

講演の書き起こしに対する統計的手法を用いた文体の整形

下岡 和也 河原 達也 奥乃 博

京都大学 情報学研究科 知能情報学専攻
〒 606-8501 京都市左京区吉田本町
e-mail: shitaoka@kuis.kyoto-u.ac.jp

あらまし 講演音声のような話し言葉の書き起こしや音声認識結果には、話し言葉特有の表現が数多く含まれて二次利用しにくいいため、文章として適した形態に整形する必要がある。本稿では、統計的機械翻訳の考え方に基づいて講演の書き起こしを自動的に整形された文章に変換する方法を提案する。本研究で扱う処理は、フィラーの削除、句点の挿入、助詞の挿入、書き言葉表現への変換、文体の統一である。これらの処理を統合的に行うようにビームサーチを導入した。以上の手法により、実際の講演の書き起こしを整形された文章に変換し、講演録編集者によって一次整形されたものを正解として、句点の挿入、助詞の挿入、文体の統一に関して定量的な評価を行った。

キーワード 話し言葉, 講演音声, 統計的機械翻訳, 自動整形, 自動要約

Automatic Transformation of Lecture Transcription into Document Style using Statistical Framework

Kazuya Shitaoka Tatsuya Kawahara Hiroshi G. Okuno

School of Informatics, Kyoto University, Kyoto 606-8501, Japan
e-mail: shitaoka@kuis.kyoto-u.ac.jp

Abstract Transcriptions and recognition results of lecture speech include many expressions peculiar to spoken language. Thus, it is necessary to transform them into document style for practical use of them. We apply the statistical approach used by machine translation to automatic transformation of the spoken language into document style sentences. We deal with deletion of the fillers, insertion of periods, insertion of particles, conversion to written word expression and unification of a text style. To apply these processings in an integrated manner, we introduce a beam search. The preliminary evaluation of the proposed method is presented using real lecture transcriptions and their archives.

keyword spontaneous speech, lecture speech, machine translation, text processing, summarization

1 緒論

音声認識研究の対象は、読み上げ音声から講演や会議などの話し言葉に移行している。このような話し言葉は日本語では特に、文章に用いる書き言葉と大きく異なるので、音声認識結果や書き起こしそのままでは、講演録や議事録として二次利用することが難しく文章として適切な形態に整形する必要がある。

これまでに、放送ニュースなどを対象とした自動要約の研究が行われている [1, 2, 3, 4, 5]。しかし、これらは人間が行うレベルで要約を行っているのではなく、主に頻出区間や重要文の抽出という処理を行っており、生成された文章が必ずしも人間にとって読みやすいものではない。話し言葉である講演を記録として残すには書き言葉への変換・修正が必要であり、実際に人間も要約を作成する際にはこのような作業を行っている。この際には、不必要な部分の削除以外に書き言葉表現への変換や必要な語の挿入、あるいは1つの文書内での「ですます調」「である調」などの文体の統一といった処理も必要である。

本研究では講演の書き起こしに対してこのような整形を自動的に行うことを考える。現在、文章を整形するソフトも存在しているが、これらはパターンマッチング的に規則ベースで変換を行っており、言語的な妥当性や前後との整合性は考慮されていない。また基本的に1対1の変換を行っているため、変換の候補が複数ある場合に対応できない。そこで、本研究では整形文章の作成において、規則に基づいて1対1の変換を行うのではなく、話し言葉と書き言葉を別の言語とみなした上で、統計的な機械翻訳の手法を適用することを考える。

2 書き起こしからの講演録作成法

2.1 講演録作成の流れ

講演録編集者は書き起こしから講演録や要約を作成する際に次の4段階の作業を行う。

1. 一次整形
2. 長文の分割、文法的チェック、ポリテイカルチェック
3. 意味的チェック
4. 要約の作成

第一段階の一次整形では、フィラーの削除や書き言葉表現への変換、助詞の挿入などを行う。第二段階の文法的チェックでは、不適切に用いられている助詞や接続詞を適切なものに修正する。ポリテイカルチェックでは、差別用語などの不適切な表現の修正を行う。第三段階の意味的チェックでは、専門用語が正しく用いられているかの確認を行う。

2.2 整形作業における処理

実際に講演録編集者が「日本語話し言葉コーパス」(CSJ)の18講演について整形・要約したものを参考にして、(a)から(i)で第一段階、(j)と(k)で第二段階の作業についてそれぞれ例を挙げながら説明する。

(a) フィラーの削除

「あのー」や「えっと」といった間投語はすべて削除する。

(b) 書き言葉表現への変換

「行ってるんですが」→「行っているのですが」
「規則合っていう方式」→「規則合成という方式」

上記の例では話し言葉と書き言葉が1対1に対応しているが、そうでない場合を以下に示す。
「このようにいたしております」→「このようにしています」
「このようにしている」
「このようにしております」

このような場合は、全体として「ですます調」か「である調」に統一されるように留意する。

(c) 助詞の挿入

「観測されたらそれ適当な分布で」→「観測されたらそれを適当な分布で」

「先程も話しありましたが」→「先程も話しがありましたが」

(d) 句点の挿入

書き起こしに句点がない場合は、適当な箇所に句点を挿入する。

(e) 倒置部の修正

「日本語は前寄りでも後寄りでも一応”あ”なんですね音韻的には。」→「日本語は前寄りでも後寄りでも音韻的には一応”あ”なんですね。」

(f) 言い淀みをしている部分の削除

「このようなスペクトルのドッ、ギャップがで

すね」

(g) 説明を付け加えている部分の削除

「モデル化を行い、そのモデル自体 HMM で
すが、そのモデルから」
このような表現は必ずしも削除されるわけでは
なく、別の表現に書き換えられることもある。
「こういう HMM、我々は MSD-HMM と呼ん
でいる」→「こういう HMM(MSD-HMM)」

(h) 独り言の部分の削除

「えー、そろそろ時間なんですけど、」

(i) 言い直しをしている部分の削除

「これは、テキストテキストからの音声合成を」
「上唇の、ごめんなさいこれ間違いです」

(j) 長文の分割

「～日本人としてある意味で嬉しいところ
あるんですけどもどうしてこういう技術が考
えられたかというのと、～」→「～日本人とし
てある意味で嬉しいところがあります。どう
してこういう技術が考えられたかというのと、～」

(k) 助詞・接続詞の修正

「どのようにして音声の合成をするかという
ことが」→「どのようにして音声の合成を
するかが」

本研究では、これらの作業のうち、(a) から (d) の
一次整形の処理について取り扱う。これらの処理だ
けでもかなり読みやすいものに整形される反面、こ
れ以上の処理については、内容の理解を含めた高度
な処理が必要であると考えられるからである。また、
助詞・接続詞の修正については取り扱わないので、正
解の講演録としては、2.1 節で述べた講演録作成にお
ける第二段階のものを使うことにした。参考のため
に図 1 に書き起こし、図 2 に整形された文章の例を
示す。

3 統計的手法による文体の整形

3.1 統計的翻訳のアプローチ

現在、音声認識や機械翻訳の研究において統計的
な手法が広く用いられている。入力系列を X、出力
系列を Y とすると、これらは X を観測した際の Y
の事後確率 $P(Y|X)$ を最大にする Y を求めるとい

まーあのーこうやってあつ話しており
ますとそれは勿論あのー言語的情報を
伝える < pause-212-msec > ということが一つ
の重要な目的なんなんですけど
< pause-278-msec > 同時にパラ言語情報そして
非言語情報が伝わっております < pause-282
-msec > まーこの三分法は藤崎先生による
ものでして < pause-597-msec > えーパラ言語
情報というのは < pause-286-msec > 要は
あのー意図的に制御できる話者が
ちゃんとコントロールして出してるんだ
けども < pause-265-msec > 言語情報と違って
連続的に変化するからカテゴライズする
ことがやや難しいそういった情報であり
ます < pause-447-msec > でそういったものが
音声生成過程の中で畳み込まれてまー
一次元の音声波として実際は出て
いる訳です < pause-551-msec >

図 1: 書き起こしの例

こうやって話をするとは、それはもちろん
言語的情報を伝えることが一つの重要な目的
ですが、同時に、パラ言語情報、非言語情報
が伝わっています。この三分法は藤崎先生に
よるもので、パラ言語情報とは、要は意図的
に制御できる。話者がちゃんとコントロール
して出しているのだけれども、言語情報と
違って連続的に変化する。それから、カテ
ゴライズすることがやや難しい。そういった情報
です。そういったものが音声生成過程の中で
たたみこまれて、一次元の音声波として実際
は出ているわけです。

図 2: 整形された文章の例

枠組みで捉えられ、ベイズ規則により以下のように
定式化される。ここで、 $P(Y)$ は系列 Y が生起する
事前確率、 $P(X|Y)$ は系列 Y から系列 X が生起する
条件付き確率である。右辺の分母 $P(X)$ は、Y の決
定に影響しないので通常無視して考える。

$$\max_Y P(Y|X) = \max_Y \frac{P(Y)P(X|Y)}{P(X)}$$

音声認識 [6, 7] の場合は、X は入力音声、Y は
出力単語列となる。この場合は、音響モデルにより
 $P(X|Y)$ を、言語モデルにより $P(Y)$ を求めている。

機械翻訳 [8, 9] の場合は、X を入力言語、Y を出
力言語として $P(Y|X)$ を最大にする Y を求めること
で、入力言語 X を出力言語 Y に変換する。この場
合、 $P(Y)$ は音声認識と同様に言語モデルにより求め
る。 $P(X|Y)$ は変換モデルを仮定しその確率を求め

る。変換モデルとは、入力単語はある出力単語系列 (null を含む) に対応づけられるという仮定の下で、どの単語に対応するかということを書全体における相対的な位置も考慮して確率で表したものである。

2章で述べたように、書き起こしの文体と整形した文章の文体は大きく異なっており、書き起こしの単語列と整形された文章の単語列を異なる言語とみなすことができる。そこで、本研究では書き起こしの文体を整形する際に、書き起こしを整形した文章に翻訳するという考え方にに基づき、機械翻訳と同様に統計的手法を適用することを検討する。

以下では、フィルターの削除、書き言葉への変換、助詞の挿入、句点の挿入、文体の統一というそれぞれの処理において、この枠組みをどのように実現するかについて説明する。

なお本研究では、書き起こしを形態素解析した結果をデータとして用いている。形態素解析には、ChaSen ver2.02 を用いている。また、話者がポーズをおいた箇所にはその情報がポーズ長とともに記録されている。

3.2 フィルターの削除

ここでは、 $P(X|Y)$ は整形文 Y に対応する話し言葉 X においてフィルターが挿入される確率とする。フィルターは、発話のどの部分にも出現する可能性があるが、特に句読点の後によく出現する傾向がある。

ただし、整形された文章にフィルターは一切出現しない。つまり、 Y をフィルターを含む単語列であると仮定すると、 $P(Y)$ の値は 0 となり、 $P(Y|X)$ の値も 0 となる。これは書き起こしの単語列 X から全てのフィルターが削除されることを意味している。

3.3 書き言葉表現への変換

ここでは、 $P(X|Y)$ は書き言葉表現の話し言葉表現への変換確率と解釈される。講演録編集者が一次整形を行う際、文章の順序を入れ替えるという処理は行わないので、書き起こしと講演録編集者により一次整形された文章を照らすことで、学習により $P(X|Y)$ の値を推定することができる。ただし本研究では、正解の講演録の数が学習には不十分であったため、あらかじめ人手により書き言葉から話し言葉への変換規則を作成しておき、それらに対してのみ確率を推定することとした。作成した変換規則数は 64 個あ

表 1: 書き言葉への変換規則・確率の一部

書き言葉 (Y)	話し言葉 (X)	$P(X Y)$
という	っていう	0.14
	という	0.86
が	けども	0.036
	けど	0.042
	が	0.922
どのよう	どういう風	0.46
	どのよう	0.54
(～し)ている	(～し)てる	0.12
	(～し)ている	0.88

り、一部を表 1 に示す。

3.4 助詞の挿入

ここでは、 $P(X|Y)$ は整形された文章 Y に含まれる助詞が話し言葉 X において脱落する確率と解釈される。書き起こしと講演録編集者により作成された講演録を比較したところ、次のようなパターンにおいて助詞が脱落していた。

- 「名詞」(助詞)「名詞」
(例)「このお話しを幹事の方から」
- 「名詞」(助詞)「動詞」
(例)「我々は作ってきたわけです」
- 「名詞」(助詞)「形容詞」
(例)「非常に能力が高くなって」
- 「名詞」(助詞)「接続詞」
(例)「これはつまりサンプルごとの」

そこで、これらのパターンの助詞が脱落する規則をあらかじめ作成しておき、それが生起する確率 $P(X|Y)$ を推定する。ただし、脱落されうる助詞の候補を人手によりあらかじめ規定しておく。そして、それらの助詞を挿入する場合としない場合との $P(Y)P(X|Y)$ の値を比較することによって助詞を挿入するかしないか、どの助詞を挿入するのかの判断を行う。

3.5 句読点の挿入

CSJ の書き起こしデータには句読点がなく、その代わりに話者のポーズ情報が記録されている。ここでは、整形された文章の単語列 Y に含まれる句読点

表 2: 句点の前後に出現する単語の候補

～ます	～ますね	～です	～ですな	～なる
～れる	～れた	～ある	～いる	～た
～だ	～と	で～		

が、音声(書き起こしの単語列 X)においてポーズ情報に変換される確率 $P(X|Y)$ を考える。なお、本研究では句点のみを扱う。なぜなら、句点を挿入する位置は人によらずほぼ一定であるが、読点を挿入する位置は様々であり定量的な評価を行うのが難しいためである。

文末には、「です」「ます」などのような典型的な表現が多い。しかし、話し言葉においては独特の表現が文の区切りになることがある。文末での「～と」と文頭での「で～」である。その例を以下に挙げる。(例)

「単位が使われていたと。あるいは」

「大きく違う。でそのままでは」

句点の前後に出現する表現の一覧を表 2 に示す。本研究では、句点を挿入する対象箇所をこれらに限定する。その上で、 $P(X|Y)=P(\text{ポーズ}|\text{句点})$ の確率において、

- (1) すべてのポーズを対象
- (2) 平均以上の長さのポーズを対象
- (3) 表現に依存してポーズ長が変わると仮定

の 3 通りのモデル化について述べる。

(1) 整形した文章 Y における句点があらゆる長さのポーズに変換されうるとする。この場合は、書き起こし X におけるポーズがすべて句点に変換されることになり、その判定を言語モデル確率 $P(Y)$ を用いて行う。我々が昨年報告した講演音声の自動インデキシングの研究 [10] では、この考え方が用いられている。

(2) 句点の挿入箇所には、ある程度長いポーズがおかれると考えられるので、 $P(X|Y)$ において、句点が各話者の平均ポーズ長以上の長さのポーズに変換されうるとする。この場合、平均ポーズ長以上のポーズに対してのみ最終的には、 $P(Y)$ を考慮して句点に変換するかどうかの判定を行う。

(3) 「です」「ます」などの典型的な文末表現にある句点はあらかゆる時間長のポーズになりうるが、書き言葉においては文の区切り表現とならない「～と」

「で～」あるいは、文中にも頻繁に使われる「～た」の部分にある句点は平均ポーズ長以上の長さのポーズになるとする。この場合、「ます」「です」などの後のポーズは時間に関係なく句点に変換されうるが、「～と」「で～」 「～た」の部分のポーズは平均長以上のポーズに限り、句点に変換されうることになる。この場合も、最終的に $P(Y)$ を考慮して判定を行う。上記 3 通りについて、4 章で比較・評価する。

3.6 文体の統一

話し言葉を書き言葉に変換する際に、その変換の候補が複数ある場合には、どの書き言葉に変換するかを判定を言語モデル確率に基づいて行う。ここで、使用する言語モデルを異なるものにとすると、変換結果も異なったものになる。

本研究では 2 種類の言語モデルを用いることによって、それぞれ文体が「ですます調」あるいは「である調」に統一された文章の生成をめざす。使用する言語モデルは、講演録のコーパスから作成されたものと、新聞記事のコーパスから作成されたものを用いる。前者の言語モデルを用いると「ですます調」の文体に、後者の言語モデルを用いると「である調」の文体に統一されると期待される。

3.7 デコーディングアルゴリズム

これまで、講演の書き起こしから整形された文章を生成する処理とモデルについて述べてきた。本研究では、尤度 $P(Y)$ の計算に単語 3-gram モデルを用いるので、これらの処理を逐次的に行うのではなく、統合的に行うように実装する必要がある。なぜなら、これらの処理を個別に行うと、すぐ前後に別の処理をする必要のある表現が存在する場合、尤度の計算に影響を与えるからである。

したがって、前後 2 単語に着目する必要がある表現が存在しなくなる範囲において、そのすべての変換パターンの尤度を比較して出力単語列を決定する。変換で複数の候補が生成されうることと考慮すると、可能な仮説の数は組み合わせ的に爆発するので、探索アルゴリズムを導入する必要がある。本研究ではビームサーチを行う。具体的には、生成したパターンの数が 100 を越えた場合は、そこまでの範囲で尤度を計算し、上位 100 個のパターンのみを選択することにした。

4 実験と評価

4.1 評価データと実験条件

以上の手法を用いて講演の書き起こしを整形し、その評価を行った。評価データとして、実際に学会で行われた4講演を用いた。評価データの概要を表3に示す。また、使用した言語モデルは、毎日新聞記事データで学習された2万語彙のもの[11]と、Web講演録で学習された1万語彙のもの[12]の2種類である。次節以降、行った実験結果と評価について述べる。ただし、本稿では変換確率 $P(X|Y)$ は学習データ量の問題から十分に推定されないため、人為的に1/0に設定して行っている。なお、3.2節のフィルターの削除は完全に機械的に行える。

表3: 評価データの概要

	講演		整形文
	時間	サイズ	サイズ
AS99SEP022 (A22)	28分	5557語	5378語
AS99SEP023 (A23)	30分	3899語	3802語
AS99SEP097 (A97)	13分	2509語	2451語
PS99SEP025 (P25)	27分	5371語	4854語

4.2 句点の挿入

まず、3.5節で述べた句点挿入の実験結果について考察する。なお、本実験は各処理を3.7節で述べたように統合的に行ったのではなく、句点の挿入についてのみ行った結果である。言語モデルは講演録コーパスを用いて作成したものを使用している。(1)の考え方に基づく場合の実験結果を表4に、(2)の考え方に基づく場合の実験結果を表5に示す。

表4において、適合率が低い主な原因は、「～と」「～た」の後と「で～」の前に誤って句点が挿入されている箇所が多いことである。「ます」「です」などの後に誤って句点が挿入されていることはなかった。これは、「～と」や「で～」は書き言葉においては文の区切りにならない表現であるためである。表5において、再現率が低い主な原因は、「ます」「です」などの、通常書き言葉で文末表現になるものの後で、ポーズ長が短い箇所が対象外になったためである。したがって、句点がポーズに変換される確率 $P(\text{ポーズ}| \text{句点})$ において、「です」「ます」などの典型的な文末表現部分にある句点はあらゆる時間長のポーズになりうるが、「～と」「で～」 「～た」の部分にある句点は平均ポーズ長以上の長さのポーズになるとす

るモデルを導入した。その結果を表6に示す。再現率・適合率とも高い精度を得ることができた。

なおA22の講演に対しては、再現率・適合率ともに低いことが原因であった。これは、この話者特有の現象であると考えられる。

表4: 句点挿入の実験結果 (ポーズ長制限なし)

講演 (ID)	再現率	適合率
A22	59/93(63.4%)	59/127(46.5%)
A23	78/86(90.7%)	78/123(63.4%)
A97	67/74(90.5%)	67/73(91.8%)
P25	108/118(91.7%)	108/128(84.4%)
平均	312/371(84.1%)	312/451(69.1%)

表5: 句点挿入の実験結果 (平均ポーズ以上)

講演 (ID)	再現率	適合率
A22	51/93(54.8%)	51/82(62.3%)
A23	61/86(70.9%)	61/69(88.4%)
A97	54/74(73.0%)	54/56(96.2%)
P25	88/118(74.6%)	88/99(88.8%)
平均	254/371(68.5%)	254/306(83.0%)

表6: 句点挿入の実験結果 (表現に依存してポーズ長を設定)

講演 (ID)	再現率	適合率
A22	57/93(61.3%)	57/83(68.7%)
A23	70/86(81.4%)	70/79(88.6%)
A97	62/74(83.8%)	62/68(91.2%)
P25	102/118(86.4%)	102/112(91.1%)
平均	291/371(78.4%)	291/342(85.1%)

4.3 文体の統一

次に3.6節で述べた文体の統一について、評価を行った。結果を表7に示す。ここで、分母は「ですます調」あるいは「である調」で統一されるべきと判断した表現の数、分子はWeb講演録言語モデルを用いた場合は「ですます調」に、新聞記事言語モデルを用いた場合は「である調」になっている表現の数である。

「である調」に関しては精度が低い。一方、「ですます調」に関しては高い精度で統一されている。「で

表 7: 文体の統一に関する実験結果

講演 (ID)	Web 講演録	新聞記事
A22	142/170(83.5%)	70/170(41.2%)
A23	94/116(81.3%)	65/116(56.0%)
A97	115/131(87.8%)	62/131(47.3%)
P25	150/166(90.4%)	61/166(36.7%)
平均	501/583(86.0%)	258/583(44.3%)

表 8: 助詞挿入の実験結果 (再現率)

正解の基準	Web 講演録	新聞記事
プロの講演録と一致	28/47 (60.0%)	29/47 (61.7%)
意味的に妥当	39/47 (83.0%)	40/47 (85.1%)

ある調」となっていない箇所が多くは、「思います」や「示します」のように、「ですます調」から「である調」に変換する際に、動詞の活用形を変化させる必要がある部分である。このような活用形を変化させる必要のある部分については変換の対象としていなかった。「ですます調」の統一に関して精度が高いのは、講演において話者は通常「ですます調」で発話するためである。

4.4 助詞の挿入

次に、3.4 節で述べた助詞の挿入に関する評価を行った。助詞が脱落している箇所は 4 講演で計 47 箇所あった。それらに対して、

1. プロの講演録編集者が作成した講演録において挿入されている助詞
2. 上記に含まれないが、意味的に妥当な助詞

を正解としてそれぞれ評価した結果を表 8 に示す。これは再現率に相当する。実際に講演録と一致したのは 60% 程度だが、意味的に妥当なものまで正解とすると 80% 以上の結果が得られた。

次に、挿入された全ての助詞に対して

- A: 意味的に妥当
B: 許容範囲である
C: 意味的におかしい

の 3 段階で評価した結果を表 9 に示す。これは適合

表 9: 助詞挿入の実験結果 (適合率)

正解の基準	Web 講演録	新聞記事
意味的に妥当	113/287 (39.3%)	101/244 (41.4%)
許容範囲	171/287 (59.5%)	142/244 (58.2%)
意味的におかしい	116/287 (40.5%)	102/244 (41.8%)

率に相当する。挿入された全ての助詞のうち、意味的に許容できる範囲までみても 60% 程度の精度しか得られていない。表 9 において、誤りであると判断された助詞が挿入されている箇所はほとんどが「名詞」「名詞」あるいは「名詞」「動詞」のパターンであった。

まず、「名詞」「名詞」のパターンについて考える。学会講演において用いられる専門用語の多くが複合名詞であるため、形態素解析を行うと「名詞」「名詞」と分解され、助詞の挿入箇所の候補になる。そこで、「名詞」「名詞」と連続しているもののうち、3 回以上出現すればその箇所は専門用語であると判断するようにした。これにより多くの専門用語は助詞の挿入箇所の候補にならなくなったが、1~2 回しか使われていない専門用語や、人名・地名などの固有名詞が複数の形態素に分解された結果、助詞が挿入されることがあった。例えば「東京大学」や「データベース」などである。あらかじめ人手により固有名詞を抽出し、それらを除外すると 75% 程度まで精度は向上する。

次に、「名詞」「動詞」のパターンについて考える。このパターンには、一見すると 1 つの動詞と考えられる表現が形態素解析によって分解されているものが多く含まれる。例えば、「記録する」という表現である。これが形態素解析によって、「記録(名詞)」「する(動詞)」と分解され、助詞の挿入候補となる。このような箇所は非常に多く、挿入される助詞の多くは「を」である。上記の例では「記録をする」となる。この場合は意味的におかしくないで誤りとは判断していない。しかし、「と」や「は」が挿入される場合がある。例えば「使用とする」「提示はする」などである。これらは、誤りと判断する。「~とする」「~はする」という表現がコーパス中に多く含まれているためにこれらの確率が高くなっている。

4.5 書き言葉表現への変換

3.3節で述べた書き言葉表現への変換についても、講演録編集者により作成された整形文章を正解として定量的な評価を行うことを検討した。

例えば、今回正解として用いた整形文章においては、「～であります」という表現は「～です」と書き換えられている。しかし、講演録のコーパスを基に作成した言語モデルを用いた場合、「～であります」の尤度が最大になった。これは、講演録において「～であります」という表現も不適切な表現でないためである。つまり、本研究で用いた講演録の作成者は、「～であります」を「～です」に書き換えるべきと判断したが、「～であります」という表現をそのまま用いて講演録を作成する人も多いと考えられる。このような表現は多数ある。このような変換は主観的要素が大きく、正解を用意して客観的に評価するのは困難であると考えた。

5 結論

本研究では、統計的な機械翻訳の考え方に基づいて文体の整形を自動的に行う手法を検討した。行った処理は、フィルターの削除、句点の挿入、助詞の挿入、書き言葉表現への変換、文体の統一である。これらを統合的に行うようにし、ビームサーチを導入した。

以上の手法により、実際の講演の書き起こしを整形された文章に変換した。正解の文章として講演録編集者によって一次整形されたものを用いて、句点の挿入、助詞の挿入、文体の統一に関して定量的な評価を行った。句点の挿入においては、再現率・適合率ともに80%程度の精度が得られた。助詞の挿入においては、上記の正解のみの基準では精度が高いとはいえないが、意味的に許容されるものを含めた場合には高い精度が得られた。文体の統一については、「～ですます調」への統一はできたが「～である調」への統一は不完全であった。全体としては、提案手法によりある程度の精度で講演の書き起こしの整形が行えたと考えられる。

今後の課題としては、変換確率 $P(X|Y)$ を大規模なコーパスを用いて推定することや、今回精度が十分でなかった「～である調」への統一について対応することが挙げられる。

参考文献

- [1] 加藤直人. ニュース文要約のための局所的要約知識獲得とその評価. 情報処理学会研究報告, 98-NL-126-10, 1998.
- [2] 若尾孝博, 江原暉将, 白井克彦. 短文分割を利用したテレビ字幕自動要約. 情報処理学会研究報告, 98-NL-126-9, 1998.
- [3] 中沢正幸, 遠藤隆, 古川清, 豊浦潤, 岡隆一. 音声波形からの音素片記号系列を用いた音声要約と話題要約の検討. 電子情報通信学会技術研究報告, SP96-28, 1996.
- [4] 堀智織, 古井貞熙. 話題語と言語モデルを用いた音声自動要約法の検討. 情報処理学会研究報告, 99-SLP-29-18, 1999.
- [5] 堀智織, 古井貞熙. 係り受け SCFG に基づく音声自動要約法の改善. 電気情報通信学会技術研究報告, SP2000-116, 2000.
- [6] 古井貞熙. 「音声情報処理」. 森北出版, 1998.
- [7] 鹿野清宏, 伊藤克巨, 河原達也, 武田一哉, 山本幹雄. 「音声認識システム」. オーム社, 2001.
- [8] I.Garcia-Varea, F.Casacuberta, and H.Ney. An Iterative, DP-Based Search Algorithm For Statistical Machine Translation. In *Proc. ICSLP*, Vol. 4, 1998.
- [9] Ye-Yi Wang and Alex Waibel. Fast Decoding For Statistical Machine Translation. In *Proc. ICSLP*, Vol. 6, 1998.
- [10] 長谷川将宏, 秋田祐哉, 河原達也. 談話標識の抽出に基づいた講演音声の自動インデキシング. 情報処理学会研究報告, 01-SLP-36-6, 2001.
- [11] 河原達也他. 日本語ディクテーション基本ソフトウェア (99年度版) の性能評価. SLP-31-2, NL-137-7, 2000.
- [12] 加藤一臣, 南條浩輝, 河原達也. 講演音声認識のための音響・言語モデルの検討. SP2000-97, NLC2000-49 (SLP-34-23), 2000.