

## 予稿とその講演書き起こしの対応付けおよび書き起こしのテキスト分割

内元 清貴 野畑 周 太田 公子 村田 真樹 馬 青 井佐原 均

総務省通信総合研究所

{uchimoto,nova,kimiko,murata,qma,isahara}@crl.go.jp

## 1 はじめに

近年、書き言葉のテキストはかなり精度良く処理できるようになってきた。これは大量のタグ付きコーパスが整備され、コーパスに基づく自然言語処理の研究が盛んに行なわれるようになってきたことに負うところが大きい。しかし、書き言葉のコーパスを基に作成されたシステムは話し言葉のテキストに対しても精度が高いとは限らない。なぜなら話し言葉と書き言葉の間には、両者では表現が異なる、話し言葉では文の単位が明確ではないなど、様々な違いがあるからである。このような状況を打開すべく、話し言葉のコーパスを作成するプロジェクト、開放的融合研究「話し言葉の言語的・パラ言語的構造の解明に基づく『話し言葉工学』の構築」プロジェクト [1] が1999年に始まった。話し言葉にはモノログやダイアログなど色々なものがあるが、このプロジェクトでは、話し言葉の初期段階の研究に相応しい対象として講演などのモノログが選ばれた。現在、学会講演などの講演を収録、その書き起こしを作成中で、2003年までには韻律情報や形態素情報などを付与したコーパスが公開される予定である。

講演には大抵予め話す内容がまとめられた予稿があり、講演そのものは予稿の言い替えあるいは要約と捉えることができる。特に学会講演は予稿の要点をまとめたものであるだけでなく、予稿の内容を詳しく読むべきかどうかを判断する材料でもある。これは要約の分類に照らすと informative な要約であるだけでなく indicative な要約であると言える。しかし、聴者が講演の途中である部分に興味を持ったとしても、予稿のどの部分を詳しく読むべきかという情報は与えられない。また、話の内容についていけなくなった場合には、あきらめるか予稿を読むしか手がない。このような場合でも、話のまとまりごとに、それと対応している予稿の節や段落が指示され、簡潔な要約あるいはキーワード、節のタイトルなどが提示されていると聴き手の理解の助けになる。また、講演には予稿にない情報も盛り込まれるので、講演を収録しておいて後で予稿をもとに検索したいというニーズもあるだろう。このような場合にも、話のまとまりごとに予稿との対応付けができていれば、対応する部分を容易に検索できるようになるし、必要に応じて予稿の対応部分の情報も利用しながら、話のまとまりごとに informative な要約を作成することも可能になる。本稿では、この話のまとまりを、予稿の節と対応する講演の各部分と定義し、講演の書き起こしとその予稿の各節とを対応付けることによって動的に同定するという方法を提案する。以降で、この話のまとまりをセグメントと呼ぶことにする。

## 2 対応付けとテキスト分割の方法

## 2.1 予稿とその講演書き起こしの対応付け

講演で予稿と同じ内容のことが話されている場合、予稿で使われている表現の多くを用いて話の内容を記述できると考えられる。特に、ある節の内容について話している場合には、その節で使われている表現を最も多く用いて記述できるだろう。この場合、予稿の各節から知識を獲得し、その知識を用いて講演の書き起こしを解析

すると、他の節の知識を用いた場合に比べて、より効率良く解析できそうである。解析の方法としては色々考えられるが、一つの方法として、各文字列に節番号のタグを付与していくことを考える。これは形態素解析において品詞などの属性を付与する代わりに節番号を付与するタスクと同じなので、典型的な形態素解析の手法を用いる。まず、解析の前に予め、予稿の各節から  $n$ -gram を抽出し、単語辞書を作成しておく。単語辞書の見出しは  $n$ -gram とし、各見出しにはどの節に現れたかという情報と単語コストを与える。単語コストは、tf-idf 法、語の長さ、節の長さ、節の表題に含まれているかどうかを考慮して単語に得点付けし、その逆数とする。予稿に現れなかった  $n$ -gram は未知語とし、どの節のタグでも付与できるものとする。そして、未知語には文字列の長さに比例した単語コストを与える。具体的には、単語コストを以下の関数  $f(s, w)$  で与える。

$$f(s, w) = \begin{cases} \frac{1}{TF(s, w) \times \frac{N}{DF(w)} \times length(w) \times \frac{l_{max}}{length(s)} \times t(s, w)} & \text{(既知語)} \\ \alpha \times length(w) & \text{(未知語)} \end{cases} \quad (1)$$

ここで、 $f(s, w)$  の  $s, w$  はそれぞれ節番号、文字列を表わす。式 (1) 中の各項が表わす意味はそれぞれ以下の通りである。

$TF(s, w)$  : 節  $s$  における文字列  $w$  の出現頻度 (単語頻度)。

$\frac{N}{DF(w)}$  :  $DF(w)$  は文字列  $w$  が出現した節の異なり数 (節頻度) を、 $N$  は予稿中の節の数を表わす。

$length(w)$  : 文字列  $w$  の長さ (文字単位)。

$\frac{l_{max}}{length(s)}$  :  $length(s)$  は節  $s$  の長さ (文字単位) を、 $l_{max}$  は最長の節の長さ (文字単位) を表わす。ただし、数式、図表は除く。

$t(s, w)$  :

$$t(s, w) = \begin{cases} k \times freq(s, w) & \text{(表題中に現れた場合)} \\ 1 & \text{(表題中に現れなかった場合)} \end{cases} \quad (2)$$

ここで、 $freq(s, w)$  は節  $s$  中の表題における文字列  $w$  の出現頻度を表わす。 $k$  は定数である。

$\alpha$  : 定数。

例えば、5 節からなる予稿で、それぞれの節の長さが同じで、「対応付け」という文字列が第1節と第2節のみに現れ、出現頻度はそれぞれ3回と5回で、そのうち1回は2節の表題に現れたものとする、それぞれの節における単語コストは以下のように表わされる。

$$f(1, \text{対応付け}) = \frac{1}{3 \times \frac{1}{5} \times 4 \times 1 \times 1} = \frac{1}{24}$$

$$f(2, \text{対応付け}) = \frac{1}{5 \times \frac{1}{5} \times 4 \times 1 \times k} = \frac{1}{40 \times k}$$

文字列  $W$  が与えられたとき、各節に現れた単語を用いて  $W$  を生成し、コストの和が最小になるような節  $s_{best}$  を  $W$  に対応する節と考える。つまり、文字列  $W$  が節  $s$

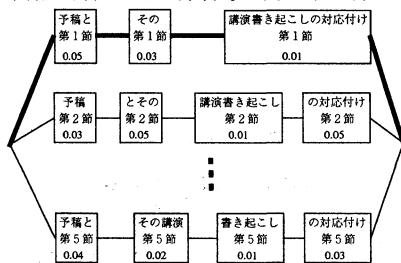
の単語  $\{w_{s,i}\}$  を用いて  $W = w_{s,1}w_{s,2}\dots w_{s,n}$  のように生成できるとき、 $W$  は

$$s_{best} = \underset{s}{\operatorname{argmin}} C(s, W) \quad (3)$$

$$= \underset{s}{\operatorname{argmin}} (f(s, w_{s,1}) + f(s, w_{s,2}) + \dots + f(s, w_{s,n})) \quad (4)$$

となる節  $s_{best}$  に対応するとする。例えば、入力文字列として「予稿とその講演書き起こしの対応付け」という文字列が与えられたとき、その文字列と最も対応付けられる節は次のような手順で求めることができる。

1. n-gram 単語辞書を参照して、入力文字列中の各文字位置から始まる語をすべて取り出す。
2. 取り出された語のうち接続規則を満たすものをつないでいく。入力文字列の最初と最後には特別なノードを設け、それぞれの節に接続可能としておく。それ以外に同じ節のみ接続可能とする。
3. それぞれの語に単語コストを与える。すると図1のようなコスト付きのグラフ構造が得られる。
4. 文字列全体で総コストが最小となるような語の並びを解とする。図1の例では、太線のリンクでつながれたノードをたどることによって得られる単語の並び(一番上のパス)がコスト最小解であり、それらの単語の属性である節番号が対応する節である。



(数字は単語コスト)

図1: 解析例

コスト最小解を効率良く求めるために、ビタビ (Viterbi) アルゴリズムを用いる。これは文頭から文末まで部分的最適解を保持しながら一文字ずつ解析を進めていくという方法である。

## 2.2 書き起こしのテキスト分割

予稿の節との対応を考慮しながら書き起こしを先頭から順に追っていくと、途中で話題が変わるところがある。この話題の変わり目がセグメントの境界であり、その前後の文字列はそれぞれ異なる節に対応する。この性質を仮定し、次の手順でセグメントの境界を推定する。

1. 先頭から順に解析し、それまでの文字列  $W$  に対し、式 (3) を用いて  $s_{best}$  を求め、 $s_0$  とする。式 (3) から、このときのコストは  $C(s_0, W)$  である。
2. 一方、 $W$  を任意の位置  $i$  (先頭からの文節の数とする) で二分割して、前後の文字列  $W_{i1}, W_{i2}$  に対し同様に  $s_{best}$  を求め、それぞれ  $s_{i1}, s_{i2}$  とする。式 (3) から、このときのコストはそれぞれ  $C(s_{i1}, W_{i1}), C(s_{i2}, W_{i2})$  である。文字列  $W_{i1}$  と  $W_{i2}$  の間の接続にはコスト  $C$  を与え、二分割したときの総コストを  $C(s_{i1}, W_{i1}) + C(s_{i2}, W_{i2}) + C$  とする。
3. 分割しなかったときのコストと二分割したときの総コストを比較したとき、 $C(s_0, W) > C(s_{i1}, W_{i1}) + C(s_{i2}, W_{i2}) + C$  となる位置  $i$  を分割位置の候補とする。このような候補が複数ある場合には、二分割したときの総コストが最小となるもののみを候補とする。

4. 後方へ  $d$  文節まで文脈を伸ばしても分割位置候補  $i$  およびその前後でコスト最小で対応付けられる節  $s_{i1}, s_{i2}$  が常に変わらない場合、その候補の位置  $i$  で分割する。このとき、分割位置  $i$  の前後の文字列  $W_{i1}, W_{i2}$  は、コスト最小で対応付けられた節  $s_{i1}, s_{i2}$  にそれぞれ対応するとする。
5. 文字列  $W_{i1}$  とその対応する節  $s_{i1}$  を出力し、分割位置  $i$  を先頭と見なして上の手順を繰り返す。

## 3 実験と考察

### 3.1 実験の材料

実験対象とした講演は、1999 年音響学会秋季大会、言語処理学会第 6 回年次大会の講演で収録に協力して頂いたものの中から無作為に抽出したもので、それぞれ 35 講演、25 講演の合計 60 講演である。実験には、それぞれの講演について作成された書き起こしと予稿を用いた。このうち、1999 年音響学会秋季大会の 10 講演分をパラメータの調整に、残りの 50 講演分をテストに用いた。

基本形	発音形
0016 00050.511-00051.143 L:	(F エ)
(F エ)	キンネン
近年	
0017 00051.425-00052.869 L:	(F エー)
(F エー)	ゲータイソカイセキ
形態素解析	
0018 00053.073-00054.503 L:	デ
で	ジューヨーナ
重要な	カタイト
課題と	ユーフー <C>
いう風 <C>	
0019 00054.707-00056.341 L:	ニ
に	カンガエラレイマスノワ
考えられていますのは	

図2: 書き起こしの例

書き起こしは図2のように基本形と発音形の部分からなる。ここで、発音形は発話者の発声を忠実に書き起こしたものであり、基本形はその仮名漢字混じり表記である。数字から始まる行はタイムスタンプを表わし、次のタイムスタンプまでの文字列が発表開始後何秒から何秒の間に発話されたかを示している。タイムスタンプ以外の各行は転記基本単位からなる。この単位の認定には明確な基準が設けられている。また、この単位は文節の概念に近いが、本実験では分割の基本単位である文節として扱った。単語が複数の転記基本単位にまたがる場合はつなげて一つの文節とみなす。このような場合、書き起こしには <C> のようなラベルが付与される。実験では発音形は用いず基本形のみを用いた。ここには括弧付きでラベルが付与されている部分がある。これには、フィラーや言い淀み、引用などいくつかのタイプのものがあり [1]、フィラーや言い淀みはラベルごと削除し、引用などの場合はラベルのみ削除して用いた。予稿は図表、数式を除いた部分を用いた。

### 3.2 実験の条件

話題のまとまりであるセグメントは予稿の各節を単位とし、節の区分としては一階層のみを考慮した。各節から抽出する n-gram の  $n$  の値は 2 以上とした。つまり、2 文字以上の文字列を抽出して辞書を作成した。対応付けの際に用いる式 (1) の  $k, \alpha$  の値はそれぞれ 5、0.1 とした。テキスト分割の際に用いる 2.2 節の接続コスト  $C$  の値は 2 とした。また、分割の条件として 2.2 節で述べた  $d$  の値は 10 とした。つまり、分割位置の候補およびその前後の文字列に対応する節が 10 文節伸ばしても変わらなければその候補の位置で分割するものとした。

### 3.3 評価方法

まず、人手で正解を作成した。二人の被験者に書き起こしをセグメントに分割してもらい、セグメントの前後には図3のように予稿の対応する節を表わすタグを

SGML 形式で付与してもらった。その際、セグメントの境界では、内容に関係のない、慣用的に用いている表現にはタグ付けしないようにお願いした。

<S4> 認識実験の... 提案法の方がより認識率が高いということが分かります </S4> まとめを行ないます <S5> 音声認識における... が挙げられます </S5>

図 3: 人間による分割の例

次に、これらを用いてシステムの出力を評価した。評価はセグメント境界の推定と対応付けの両方について行なった。ここで、セグメント境界とはタグ付けされなかった範囲、つまりセグメントの終りから次のセグメントの始まりまでを意味するものとし、評価の際には、システムがその範囲のどこを分割位置と推定しても正しく推定できたものとする。

セグメント境界の推定は再現率、適合率で評価した。それぞれ多少のずれを許容し、以下の式によって計算する。このような評価方法を用いたのは、人間の間でも分割位置の判断がずれることが多いためである。

$$\text{再現率} = \frac{\left( \begin{array}{l} \text{正解の境界のうち、その前後 } g \text{ 文節以内にシ} \\ \text{ステムが推定した境界が含まれていたものの数} \end{array} \right)}{\text{正解の境界の総数}} \quad (5)$$

$$\text{適合率} = \frac{\left( \begin{array}{l} \text{システムが推定した境界のうち、正解の境界} \\ \text{の前後 } g \text{ 文節以内に含まれていたものの数} \end{array} \right)}{\text{システムが推定した境界の総数}} \quad (6)$$

$g = 0$  のときは標準的な再現率と適合率の計算式になる。対応付けは二種類の方法で評価した。一つは正しいセグメントの区切りを与えたときの対応付けの精度を調べる方法であり、セグメントの境界が正しく推定できた場合にどの程度正しく対応付けができるかを知る一つの指標になるものである。精度は以下の式で計算する。

$$\text{精度} = \frac{\left( \begin{array}{l} \text{システムが、対応する節番号を正しく推定で} \\ \text{きたセグメントの数} \end{array} \right)}{\text{正解のセグメントの総数}} \quad (7)$$

もう一つは文節単位での対応付けの精度を調べる方法であり、セグメント境界および予稿の対応する節の両方を同時に評価するためのものである。この評価方法では、まずタグが付与されたセグメントを文節に分割し、各文節にそのセグメントが対応していた節をタグ付けする。そして、タグ付けされた全文節に対し、対応する節を正しく推定できた文節の割合を計算して評価する。つまり、精度は以下の式で計算する。

$$\text{精度} = \frac{\left( \begin{array}{l} \text{システムが、対応する節番号を正しく推定で} \\ \text{きた文節の数} \end{array} \right)}{\text{正解においてタグ付けされた文節の総数}} \quad (8)$$

### 3.4 実験結果

テスト用の 50 講演の書き起こしに対し、セグメント境界の推定と予稿との対応付けを行なった。そして、二人の被験者によるタグ付け結果をそれぞれ正解 1、正解 2 とし、それぞれに対する精度を調べた。その結果を表 1～表 3 にあげる。

表 3: 対応付けの精度

	正解 1		正解 2	
	正しい区切り	文節単位	正しい区切り	文節単位
closed (10 講演)	89.9%	81.6%	88.8%	80.6%
open (50 講演)	71.0%	69.3%	70.6%	67.6%
正解 1 (60 講演)			88.4%	
正解 2 (60 講演)				
set1 (23 講演)	63.2%	62.6%	65.4%	65.3%
set2 (27 講演)	77.5%	69.6%	75.1%	75.0%
set3 (22 講演)	62.9%	64.3%	63.1%	61.7%
set4 (28 講演)	77.3%	73.2%	76.6%	72.3%

ちなみに、正しい区切りが与えられたとき、1 から昇順に番号を付与していくという方法では、正解率は正解 1、正解 2 それぞれに対し、90.3%と 88.0%であった。つまり、全体の約 9 割が予稿の節の順番そのままに講演していたことになる。

### 3.5 考察

3.4 節の実験結果から分かるように、パラメータ調整に用いた 10 講演に比べてテストに用いた 50 講演に対する精度はかなり低くなっている。50 講演に対する精度を詳しく調べたところ、対応付けの精度が文節単位で 100%に近いものもあれば、30%程度のももあった。これは学習データが少なく、過学習になったためであると考えられる。過学習を避けるためには、60 講演を 6 分割して 10 講演分ずつテストに使い、残りの 50 講演でパラメータ調整を行なうというクロスバリデーションで評価した方がよいだろう。今回の実験では、少ない学習データからどの程度パラメータ調整が可能かを調べるためにこのような方法はとらなかった。もしクロスバリデーションにより評価すれば、表 1～表 3 における closed と open の中間くらいの値が得られるのではないかと予想している。

対応付けの誤りとしては、システムが細かく対応付けし過ぎていた部分が目立った。明らかな間違いもあるが、細かく対応付けするとしてもよいと思われる部分もいくつか見受けられた。例えば、講演の冒頭は全体の流れについて説明されることが多く、講演の途中では、一部分だけ別の節の内容が言及されるときもあり、さらにその内容がどちらの節の内容とも言えそうな場合もある。システムはこのような部分を細かく分割する傾向にある。被験者にはまとめて一つのセグメントとしてももらったが、実際はシステムが分割したように、もう少し細かく分割してもよいと考えている。今後、細かい節の対応付けを行なうことも考えており、その際このような問題も考慮するようにしたい。

セグメント境界の推定誤りとしては、上に述べたように対応付けが微妙だと思われるところの他、意味的に不自然なところで区切られている場合が多かった。このようなずれに対しては、名詞連続の途中では区切らないようにする、係り受けの関係にある部分を推定し係り受けの内側では区切らないようにするなどの制約を加える必要があると考えている。また、韻律情報やポーズなど話し言葉に特有の現象を積極的に利用する必要もあると考えており、現在実験を進めているところである。

今回実験対象とした講演には、予稿と完全に対応のとれないようなものはなかったが、人間が対応の付け難かったものと付け易かったもの、二人の被験者の間で食い違いが顕著であったものとそうでなかったものがあつた。そこで、そのそれそれに対する精度を調べた。まず、被験者に、テストに用いた 50 講演を、対応を付け難かったものと付け易かったものに分けてもらったところ、それぞれ 23 講演、27 講演となった。それぞれに対する精度を表 1～表 3 の set1 および set2 にあげる。次に、正解 1 と正解 2 の間で文節単位での対応が 90%以下のものと 90%を越えるものを調べたところ、それぞれ 22 講演、28 講演であった。それぞれに対する精度を表 1～表 3 の set3 および set4 にあげる。ちなみに set1 と set3、set2 と set4 に共通する講演はそれぞれ、11 講演、16 講演であった。表の結果から分かるように、人間にとっても対応付けが易しかったものや人間の間で食い違いが少ないものに対しては、対応付けが難しかったものや食い違いが多かったものに比べてかなり良い精度でセグメント境界の推定や対応付けができていた。

予稿と対応していない場合でも、使われる単語特に専門用語などは予稿のものと同相があると考えられる。また、予稿の各節はまとまった内容を持っているため、各節における単語の共起の情報は語彙的な結束性を計算する上でも有効に働くと考えられる。こういった情報を従来のテキスト分割の手法に採り入れることにより、予稿と対応していない場合でもより良い精度のものができると考えている。他の手法との併用は今後の課題としたい。

表 1: セグメント境界の推定精度: 正解 1 に対して (再現率 (%), 適合率 (%), F-measure)

	$g=0$	$g=5$	$g=10$	$g=15$	$g=20$
closed (10 講演)	19.0, 16.0, 17.3	45.0, 38.7, 40.9	47.0, 40.1, 42.6	65.7, 56.3, 59.8	69.0, 61.3, 63.8
open (50 講演)	15.1, 12.8, 13.4	36.9, 31.5, 32.4	47.5, 40.7, 41.4	51.1, 43.7, 44.5	54.8, 47.0, 47.9
正解 2 (60 講演)	59.5, 62.0, 60.3	67.6, 70.1, 68.4	71.2, 73.8, 72.1	73.7, 76.3, 74.6	75.4, 78.6, 76.4
set1 (23 講演)	8.7, 8.2, 7.9	32.0, 30.9, 29.0	39.1, 38.7, 35.0	40.4, 41.4, 37.2	45.4, 45.1, 41.3
set2 (27 講演)	20.5, 16.8, 18.1	41.0, 32.1, 35.3	54.7, 42.4, 46.8	60.1, 45.7, 50.8	62.9, 48.5, 53.6
set3 (22 講演)	8.9, 6.3, 7.2	35.0, 25.5, 28.7	43.5, 35.1, 36.4	47.7, 40.4, 41.4	48.4, 42.6, 43.0
set4 (28 講演)	19.9, 18.0, 18.3	38.4, 36.3, 35.4	50.7, 45.2, 45.3	53.7, 46.3, 47.0	59.9, 50.4, 51.7

表 2: セグメント境界の推定精度: 正解 2 に対して (再現率 (%), 適合率 (%), F-measure)

	$g=0$	$g=5$	$g=10$	$g=15$	$g=20$
closed (10 講演)	19.0, 14.5, 16.2	42.5, 34.2, 36.8	50.3, 42.0, 44.3	70.7, 58.2, 62.0	70.7, 58.2, 62.0
open (50 講演)	13.1, 11.2, 11.8	32.2, 27.1, 28.0	41.3, 35.4, 35.8	49.0, 40.8, 42.0	54.7, 44.6, 46.3
正解 1 (60 講演)	58.4, 56.8, 57.3	72.2, 69.8, 70.6	74.2, 71.6, 72.4	76.1, 73.5, 74.4	77.8, 76.0, 76.5
set1 (23 講演)	8.0, 7.4, 7.4	25.7, 24.6, 23.0	33.6, 32.9, 29.4	37.9, 36.6, 33.3	47.0, 41.6, 39.4
set2 (27 講演)	17.5, 14.5, 15.5	37.8, 29.2, 32.3	47.9, 37.6, 41.3	58.5, 44.4, 49.4	61.3, 47.2, 52.2
set3 (22 講演)	6.6, 4.5, 5.2	28.4, 19.8, 22.8	41.7, 32.4, 33.9	49.1, 38.6, 40.5	51.7, 42.0, 43.5
set4 (28 講演)	18.3, 16.5, 16.9	35.2, 32.8, 32.2	41.0, 37.9, 37.4	49.0, 42.5, 43.2	57.0, 46.6, 48.5

## 4 関連研究

我々は本稿で予稿と書き起こしを対応付けながら同時にテキスト分割をする方法を提案した。李ら [2] は、セグメントとそのセグメントを代表するトピックの同定を同時に行なう統計モデルを提案しており、トピックを予稿の節と考えれば、彼らのタスクは我々のものと近い。しかし、彼らの手法では統計モデルを採用しているため大量の学習コーパスを必要とし、各講演に対し一つの予稿の情報からしか学習できない我々のタスクにそのまま適用するのは難しいと考えられる。また、彼らのタスクは書き言葉におけるテキスト分割であり、文の区切りなど様々な条件が我々のものとは異なる。

対応付けの研究としては、対訳文の対応付け [3][4] や TV ニュースと新聞記事との対応付け [5] などがある。対訳文の対応付けの研究では、対応付けすべき文がほぼ対応していることが前提となっている。一方、予稿と書き起こしは、内容の一部が対応することがあるが、文の対応がとれることは稀で、順序も変わることが多い。このように対応がとれないものについてもある程度のまとまりで区切ることができる点が本手法の利点であると考えている。TV ニュースと新聞記事の関係は予稿と書き起こしの関係と似ているので、それらの対応付けに本稿で提案した手法が利用できると考えられる。渡辺らの研究 [5] では TV ニュースの映像中の短いテキストのみを用いており、我々が対象にしたような長いテキスト同士の対応付けは行っていない。

テキスト分割の研究としては、文と文あるいはテキストとテキストの類似度を計算する方法 [6][7][8][9][10] やテキストの境界を学習する方法、[11] 表層的な手がかりを用いる方法 [12] などがある。しかし、類似度の計算や学習にはいずれも大規模な辞書や学習データが必要であり、今回のタスクにそのまま適用するのは難しいと考えられる。中でも辞書を用いる方法は今回のタスクに有効であると思われるが、今回のような専門用語が多く見られるような講演が対象の場合には、未知語が多くてその方法のみでは良い精度は期待できないと思われる。表層的な手がかりは今回我々が対象とした講演に対しても非常に有効に働くと考えられる。こういった辞書や表層的な手がかりを我々の手法にうまく取り込むことができれば精度向上が期待できるだろう。これは今後の課題である。

## 5 まとめ

本稿では、予稿の各節に含まれる文字列を辞書情報として用いて講演の書き起こしを解析することにより、書き起こしと予稿の各節とを対応付けながら同時にその対応付けの結果を基にして書き起こしをセグメントに分割する方法について述べた。予稿と書き起こしの間では、

予稿ではアルファベットなのに書き起こしではカタカナになっている部分があるなど表記の違いがある。それにも関わらず、実験では対応付けがうまくできるものが多かった。この結果から、入力となる音声認識結果が多少誤りを含んでいてもある程度対応付けることができると推測される。また逆に、このように予稿との対応をとり、対応している予稿の節に含まれる単語を優先的に認識単語の候補とするような仕組みを音声認識の手法に組み込めば、音声認識における精度向上も期待できると考えている。

講演では OHP やスライドなどが補助に使われることが多い。OHP やスライドなどは予稿の内容を反映していることが多いので、これらの情報と我々の手法とを併用すればより精度良く対応付けやセグメントへの分割ができるようになるだろう。

講演は韻律情報やポーズの情報など書き言葉には現れない現象を伴う。また、講演では、「それでは」、「次に」、「続きまして」、「で」、「最後にまとめます」など特有の表現が使われ、その前後で異なる内容が話されることが多い。これらは本稿では扱わなかったが、かなり有効な情報として利用できると考えている。今後、このような話し言葉に特有の現象を積極的に利用したい。

## 参考文献

- [1] 開放的融合研究制度. 「話し言葉の言語的・パラ言語的構造の解明に基づく話し言葉工学」の構築. <http://www2.crl.go.jp/pub/orc-speech/>, 1999.
- [2] Hang Li and Kenji Yamanishi. Topic Analysis Using a Finite Mixture Model. In *Proceedings of EMNLP/VLC*, 2000.
- [3] Takehito Utsuro, Hiroshi Ikeda, Masaya Yamane, Yuji Matsumoto, and Makoto Nagao. Bilingual Text Matching Using Bilingual Dictionary and Statistics. In *Proceedings of COLING*, pp. 1076-1082, 1994.
- [4] Masahiko Haruno and Takefumi Yamazaki. High-Performance Bilingual Text Alignment Using Statistical and Dictionary Information. In *Proceedings of ACL*, pp. 131-138, 1996.
- [5] 渡辺増彦, 岡田至弘, 角田達彦, 長尾真. Tv ニュースと新聞記事の対応付け. 人工知能学会, Vol. 12, No. 6, pp. 921-927, 1997.
- [6] Jane Morris and Graeme Hirst. Lexical Cohesion Computed by Thesaural Relation as an Indicator of the Structure of Text. *Computational Linguistics*, Vol. 17, No. 1, pp. 21-48, 1991.
- [7] Hideki Kozima. Text Segmentation Based on Similarity between Words. In *Proceedings of ACL*, pp. 286-288, 1993.
- [8] Marti A. Hearst. Multi-Paragraph Segmentation of Expository Text. In *Proceedings of ACL*, pp. 9-16, 1994.
- [9] 仲尾由雄. 語彙的結束性に基づく話題の階層構成の認定. 自然言語処理, Vol. 6, No. 6, pp. 83-112, 1999.
- [10] Jay M. Ponte and W. Bruce Croft. Text Segmentation by Topic. In *Proceedings of the First European Conference Research and Advanced Technology for on Digital Linguistics*, pp. 120-129, 1997.
- [11] Doug Beeferman, Adam Berger, and John Lafferty. Statistical Models for Text Segmentation. *Machine Learning*, Vol. 34, pp. 177-210, 1999.
- [12] 望月源, 本田岳夫, 奥村学. 複数の表層の手がかりを統合したテキストセグメンテーション. 自然言語処理, Vol. 6, No. 3, pp. 43-58, 1999.