

# 読みやすい字幕生成のための独話文への改行挿入

村田 匡輝†

大野 誠寛‡

松原 茂樹§

柏岡 秀紀¶

†名古屋大学工学部 ‡名古屋大学大学院国際開発研究科 §名古屋大学情報連携基盤センター

¶情報通信研究機構 †ATR 音声言語コミュニケーション研究所

## 1 はじめに

聴覚障害者や高齢者、外国人等による講演や解説などの独話音声の理解を支援するための環境として、リアルタイム字幕生成システムの開発が望まれている。1文が長い傾向にある独話文を、そのまま字幕として生成し表示しても必ずしも読みやすくない、テキストを適切な位置で改行することが必要となる。

本稿では、独話音声に対する読みやすい字幕を生成するための要素技術として、独話文の改行手法を提案する。本手法では、解説や講演などの現場における字幕情報の提供手段として、字幕のみが複数行表示されるディスプレイを想定し、形態素、係り受け、節境界等の出現パターンに基づく改行挿入ルールの適用により、適切な位置で改行された字幕を生成することを試みる。

本研究では、形態素情報、文節境界情報、係り受け情報、節境界情報が付与されたデータ [1] に対して、改行位置を手で付与し、独話文に対する改行データを作成した。次に、改行データの詳細な分析に基づき、改行位置を決定するための改行ルールを定めた。NHK の解説番組「あすを読む」の書き起こしテキストを用いて改行挿入実験を実施した結果、手で改行位置を付与した正解データに対して 86.8% の挿入精度を達成し、本手法の利用可能性を確認した。

## 2 独話音声の字幕表示

本研究では、解説や講演などが遂行される現場において、字幕が提示される環境の実現を目指しており、字幕だけが表示されるディスプレイの設置を想定している。図 1 の独話音声に対して提示する字幕例を図 2 に示す。テレビ番組のクロズドキャプションとは異なり、テキストが行単位で入れ替わり、スクロールされながら表示される字幕の生成を目指す。

### 2.1 関連研究

字幕生成のための改行挿入に関する従来研究として、門馬らは、改行位置を形態素列のパタンにより改行位置を決定する手法を提案している [2]。この研究では、テレビ番組におけるクロズドキャプションの改行位置を分析し、形態素列パタンに基づいて改行位置を定めている。しかし、日本のテレビ番組におけるクロズドキャプションでは、1画面 2行の字幕を一度に切り替える表示方式が標準となっており、複数行の字幕の段階的な表示環境とは、挿入すべき改行の位置は異なる。

アメリカでは消費者教育を受ける権利つまり合理的で賢い意思表示をして市場に参加できるための教育を誰もが受けられることこの消費者教育を受ける権利が消費者の権利の一つとして付け加えられたのが千九百七十五年です

図 1: 独話音声

アメリカでは  
消費者教育を受ける権利  
つまり合理的で賢い意思表示をして  
市場に参加できるための教育を  
誰もが受けられること  
この消費者教育を受ける権利が  
消費者の権利の一つとして  
付け加えられたのが千九百七十五年です

図 2: 字幕例

また、西光らは、多段階にチャンキングした言語的なまとまりに基づいて字幕を生成する手法を提案している [3]。この手法では、文の主題や述語、格要素などに対応する「構成要素」と複数の構成要素からなる「フレーズ」にチャンキングする。構成素単位で 15 文字になるまで文字列を連結し、異なるフレーズに含まれる構成要素を連結させないことにより、言語的なまとまりでの改行を実現している。しかし、人手による改行位置と「構成要素」や「フレーズ」との関係性は検証されていない。

### 2.2 独話文における改行の挿入位置

複数行の字幕が表示される画面では特に、改行を考慮することなく字幕を生成し表示すると読みにくいテキストとなるため、適切な位置で改行する必要がある。本研究では、字幕生成における適切な改行挿入位置として、以下の考え方を設けた。

- 各行が意味的なまとまりを形成するように改行する。ただし、連体節については意味的にまとまっているものの、そこで改行すると、行末を文末と誤解される恐れがあるため、改行の対象外とする。
- ディスプレイの大きさと文字の可読性を考慮した行の最長文字数を設定し、それ以下となるように改行する。

表 1: 改行データの規模

番組数	6
文数	410
文節数	4,386
付与した改行数	1,011

- 行によって長さが大きく異なると閲覧性が低下すると考えられるため、各行の文字数がおおよそ均等となるように改行する（すなわち、最長文字数に近くなるような位置で改行する）。

なお、本研究では、改行の挿入位置は文節境界に限定する。

### 3 独話文における改行位置の検出

提案手法では、形態素解析、文節まとめ上げ、節境界解析、係り受け解析が施された文を入力とし、改行位置（以下、**改行点**）が付与された文を出力とする。

一行が最長文字数以下となるように、意味的なまとまりの切れ目を手がかりに改行点を特定する。本手法における改行点の検出プロセスは以下の通りである。

- (1) **改行点の候補選定** 改行点検出ルールを適用することにより、一行が最長文字数を超えない程度の数の改行点を挿入し、それを改行点候補とする。
- (2) **改行点の決定** 一行が最長文字数を超えない範囲で、可能であれば改行点候補を削除し、最終的に残った候補を出力する改行点として決定する。

なお、本研究では、2.2節で示した考え方に基づいて、独話文の改行点を検出するルール（以下、**改行ルール**）を作成した。改行ルールは、入力文における意味的なまとまりを捉えるという観点から、節境界、及び、係り受けに着目して定めた。ルールの作成は、NHKの解説番組「あすを読む」の書き起こしテキスト<sup>1</sup>に対して、人手で改行を施したデータ（以下、**改行データ**）を調査することによって行った。使用したデータの規模を表1に示す。

以下ではまず、作成した改行ルールについて説明し、次に最終的な改行点決定法について述べる。

#### 3.1 節境界に基づく改行

節境界とは、隣接する節間の境界である。一般に、節は、意味的なまとまりを形成するため、節境界は改行点の有力な候補となる。ただし、あらゆる種類の節境界が必ずしも改行点になりやすいわけではない。我々の調査では、最も改行点になりやすい節境界の種類<sup>2</sup>は表2に示す通りである。すなわち、

<sup>1</sup>形態素情報、文節境界情報、係り受け情報、節境界情報が付与されたデータが構築されている [1]。

<sup>2</sup>本研究では、節境界の種類として節境界解析ツール CBAP[4] の定義を用いた。

表 2: 最も改行点になりやすい節境界の種類

並列節	トカ、ガ、シ、デ、ケレドモ、タリ
条件節	カギリ、バ、タラ、ケッカ、トコロ
時間節	トキニ、アトニ、イマ、アト、トキノ、その他
理由節	ノデ、カラ
その他	体言止、間投句、連用節その他、タメニハ節、タメ節、ナガラ節、ナド節、テハ節、ヨウ節

表 3: 改行点になりやすい節境界の種類

間接疑問節、補足節、譲歩節テモ、連用節

**改行ルール 1:** 表2に示した種類の節境界に改行点を挿入する<sup>3</sup>

を設ける。図3に、表2に該当する節境界での改行の例を示す。改行データに出現した節境界は、全部で59種類存在しており、改行点になる節境界は49.1%を占めている。

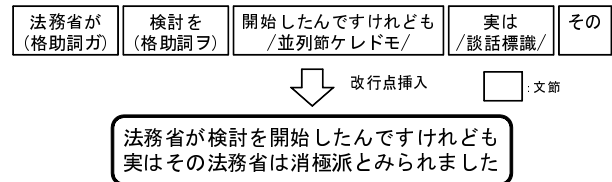


図 3: 「並列節ケレドモ」での改行例

ただし、ルール1だけで、「一行が最長文字数を超えない」という条件を満たさない場合には、上述の節境界の次に改行点になりやすい節境界での改行を実施する。すなわち、

**改行ルール 2:** 表3に示した種類の節境界に改行点を挿入する<sup>4</sup>

を設ける。図4に表3に該当する節境界での改行の例を示す。

#### 3.2 係り受け関係に基づく改行

係り受け関係とは、係り受け文節と受け文節の間関係であり、ある種の意味的なまとまりを形成する。そのため、係り受け関係の直後には、節境界に準じて改行が挿入されやすい。同時に、係り文節と受け文節の間に改行が挿入される場合、文節の係り先の解釈を読み手が誤る可能性があり、それを考慮した改行点を選ぶ必要がある。

<sup>3</sup>ただし、「時間節トキニハ」及び「並列節デ」については、節境界の直前の文節が直後の文節に係らない場合のみ改行点となる。また、「体言止」については、「体言止」の直後の文節が「あるいは」または「そして」であり、名詞もしくは名詞句が接続されている場合でないとき改行点となる。

<sup>4</sup>ただし、「補足節」については「補足節」の直前の文節が、格助詞ガまたは係助詞ハを持ち、直後の文節に係らない場合のみ、また「連用節」については「連用節」までの文節列の文字数が4文字以上の場合のみ改行点となる。

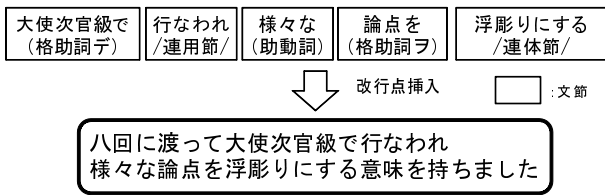


図 4: 「連用節」での改行例

係り受け関係のうち、最も改行が挿入されやすい場合は、連体節の係り先の直後である。加えて、連体節で改行を行うと、行末を文末と誤解される恐れがあるため、

**改行ルール 3:** 連体節の最終文節の受け文節の直後に改行点を挿入する

を設ける。図 5 に連体節の係り先の直後での改行の例を示す。

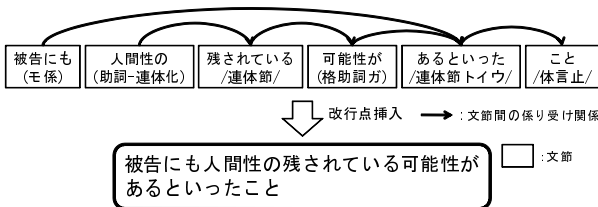


図 5: 連体節の係り先の直後での改行例

また、文の冒頭に節境界「主題ハ」が存在し、かつ、述語が複数存在する場合には、「主題ハ」の文節は文末に係ることが多い。その場合、複数の述語の間で行を分割してしまうと、「主題ハ」の文節の係り先が前者の述語であると、読み手が誤って解釈する恐れがある。これを避けるために、節境界「主題ハ」の直後を改行の候補とする。すなわち、

**改行ルール 4:** 文の冒頭に、文末に係る節境界「主題ハ」が存在し、かつ、述語が複数存在する場合には、「主題ハ」の直後に改行点を挿入する。

を設ける。図 6 に「主題ハ」の直後での改行の例を示す。

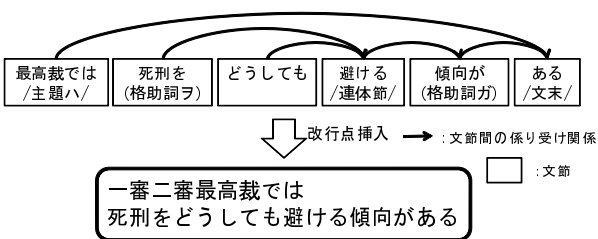


図 6: 主題ハの直後での改行例

その他に、係り文節と受け文節が離れて位置する場合などにおけるルールとして、

**改行ルール 5:** 係り文節と受け文節の間に改行が挿入される場合、係り文節の直後に改行点を挿入する。

を設ける。この場合の改行の例を図 7 に示す。

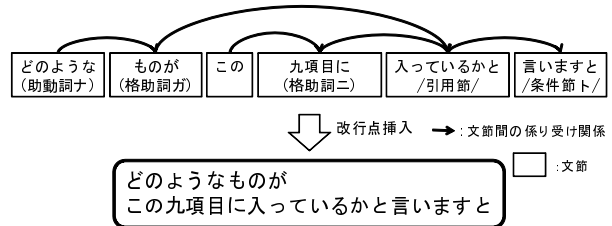


図 7: 係り文節と受け文節の間に改行が挿入される例

以上の改行ルールを手がかりに改行点を検出する。なお、番号は、ルールの優先性を示しており、例えば、改行ルール 1 は最も強力であり、該当の位置には無条件に改行が挿入される。

### 3.3 改行点の決定

改行点の候補選定では一行の文字数を最長文字数に近づけることはあまり考慮されていないため、改行点候補をそのまま改行点として確定すると短い行が連続して生じる場合がある。改行点の決定処理では、改行点候補をそのまま改行点として確定した場合に生じる連続した短い行を一行に連結できるか否かを判定し、結合できる場合には、連結する行間の改行点候補を削除する。

まず、連続する二行を結合した文字列が最長文字数を超える場合は、これら二行は結合できないと判定する。次に、たとえ、結合した文字列が最長文字数以内であっても、改行点候補の位置が意味的に強い切れ目であれば、必ず改行される必要があり、そこでは結合できないと判定する。さらに、改行点候補の位置が意味的に強い切れ目でない場合でも、一行が意味的なまとまりとなることが望ましいため、「連続する二行の最終文節が共に同じ文節に係る」、もしくは、「結合した文字列内で係り受け構造が閉じている」場合のみ、結合できると判定する。図 8、図 9 に、結合できる場合の例を、「連続する二行の最終文節が共に同じ文節に係る」場合と「結合した文字列内で係り受け構造が閉じている」場合に分けてそれぞれ示す。

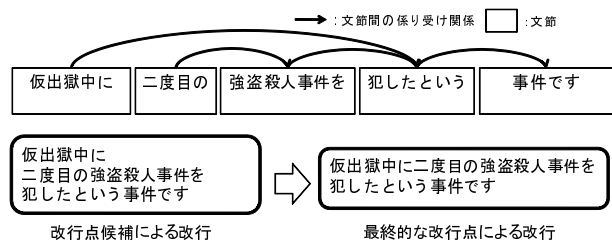


図 8: 連続する二行の最終文節が共に同じ文節に係る例

なお、以下の条件を全て満たす場合、改行点候補の位置は意味的に強い切れ目でないとする。

- その箇所が節境界でない
- その箇所の直前の形態素が、「係助詞モ」もしくは「名詞」、「副詞」、「場合 (名詞-副詞可能)」でない

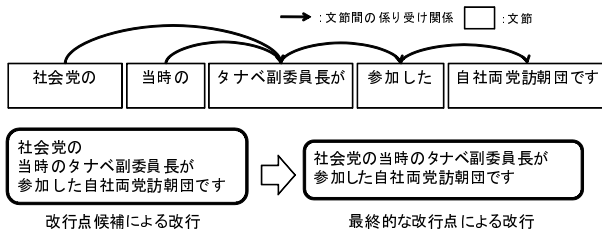


図 9: 結合した文字列内で係り受け構造が閉じている例

## 4 実験

本手法の利用可能性を確認するため、実際の独話データに対して改行挿入実験を実施した。

### 4.1 実験の概要

実験データとして、NHK の解説番組「あすを読む」3 番組の書き起こしテキスト (219 文, 2121 文節) を使用した。データには、節境界情報、形態素情報、係り受け情報を人手で付与している。

実験データに対して、本手法による改行挿入実験、また、文節単位ごとに文字列を連結していき、文字数が最長文字数に最も近づいた時点で改行を挿入するという手法をベースラインとして改行挿入実験を実施し、結果を比較した。

改行挿入結果に対する評価は、正解の改行位置に対する精度 (挿入した改行のうち正解と一致する割合)、及び、再現率 (正解の改行位置に改行が挿入された割合) を測定することにより実施した。

正解データは、二人の作業者が共同で作成した。すなわち、まず二人が独立に正解データを作成し、それぞれのデータをもとにした協議を経て、正解データを作り上げた。生成する字幕として容認可能な改行位置が正解データ以外に存在することを考慮し、正解データの作成前に二人の作業者が別々に作成した 2 つの改行データを使用し、それらとの一致に基づく精度も併せて測定した。

### 4.2 実験結果

実験結果を表 4 に示す。精度で 78.4%、再現率で 78.7% と、ベースラインと比べて高い性能を示した。また、正解とは異なる 2 つの改行データを対象に評価したところ、86.8% の精度を達成した。以上より、本手法の利用可能性を確認した。

挿入誤りを分析した結果、

- 隣接する文節「考えてみたい」と「思います」の間のように、改行点挿入の可能性が、文法とは別に、語彙の種類によって決まる場合が存在する。
- 改行決定プロセスでは、削除する改行点候補の順序により、最終的な改行点が異なってくる。

などの原因が明らかになった。

表 4: 実験結果

	本手法	ベースライン
再現率	78.7% (406/515)	30.1% (155/515)
精度	78.4% (406/517)	37.9% (155/409)

## 5 おわりに

本稿では、聴覚障害者、高齢者、外国人等による独話音声の理解を支援することを目指し、字幕生成における独話文への改行挿入手法を提案した。本手法では、形態素、係り受け、節境界、ポーズ・フィラー等の出現に基づく改行挿入ルールの適用することにより、字幕テキストを読みやすくするための改行を実現する。解説番組を用いた改行挿入実験では、精度で 78.4%、再現率で 78.7% であり、本手法の利用可能性を確認した。

人手により改行挿入ルールを作成する方式では、ルールの大規模化、詳細化、体系化の面で限界がある。現在、改行挿入データの作成を進めており、今後、本研究により得られた知見をもとにルール獲得方式を考案することを検討している。また、本研究で対象とした改行の挿入位置は、音声合成における適切なポーズ挿入位置 [5] と関連することが予想される。両者の言語的特徴の共通性について明らかにすることも今後の課題である。

**謝辞** 本研究は、科学研究費補助金 (特別研究員奨励費) 「大規模音声言語コーパスを用いた独話データの構造化とその応用に関する研究」 (課題番号 18・6433) により実施したものである。

## 参考文献

- [1] T. Ohno, S. Matsubara, H. Kashioka, N. Kato, Y. Inagaki: A Syntactically Annotated Corpus of Japanese Spoken Monologue, Proc. of 5th LREC, pp.1590-1595 (2006) .
- [2] 門馬 隆雄, 沢村 英治, 福島 孝博, 丸山 一郎, 江原 暉政, 白井 克彦: 聴覚障害者向け字幕付きテレビ番組の自動制作システム, 電子情報通信学会論文誌, Vol.J84-D-II, No.6, pp.888-897 (2001).
- [3] 西光 雅弘, 高梨 克也, 河原 達也: 係り受けとポーズ・フィラーの情報をを用いた話し言葉の段階的チャンキング, 情報処理学会研究報告, 2006-SLP-61, pp.19-23 (2006)
- [4] 丸山 岳彦, 柏岡 秀紀, 熊野 正, 田中 英輝: 日本語節境界検出プログラム CBAP の開発と評価, 自然言語処理, Vol. 11, No. 3, pp. 39-68 (2004).
- [5] 佐藤 奈穂子, 小島 裕一, 望主 雅子, 亀田 雅之: テキスト音声合成における係り受け解析結果を用いたポーズ挿入処理, 自然言語処理, Vol.6, No.2, pp.117-133 (1999).