

# 日本語生コーパスから自動獲得した未知語と 言語モデルによる大語彙連続音声認識

倉田 岳人 森 信介 西村 雅史

日本アイ・ビー・エム東京基礎研究所

〒 242-8502 神奈川県大和市下鶴間 1623-14

{gakuto,forest,nisimura}@jp.ibm.com

あらまし

大語彙連続音声認識 (LVCSR) においては、言語モデル (LM) が必要である。我々は新聞などの一般的な大規模コーパスから構築した LM を利用することができる。しかし、このような LM では様々な分野の音声を正確に認識することができない。これは、その分野特有の単語、言い回しが存在するからである。これらに対応するための理想的な方法は、その分野のコーパスを収集し、専門家がそれを人手で単語分割し、単語分割されたコーパスから構築した LM を利用することである。しかし、このような処理には多大なコストがかかる。本報告では、生コーパスから直接 LM を構築する方法を提案し、その LM を LVCSR の性能で評価した。実験の結果、少量のコーパスを多大なコストをかけて単語分割するよりも、大量のコーパスを提案手法で処理して LM を構築する方が、LVCSR において高い性能を得られることが確認された。

キーワード 大語彙連続音声認識 言語モデル 未知語

## Large Vocabulary Continuous Speech Recognition with a Japanese Language Model from a Raw Corpus

Gakuto KURATA, Shinsuke MORI, Masafumi NISHIMURA

IBM Research, Tokyo Research Laboratory, IBM Japan, Ltd.

1623-14 Shimotsuruma Yamatoshi Kanagawaken 242-8502 Japan

{gakuto,forest,nisimura}@jp.ibm.com

### Abstract

A large segmented corpus is essential for building a Japanese language model (LM), which is used in large-vocabulary continuous speech recognition (LVCSR). We have large segmented corpora such as newspapers. However, recognition of speech in a specific domain is difficult with an LM based on that kind of corpora because it doesn't cover the Out-of-Vocabulary words. An ideal approach to this problem is to prepare a manually segmented corpus. This requires, however, a lot of time and money. In this paper, we propose a method to build a Japanese LM from a raw corpus and evaluate the LM built by our method in the accuracies of LVCSR. The experiments showed that our LM from a raw corpus was superior to one from a corpus segmented by an automatic word segmenter. Also, the LM from a large raw corpus with our method outperformed one from a small correctly segmented corpus.

**Key Words** Large Vocabulary Continuous Speech Recognition, Language Model, Unknown Words, Out-Of-Vocabulary

# 1 はじめに

大語彙連続音声認識 (LVCSR : Large Vocabulary Continuous Speech Recognition) では、探索空間の絞込みのために、言語モデル (LM : Language Model) が利用される。現在、LM としては単語  $n$ -gram モデルが最もよく利用されている。これは、コーパス中の単語、および単語の接続の頻度に基づき、単語の出現確率を予想するモデルである。つまり、単語  $n$ -gram モデルの構築のためには、単語分割された大規模コーパスが必要となる。しかし、日本語や中国語のようなアジア系言語においては、コーパスは空白によって単語分割されていない。

日本語の場合、様々な新聞記事を単語分割した大規模コーパスが利用可能である。このようなコーパスに基づく単語  $n$ -gram モデルを用いた場合、新聞記事に近い内容の音声の認識は精度よく実現することができる。

最近では、LVCSR は、コールセンターの書き起こし、裁判所での自動調書作成、大学での講義での字幕作成など様々な分野で利用されるようになってきている [1]。LVCSR をある特定の分野に導入する場合、その音声には、新聞に出現するような一般的語彙に含まれない、その分野特有の語彙 (OOV: Out-of-Vocabulary) や表現が出現する。これらは、新聞コーパスから構築した LM では適切に扱うことができず、結果として、LVCSR において高い認識精度は期待できなくなる。このような場合、導入分野のコーパスに基づく LM を構築し、それと新聞コーパスなどから構築された一般的な LM とを組み合わせて利用することで、導入分野特有の OOV、表現を考慮した LVCSR が実現される [2]。上述したように、単語  $n$ -gram モデルを構築するためには、単語分割されたコーパスが必要であり、導入対象分野の正確に単語分割されたコーパスが存在する場合、それから構築された LM を利用することが、最も理想的な手法である。しかし、実際に LVCSR を導入するすべての分野について、その分野の単語分割済みのコーパスが存在するわけではない。その分野の単語分割されていないコーパスに対して、人手で正しく単語分割し、それに基づいた LM を構築することが理想的であるが、このような処理には莫大なコストと時間がかかり、現実的ではない。また、正確な単語分割を行うためには、その分野に対する専門的知識を有する人が必要となる。

一方、最近の自動単語分割システムは非常に高精度に動作し、これを用いて導入分野のコーパスの単語分割を行う、ということも考えられる [3]。しかし、自動単語分割システムも、導入分野特有の OOV を適切に扱うことができない。これは、単語自動分割システム自体も、一般的な新聞コーパスなどに基づく統計情報を利用しているためである。このような観点から、新しい導入分野特有の OOV を、既存の統計情報に基づくシステムで処理する、ということには根本的な難しさがある、ということができる。

また、OOV に対応する方法として、サブワードモデル

なども提案されている [4]。この手法を用いると、OOV の問題の一部は解決される。しかし、この手法を用いるためには、単語分割されたコーパスが必要となる。

本報告では、対応する LM が存在しない分野に LVCSR を導入する場合に、OOV を適切に取り扱い、高精度の認識精度を実現する方法を提案する。提案手法では、LVCSR の対象分野の単語分割されていないコーパス (以下、「生コーパス」と称する) から直接 LM を作成する。従来は、生コーパスに対して、自動単語分割システムにより単語分割を行い、その結果に基づき LM を構築していたが、自動単語分割の時点で、OOV などに起因する分割誤りが生じ、その結果、OOV を正しく認識することができなかった。しかし、提案手法では、自動単語分割を行わないため、このような間違いが混入することを防ぐことができる。

提案手法を利用する場合、LVCSR を導入しようとしている分野に対する生コーパスは豊富に存在する、ということ的前提としている。例えば、テレビニュースの LVCSR を考えた場合、最新のニュースの記事を Web から得ることができる。また、大学の講義の LVCSR を考えた場合、講義の教科書を利用することができる。このようなことから、我々が前提としている状況は、妥当なものである、と考えられる。

実験の結果、提案手法に基づく LM を利用した場合、自動単語分割システムの結果に基づく LM を利用する場合よりも、LVCSR において高い認識精度を得ることができた。また、LVCSR の対象分野の少量のコーパスが人手により正確に単語分割されており利用可能な場合でも、それに基づき作成した LM よりも、大量の対象分野の生コーパスに対して、提案手法を適用して作成した LM の方が、LVCSR において高い性能を示した。

## 2 提案手法

本節では、生コーパスから単語  $n$ -gram 確率を推定する方法を説明する。また、LVCSR に必要な情報の準備方法についても説明する。

提案手法の処理の流れは、以下の通りである。

1. 生コーパスから OOV の候補 (以下では「OOV 候補」と称する。) を抽出
2. 最も確率の高い読みを OOV 候補に付与
3. 生コーパス中での、OOV 候補と一般的語彙に対する単語  $n$ -gram 確率を推定

以下では、各処理について説明する。

### 2.1 OOV 候補の抽出

生コーパスから OOV 候補となる文字列を抽出する。この処理の最大の目的は、後に続く  $n$ -gram 確率の推定にかかる計算量を削減することである。そして、この処理で抽

出された OOV 候補は、次の処理で最も確率の高い発音を割り当てられて、認識語彙に加えられる。この観点から、この処理においては、精度 (Precision) よりも再現率 (Recall) が重要となる。

日本語文章中から単語、特に OOV を同定するのは非常に難しいタスクである。このタスクにおいては、文字レベルでの処理が有効であることが示されている [5]。

ここでは、我々は再現率を重視した大まかな抽出方法を利用した。以下に処理の流れを示す。

1. ストップワードで生コーパスを分割する。
2. 全ての部分文字列の頻度を計数する。
3. ある部分文字列の頻度に対して、左右どちらかに一文字を追加した部分文字列の頻度が小さくなった場合、その部分文字列を抽出する。

図 1 に、上記の処理の概要を示した。

部分文字列	頻度
アデニル酸シクラゼ	10
アデニル酸シクラゼは	6
アデニル酸シクラゼが	4
のアデニル酸シクラゼ	7
にアデニル酸シクラゼ	3

↓

アデニル酸シクラゼ
-----------

図 1: 部分文字列の頻度の例

この処理により抽出された部分文字列の中で、3.2 で示す一般的な語彙に含まれない部分文字列を OOV 候補として扱うこととした。

## 2.2 OOV 候補に対する音素の割り当て

2.1 で抽出された OOV 候補に対して、発音、つまり適切な音素列を割り当てる方法を説明する。日本語の多くの文字には複数の読みが存在するため、一つの OOV 候補には、非常に多くの読みの可能性がある。これらの中から最も適切な読みを選択するため、我々が開発している Text-to-Speech システムの未知語読みモジュールを利用した。この未知語読みモジュールは、文字とそれに対応する読みの組み合わせを単位とする  $n$ -gram に基づいており、その精度はおよそ 80% 程度となっている。

2.1 で抽出された OOV 候補は、この処理で適切な読みを割り当てられて、LVCSR の認識語彙に組み込まれるようにした。

## 2.3 生コーパスに基づく OOV 候補の単語 $n$ -gram 確率の推定

一般的な認識語彙と、2.1 で抽出した OOV 候補が与えられた場合、これらに対する LM が、LVCSR には必要と

なる。日本語は、単語間が空白で分割されていないため、生コーパスからこれらの単語に対する単語  $n$ -gram 確率を推定しないと行けない。

生コーパス中のすべての部分文字列を単語セットと考え、それに対する単語  $n$ -gram 確率を推定する方法が提案されている [6]。ここでは、 $n_r$  文字からなる生コーパスは文字列  $x = x_1x_2 \cdots x_{n_r}$  とみなされる。次に、 $i$  番目の文字  $x_i$  のあとの単語境界が存在する確率  $P_i$  がすべての  $i \in \{1, 2, \dots, n_r - 1\}$  について計算される。このように、単語間に単語境界が存在するかどうか、ということが示されている生コーパスを、「確率的に単語分割されているコーパス」と呼ぶ。このコーパス中に存在する単語数 (確率的単語 zero-gram 頻度) は、 $f_r(\cdot) = 1 + \sum_{i=1}^{n_r-1} P_i$  として求まる。

文字列  $x_{i+1}^{i+k}$  は、この文字列の前後に単語境界が存在し、この文字列中に単語境界が存在しない場合に、単語  $w = x_{i+1}^{i+k}$  として成立する。そのため、単語  $w$  の生コーパス中での確率的頻度  $f_r$  は、単語  $w$  の表記のすべての出現  $O_1 = \{i \mid x_{i+1}^{i+k} = w\}$  に対する期待頻度の和として、次のように定義される。

$$f_r(w) = \sum_{i \in O_1} P_i \left[ \prod_{j=1}^{k-1} (1 - P_{i+j}) \right] P_{i+k}$$

ある単語の uni-gram 確率  $P_r(w)$  は、その単語の確率的 uni-gram 頻度  $f_r(w)$  と確率的 zero-gram 頻度  $f_r(\cdot)$  から、 $P_r(w) = f_r(w)/f_r(\cdot)$  のように求めることができる。

単語 uni-gram 確率と同様に、単語  $n$ -gram 確率も、確率的単語  $n$ -gram 頻度を確率的単語  $(n-1)$ -gram 頻度で割ることにより算出することができる。

本報告では、上述の単語セット (一般的な認識語彙と抽出した OOV 候補) に対する単語 bi-gram モデルが必要となる。単語セットに含まれる単語列の、生コーパス中での確率的頻度は、すでに述べた方法で求めることができる。しかし、この単語セットは生コーパス中のすべての部分文字列を含んでいるわけではない。よって、この単語セットに含まれない部分文字列 (以下では、これらを「未知語」と称する。また数式中では  $\langle \text{UNK} \rangle$  と表す。) を含む単語列の確率的頻度を求める必要がある。

単語分割されたコーパスにおいては、未知語の頻度は、すべての未知語の頻度の合計として定義される。同様に、確率的に単語分割されたコーパスにおいても、既知の単語セットに含まれないすべての部分文字列の確率的頻度の合計をもって、未知語の頻度  $f_r(\langle \text{UNK} \rangle)$  とすることができる。

$$f_r(\langle \text{UNK} \rangle) = \sum_{w \in \mathcal{W}_u} f_r(w)$$

ここで、 $\mathcal{W}_u$  は、すべての未知語を表している。この計算を実現するためには、生コーパス中のすべての部分文字

列を数えなければならず、非常に多くの計算量を要する。これを避けるために、以下の等式を使用した。

$$f_r(\cdot) = \sum_{w \in \mathcal{X}^+} f_r(w) = \sum_{w \in \mathcal{W}_k} f_r(w) + \sum_{w \in \mathcal{W}_u} f_r(w)$$

ここで、 $\mathcal{W}_k$  は既知の単語セットを表している。また、 $\mathcal{X}^+$  は、生コーパス中のすべての文字列を表している。これにより、未知語の確率的頻度は、 $f_r(\langle \text{UNK} \rangle) = f_r(\cdot) - \sum_{w \in \mathcal{W}_k} f_r(w)$  のように求めることができる。同様に、未知語を含む bi-gram の確率的頻度も以下のように求めることができる。

$$\begin{aligned} f_r(w_1, \langle \text{UNK} \rangle) &= f_r(w_1) - \sum_{w \in \mathcal{W}_k} f_r(w_1, w) \\ f_r(\langle \text{UNK} \rangle, w_2) &= f_r(w_2) - \sum_{w \in \mathcal{W}_k} f_r(w, w_2) \\ f_r(\langle \text{UNK} \rangle, \langle \text{UNK} \rangle) &= f_r(\cdot) - f_r(w_1, \langle \text{UNK} \rangle) \\ &\quad - f_r(\langle \text{UNK} \rangle, w_2) - \sum_{(w_1, w_2) \in \mathcal{W}_k} f_r(w_1, w_2) \end{aligned}$$

以上により、未知語を含む場合でも、確率的  $n$ -gram 頻度を確率的  $(n-1)$ -gram 頻度で割ることにより、生コーパスから単語  $n$ -gram 確率 ( $n=1, 2$ ) を推定できることを示すことができた。

### 3 実験条件

本節ではまず、認識実験に利用した音響モデル (AM : Acoustic Model) について説明する。次に、利用することができる新聞記事に基づく一般的な LM について述べる。

#### 3.1 音響モデル

表 1 に、今回の実験で利用した AM を構築する際に利用した、自然発話音声コーパスのサイズを示す。

表 1: 自然発話音声コーパスの概要

話者数	97 人
発話時間	83 時間
文数	27,135 文
単語数	1,098,888 単語
異なり単語数	23,929 単語

各音素は環境依存で 3 状態の left-to-right 型の HMM として表現している。HMM の各状態は音素環境決定木によりクラスタリングされており、決定木のリーフ数は 2,728 である。また、HMM の各状態は、11 混合の混合正規分布でモデル化されている。

#### 3.2 一般的な言語モデル

表 2 に、基準となる一般的な LM を構築するのに利用したコーパスの概要を示した。このコーパスは、主に新聞

記事から構成されている。単語分割に関しては、一部分は専門家により正確に処理されている。残りの部分に関しては、自動単語分割システムにより分割された後、専門家により大まかに点検されている。

表 2: 新聞 LM を構築するためのコーパスの概要

コーパス中の単語数	24,442,503 単語
一般的語彙の語彙数	45,402 単語

このコーパスから構築した LM を、以下では「新聞 LM」と呼ぶ。また、このコーパスに出現する単語の頻度上位 95% に対して、正確な音素記号を割り当てて、LVCSR の認識対象語彙とした。以下では、この語彙を「一般的語彙」、この語彙に含まれない語彙を「OOV」と称することとする。

実験では、単語 bi-gram モデルを用いた。単語 tri-gram を用いた場合、生コーパスからの単語  $n$ -gram 確率の算出に多大な計算が必要となる。しかし、予備実験において、単語 bi-gram モデルを用いた場合と単語 tri-gram モデルを用いた場合で、本質的な精度の差はなかった。よって、本実験では、単語 bi-gram モデルを利用することとした。

## 4 実験

本節では、提案手法に基づく LM の性能を検証するために行った実験について説明する。

### 4.1 3 種類の言語モデル

LVCSR の適用分野の生コーパスが利用可能である場合、その生コーパスに基づいて以下の 3 種類の手法で LM を構築した。LVCSR の際に、OOV、OOV 候補に対して発音が割り当てられていないといけませんが、その方法についても以下に示した。

従来手法 M 専門家が生コーパスを正確に単語分割し、それに基づき LM を作成する。OOV の発音も専門家が正確に与える。

従来手法 A 生コーパスを自動単語分割システムによって単語分割し、それに基づき LM を作成する。OOV の発音は 2.2 の方法で自動的に割り当てた。

提案手法 P 生コーパスから 2.1 に示した方法で OOV 候補を抽出した。そして、2.3 で示した方法を用いて、一般的語彙と OOV 候補の単語  $n$ -gram 確率を算出した。OOV 候補の発音は、2.2 の方法で自動的に割り当てた。

手法 M が最も理想的な手法であるが、この手法には多大な時間とコストがかかり、大量の生コーパスを手法 M で処理して LM を構築することは現実的ではない。

これに対して、手法 *A*、および手法 *M* では人手による処理はなく、すべての処理が自動的に行われる。よって、これらの手法で大規模な生コーパスを処理し LM を構築することは実現可能である。

いま、3.2 で示した主に新聞記事から構築されている LM を利用することができる。今回の実験では、上で示した 3 種類の手法で構築した LM のうちの 1 つと、この新聞 LM の補間モデルを構築した。そして、3 種類の補間モデルを用いた LVCSR の精度をもって、上の 3 種類の LM 構築手法の評価を行った。

図 2 に今回行った実験の流れを示した。この図の中で、「New Words」として現されているものは、手法 *M* での OOV、手法 *A* での自動単語分割の誤りに基づく誤りを含む可能性がある OOV、および、手法 *P* での OOV 候補を表している。

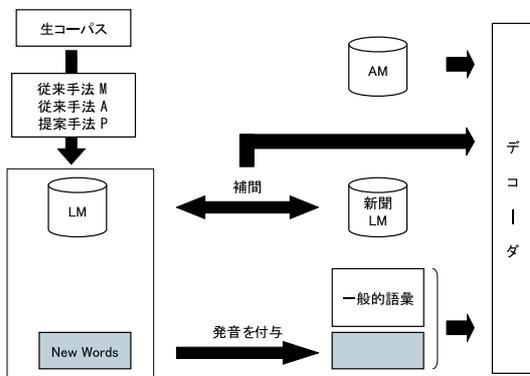


図 2: 実験の流れ

## 4.2 評価用音声データ

放送大学はテレビとラジオを通じて講義を配信している。講義の内容は多岐に渡り、一般的語彙に含まれない単語も頻出する。よって、LVCSR を放送大学の各々の講義に導入する場合を考えて、我々の提案手法の有効性を検証することとする。

今回の実験では、放送大学の 3 種類の講義を対象とした。そして、講義中の講師の自然発話をテストデータとして用いた。表 3 に、テストデータの概要を示した。

## 4.3 関連する生コーパス

テストデータとした 3 種類の講義それぞれに関連する生コーパスとして、講義の教科書や関連する文書を利用した。各々の講義に対する生コーパスとして、2 種類のサイズのコーパスを用意した。各々を以下では、「小規模コーパス」、「大規模コーパス」と称する。ここで、小規模コーパスのサイズは、手法 *M* で処理することができる生コーパスのサイズの常識的な上限のサイズとして設定した。実際に 3 種類の小規模コーパスを正確に単語分割するために約 6 日がかかった。

## 5 評価実験の結果と考察

本節では、行った実験の結果を示し、それに対して考察を加える。また、提案手法により、適切に扱うことができた OOV の例も示す。

### 5.1 結果と考察

放送大学の 3 回の講義を対象とした LVCSR の結果を、表 4 に示した。

左から 2 列目の結果が、新聞 LM を単独で用いた場合の文字誤り率 (CER: Character Error Ratio) である。そして、他の列の結果が、新聞 LM と 4.1 で示した 3 種類の手法で構築した LM のうちの 1 つとの補間モデルを用いた場合の CER である。

左から 2 列目と他の列を比較すると、新聞 LM のみを用いた場合は、大学の講義のように、専門性の高い用語を含む発話を適切に扱うことができない、ということがわかる。実際に認識結果を見てみると、我々の予想通り、OOV の部分での認識誤りが多く見受けられた。

#### 5.1.1 関連する小規模コーパスが利用可能な場合

左から 3,4,5 列目は、関連する小規模コーパスが利用できるという条件で、3 種類の方法で LM を構築した場合の、認識実験の結果を示している。

2 列目と 4 列目を比較すると、小規模コーパスを利用した場合に、認識性能が向上していることがわかる。また、5 列目を見ると、人手で小規模コーパスを正確に単語分割した場合は、さらに認識性能が向上していることがわかる。このように、たとえ少量であっても、関連するコーパスを利用することで、LVCSR の性能向上につながる、ということが確認された。これは、先行研究の知見とも合致する [2]。

さらに、4 列目と 5 列目を比較すると、提案手法 *P* を利用した場合、従来の自動単語分割システムによる分割結果を利用して LM を構築する場合よりも、高い認識性能を得ることができる、ということがわかる。これは、これらの 2 つの処理がすべて自動で行われる、という点から、提案手法 *P* を利用することにより、人手の作業に伴うコストの増加なしに、従来手法 *A* よりも高い認識性能を得ることができる、ということを示している。また、提案手法 *P* を用いた場合の認識性能は、従来手法 *M*、つまりコストがかかる人手による単語分割の結果に基づく LM を利用する場合の認識性能に肉薄するものであった。

#### 5.1.2 関連する大規模コーパスが利用可能な場合

左から 6,7,8 列目は、関連する大規模コーパスが利用できるという条件で、3 種類の方法で LM を構築した場合の、認識実験の結果を示している。

大規模コーパスが存在しても、それをすべて人手で正確に単語分割することは、コストの面からも、時間の面から

表 3: テスト用音声データおよび関連する生コーパスの概要

講義 ID	テスト用音声データ		関連する生コーパス (総文字数)	
	講義のタイトル	単語数	小規模コーパス	大規模コーパス
$B$	生物	2,260	10,641	73,437
$M$	音楽	2,679	16,251	88,996
$G$	地球科学	2,270	10,892	69,617

表 4: 文字誤り率 (%)

講義 ID	新聞	生コーパス					
		小規模コーパス			大規模コーパス		
	LM	$M$	$A$	$P$	$M$	$A$	$P$
$B$	27.0	11.5	12.9	<b>11.6</b>	N/A	13.7	<b>11.0</b>
$M$	24.9	17.3	17.9	<b>17.6</b>	N/A	18.0	<b>17.0</b>
$G$	28.0	23.3	25.0	<b>23.1</b>	N/A	23.4	<b>22.9</b>

も現実的ではない。そのため、手法  $M$  で大量の生コーパスから LM を構築することは不可能である。

5 列目と 8 列目を比較した結果、大規模コーパスに基づき、提案手法  $P$  を用いて LM を構築した方が高い認識性能を得ることができることがわかる。つまり、提案手法  $P$  を用いる場合、コーパスの量が増えれば、認識性能の向上が期待できる、ということが言える。

次に、3 列目と 8 列目を比較した場合、8 列目の結果の方が優れている。これは、大規模コーパスから提案手法  $P$  を用いて LM を構築する方が、小規模コーパスを人手により正確に単語分割し、その結果を利用して LM を構築した場合 (従来手法  $M$ ) よりも、高い認識性能を得ることができる、ということを示している。前者の処理がすべて自動で行われるのに対して、後者の処理が単語の分割という非常にコストのかかる処理を伴う、ということを経験すると、この結果は非常に望ましいものである、と言える。

## 5.2 提案手法により認識できた OOV

図 3 に、提案手法を用いることにより正確に認識することができた OOV を示した。これらの単語は、手法  $A$  を用いた場合、単語分割誤りが原因となり、正しく認識することができなかった単語である。

正しく認識された OOV	自動単語分割の誤りの例
先カンブリア紀 三葉虫 Gたんぱく質	先 / カンブリア / 紀 三 / 葉 / 虫 G / たんぱく質

図 3: 適切に認識された OOV

## 6 おわりに

新聞記事から構築された LM だけを利用して、様々な分野に LVCSR を導入しようとした場合、OOV やその分野特有の表現などを一般的な LM では扱うことができないた

め、高い認識精度は期待できない。そのため、従来は、導入する分野のコーパスを収集し、それを人手により正確に単語分割する、という非常にコストのかかる作業が必須となっていた。

本報告では、生コーパスから直接 LM を構築する手法を用いて、様々な分野に LVCSR を導入した場合に、高い認識性能を得ることができる方法を提案した。そして、従来のように、導入分野のコーパスを人手で分割して LM を構築するよりも、単純に導入分野のコーパスを大量に集め、提案手法を用いる方が高い認識性能を得ることができることを示した。これは、今後様々な分野に LVCSR を導入する際のコストを抑えることができる可能性を示唆している。

## 7 謝辞

放送大学の番組制作に携わっておられる方々に深謝します。

## 参考文献

- [1] K. Miyamoto. “Effective Master-Client Closed Caption Editing System for Wide Range Workforces”. In *Proc. of HCI International 2005 (to appear)*.
- [2] R. De Mori D. Janiszek and F. Bechet. “Data Augmentation and Language Model Adaptation”. In *Proc. of ICASSP 2001*, pages 549–552.
- [3] M. Nagata. “A Stochastic Japanese Morphological Analyzer Using a Forward-DP Backward-A\* N-Best Search Algorithm”. In *Proc. of COLING 94*, pages 201–207.
- [4] Y. Ogawa, H. Yamamoto, Y. Sagisaka and G. Kikui. “Word Class Modeling for Speech Recognition with Out-of-Task Words Using a Hierarchical Language Model”. In *Proc. of EUROSPEECH 2003*, pages 221–224.
- [5] M. Asahara and Y. Matsumoto. “Japanese Unknown Word Identification by Character-based Chunking”. In *Proc. of COLING 2004*, pages 459–465.
- [6] S. Mori and D. Takuma. “Word  $N$ -gram Probability Estimation From A Japanese Raw Corpus”. In *Proc. of ICSLP 2004*, pages 201–207.