独話音声コンテンツ化のためのトピックセグメンテーション

伊藤　正詩 †　　大野　誠寛 ††　　松原　茂樹 ††

† 名古屋大学大学院情報科学研究科　†† 名古屋大学情報連携基盤センター

Topic Segmentation for Structuring Monologue Speech

Masashi Ito　　Tomohiro Ohno　　Shigeki Matsubara

†Graduate School of Information Science, Nagoya University
††Information Technology Center, Nagoya University

1　はじめに

近年，講義や講演などの独話音声データの蓄積が盛んであり，蓄積されたデータに対する効率的なアクセスや効果的な再利用が望まれている．しかし現状では，音声データがそのままの形で蓄積されていることがほとんどである．効率的なアクセスや効果的な再利用を実現するために，独話音声データに対して，書き起こしはもちろん，構造化や文体整形などにより，独話音声データをコンテンツ化することが望まれる．

独話音声の自動コンテンツ化に不可欠な要素技術としてトピックセグメンテーションがある．トピックセグメンテーションとは，話題ごとに文書を分割することであり，これまでに数多くの研究が行われている．しかし，提案された手法の多くは書き言葉に対するものであり [1,2,3]，改行や段落などいった話し言葉では明示的に示されない情報を利用しており，独話音声に対する手法としては適さない．一部，話し言葉を対象とした従来研究として，塩崎ら [5] が対話を，長谷川ら [4] が学会講演を対象に，トピックセグメンテーションを行っている．しかし，対話では，一文が短いなど，独話に見られない特徴が存在する．また，学会講演は，一般の講演に比べ整った発話になる傾向があり，さらに,「背景」や「手法」などといった，比較的明瞭なトピックが存在するという特徴がある．このため，従来の提案方式が一般の講演データに適用できる可能性は明らかでない．

そこで本論文では，一般的な講演音声の書き起こしテキストを一定の単位に分割し，それらを結合することによりトピック境界を特定する手法を提案する．本手法では，トピック境界の前後に現れる特徴的な表現を手掛かりに，トピックセグメンテーションを行う．さらに，トピックを特徴付けるキーワードも利用する．講演中のある一部分に集中的に出現する語がトピックを特徴付ける語であると仮定し，そのような語を自動的に抽出する．本手法を実データに適用したところ，精度 0.357，再現率 0.642，F 値 0.499 を得ることができた．

本論文の構成は以下の通りである．2 章では，著者らが提案する独話音声コンテンツの概要を示す．3 章では，コンテンツ化のためのトピックセグメンテーション手法について述べる．4 章では，評価実験について報告する．

で後もう一個メリットがあったのはその会社自体がサイパンとかその近くのココス島っていうところでホテルをやってたんですね

だから現地の駐在員ていうのが会社の社員がいるんですね

だから結構ちょっとローカルスポットに車で案内してくれたりとか後現地の社員もいるんで現地の社員と一緒に遊んだりとか盛りだくさんだったんですね

でまず海が全然大磯と違うっていうことごみが浮いてないっていうことと後はもう本当に透明度がいいっていうのともう本当に色とりどりのピンクとか何かレモンイエローの魚とかがいてそれにこう餌付けができたりとか凄く楽しかったんですね

で後はドライブなんか行っても四駆でこうジャングルみたいなとこに入ってくんですよね

何か日本軍の何かね戦車とかの残骸があるようなところに入っていっちゃったりとか何か本当にOLの貧乏旅行だったんだけれども凄く現地の社員がいたお陰で楽しく過ごせた

と後はもう一個良かったのが社長の別荘っていうのがあるんですね

そこに別邸がで勿論社長は普段そこにいないんでその現地の方のメードさんお手伝いさんと後うちの現地のスタッフが鍵を持っていてもう貧乏旅行なんだけどそこの社長の別邸でちょっと遊べるっていう企画があったんですね

図 1: 講演の音声書き起こしテキスト（一部）

2　独話音声のコンテンツ化

2.1　独話音声の構造化

近年，講義や講演の音声を録音し，蓄積することが多くなり，蓄積したデータを有効に再利用することが望まれている．そのようなデータの多くは音声データであるが，全文検索によるアクセスを想定し，書き起こしテキストが与えられている場合もある．例として，講演の書き起こしテキストの一部を図 1 に示す．図 1 のようなデータは，話者の発言を時系列で記録したものであり，文の列にすぎない．しかし，講演自体は構造を有している．構造化を行うことにより，データが持つ潜在的な構造を明示化し，ユーザによる内容の理解を容易にすることが可能になる．

2.2　独話音声コンテンツの概要

コンテンツ化は，講演音声の書き起こしテキストを構造化・整形し，表示することによって行う．図 2 に，独話音声の構造を示す．これは，トピックをノードとする木構造であり，各トピックには，トピックの内容を示す見出しが付与される．講演

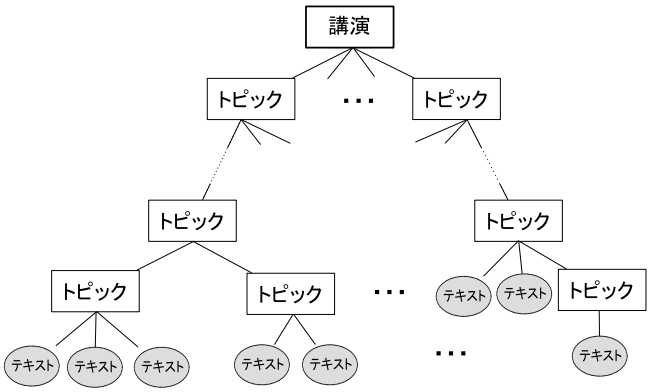

図 2: 独話音声コンテンツの構造

ノードは，いくつかのトピックノードによって形成される．トピックノードは，トピックノードあるいはテキストノードからなる．テキストノードは講演の書き起こしテキストの一部に対応している．

図 1 の書き起こしテキストをコンテンツ化した例を図 3 に示す．本研究では，コンテンツを公開することを前提としているため，コンテンツは Web 上で表示可能な形式で作成する．トピックノードは下線を付与して表示し，テキストノードと区別する．テキストノードは，講演音声の書き起こしテキストを整形したものであり，箇条書きで記述することにより効率的な内容理解を可能にする．図 3 では，図 1 において点線で囲まれたトピックとその子ノードを表示している．

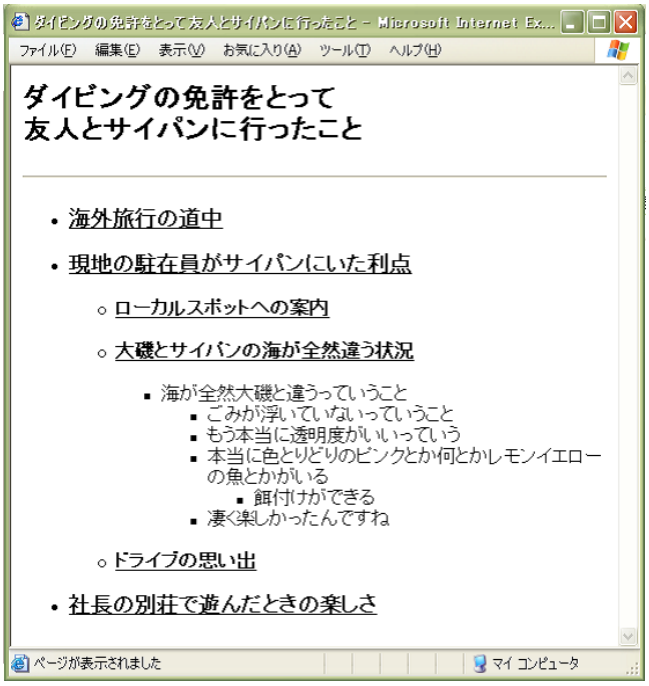

図 3: 独話音声コンテンツの例

3　トピックセグメンテーション

3.1　アルゴリズム

本研究では，講演を文の系列とし，$D = s_1 \cdots s_n$（$s_i (1 \leq i \leq n)$ は文）で表わす．トピックは，講演に含まれる文の列であり，$T_j = s_{j-k} \cdots s_j$（$1 \leq j - k \leq j \leq n$）と表わす．文 s_i が，直前の文 s_{i-1} を含む T_{i-1} と結合するかどうかを判定することにより，ボトムアップにトピックセグメンテーションを行う．

本手法では，次に示す 5 つのルールに基づいて判定する．ルールの適用は上位のルールから順に行う．

セグメンテーションルール

1. T_{i-1} と文 s_i に同一のキーワードが含まれる場合，文 s_i をトピック T_{i-1} と連接し，それを T_i とする．
2. 文 s_i がトピック境界に関する手掛かり表現を含む場合，
   - 手掛かり表現が直前のトピックと結合することを示す表現の場合，文 s_i をトピック T_{i-1} と連接し，それを T_i とする．
   - 手掛かり表現がトピック境界となることを示す表現の場合，T_{i-1} をトピックとして決定し，新たな T_i を設け，$T_i = s_i$ とする．
3. 文 s_i 中の最初の節に指示語が含まれている場合，文 s_i をトピック T_{i-1} と連接し，それを T_i とする．
4. トピック T_{i-1} が $T_{i-1} = s_{i-1}$ の場合，文 s_i をトピック T_{i-1} と連接し，それを T_i とする．
5. 文 s_i とトピック T_{i-1} が上記のルールに 1 つも該当しない場合，トピック T_{i-1} をトピックとして決定し，新たなトピック T_i を設け，$T_i = s_i$ とする．

なお，それぞれのルールは，次の考えに基づいて定めた．

1. あるトピック内に頻出し，かつ，他のトピックでの出現頻度が低い語をキーワードとしたとき，トピック T_{i-1} と文 s_i に同一のキーワードが含まれている場合，それらは同一のトピックとなる可能性が高い．
2. 例えば，文 s_i に「第一に」や「次に」といった表現が含まれていれば，s_i はトピックの先頭となる文であることが予想される．また,「そこで」や「もしくは」などの表現が含まれる場合は，トピック T_{i-1} と結合することが予想される．
3. 文 s_i に「それ」や「これ」といった指示語が含まれ，指示語がトピック T_{i-1} に含まれる文中の語を参照している場合，s_i は T_{i-1} との関連性が大きいと考えられる．しかし，指示語が s_i 中の語を参照している可能性も考えられるため，s_i 中の最初の節に含まれる指示語に限り，s_i は T_{i-1} と結合することとした．
4. 1 つのトピックは通常，ある程度の大きさを持って存在する．本研究では，1 つのトピックは 2 つ以上の文からなると仮定している．
5. トピック T_{i-1} と文 s_i が上記の全てのルールに該当しない場合，T_{i-1} と s_i との関係は小さいと考えられる．

3.2 キーワードの自動抽出

キーワードとなる語は，ある 1 つのトピックにおいてのみ集中的に出現する語である．本手法では，次の式によって定義される語 w のスコア $score_w$ に基づいて，キーワードを抽出する．

$$score_w = \frac{\sum_{i=0}^{n}(x_i - \mu)^2}{n} \quad (1)$$

ここで，n は w の出現回数である．x_i は i 番目の w の出現位置である．出現位置とは，講演を構成する単語列上の位置である．μ は，x_i の出現位置の重心である．

式（1）は，語の位置の標準偏差を表わしている．標準偏差が小さいとは，その語が講演中の一部分に集中的に出現していることを意味する．本手法では，講演中に 2 回以上出現する名詞を対象に式（1）によってスコアを計算している．

3.3 トピック境界に関する手掛かり表現の取得

手掛かり表現は，日本語話し言葉コーパス（CSJ）[6] を利用して取得した．利用したデータは，CSJ に含まれる模擬講演のうち，談話境界タグが付与された 13 講演である．談話境界タグは，トピックが切り替わる所に付与されている．取得した手掛かり表現の例を以下に示す．

- 直前のトピックと結合することを示す表現

  また，なので，そこで，よって，もしくは，故に

- トピック境界であることを示す表現

  次に，最後に，第一点，後は，さて，ところで

直前のトピックと結合することを示す表現とは，その表現が含まれる文 s_i が直前のトピック T_{i-1} と結合することを示す．トピック境界であることを示す表現とは，その表現が含まれる文 s_i が直前のトピック T_{i-1} と結合せず，$T_i = s_i$ となり，トピックの始端となることを示す．

手掛かり表現は，トピック境界に関して特徴的であると思われる表現を人手によって選択した．計 74 種類の表現を取得した．そのうち直前のトピックと結合することを示す表現は 13 種類，トピック境界であることを示す表現は 61 種類であった．

3.4 実行例

トピックセグメンテーションの例を図 4 に示す．図 4 では，文 s_{10} から文 s_{14} が記されている．以下では，$T_{10} = s_{10}$ のときの s_{11}〜s_{14} の処理について記す．3.2 節で述べたキーワードの自動抽出の結果，例に挙げた講演でのキーワードは，「十ページ」「フラタニティー」であると仮定する．

s_{11} の処理

T_{10} と s_{11} には，共通のキーワード「十ページ」が含まれているため，s_{11} を T_{10} と連接し，$T_{11} = s_{10}s_{11}$ とする．

s_{12} の処理

s_{12} には，最初の節に指示語「それ」が含まれているため，s_{12} を T_{11} と連接し，$T_{12} = s_{10}s_{11}s_{12}$ とする．

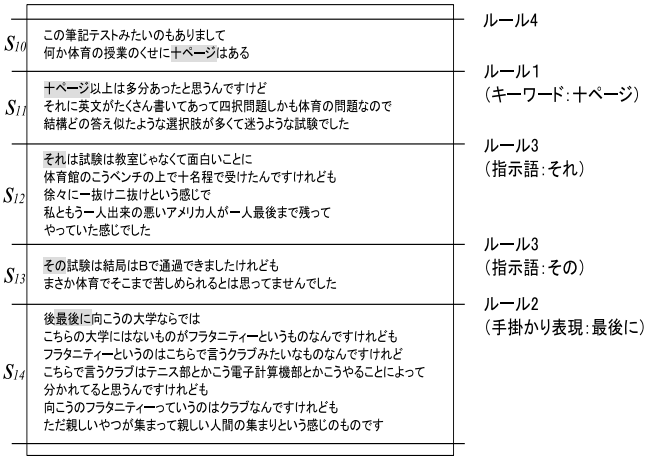

図 4: トピックセグメンテーションの実行例

表 1: トピックセグメンテーションの結果

講演 ID	精度	再現率	F 値
S01F0157	0.667	0.800	0.733
S01F0183	0.563	0.900	0.731
S00F0209	0.500	0.824	0.662
S00M0117	0.533	0.727	0.630
S04F0013	0.318	0.778	0.548
S03M0098	0.350	0.636	0.493
S05M0412	0.300	0.667	0.483
S03F0214	0.261	0.600	0.430
S02F0189	0.240	0.429	0.334
S01M0227	0.154	0.500	0.327
S00M0065	0.214	0.429	0.321
S02M0161	0.179	0.417	0.298
平均	0.357	0.642	0.499

s_{13} の処理

s_{13} には，最初の節に指示語「その」が含まれているため，s_{13} を T_{12} と連接し，$T_{13} = s_{10}s_{11}s_{12}s_{13}$ とする．

s_{14} の処理

s_{14} には，手掛かり表現「最後に」が含まれているため，$T_{13} = s_{10}s_{11}s_{12}s_{13}$ をトピックとして決定する．さらに，新たなトピック T_{14} を設け，$T_{14} = s_{14}$ とする．

4 評価実験

4.1 実験方法

本手法を評価するために，実際の講演データを用いてセグメンテーション実験を行った．講演データには，CSJ[6] に含まれる談話構造タグ付き模擬講演のうち，12 講演を用いた．これは，

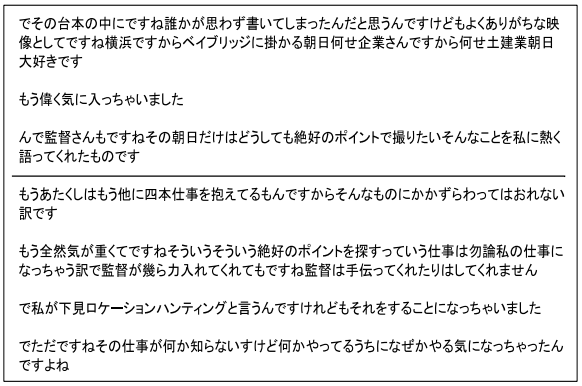

図 5: トピックセグメンテーションの失敗例

3.3 節で手掛かり表現の取得に用いた 13 講演とは異なる．本手法では，文に分割された講演データを前提とする．文を決定するために，CSJ に付与されている節境界情報を利用した．節が切れる強さによって,「絶対境界」「強境界」「弱境界」の 3 段階のレベルが設定されており，このうち,「絶対境界」が文末表現に相当するため，絶対境界によって分割される節を文とみなした．

4.2 実験結果

結果を表 1 に示す．平均で，精度 0.357，再現率 0.642，F 値 0.499 という値が得られた．本手法のように，いくつかのシンプルなルールに基づく手法でこのような高い値を得ることができ，作成したルールの有効性が示された．

実験結果では，適用する講演によって精度や再現率に大きな差が見られた．これは，話者の発話に関する特徴によっては，現状で制定されているルールや手掛かり表現では，対応できない講演があることを示している．本手法で特定することは難しいトピック境界の例を図 5 に示す．この例では 2 つのトピックが存在する．図の上の部分には,「台本が気に入られた理由」が述べられており，下の部分には,「私の気持ちの変化」が述べられている．これらは別のトピックであるが，本手法で用いる手掛かり表現やキーワード，指示語といった表層的な情報は含まれていない．

5 おわりに

本論文では，講演を対象とした独話音声のコンテンツ化について述べた．さらに，独話音声のコンテンツ化に向けたトピックセグメンテーションの手法を提案した．トピックセグメンテーションの手法の有効性を確認するために，CSJ の模擬講演を用いて実験を行った．その結果，ルールのみに基づく手法によるトピックセグメンテーションの実現可能性を示した．

本手法のルールを改善することにより，性能の向上が期待できる．例えば，ルールや手掛かり表現に重みを付与し，スコアリングを行うことなどが考えられる．また，現時点では，直前のトピックに対してのみルールを適用しているが，この点についてもさらなる検討の余地がある．

参考文献

[1] 望月源, 本田岳夫, 奥村学：複数の知識の組合せを用いたテキストセグメンテーション, 情報処理学会研究報告，Vol. 95, No. 89, pp. 47-54 (1995).

[2] 竹下敦, 井上孝史, 田中一男：テキストの概要把握支援のための話題構造抽出, 情報処理学会論文誌, Vol. 37, No. 11, pp. 1941-1949 (1996).

[3] 中野滋徳, 足立顕, 牧野武則：語の近接性に基づいた意味段落境界の判定手法, 情報処理学会研究報告, Vol. 2005, No. 22, pp. 23-30 (2005).

[4] 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, 情報処理学会論文誌, Vol. 43, No. 7, pp. 2222-2229 (2001).

[5] 塩崎敏也，鷹合基行，武田英明，西田豊明：設計における談話の分析と構造化，電子情報通信学会技術研究報告, Vol. 97, No. 632, pp. 41-48 (1998).

[6] 前川喜久雄, 籠宮隆之, 小磯花絵, 小椋秀樹, 菊池英明:日本語話し言葉コーパスの設計, 音声研究, Vol. 4, No.2, pp. 51-61(2000).