

談話標識の統計的選択に基づいたCSJの講演からの重要文抽出

南條 浩輝[†] 北出 祐[†] 河原 達也[†]

[†] 京都大学 情報学研究科 知能情報学専攻
〒606-8501 京都市左京区吉田二本松町
E-mail: †nanjo@ar.media.kyoto-u.ac.jp

あらまし 講演の自動インデキシングを目的として、談話標識の統計的選択に基づく手法をCSJ(日本語話し言葉コーパス)の重要文セットに適用した結果について報告する。本手法は、学会講演特有の談話構造の境界に類出する談話標識を自動的に求めた上で、これに基づく統計的な重要度尺度を定義するものである。さらに話題語(キーワード)の統計量の基づく重要度尺度と統合も行う。CSJの19件の学会講演を対象に重要文抽出精度の評価を行ない、(1)談話構造に基づく手法が有効であること、(2)話題語に基づく手法と統合することで相乗効果が得られること、(3)提案手法が音声認識誤りに対して頑健であること、(4)人間による抽出精度と比較して10-15%の低下であること、を確認した。また、文区切りの精度が重要文抽出に影響が大きいことが判明したので、韻律情報を用いることで文区切り精度および重要文抽出精度の改善を行なった。

キーワード 講演音声, 重要文抽出, 談話標識, 教師なし学習

Automatic Extraction of Key Sentences from CSJ Lectures using Statistics of Presumed Discourse Markers

Hiroaki NANJO[†], Tasuku KITADE[†], and Tatsuya KAWAHARA[†]

[†] Graduate School of Informatics, Kyoto University,
Yoshida-Nihonmatsu-cho, Sakyo-ku, Kyoto 606-8501 Japan
E-mail: †nanjo@ar.media.kyoto-u.ac.jp

Abstract Automatic extraction of key sentences from lecture audio archives is addressed. The method makes use of the characteristic expressions used in initial utterances of sections, which are defined as discourse markers and derived in a totally unsupervised manner based on word statistics. The statistics of the presumed discourse markers are then used to define the importance of the sentences. It is also combined with the conventional tf-idf measure of content words. Experimental results using lectures of CSJ (Corpus of Spontaneous Japanese) confirm the effectiveness of the method based on the discourse markers and its combination with the keyword-based method. It is also shown that the method is robust against ASR errors and sentence segmentation accuracy is more vital. Thus, we enhance the segmentation using prosodic information.

Key words spontaneous presentation speech, automatic key sentence extraction, discourse marker, unsupervised training

1. はじめに

近年の計算機性能の向上やメディア処理技術の進展に伴い、講演や討論などの音声をデジタルアーカイブとして保存できる環境が整ってきている。しかし、音声のデジタルアーカイブはテキストとは異なり、そのままでは目的とする情報を得ることや短時間で全体の内容を把握することが困難である。したがって、内容を把握する上で、インデックスや要約などの二次

情報をアーカイブに付与することが必要不可欠である。これらの情報を人手で付与するには膨大な時間と手間を要するため、自動的に付与できることが望まれている。

テキストを対象とした重要文抽出や要約の研究はこれまで数多く行なわれている。このような新聞記事や論文を対象にした重要文抽出においては、セクション構造などの位置情報が有用であることが知られている。これは、それぞれの記事やパラグラフの冒頭にはその話題を説明する重要文が含まれていること

表1 重要文抽出における文の一致度

Table 1 Agreement among subjects in extraction of key sentences

	任意の2名	3名
50%抽出 (全体における抽出比率)	75% (37.5%)	60% (30.0%)
10%抽出 (全体における抽出比率)	45% (4.5%)	30% (3.0%)

が多いためである。しかし、音声ではパラグラフやセクションの境界が明確でないため、このようなアプローチを適用することは容易ではない。また、日本語の話し言葉を対象として重要文抽出を試みた研究自体あまり行なわれておらず [1] [2]、各研究機関での評価データや評価方法が異なるため、比較が困難であった。その要因として、利用可能な大規模コーパスがこれまで整備されていなかったことが挙げられる。これに対して「話し言葉工学」プロジェクトにより整備が進められている「日本語話し言葉コーパス (CSJ)」[3] [4] の一部には重要文の抽出作業が行われており、3名による重要文タグが付与されている。本稿では、この重要文タグが付与された CSJ の学会講演を用いて、人間同士の重要文選択の一致度について調べた結果を報告し、我々の提案する重要文の自動抽出手法 [5] を適用した結果について述べる。

2. コーパスとシステムの概要

2.1 日本語話し言葉コーパス (CSJ) における重要文セット

CSJ は主に学会講演と模擬的な講演からなるコーパスであり、音声データと人手による書き起こしテキストから構成される。CSJ の講演のうち、コアとよばれる一部の講演には複数人による重要文タグが付与されている。講演ごとに3人の作業者が割り当てられており、全体の文の50%と10%を目安に重要文の抽出が行われている。作業者は研究者であり、学会講演の発表スタイルに精通しているが、必ずしもそれぞれのテストセットの分野における専門家ではない。

本稿では、2003年8月時点で重要文タグが付与された CSJ の学会講演 19 講演を用いて実験・評価を行う。まず、作業者間での重要文抽出の一致率を調べた。結果を表1に示す。50%抽出では比較的高い一致率がみられるが、10%抽出での一致率は半分以下であり、人間同士でもあまり一致しないことがわかる。本研究では、人間の作業者間で比較的高い一致率がみられた50%抽出で重要文の自動抽出の実験を行い、提案手法の評価を行う。

2.2 処理の概要

図1に重要文抽出の処理の概要を示す。まず、講演の音声を認識して書き起こしを生成する。次に、話し言葉の音声（の認識結果）は読み上げ音声とは異なり、ポーズで区切るだけでは文単位にならないため、言語的情報とポーズ情報を用いて文に分割する。次に分割された各文に対し、重要度を付与する。最後に各文の重要度に基づいて、指定した抽出率（全文に対する割合）または文数に達するまで重要文を抽出する。これらの処理を以下に詳細に説明する。

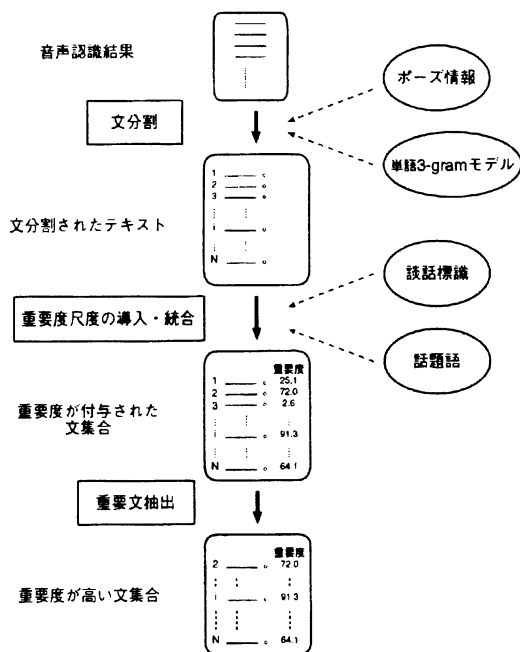


図1 重要文抽出処理の概観

Fig. 1 Flowchart of key sentence extraction

2.2.1 音声認識システム

本研究では、以下に示す音響・言語モデルと認識エンジン Julius rev.3.3p3 で構成した音声認識システムを用いる。音響モデルには、CSJ の 2496 講演 (486 時間) で学習した状態共有型 triphone モデル (3000 状態) を用いる。言語モデルには、CSJ の 2592 講演 (6.7M 形態素) から学習した単語 3-gram モデル (語彙サイズ: 24437) を用いた [6]。

本研究では形態素単位として、国立国語研究所で定義された短単位 [7] を使用し、形態素解析システムは、通信総合研究所で最大エントロピー法により CSJ を用いて統計的に学習されたものを用いる [8]。

2.2.2 文単位への分割

日本語の話し言葉においては、文の定義・境界が曖昧であるためこれを定義する必要がある。「話し言葉工学」プロジェクトでは、原則として「です」「ます」のような文末表現と「～けれども」「～が」「～まして」「～でして」「～し」などの箇所のうち機能的に文を区切っていると判断できる箇所を区切ると定められている。本研究でも、これを文と定義する。

従来の音声認識では、長いポーズを句点に、短いポーズを読点に割り当てるのが一般であったが、話し言葉においては、必ずしもポーズが句読点と一致しないため、文への分割が大きな問題となる。また、話し言葉には「～と」のような独特の文末表現が存在するため、パターンマッチング的に規則を適用して句読点を挿入することも難しい。

本研究では、話し言葉と書き言葉を別の言語とみなした上で統計的機械翻訳により句読点挿入を行う手法 [9] を採用する。こ

これは、式(1)に示すように、ポーズを含む話し言葉の文字列 X に対して、それを最もよく説明する書き言葉の(句読点を含む)文字列 Y を求める問題として定式化するものである。具体的には、ポーズが句読点に変換されうる ($P(X|Y) = 1$ となる) 全ての箇所に対して、句読点を挿入する場合としない場合の言語モデルのゆう度 $P(Y)$ を比較し、句読点挿入の判定を行う¹⁾。

$$\max_Y P(Y|X) = \max_Y \frac{P(Y)P(X|Y)}{P(X)} \quad (1)$$

言語モデルのゆう度 $P(Y)$ の計算には、書き言葉のコーパスから学習された単語 3-gram モデルを使用する。すなわち、 $Y = (y_1 \dots y_N)$ について、 $P(Y) = \prod_{i=1}^N P(y_i | y_{i-1}, y_{i-2})$ として求める。変換モデル $P(X|Y)$ には、ポーズ前後の表現とポーズ長に依存するモデルを用いた。具体的には、典型的な文末表現(「～ます」など)では、短いポーズでも句点が挿入しうる ($P(X|Y) = 1$) が、話し言葉特有の文末表現「～と」「～ない」「で～」及び「～た」においては、平均ポーズ長以上の場合のみに挿入しうるとした。

3. 談話標識の単語頻度及び文頻度に基づく重要度

3.1 学会講演の談話構造と区分化

一般に学会講演では、論文などにみられる章構造と同様の談話構造が観察される。すなわち、いくつかの論点が論理的に順序付けられて展開していくという特徴がある。学会講演は話題を問わず、大きく緒論、本論、結論に分けられ、さらに多くの場合、緒論は背景と目的に、本論は手法の説明と実験・評価に分けることができる。これらは、講演においては1～数枚程度のスライドに対応するまとまった話題の単位に該当し、これをセクションと定義する。これらのセクションの冒頭では、典型的な表現(例:「次に本研究の手法を説明いたしますが、これは…」、「今回の評価実験としまして…」)が多く用いられる。これは発表者は次に話すことを簡潔に伝え、聴取者の注意を引こうとするためと考えられる。このことは、セクション冒頭部において、重要な発話がなされることが多いことを示している。

このようなセクション境界情報は、テキストベースの自然言語処理において重要文を抽出する上で有効であることが知られている。テキストベースでの処理では、改行や字下げなどの明示的な情報を利用してセクション境界を容易に決定できるが、音声にはそれらに相当するものが存在しない。これに対して、本研究では、このようなセクション冒頭における典型的な表現を談話標識と定義し、これらに着目してセクション境界を自動検出することで、重要文を抽出する。従来研究では、談話標識は言語的分析に基づいて人手により設定されるのが一般的であったが[10]、本研究では、談話標識をコーパスから自動的に抽出する。

(注1) 実際には、一般的に単語数が増加するゆう度が低下するため、単語数に対する重み(音声認識のデコーダで用いられる単語挿入ハザルティに相当)を考慮している。

表2 学会講演から推定された談話標識(上位60)

Table 2 Discourse Markers of Academic Presentation Speech (Top 60)

こと	もの	次	結果	研究	語
方	実験	場合	説明	訳	先程
音声	モデル	問題	学習	ところ	今
データ	中	情報	時	実際	評価
文	単語	認識	システム	関係	発表
例	音	為	意味	方法	後
分析	形	以上	大学	数	発話
表現	部分	今回	時間	話	構造
人	話者	目的	点	特徴	最後
言語	率	凶	手法	解析	最初

3.2 談話標識に基づく重要度

談話標識の学習は、以下のように行う[11]。この談話標識の抽出は完全に教師なしで行われ、人手による事前のタグづけは一切必要としない。

(1) 学習データからセクション境界候補を抽出する。セクション境界候補は、各講演において平均ポーズ長より長いポーズが存在する箇所とする。これは、話題の転換点では、スライドを変えろといった作業や息継ぎなどの目的で、比較的長めのポーズが挿入されるという仮定に基づいている。

(2) 得られたセクション境界候補の直後の文を選択し、セクション冒頭の文集合を抽出する。

(3) セクション冒頭文に特徴的な単語を談話標識として抽出する。具体的には、以下の単語頻度と文頻度に基づく統計量 DM に基づいて選択する。

$$DM_m = wf_m * \log \left(\frac{N_s}{sf_m} \right) \quad (2)$$

ここで、単語頻度 wf_m は、セクション冒頭の文の集合において単語 m が出現する回数であり、文頻度 sf_m は、学習セットの全ての文(総数 N_s)で、単語 m が出現する文数である。セクション境界の先頭部分に頻出し、それ以外の箇所ではあまり出現しない談話標識に対して、この値 DM は大きな値となる。ここでは、CSJの学会講演930講演を用いて統計値を学習した。談話標識に選定する対象として、予備実験を行った結果、名詞(23170種)に限定することにした。

この談話標識の統計値を利用して各文 s_j に重要度を付与する。すなわち各文に出現する異なる全ての談話標識の統計値の合計値(式(3))はセクション境界らしさ(=重要度)を表すので、この評価値が高いものを重要文として抽出する。

$$S_{DM}(j) = \sum_{m_i \in s_j} DM_{m_i} \quad (3)$$

表2に学会講演から推定された談話標識(DMの値が高いもの上位60)を示す。「次」や「結果」、「研究」、「実験」、「先程」、「評価」、「発表」などの学会講演のセクションの先頭に頻出する単語が多く含まれていることがわかる。これは、談話標識を用いることでセクション境界を推定できることを示している。

3.3 話題語に基づく重要度との統合

次に、話題語の統計量を用いた重要度について述べる。これは、話題と関連のある単語（話題語）は当該講演において繰り返し出現すると仮定して、話題語を多く含む文を重要文として抽出するものである。

話題語の統計量としては、式(4)で定義される単語（名詞）の $tf \cdot idf$ 値を用いる。

$$KW_m = tf_m * \log \left(\frac{N_d}{df_m} \right) \quad (4)$$

tf_m は名詞 m の当該講演内での出現回数を表わし、 df_m は名詞 m が出現した講演数を表す。全講演数 N_d を df 値で除いたものが idf 値である。各名詞に対する統計値 df_m は、談話標識の統計値を求めたものと同一の 930 講演から求めている。当該講演を特徴づける語（話題語）に対しては、この $tf \cdot idf$ 値は大きくなる。ここでは、単純に単語そのものを扱うのではなく、当該講演において連続して 3 回以上出現する名詞列を複合語として扱う。これは、より話題を明確に示す語を定義するためである。例えば、「音声」「認識」という 2 単語が連続して出現した場合、これらを複合名詞「音声認識」の 1 単語とみなすことで、「音声」や「認識」を単独で用いる場合に比べて、話題をより明確に示す単語となる。

話題語に基づく重要度は各文 s_j に含まれる名詞の $tf \cdot idf$ 値の合計値（式(5)）とする。

$$S_{KW}(j) = \sum_{m_i \in s_j} KW_{m_i} \quad (5)$$

次に、話題語に基づく重要度と談話標識に基づく重要度とを統合する。具体的には、それぞれの手法により得られる各文の重要度（ $S_{KW}(j)$ 、 $S_{DM}(j)$ ）の重みつき幾何平均をとり、新たな重要度とする。

$$S_{KW}(j)^\alpha * S_{DM}(j)^{(1-\alpha)} \quad (6)$$

本実験では、この重み係数 α の値を事後的に 0.6 に決定しているが、この値を極端な値に設定した場合を除いて、最終的な重要文抽出の結果に大きな差はない。

4. 評価実験

2.1 節で述べた重要文タグが付与された CSJ の学会講演 19 講演を用いて、重要文の自動抽出の評価（抽出率 50%）を行う。

まず、評価を行う際に正解とする重要文の設定と評価法について述べる。作業員間で重要文として抽出される文が異なるため、正解とする重要文は、3 人から任意の 2 人を選び、その 2 人がともに 50% の重要文抽出において抽出した文とする。評価は、3 通りの正解それぞれに対する再現率（recall）、適合率（precision）、F 値（F-measure）を算出し、その平均で行う。ここで、F 値は以下の式(7)で定義する。

$$F\text{-measure} = \frac{2 * recall * precision}{recall + precision} \quad (7)$$

この 2 人一致による正解重要文を用いて、他の 1 人の重要文

表 3 各手法による書き起こしからの重要文抽出

Table 3 Extraction of key sentences using 3 methods for manual transcriptions

	再現率	適合率	F 値
冒頭・末尾	0.521	0.391	0.447
談話標識 (DM)	0.710	0.533	0.609
話題語 (KW)	0.717	0.538	0.614
KW+DM	0.741	0.556	0.635
人間	0.832	0.627	0.715

表 4 音声認識結果からの重要文抽出

Table 4 Extraction of key sentences for ASR results

	テキスト	文分割	再現率	適合率	F 値
(1)	書き起こし	人手	0.740	0.555	0.635
(2)	書き起こし	自動	0.731	0.458	0.563
(3)	音声認識結果	自動	0.727	0.456	0.561

抽出の精度（人間の抽出精度）を算出してその平均を求めたところ、再現率 83.2%、適合率 62.7%、F 値 0.715 であった。

談話標識に基づく提案方法（DM）と話題語に基づく方法（KW）、およびそれらの統合手法（DM+KW）による書き起こしからの重要文抽出の結果を表 3 に示す。談話標識に基づくセクション境界情報を用いた手法（DM）により F 値 0.609 が得られた。一方、講演全体の冒頭と末尾を抽出した場合（表 3 の冒頭・末尾）は、F 値は 0.447 であった。これは、セクション境界を用いることの有効性を示すものである。提案手法（DM）は話題語を用いた手法（KW）と同程度の重要文抽出精度であったが、両者の統合（DM+KW）の効果が見られ、再現率 74.1%、適合率 55.6%、F 値 0.635 となった。ただし人間同士の一致率に比べると 10% 程度精度が低く、改善の余地があるといえる。

4.1 音声認識結果への適用

次に、音声認識結果に対して適用を行った。音声認識結果に対しては、2.2.2 章で述べた手法で文に分割を行った上で、重要文を抽出する。

重要文抽出の評価は文の区切りが一致しないと行えないため、ここでは、ある文を抽出した場合に、それが対応する（複数の）人手文区切りにおける文を抽出したとみなして評価を行う。

音声認識結果への適用に先だって、書き起こしを用いて自動文区切りの影響を調べた。結果を表 4 に示す。人手で文に分割した場合（表 4 の (1)）と比較すると、自動文区切りを用いた場合（表 4 の (2)）では、適合率が 10% 程度低下している。これは、正しく句点が挿入されなかったために長い文が生成され、本来の非重要文とあわせて抽出した結果と考えられる。一方、再現率には大きな差が見られず、抽出すべき箇所は正しく抽出できていることを示していると考えられる。

音声認識結果からの重要文抽出の結果も表 4 の (3) に示している。なお、テスト講演ごとの結果は表 5 にまとめてある。書き起こしに対して自動で文区切りを行った場合（表 4 の (2)）と精度の差がみられず、提案手法は認識誤りに対して頑健であることと、文分割の精度が重要文抽出に影響が大きいことがわ

表5 テスト講演ごとの音声認識率(%)、文分割精度(F値)、重要文抽出精度(F値)

Table 5 List of test-set lectures with speech recognition accuracy, segmentation performance, and indexing performance

講演ID	認識率	文分割	重要文抽出
A01M0056	85.15%	0.821	0.458
A01M0096	91.21%	0.812	0.567
A01M0151	92.21%	0.920	0.656
A01M0035	64.95%	0.505	0.529
A01M0007	78.32%	0.613	0.533
A01F0001	77.56%	0.851	0.559
A01M0025	92.18%	0.878	0.671
A01M0110	86.15%	0.915	0.598
A01F0132	87.15%	0.794	0.495
A01M0083	91.35%	0.822	0.580
A01M0137	72.74%	0.740	0.561
A01M0074	80.54%	0.745	0.484
A01M0097	84.76%	0.844	0.536
A03M0112	81.41%	0.912	0.630
A03M0106	61.37%	0.720	0.489
A03F0072	71.31%	0.735	0.591
A05M0031	74.68%	0.783	0.629
A06M0134	68.58%	0.643	0.606
YG99JUN001	69.17%	0.512	0.501
total	76.99%	0.740	0.561

かった。

5. 韻律的特徴の利用

文分割の精度が重要文抽出に影響が大きいので、韻律情報を利用して文分割の精度の改善を図った。また、韻律情報を用いたセクション境界検出についても検討した。

2.2.2章で述べたように、本研究で用いている句点挿入アルゴリズムは、基本的にポーズを句読点に変換するものである。したがって、文末であってもポーズがない場合は、句点を挿入できなかった。これに対して、文末候補はポーズが存在しなくても区点に変換されるという変換モデルを用いることを考える。このモデル化は湧き出し誤りを増加させて句読点挿入の適合率の低下を招くため、韻律的特徴である基本周波数(F0)を導入する。

ベースラインでは区点を挿入する場合としない場合のそれぞれの言語スコア(式(1)の $P(Y)$ の対数をとったもの)を比較して区点挿入を判定していたが、韻律情報を利用する場合は、それぞれの言語スコアに式(8)で定義する韻律スコアを加えた上で比較して区点挿入を判定する。

$$a * F0_{avr_diff} - b \quad (8)$$

本研究で用いる韻律スコアは、文境界候補の前後でのF0平均の差に重みを乗じて定数を加えたものである。これは、日本語では文の先頭でF0が大きく立ち上がるという知見に基づくものである。ただし、ベースラインの句点挿入アルゴリズムと同様に、話し言葉特有の文末・文頭パターン「〜と」「〜ない」

表6 韻律情報を用いた重要文抽出の結果(音声認識結果からの重要文抽出)

Table 6 Results of key sentence extraction incorporating prosodic features (ASR results)

重要文抽出手法	文分割手法	文分割精度	重要文抽出精度
DM+KW	ベースライン	0.740	0.561
DM+KW	提案手法	0.759	0.583
DM+KW+PROSODY	提案手法	0.759	0.592

「で〜」及び「〜た」の箇所に対しては、特別な処理を行なう。具体的には、これらの箇所でポーズ長が平均値以下の場合には、韻律スコアを2倍している。

次に、韻律情報を利用したセクション境界の検出に基づく重要度についても検討を行った。本研究では、韻律情報としてF0およびパワーの立ち上がりを用いた。これは、発表者は聴衆の注意を引くために、セクション境界では比較的強調した発話をするという仮定に基づくものである。具体的には以下に従って重要度を計算する。

(1) 文頭でのF0とパワー(PW)の立ち上がりを計算し、平均と分散の値で正規化する(式(9),(10))。

(2) それらの和を0から1の間に正規化し(式(11),(12))、各文のセクション境界ゆう度 P_j とする。

(3) 重要文はセクション境界前後にあると仮定して、前後の文のセクション境界ゆう度の重みづけ線形和を取り、各文 s_j の重要度 $S_{PR}(j)$ を求める(式(13))。

$$F0_j = (F0_{onset_j} - F0_{ave}) / F0_{sd} \quad (9)$$

$$PW_j = (PW_{onset_j} - PW_{ave}) / PW_{sd} \quad (10)$$

$$P_j = \text{sigmoid}(F0_j + PW_j) \quad (11)$$

$$\text{ただし、} \text{sigmoid}(x) = \frac{1}{1 + \exp(-x/4)} \quad (12)$$

$$S_{PR}(j) = \frac{P_{j-2}}{3} + \frac{P_{j-1}}{2} + P_j + P_{j+1} + \frac{P_{j+2}}{2} + \frac{P_{j+3}}{3} \quad (13)$$

この重要度と前述の言語情報(談話標識と話題語)に基づく重要度とを統合(重みつき幾何平均)し、新しい重要度と定義する。

結果を表6に示す。韻律的特徴を用いることで文分割精度が向上したのに伴い(0.740→0.759)、重要文抽出の精度も0.561から0.583に改善された。さらに、韻律情報に基づいた重要度を統合することで重要文抽出精度の改善が得られ、音声認識結果からの重要文抽出においてF値0.592を達成した。

6. 結論

CSJにおいて重要タグが付与されている講演セットを用いて重要文抽出の実験・評価を行った。学会講演のセクション冒頭発話に特有な表現(談話標識)を用いた重要度を定義し、話題語に基づく重要度と統合した。人間の重要文抽出の精度が50%抽

出において 0.715 (F 値) であったのに対し, 自動で行った場合は, 書き起こしからの抽出で F 値 0.635, 音声認識結果からの抽出で F 値 0.561 であった. さらに韻律的特徴の利用を検討し, 言語的情報を用いた手法と組み合わせることで改善が得られた.

謝辞 本研究は, 開放的融合研究「話し言葉工学」プロジェクトの一環としておこなわれた. アドバイスを頂きました東京工業大学の古井貞熙教授をはじめとして, ご協力を頂いた関係各位に感謝いたします.

文 献

- [1] 伊藤山彦, 松本賢司, 谷田泰郎, 柏岡秀紀, 田中英輝. 講演文を対象にした重要文抽出実験. 「話し言葉の科学と工学」ワークショップ講演予稿集, pp.157-164, 2001.
- [2] 笠原力弥, 山下洋一. 講演音声における重要文と韻律的特徴の関係. 情処学研報, 2001-SLP-35-5, 2001.
- [3] 古井貞熙, 前川喜久雄, 井佐原均. 科学技術振興調整費開放的融合研究推進制度—大規模コーパスに基づく「話し言葉工学」の構築—. 日本音響学会誌, Vol.56, No.11, pp.752-755, 2000.
- [4] 前川喜久雄. 言語研究における自発音声. 音講論, 1-3-10, 春季 2001.
- [5] 北出祐, 南條浩輝, 河原達也, 奥乃博. 談話標識と話題語に基づく統計的尺度による講演からの重要文抽出. 情処学研報, 2003-SLP-46-2, 2003.
- [6] T.Kawahara, H.Nanjo, T.Shinozaki, and S.Furui. Benchmark test for speech recognition using the Corpus of Spontaneous Japanese. In *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pp. 135-138, 2003.
- [7] 小椋秀樹. 話し言葉コーパスの単位認定基準について. 「話し言葉の科学と工学」ワークショップ講演予稿集, pp. 21-28, 2001.
- [8] 内元清貴, 井佐原均. 話し言葉コーパスの形態素解析. 「話し言葉の科学と工学」ワークショップ講演予稿集, pp. 33-38, 2002.
- [9] 河原達也, 下岡和也. 統計的手法を用いた講演書き起こしの自動整形. 音講論, 3-9-10, 秋季 2002.
- [10] 小林聡, 吉川裕規, 中川聖一. 表層情報と韻律情報を利用した講演音声の要約. 情処学研報, 2002-SLP-43-7, 2002.
- [11] 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. 情処学論, Vol. 43, No. 7, pp. 2222-2229, 2002.