度で解析することが可能である。また、単語や文法の定義、単語間の接続関係の定義を使用者が追加したり、変更できるように配慮されている。

ChaSen では、文章が入力されると、まずその文字列と部分一致する単語を検索する。そして、前の単語との接続関係を調べ、一番信頼性の高い単語と各語の品詞、活用形、読み等を入力する。解析は、単語ひとつひとつの情報が定義されている辞書定義ファイル、品詞の種類が定義されている品詞定義ファイル、活用型の種類が品詞ごとに定義されている活用型定義ファイル、活用型定義ファイルに書かれている活用型がどのような活用をするかが定義されている活用形定義ファイル、単語情報や活用情報に関する接続の規則が定義されている接続表定義ファイルの 5 つの定義ファイルの情報を基に行われる。

3 形態素解析システム ChaSen の改良

まず、広島弁の対話文を広島弁の書籍や広島弁で書かれたブログ 658 文から収集した。そして、広島弁を形態素解析するために必要となる単語情報や活用情報を収集し、それらの情報を ChaSen の辞書を構成するための 5 つの定義ファイルにそれぞれ追加した (単語情報 1,376 語、品詞情報 3 個、活用型情報 11 個、活用形情報 15 個、接続情報 26 個)。

単語情報を追加する際、形態素生起コストの値を意味が類似する標準語の単語と同じ値に設定した。活用情報を追加する際は、単語情報を定義する際や広島弁特有の活用に対応するために必要となる、品詞や活用型、活用形の定義を新たに決めて追加した。さらに、追加した定義に対して、それらに類似する定義に関する接続の定義があればそれを参照し、追加した。また、最初に定義した広島弁の単語 418 語のうち、品詞が間違っていた 12 語を削除し、残りの 406 語を定義しなおした。追加した活用形定義の例を図 1 に示す。

(サ変・スル広島弁			
(
(基本形	する	スル)
(未然又接続	せ	セ)
(命令 e	せえ	セエ	セー)
(命令 e	せえ	セエ	セー)
(命令 y a	しろや	シロヤ)
(命令 y a	せえや	セエヤ	セーヤ)
(命令 y o	せえよ	セエヨ)
)			

図 1：追加した活用形の定義の例

4 音声認識装置 Julius の改良

まず、広島弁の対話文を広島弁の書籍や広島弁が紹介されている Web サイト、Twitter から 1,993 文収集した。また、標準語の対話文を書籍とデータベース（京都観光案内対話コーパス[3]）から 44,967 文収集した。そして、言語モデルの作成を行い、広島弁認識のために適したコーパスの組み合わせを確かめる実験を行った。

この時、音声実験を行うデータとして言語モデル作成に利用したコーパスに含まれる広島弁 (closed 広島弁) と標準語 (closed 標準語)、また、コーパスに含まれていない広島弁 (open 広島弁) と標準語

(open 標準語) の文章を合計 165 文を用意し、広島出身の 22 歳から 25 歳までの男女 12 名 (男子 11 名、女子 1 名) に読んでもらい音声データを収集した。そして、そのうち雑音が入っていない 147 個のデータを実験に用いた。その内訳はそれぞれ closed 広島弁 20 文、open 広島弁 55 文、closed 標準語 27 文、open 標準語 54 文である。

5 広島弁コーパスからの言語モデルの作成

広島弁対話文データと標準語対話文データの比率を確かめるため一定数の広島弁対話文データに標準語対話文データを適宜足して言語モデルを作成し認識率を比較する音声認識実験を行った。

この時、認識率は以下の式によって算出する。

$$\text{認識率} = \frac{\text{正解単語数}}{\text{全単語数}} \times 100 \quad (1)$$

図 2 に認識結果を示す。図中の縦軸は認識率、横軸は言語モデル作成に使用した標準語の数を示す。また、標準語のみとは元の Julius の言語モデルの認識結果である。この時、一定の広島弁対話文データ

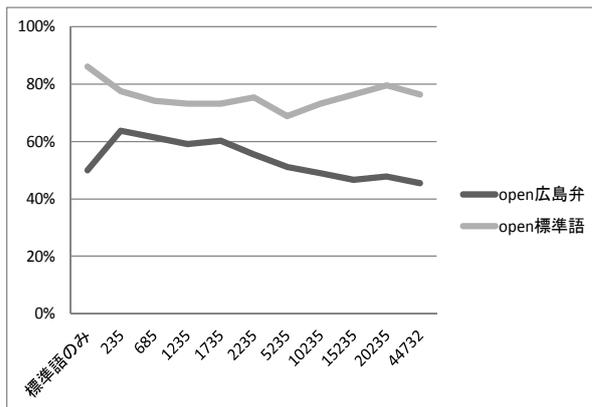


図 2：標準語データを増やした場合の認識率

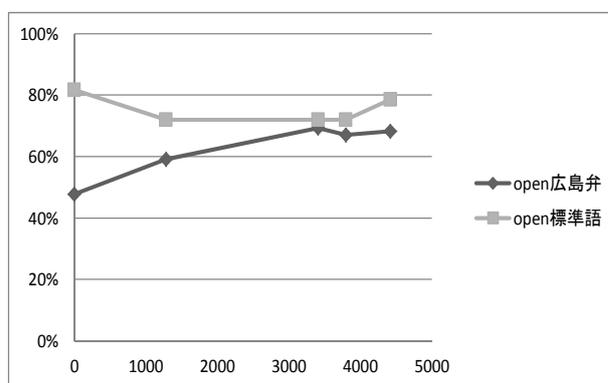


図 3：広島弁データ数を増やした場合の認識率

として 343 文のデータを用いた。また、評価のための音声データは収集したデータの内 open データをそれぞれ 20 文ずつ用いた。

図 2 より、標準語の比率が増すにつれて広島弁の認識率の低下がみられる。特に 1,735 文から 2,235 文以降、低下の幅が大きくなっている。標準語では 5,235 文以降で認識率に大きな上昇が見られる。このことより広島弁対話文データの約 6 倍の量の標準語対話文データを含んだ言語モデルなら広島弁の認識率があまり下がらず標準語も認識できるといえる。

次に、広島弁対話文データの文数を少しずつ増やして言語モデルを作成し認識率がどのように変化するか実験を行った。評価用音声データとして収集した open データをそれぞれ 20 文ずつ用いた。

結果を図 3 に示す。横軸は単語辞書の語彙数となっている。また、0 の位置にある数値は元の Julius の言語モデルを用いた際の認識結果である。

図 3 では、最初広島弁の認識率は上昇し、3500 語を超えた辺りから横ばいになっている。これより、単語辞書の語彙数が 3500 語以上あればある程度の認識率が得られることがわかる。

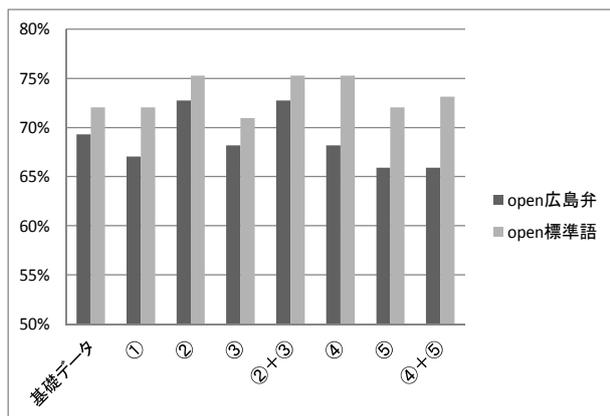


図4: 学習データの種類を変えた場合の認識率

また、どのデータが有効であるか言語モデルを作成するコーパスに追加するデータの種類の種類に関する実験を行った。評価用音声データは open データをそれぞれ 20 文ずつ用いた。比較する値として式 (1) の認識率を用いる。

この時、使用したコーパスは、まず、基礎データとして収集したデータから 1963 文を使用し、それに次に示す対話文を加えたもので比較した。

- ① Twitter から収集したデータ (2192 文)
- ② Web サイト A から収集したデータ (2524 文)
- ③ Web サイト B から収集したデータ (2041 文)
- ②+③ Web サイト A・B から収集したデータ (2602 文)
- ④ 広島弁で書かれた昔話の会話部分 (2159 文)
- ⑤ 広島弁で書かれた昔話の会話以外 (2141 文)
- ④+⑤ 広島弁で書かれた昔話の全て (2337 文)

図4より、Web サイトから収集したデータにおいて、Web サイト A から収集したデータでは認識率が上昇するが、Web サイト B から収集したデータでは認識率が低下している。二つを足したものにおいても Web サイト A から収集したデータより認識率が低下することから、Web サイト B から収集したデータからとってきたデータは有用ではないといえる。

広島弁で書かれた昔話から収集したデータでは、図4においてどのデータにおいても認識率は低下するものの、会話では少しの低下となっているが会話以外では大きく低下している。二つを足したものにおいてはより認識率は低下している。これより、会話以外は有用ではないと考える。

これらより、それぞれ有用であると考えられる

データの特徴として、Web サイト A から収集したデータでは①話し言葉で書かれた例が多いこと、②より多くの広島弁に対応した文章が掲載されていること、また、広島弁で書かれた昔話では、①話し言葉で書かれていること、があげられる。これより、より多くの広島弁に対応する広島弁の話し言葉コーパスを作成することが必要であるといえる。

6 評価実験

6.1 広島弁形態素解析システムの評価

改良した ChaSen を評価するため、元の ChaSen と、単語情報だけを追加した ChaSen、単語情報と活用情報の両方を追加した ChaSen の 3 パターンを使用して、広島弁の形態素解析の正解率がどの程度上がるかを調べた。解析には収集した広島弁の対話文と標準語の対話文をそれぞれ 30 文ずつ使用した。また、正解率は以下の式で求めた。

$$\text{正解率} = \frac{\text{正解した形態素数}}{\text{全形態素数}} \times 100 \quad (2)$$

表1: 広島弁の対話文の解析結果

	正解率	
	形態素の切り方	形態素の切り方と品詞
従来手法	79.4%	75.0%
単語情報だけ	81.3%	76.9%
単語情報と活用情報	92.3%	89.7%

表2: 標準語の対話文の解析結果

	正解率	
	形態素の切り方	形態素の切り方と品詞
従来手法	98.5%	97.5%
単語情報だけ	98.5%	97.5%
単語情報と活用情報	98.5%	97.5%

その結果、広島弁の形態素解析実験では形態素の切り方に着目した場合、正解率は従来手法から単語情報だけを追加した ChaSen で上がり、単語情報と活用情報の両方を追加した ChaSen ではさらに上がっている。また、形態素の切り方と品詞に着目した場合も正解率は上がった。よって、単語情報だけを追加した ChaSen よりも、単語情報と活用情報の両方を追加した ChaSen の方が、正解率が上がると言える。なお、標準語の形態素解析実験の結果は、

どの ChaSen を使用しても形態素の切り方に着目した場合の正解率、形態素の切り方と品詞に着目した場合の正解率ともに解析結果も含め全て同じであり、提案手法による標準語の形態素解析への悪影響は出ていないといえる。

6.2 広島弁音声認識システムの評価

次に、単語情報と活用情報の両方を追加した ChaSen を用いて、5 節を参考に元の Julius の言語モデル (元 Julius) と、データを選んで作成した広島弁コーパスを用いて作成した言語モデル (広島弁)、それに、6 倍の標準語対話文データを追加して作成した言語モデル (広島弁+標準語) の 3 パターンの言語モデルで音声認識実験を行った。音声データは open データを用いた。結果を表 3 に示す。

表 3 : 認識結果

	広島弁	標準語
元 Julius	33.46%	96.48%
広島弁	89.85%	78.41%
広島弁+標準語	75.66%	77.97%

これより広島弁の認識率は元 Julius、広島弁+標準語、広島弁と上昇していることがわかる。標準語の認識率の低下は広島弁、広島弁+標準語ともに 18% 程度の低下にとどまった。今回は全体的なコーパスの量が少ないため標準語データ追加による標準語認識率の向上は見られなかったが、音声認識結果のバリエーションは増えていたため、今後より大きなコーパスを使用することで認識率が向上する可能性がある。

7 おわりに

本研究では、Julius の言語モデル作成に使用する日本語形態素解析システム ChaSen を広島弁の活用情報に対応できるよう改良した。さらに、改良された ChaSen によって単語辞書と言語モデルを作成し、広島弁認識用に Julius を改良することにより、広島弁の認識率の向上を目指した。

そのために、まずは広島弁の対話文を 658 文収集し、広島弁の形態素解析に必要な単語情報と活用情報を得た。そして、それらの情報を基に、辞書定義ファイルに単語の定義を 1,376 語追加した。また、最初に定義した広島弁の単語 418 語のうち、品詞が間違っていた 12 語を削除、残りの 406 語のコスト値

を変更した。さらに、品詞定義辞書ファイルに品詞の定義を 3 種類、活用型定義ファイルに活用型の定義を 11 種類、活用形定義ファイルに活用形の定義を 15 種類、接続表定義ファイルに接続の定義を 26 種類、それぞれ追加した。次に、広島弁の対話文を 1993 文、標準語の対話文を 44,967 文収集し、広島弁コーパスを作成した。その広島弁コーパスを上記 ChaSen を用いて形態素解析し、Julius の単語辞書・言語モデルを作成した。この時、様々な条件で言語モデルを作り音声認識実験を行った。

評価実験として、形態素解析実験では、広島弁の形態素解析の正解率は、元の Chasen より単語情報だけを追加した ChaSen、それよりも単語情報と活用情報の両方を追加した ChaSen と上がった。音声認識実験では、広島弁の対話文データ 2741 文と標準語の対話文データ 235 文から作成した言語モデル (広島弁)、そのデータに標準語のデータ 16049 文を追加した言語モデル (広島弁+標準語) の音声認識率を元の Julius (元 Julius) と比較する実験を行った。その結果、広島弁の認識率は元 Julius で 33.46%、広島弁+標準語で 75.66%、広島弁で 89.85% と認識率が向上した。標準語の認識率では、元 Julius で 96.48%、広島弁で 78.41%、広島弁+標準語で 77.97% と認識率がやや減少した。

今後の課題として、形態素解析実験では表記揺れを考慮した単語の追加や、追加した単語に対する必要な活用形や接続の定義の追加、そして助詞との融合における単語の変化に対応した新たな手法の提案が挙げられる。さらに、多くの広島弁特有の単語が含まれる広島弁の話し言葉コーパスを作成すること、形態素解析システムの辞書に新たな単語やそれについての活用情報を追加することが挙げられる。

謝辞

本研究において、本研究の基となった研究を行われた河野静香氏、また、本研究での形態素解析システムの改良を行われた松本優氏に深く感謝致します。

参考文献

- [1] A.Lee, et al., "Julius---an open source real-time large vocabulary recognition engine", Proc. of the 7th European Conference on Speech Communication and Technology, pp.1691-1694 (2001)
- [2] "chasen legacy", <http://chasen-legacy.sourceforge.jp/>
- [3] 大竹清敬ほか, "京都観光案内対話コーパスにおける対話行為の分析", 言語処理学会年次大会, pp.159-162 (2008)