

読みやすい字幕生成のための講演テキストへの改行挿入

村田 匡輝[†] 大野 誠寛^{††} 松原 茂樹^{†††}

Linefeed Insertion into Lecture Transcription for Automatic Captioning

Masaki MURATA[†], Tomohiro OHNO^{††}, and Shigeki MATSUBARA^{†††}

あらまし リアルタイム字幕生成とは、講演や解説などの音声をテキストで提示するものであり、聴覚障害者や高齢者、外国人らによる講演音声の理解を支援するための技術である。講演では一文が長くなる傾向にあり、多くの文がスクリーン上で複数行にまたがって表示されることになるため、テキストが読みやすくなる位置に改行が挿入される必要がある。本論文では、読みやすい字幕を生成するための要素技術として、日本語講演文への改行挿入手法を提案する。本手法では、係り受け、節境界やポーズ、行長などの情報に基づき、統計的手法によって改行位置を決定する。日本語講演データの 1,714 文を使用した実験では、改行挿入の再現率で 82.66%、適合率で 80.24% を達成し、本手法の有効性を確認した。

キーワード 音声言語、リアルタイム字幕生成、係り受け構造、節境界、音声コーパス

1. ま え が き

リアルタイム字幕生成とは、講演などの音声をテキストで提示するものであり、聴覚障害者や高齢者、外国人らによる音声理解を支援することを目的とする。近年、字幕の自動生成の実現を目指した研究がいくつか行われており [1]、字幕生成のための音声認識技術について検討が進んでいる [2] ~ [7]。しかしながら、読みやすい字幕を生成するためには、音声を精度よく文字化することだけでなく、文字化されたテキストをどのように提示するかということもまた重要となる [8], [9]。特に、講演では文が長くなる傾向にあり、一文が字幕スクリーン上で複数行にまたがって表示されることになるため、提示されたテキストが読みやすくなるように、適切な箇所に改行が挿入されていることが望まれる。

これまで、字幕の自動生成におけるテキストの提示方法に関する研究はほとんどない。字幕への改行挿入に関する研究として、門馬らは、形態素列のボタンにより改行位置を決定する手法を提案している [10]。しかし、この研究は、テレビ番組におけるクローズドキャプションを対象としている。日本のテレビ番組におけるクローズドキャプションは、1 画面 2 行の字幕を一度に切り換える表示方式が標準であり、講演会場の字幕提示環境とは、挿入すべき改行の位置は異なる。

本論文では、読みやすい字幕を生成するための基盤技術として、日本語講演音声の書き起こし文への改行挿入手法を提案する。本研究では、講演会場での聴衆への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の設置を想定している。本手法では、文節境界を改行挿入位置の候補とし、節境界、係り受け関係、ポーズ、行長などの情報に基づいて、統計的手法により改行位置を決定する。

本手法は、文を複数の行に分割することを目的とする。この点で、文を複数の節に分割する節境界解析（例えば、[11] ~ [13]）と関連する。しかし、節は言語単位の一つであり、その境界は文法的に定めることができるのに対して、適切な改行位置は読みやすさという観点から定まるものである。本研究は、このような主観的要因に基づく文の分割を機械的に実現する点に特徴がある。

[†] 名古屋大学大学院情報科学研究科社会システム情報学専攻, 名古屋市

Department of Systems and Social Informatics, Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464-8603 Japan

^{††} 名古屋大学大学院国際開発研究科, 名古屋市

Graduate School of International Development, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601 Japan

^{†††} 名古屋大学情報基盤センター, 名古屋市

Information Technology Center, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601 Japan

日本語講演データを用いて実験を行った．1,714 文に対して改行挿入を実行した結果，人手で改行位置を付与した正解データに対して，再現率で 82.66%，適合率で 80.24% を達成した．比較のために設定した四つのベースライン方式と比べ，大幅に性能が向上しており，本手法の有効性を確認した．

本論文の構成は，以下のとおりである．2. では，本研究が目指す字幕テキストにおける改行について述べる．3. で適切な改行位置について分析し，4. で本論文で提案する改行挿入手法を説明する．5. では，改行挿入実験について報告し，6. でその考察を与える．

2. 講演テキストへの改行挿入

本研究では，講演会場における字幕提示環境として，プレゼンテーションスライドを表示するスクリーンに併設された，字幕テキスト表示専用のディスプレイの利用を想定する．テキストは行単位で入れ換わり，スクロールしながら常に数行表示される．図 1 に，想定する字幕提示環境を示す．

図 2 に示すように，音声の書き起こしテキストを，改行位置を考慮することなくディスプレイの幅に合わせて表示すると，読みにくいテキストとなる．特に，字幕テキストでは，話者の発声スピードに合わせて読むことが強いられるため，図 3 に示すように読みやすい位置で改行されていることは重要である．

テキストを読みやすくするための改行挿入の効果を明らかにするために，講演音声の書き起こしテキストを用いて調査した．名古屋大学同時通訳データベース [14] に収録された日本語講演の書き起こしテキストからランダムに選択した 50 文に対して^(注1)，

(1) 行頭から 20 文字の位置に改行を挿入したテキスト

(2) 適切な位置に人手で改行を挿入したテキストを用意した．図 2 は (1) のテキストに，図 3 は (2)

のテキストにそれぞれ相当する．なお，(2) のテキストは，3 人の作業者による協議に基づきテキストに対する適切な改行位置を定めることにより作成した．被験者 10 名はどちらのテキストが読みやすいかを選択した．図 4 に調査結果を示す．50 文のうち，(2) の方が読みやすいと評価された文の割合は，被験者平均で 87.0% であった．また，(1) の方が読みやすいと評価した被験者が過半数に至った文は存在しなかった^(注2)．これらのことは改行挿入によってテキストが読みやすくなることを示している．

そこで本論文では，講演テキストに対して，それが

例えば環境の問題あるいは人口の問題エイズの問題など地球規模の問題たくさん生じておりますが残念ながらこれらの問題は二十一世紀にも継続しあるいは悲観的な見方をすればさらに悪くなるという風に思われます

図 2 講演音声の書き起こしテキスト
Fig. 2 Transcription of lecture.

例えば環境の問題
あるいは人口の問題
エイズの問題などなど
地球規模の問題たくさん生じておりますが
残念ながらこれらの問題は
二十一世紀にも継続し
あるいは悲観的な見方をすれば
さらに悪くなるという風に思われます

図 3 適切な位置に改行が挿入されたテキスト
Fig. 3 Transcription into which linefeeds were properly inserted.

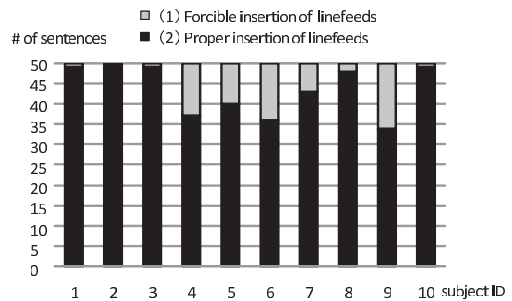


図 4 講演テキストへの改行挿入の効果
Fig. 4 Effect of linefeed insertion into transcription.

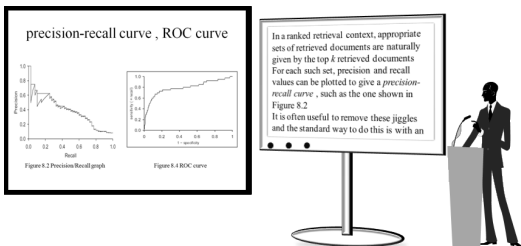


図 1 講演音声の字幕提示環境
Fig. 1 Caption display of lecture transcription.

(注1): 50 文の平均文字数は 71.0 文字である．なお，名古屋大学同時通訳データベース [14] の書き起こしテキストには文末タグが付与されており，本研究における「文」はそれに基づいている．

(注2): 文によっては，(1) の方が読みやすいと評価した被験者が存在したが，これは強制的に 20 文字ごとに挿入した改行が偶然，不適切ではない位置に挿入され，(2) のテキストと評価が分かれたことによる．

読みやすくなる位置に改行を挿入する手法について述べる．改行挿入の手法を提案するにあたり，まず，想定する入力テキストを定める必要がある．というのも，字幕提示システムでは，講演音声文字化するのに方法がいくつか考えられるためである [15]．代表的な方法として，音声認識システムの利用やパソコン要約筆記があるが，例えば，音声認識を用いる場合でも，音声直接認識する方式，復唱音声を認識する方式などがあり，更に認識誤りを人手で修正するかどうかの選択もある．また，パソコン要約筆記においても，その文字化スタイルは筆者や使用するツールによって異なる．このような文字化方式の違いにより，入力テキストに認識誤りが含まれるかどうか，含まれるとしたらどの程度含まれるか，また，テキストがどのような単位で，そして，どのようなタイミングで入力されるか，が異なることとなる．

そこで本研究では，特定の文字化方式に依存しないという観点に立ち，正しく書き起こされた文を入力とする．また，読点の存在は前提としない．読点の位置は改行挿入において有力な情報となることが予想されるが，それにもかかわらず前提としないのは，句点の位置が人によらずほぼ一定なのに対して，読点の挿入位置は人による違いが小さくないためである．

次に，想定する出力テキストとして，本研究では，字幕生成における改行挿入位置について，以下の前提を設けた．

- ディスプレイの大きさを考慮した行の最長文字数を設定し，各行の文字数をそれ以下とする．
- 日本語では，文節は意味のまとまりの基本単位であることを考慮し，文節境界を改行位置の候補とする．

なお，本論文の以下では，改行が挿入される文節境界を改行点 (linefeed point) という．

3. 改行点の分析

読みやすい講演テキストのための適切な改行挿入位置とは，いくつかの要因のバランスのもとに定まると考えられるため，本研究では，改行点を同定するために統計的アプローチを採用する．そのための有効な素性について検討するため，音声言語コーパスを用いて事前分析を与えた．コーパスには，名古屋大学同時通訳データベース [14] の日本語講演音声データを用いた．書き起こしテキストには，形態素，文節境界，係り受け構造，節境界等の構文的情報，並びに，改行点

表 1 分析データのサイズ
Table 1 Size of analysis data.

文数	221
文節数	2,891
文字数	13,899
改行挿入数	833
1 行当りの文字数	13.2

が，人手により付与されている．改行点は 2. で調査に使用したデータと同様の手順で付与した．分析に使用したデータの規模を表 1 に示す．

なお改行は，行当りの文字数が最大 20 文字であるとして，人手で付与した．20 文字という設定は，ディスプレイ上での文字の可読性を考慮して設定した^(注3)．文節境界 (すなわち，改行点候補) 2,670 箇所に対して，833 箇所改行が挿入されており，改行挿入率は 31.2% である．分析では，構文情報として，節境界や係り受け構造，行長，ポーズ，行頭の形態素に注目し，それらと改行挿入との関係について調査した．

3.1 節境界と改行点

節 (clause) は構文的かつ意味的なまとまりを形成する言語単位の一つであり，節の境界は改行点として有力である．分析データのうち，文末を除く節境界は 969 箇所あり，そのうち 490 箇所改行が挿入されており，挿入率は 51.1% であった．文節境界に対する挿入率よりも高いことから，節境界には改行が挿入されやすいといえる．

分析データに出現した 42 種類の節境界^(注4)について，その改行挿入率を調査した．出現数にして上位 10 種類の節境界とその改行挿入率を表 2 に示す．節境界「並列節ガ」「並列節ケレドモ」の改行挿入率は 100% であるのに対して，「引用節」「連体節」などは 30% 以下であった．これらは，節境界の種類によって改行の挿入されやすさが異なることを示している．図 5 の改行挿入の例では，節境界「連体節」には改行が挿入されていないが，節境界「並列節ガ」には改行が挿入されている．

3.2 係り受け構造と改行点

隣接する文節間に直接的な係り受け関係が存在する場合，それら二つの文節で意味的なまとまりを形成す

(注3): スライドを用いたプレゼンテーションでは，スライド本文のフォントサイズは 26 ポイント以上が望ましいとされている [16]．代表的なプレゼンテーションソフトウェアである PowerPoint においてフォントサイズを考慮し，20 文字と設定した．

(注4): 節境界の種類として，節境界解析ツール CBAP [12] で定義されたものを用いた．

表 2 節境界への改行挿入率

Table 2 Ratio of linefeed insertion into clause boundary.

type of clause boundary	ratio of linefeed insertion (%)
主題八	50.8 (93/183)
談話標識	12.0 (21/175)
引用節	22.1 (21/95)
連体節	23.3 (20/86)
テ節	90.2 (74/82)
補足節	68.0 (34/50)
並列節ガ	100.0 (38/38)
並列節ケレドモ	100.0 (35/35)
条件節ト	93.5 (29/31)
連体節トイウ	27.3 (6/22)

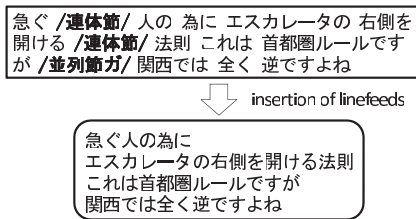


図 5 節境界の種類と改行点の関係

Fig. 5 Relation between clause boundary and linefeed point.

るため、そのような文節境界には改行が挿入されにくいと思われる。図 6 に隣接文節間の係り受け関係と改行挿入の関係を示す。図 6 では、係り受け関係にある隣接文節（「私の」と「車を」など）の境界には改行が挿入されていない。一方、係り受け関係にない隣接文節（「記者が」と「私の」）の間には改行が挿入されている。実際、分析データを調査したところ、係り受け関係にある隣接文節間 1,459 箇所に対して、改行が挿入されたのは 192 箇所であった。このような挿入例を図 7 に示す。挿入率は 13.2%であり、これは、文節境界に対する挿入率の半分以下である。一方、係り受け関係にない隣接文節間への挿入率は 52.7%であった。

上述の分析は、係り受け関係にある文節間の距離に注目した分析であるが、係り受け関係のタイプによっても改行の挿入されやすさが異なる。例えば、係り文節が連体節の節末であるとき、その受け文節の直後への改行挿入率は 43.1%であり、文節境界一般の挿入率よりも高い。

また、係り受け構造と改行点との関係、すなわち、行内で係り受けが閉じているかどうかを調べた。ここで、係り受けが閉じている行とは、行外の文節に係る文節が、行末の文節以外に存在しない行のことをいう。図 8 にそのような改行挿入の例を示す。図 8 の例で

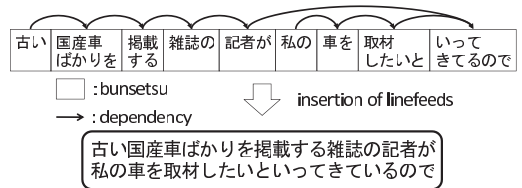


図 6 隣接文節間の係り受け関係と改行点の関係

Fig. 6 Relation between local dependency relation and linefeed point.

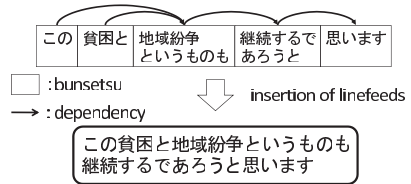


図 7 係り受け関係にある隣接文節間への改行挿入例

Fig. 7 Example of linefeed insertion between bunsetsus in local dependency relation.

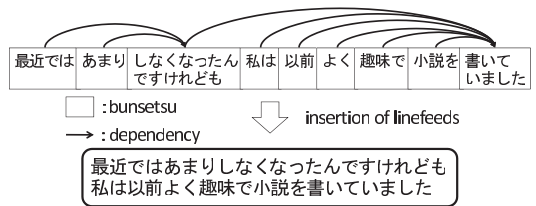


図 8 行内の係り受け構造と改行点の関係

Fig. 8 Relation between dependency structure in a line and linefeed point.

は、文節「しなくなっただすけれども」までで一つの行を形成しており、行を構成する文節「最近では」「あまり」が「しなくなっただすけれども」に係るため、行内で係り受けが閉じている。分析データの 833 行のうち、599 行で係り受けが閉じていた。この結果は、意味的なまとまりの多くが、係り受けが閉じている文節列で形成される傾向を反映している。

3.3 行長と改行点

行によって長さのばらつきが大きいと字幕の読みやすさが低下するため、極端に短い行は生成されにくいと考えられる。分析データの行長を調べたところ、長さが 6 文字以下の行は少なく、全体の 7.59%にすぎなかった。これは、行頭から行末までの文字列がある程度の長さ（本分析では 7 文字以上）をもつような文節境界に改行が挿入されやすいことを示している。

3.4 ポーズと改行点

ポーズは構文的区切りと一致しやすいという知見が

私は特に混雑することで有名な<ポーズ>小田急線という電車を利用しています

↓ insertion of linefeeds

私は特に混雑することで有名な小田急線という電車を利用しています

図 9 ポーズと改行点の関係

Fig. 9 Relation between pause and linefeed point.

表 3 行頭での出現率が低い形態素
Table 3 Morphemes which are unlikely to become a line head.

基本形	必要 (1/21), なる (2/32), する (3/33) 問題 (0/42), 思う (2/70)
品詞	名詞-非自立-一般 (0/40) 名詞-ナイ形容詞語幹 (0/40) 名詞-非自立-副詞可能 (0/27)

得られている [17]. すなわち, 文節境界におけるポーズの存在が改行の挿入と関連する可能性がある. 図 9 において, 文節「有名な」と「小田急線という」の間に改行が挿入されているが, その文節境界にはポーズが存在している.

本研究では, 200 ms 以上の連続する無音区間をポーズとして定義した. 分析データのうち, ポーズが存在する文節境界は 748 箇所あり, そのうち 471 箇所に改行が挿入されていた (改行挿入率 62.97%). 文節境界に対する改行挿入率よりも高く, ポーズが存在する文節境界には改行が挿入されやすいことが分かった.

3.5 行頭の形態素と改行点

形態素によっては, 行の先頭に出現すると読みにくいテキストになるものが存在する. そこで, 分析データ中の全文節の先頭の形態素に対する行頭での出現率を調査した. ここでは, 形態素の基本形と品詞を調査対象とした. 出現頻度が 20 回以上であり, かつ, 行頭での出現率が 10% 以下だった形態素を表 3 に示す^(注5). 括弧中の数字は, 分析データ全体での出現頻度に対する行頭での出現頻度の割合を示している. これらの形態素を第 1 形態素にもつ文節の直前の文節境界には改行が挿入されにくいと考えられる.

4. 改行挿入手法

本手法では, 形態素解析, 文節まとめ上げ, 節境界解析, 係り受け解析が与えられた文を入力とし, 入力文中の各文節境界に対して, その位置に改行を挿入するか否かを同定する. 入力文に対する適切な改行点を同定するために, 1 行当りの文字数が最長文字数を超

えないという条件のもと, 1 文中に挿入され得る改行点のすべての組合せの中から, 最適な組合せを確率モデルを用いて決定する.

以下では, n 個の文節からなる入力文を $B = b_1 \cdots b_n$ とするとき, 改行結果を $R = r_1 \cdots r_n$ と記す. ここで, r_i は, 文節 b_i の直後に改行が挿入されるか ($r_i = 1$) 否か ($r_i = 0$) のいずれかの値をとる. なお, $r_n = 1$ である. 入力文を m 行に分割した j 行目の文節列を $L_j = b_1^j \cdots b_{n_j}^j$ ($1 \leq j \leq m$) とした場合, $1 \leq k < n_j$ のとき $r_k^j = 0$, $k = n_j$ のとき $r_k^j = 1$ となる.

4.1 改行挿入のための確率モデル

本手法では, 入力文の文節列を B とするとき, $P(R|B)$ を最大にする改行挿入結果 R を求める. 各文節境界に改行が挿入されるか否かは, 直前の改行点を除く, 他の改行点とは独立であると仮定すると, $P(R|B)$ は次のように計算できる.

$$\begin{aligned}
 &P(R|B) \\
 &= P(r_1^1 = 0, \dots, r_{n_1-1}^1 = 0, r_{n_1}^1 = 1, \dots, \\
 &\quad r_1^m = 0, \dots, r_{n_m-1}^m = 0, r_{n_m}^m = 1|B) \\
 &\cong P(r_1^1 = 0|B) \times \dots \\
 &\quad \times P(r_{n_1-1}^1 = 0|r_{n_1-2}^1 = 0, \dots, r_1^1 = 0, B) \\
 &\quad \times P(r_{n_1}^1 = 1|r_{n_1-1}^1 = 0, \dots, r_1^1 = 0, B) \times \dots \\
 &\quad \times P(r_1^m = 0|r_{n_m-1}^m = 1, B) \times \dots \\
 &\quad \times P(r_{n_m-1}^m = 0|r_{n_m-2}^m = 0, \dots, r_1^m = 0, r_{n_m-1}^m = 1, B) \\
 &\quad \times P(r_{n_m}^m = 1|r_{n_m-1}^m = 0, \dots, r_1^m = 0, r_{n_m-1}^m = 1, B)
 \end{aligned} \tag{1}$$

ここで, $P(r_k^j = 1|r_{k-1}^j = 0, \dots, r_1^j = 0, r_{n_j-1}^j = 1, B)$ は, 1 文の文節列 B が与えられ, $j-1$ 行目の行末位置が同定されているときに, 文節 b_k^j の直後に改行が挿入される確率を表す. 同様に, $P(r_k^j = 0|r_{k-1}^j = 0, \dots, r_1^j = 0, r_{n_j-1}^j = 1, B)$ は, 文節 b_k^j の直後に改行が挿入されない確率を表す. これらの確率を最大エントロピー法により推定した. 最優秀の改行結果は, 式 (1) の確率を最大とする改行結果であるとして動的計画法を用いて計算する.

4.2 最大エントロピー法で用いた素性

本研究では, $P(r_k^j = 1|r_{k-1}^j = 0, \dots, r_1^j = 0,$

(注5): 形態素体系は日本語辞書 IPADIC の品詞体系 [18] に準拠した. 表 3 の品詞「名詞-非自立-一般」に属する形態素として「こと」「もの」, 「名詞-ナイ形容詞語幹」として「申し訳」「仕方」, 「名詞-非自立-副詞可能」として「ところ」「以外」などがある.

$r_{n_{j-1}}^{j-1} = 1, B)$ 並びに $P(r_k^j = 0 | r_{k-1}^j = 0, \dots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B)$ を最大エントロピー法により推定する際、3. の分析に基づき、以下に示す素性を用いた。

[形態素情報]

- 文節 b_k^j の主辞（品詞、活用形）と語形（品詞）

[節境界情報]

- b_k^j の直後に節境界があるか否か
- b_k^j の直後の節境界の種類（節境界がある場合）

[係り受け情報]

- b_k^j が直後の文節に係るか否か
- b_k^j が節末文節に係るか否か
- b_k^j が行頭からの文字数が最長文字数以内の位置にある文節に係るか否か
- b_k^j が連体節の節末文節から係られるか否か
- b_k^j が直前の文節から係られるか否か
- 行頭文節 b_1^j から b_k^j までの間で係り受けが閉じているか否か

- b_k^j の右側で、かつ、行頭からの文字数が最長文字数以内の位置にある文節の中で、 b_k^j と同じ係り先をもつ文節があるか否か

[行長]

- 行頭から b_k^j までの文字数が以下の3分類のいずれであるか^(注6)
 - 2文字以下
 - 3文字以上6文字以下
 - 7文字以上

[ポーズ情報]

- b_k^j の直後にポーズがあるか否か

[文節の第1形態素]

- b_k^j の直後の文節の第1形態素の表層文字が「する、なる、思う、問題、必要」のいずれか、若しくはその品詞が「名詞-非自立-一般、名詞-非自立-副詞可能、名詞-ナイ形容詞語幹」のいずれかであるか否か

5. 実験

本手法の有効性を評価するため、日本語講演データを用いて改行挿入実験を実施した。

5.1 実験概要

実験データとして、名古屋大学同時通訳データベース[14]に収録されている日本語講演音声の書き起こしデータを使用した。すべてのデータに、形態素情報、係り受け情報、節境界情報が人手で付与されている。実験は、全16講演を用いた交差検定により実施した。すなわち、1講演をテストデータとし、残りの15講演

を学習データとして改行点の同定処理を実行した。ただし、16講演のうち2講演は事前分析データとして使用したため評価データから取り除き、残りの14講演(1,714文、20,707文節)に対する実験結果に基づいて評価した。なお、実験のための最大エントロピー法のツールとしては、文献[20]のものを利用した。オプションに関しては、学習アルゴリズムにおける繰り返し回数を2,000に設定し、それ以外はデフォルトのまま使用した。

評価は、正解の改行点に対する再現率及び適合率により行った。再現率、適合率はそれぞれ、

$$\text{再現率} = \frac{\text{正しく挿入された改行数}}{\text{正解の改行数}}$$

$$\text{適合率} = \frac{\text{正しく挿入された改行数}}{\text{挿入された改行数}}$$

を測定した。

比較のために、以下の四つのベースラインを設けた。

- ベースライン1: 最長文字数を超えない最右の文節境界を改行点とする(文節境界に基づく改行)。
- ベースライン2: 節境界を改行点とする。ただし、最長文字数内に節境界がなければ、その最右の文節境界を改行点とする(節境界に基づく改行)。
- ベースライン3: 係り受け関係のない隣接文節間を改行点とする。ただし、最長文字数内に係り受け関係のない隣接文節がなければ、その最右の文節境界を改行点とする(係り受け関係に基づく改行)。
- ベースライン4: ポーズが存在する文節境界を改行点とする。ただし、最長文字数内にポーズが存在する文節境界がなければ、その最右の文節境界を改行点とする(ポーズに基づく改行)。

実験では、1行の最長文字数を20文字とした。正解の改行データは、2.の調査に使用したデータと同様の手順で作成した。正解データの例を図10に示す。評価データ全体で改行点は5,497箇所存在した。

5.2 実験結果

提案手法並びに各ベースラインの適合率と再現率を表4に示す。提案手法は、再現率で82.66%、適合率で80.24%を達成した。これらの調和平均であるF値の比較において最も高い性能を示しており、提案手法の有効性を確認した。

再現率においては、ベースライン3が最も高かった。

(注6): 3.3の分析において、2文字以下の行はほとんど観察されなかったことから、このような3分類を採用した。

それから二番目に
先程伊藤さんからもお話ございましたように
今年は何戦五十年ということで
特別な年でございますので
それに関する事を
若干話させて頂きたいと思ます

それから現在我々が住んでおります
冷戦後の世界というものは
どういふものかという点につきまして
私の考えを述べさせて頂きたいと思ます

そして最後に
二十一世紀の日本外交なんて言ってしまうと
若干後悔しているんですが
二十一世紀といっても
五十年百年後というところは
予測が不可能でございますが
二十一世紀の初めの方は
どうなるのだろうか
またその二十一世紀に入って
我々としては
どうすべきかということについて
私なりの考えを話させて頂きたいと思ます

図 10 正解データの例

Fig. 10 Example of the correct data.

表 4 実験結果

Table 4 Experimental results.

	recall	precision	F-measure
our method	82.66% (4,544/5,497)	80.24% (4,544/5,663)	81.43
baseline 1	27.47% (1,510/5,497)	34.51% (1,510/4,376)	30.59
baseline 2	69.35% (3,812/5,497)	48.66% (3,812/7,834)	57.19
baseline 3	89.49% (4,919/5,497)	53.73% (4,919/9,155)	67.14
baseline 4	69.84% (3,893/5,497)	55.60% (3,893/6,905)	61.91

これは、正解データにおいて、互いに係り受け関係にある隣接文節間には改行が挿入されにくいという事実を反映している。しかし、その一方で、係り受け関係がないあらゆる文節間に改行を挿入することになるため、他の手法に比べて挿入される改行数が多く、その分、適合率が低いという結果になった。

ベースライン 2 及び 4 については、再現率、適合率ともに、ベースライン 1 を上回ったものの、提案手法と比べると低い値であった。このことは、節境界やポーズの出現位置などは、改行点の同定に有効な情報であるものの、それらを単独で利用するだけでは適切な位置に改行を挿入することは難しいということを示唆している。

5.3 改行挿入誤りの分析

実験における改行挿入の誤りについて分析した。誤りが生じる理由は複合的であり、必ずしも単一の原因として特定できるわけではないが、誤りの箇所に関して以下の傾向が観察された。

まず、提案手法によって挿入された改行のうち、正解の位置と異なるものは 1,119 箇所存在し、そのうち、特に、節境界「連体節」に誤って改行を挿入する傾向が観察された。3.1 の分析でも示したように、「連体節」は本来、改行が挿入されにくい箇所であるが、実験結果では改行が多く挿入されていた。例えば、

- 最近ではカセットよりもミニディスクを持つ人が多くなってきました

「持つ」と「人が」の間に改行が挿入されており、「ミニディスクを持つ人が」というまとまりがとらえられていない。このような改行挿入誤りが全体の 10.19% を占めている。

一方で、正解の改行点のうち、提案手法によって改行が挿入されなかった文節境界は 953 箇所であり、その約 11.1% は、節境界「主題八」であった。例えば、提案手法による改行挿入結果

- 八年位でたぶん私は隠居することになると
思いますので

では、「私は」の直後で改行が挿入されなかったものの、正解データでは「私は」の直後が改行点であった。節境界「主題八」は他の種類の節境界に比べ、出現数が多い。また、表 2 に示したとおり改行が挿入される割合も大きい。このため、「主題八」において正しく改行を挿入できることが高い再現性の実現に寄与するといえる。

6. 考 察

本章では、提案手法の有効性について、より詳細に検討するために、5. の実験結果について考察する。

6.1 改行挿入結果の主観的評価

本研究の目的は、改行を挿入することにより講演テキストを読みやすくすることにある。そこで、被験者によるテキストの主観的評価を実施した。

評価では、改行点のみ異なる 2 種類のテキストを提示し、被験者が読みやすい方を選択することにより行った。提案手法の比較対象として、前章の実験で設定したベースラインのうち、F 値が最も高かったベースライン 3 を使用し、ランダムに選んだ 50 文に対する改行挿入結果を並べて提示した。評価は 10 人の被

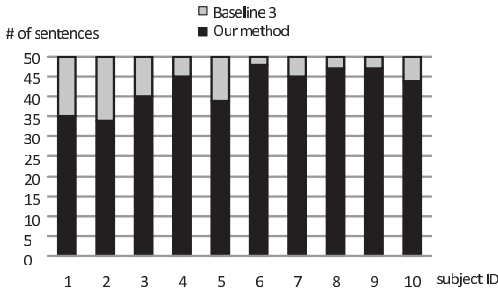


図 11 被験者による主観的評価の結果
Fig. 11 Result of subjective evaluation.

実は私もですねかくゆう私も
大学生の頃はよくキセルをしております
捕まったものです

図 12 平仮名が連続する場合の例
Fig. 12 Example of succession of hiragana.

私は残り少なくなったエネルギー資源を
巡って
過去と未来の人間たちが戦いを繰り広げる
エスエフ小説を書いています

図 13 極端に長さが違う行の出現
Fig. 13 Lines that have extremely different length.

験者が行った。

結果を図 11 に示す。グラフは、選択されたテキストの、被験者ごとの内訳を表している。提案手法によって改行されたテキストを選択した割合は、最も高い人で 94%、最も低い人でも 68%であり、読みやすい講演テキストの生成における提案手法の効果が示された。

一方で、半数以上の被験者が提案手法よりベースライン 3 による改行結果の方が読みやすいと判定した文が 3 文存在した。その 3 文について調べたところ、以下の現象の出現が、テキストが読みにくくなる要因となることが分かった。

- 平仮名が文節にまたがって連続して出現する。
- 隣り合う行との間で長さが著しく異なる行が出現する。

これらの要因を含む例を、図 12、図 13 にそれぞれ示す。図 12 では、1 行目に「ですね」と「かくゆう」という、それぞれ異なる文節に属する平仮名列が同一行に連続して表示されており、また、図 13 では、2 行目の行長が 1 行目や 3 行目と比べて極端に短くなっており、いずれも読みにくいテキストとなっている。

表 5 正解データの作成に携わっていない作業者による改行挿入

Table 5 Linefeed insertion by another person.

recall	precision	F-measure
89.82%	89.82%	89.82
(459/511)	(459/511)	

表 6 自動的に言語解析されたデータに対する実験結果

Table 6 Experimental results on automatically annotated data.

	recall	precision	F-measure
our method	77.37%	75.04%	76.18
	(4,253/5,497)	(4,253/5,668)	
baseline 1	27.47%	34.51%	30.59
	(1,510/5,497)	(1,510/4,376)	
baseline 2	69.51%	48.63%	57.23
	(3,821/5,497)	(3,821/7,857)	
baseline 3	84.01%	52.03%	64.26
	(4,618/5,497)	(4,618/8,876)	
baseline 4	69.84%	55.60%	61.91
	(3,893/5,497)	(3,893/6,905)	

6.2 人間による改行挿入の一致率

5. では、正解データとの比較によって、改行挿入の結果を評価したが、テキストが読みやすくなるための適切な改行位置は、人ごとに必ずしも一致するわけではない。そこで、人間による改行点の一致の程度を測定し、実験結果と比較することにより、提案手法の改行性能を評価した。

テストデータの 1 講演に相当する 128 文 (511 文節) に対して、正解データの作成に携わっていない作業者 1 名が改行挿入を行った。正解データに対する再現率、適合率とその F 値を表 5 に示す。提案手法は F 値において人間による改行の 90.65% (81.43/89.82) を達成している。

6.3 テキストの自動解析に基づく改行挿入性能

実験で比較した各手法の間で、利用している言語情報に差があるため、より公平に比較するためには、共通して利用されているわけではない言語情報については自動解析によって付与し、用いるべきである。そこで、節境界解析並びに係り受け解析を機械的に実行した結果に基づいて、5. に記した実験を実施し、提案手法とベースラインとの性能比較を行った。なお、係り受け解析には CaboCha [19] を、節境界検出には節境界解析ツール CBAP [12] をそれぞれ用いた。

実験結果を表 6 に示す。人手で付与した情報を利用した実験結果 (表 4) と比べ、提案手法は、再現率、適合率とも約 5% 低下したものの、F 値による比較に

表 7 西光らの手法による実験結果

Table 7 Experimental results by Saiko's method.

recall	precision	F-measure
73.08%	58.28%	64.84
(4,017/5,497)	(4,017/6,893)	

においてベースラインとの差は著しく、改行挿入手法としての利用可能性を確認した。

6.4 チャンキングに基づく改行挿入との比較

講演などの話し言葉を適当な単位に分割する手法として、西光らは、段階的チャンキング方式を提案している [21]。この手法では、音声言語の係り受け解析が、隣接文節間の係り受け関係の検出については音声認識結果に対して頑健に実行できることに着目し、そのような係り受け関係、更には、述語要素、ポーズ、フィラーなどを考慮して、言語的まとまり（チャンク）を階層的に生成する。この手法は、生成されたチャンクに基づいて最長文字数を超えない長さまで文字を連結するという方式を導入することにより、字幕テキストにおける改行挿入に利用することができる [22]。

上記の先行研究は、音声言語の単位としてのチャンクの検出を、音声認識結果に対して頑健に実行するための手法の提案であり、書き起こされた音声言語テキストにおける最適な改行位置を同定することを目的とする本研究との間で単純に比較できるわけではないが、本手法による改行挿入の性能について更に考察を与えるために、段階的チャンキングに基づく改行挿入手法 [22] を実装し、5. と同様のテストデータを用いて比較実験を行った。なお、最長文字数はいずれも 20 文字で統一した。

西光らの手法に基づく改行挿入の実験結果を表 7 に示す。提案手法による実験結果（表 4）と比べて、再現率で約 10%、適合率で約 22%低下するという結果になった。特に、適合率において差が大きく、これは、西光らの手法では、ポーズやフィラーが存在するチャンクの境界に必ず改行を挿入していることが理由として考えられる。テストデータにおいて、チャンク境界に存在したポーズとフィラーは合わせて 4,851 個あったが、正解データではそのうち 64.05%にしか改行が挿入されておらず、そのようなチャンク境界が改行点と単純には一致しないことが分かる。また、チャンクは隣接文節間の係り受けに着目して生成されているため、単純にチャンクを結合するだけでは、不自然な位置に改行が挿入されることがあった。

それから千九百六十年には日米の新しい安保条約が締結されまして安保条約の上でわが国の発言権がより強くなったということがいえると思います

図 14 西光らの手法による改行挿入の例

Fig. 14 Example of linefeed insertion by Saiko's method.

それから千九百六十年には日米の新しい安保条約が締結されまして安保条約の上でわが国の発言権がより強くなったということがいえると思います

図 15 提案手法による改行挿入の例

Fig. 15 Example of linefeed insertion by our method.

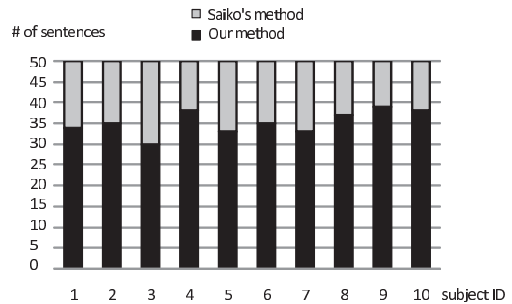


図 16 西光らの手法と提案手法の主観的評価の結果

Fig. 16 Result of subjective evaluation by comparison between Saiko's method and our method.

一方、提案手法は、多くの言語情報を用いた統計的手法に基づくため、よりきめ細やかな改行が実現されている。段階的チャンキングによる改行挿入の例（図 14）では、「日米の」や「より」の直後に改行が挿入されているものの、提案手法による改行挿入結果（図 15）では、「より強くなったということが」のような意味的なまとまりをとらえた改行挿入が実現できている。

主観的評価による比較を行った。5.3 と同様の手順で、名古屋大学同時通訳データベース [14] からランダムに選択した 50 文に対して、提案手法により改行が挿入されたテキストと、西光らの手法によるテキストを用意し、被験者は読みやすい方を選択した。評価結果を図 16 に示す。いずれの被験者においても提案手法によるテキストの方を読みやすいとした文の方が多く、提案手法の改行挿入性能の高さが示された。

7. む す び

本論文では、聴覚障害者、高齢者、外国人等による音声理解の支援を目的に、日本語講演データへの改行挿入手法を提案した。本手法では、係り受け、節境界、ポーズ、行長等の情報に基づき、統計的手法によって読みやすい位置への改行挿入を実現する。日本語講演の書き起こしデータを用いた改行挿入実験では、再現率で82.66%、適合率で80.24%を示しており、本手法の有効性を確認した。

本論文では、講演の書き起こしテキストに対して、適切な位置に改行を挿入する手法について述べたが、実際のリアルタイム字幕生成に応用するためには、音声認識の利用を前提とした、より実践的な方式を検討する必要がある。そのために、まず、認識誤りを含む文に対しても適切に改行を挿入し提示する必要がある。また、字幕提示のリアルタイム性を高めるためには、音声認識によって順次生成される、文の途中までの入力に対して、改行点を動的に決定することが必須である。本論文で提案した手法を拡張し、音声入力に対して漸進的かつ頑健な改行挿入手法を実現することは今後の課題である。

また本論文では、改行挿入によりテキストを読みやすくすることについて検討したが、このほかに、読点を適切な位置に挿入することが考えられる。適切な改行位置と読点位置は互いに関係するものの、それらの位置が必ず一致するというわけではない。読みやすさという観点では、読点と改行を併用することが効果的であると考えられ、それらを組み合わせる挿入手法について、今後検討していきたい。

謝辞 本研究は、一部、科学研究費補助金（基盤研究（B））（No. 20300058）、並びに、財団法人旭硝子財団研究助成（若手継続 grant）により実施したものである。

文 献

- [1] 今井 亨, 宮本晃太郎, “放送・教育における音声を利用した障害者支援,” 信学誌, vol.91, no.12, pp.1024–1029, Dec. 2008.
- [2] G. Boulianne, J.-F. Beaumont, M. Boisvert, J. Brousseau, P. Cardinal, C. Chapdelaine, M. Comeau, P. Ouellet, and F. Osterrath, “Computer-assisted closed-captioning of live TV broadcasts in French,” Proc. 9th International Conference on Spoken Language Processing, no.Mon2A20-1, pp.273–276, Pittsburgh, Pennsylvania, USA, Sept. 2006.
- [3] J. Xue, R. Hu, and Y. Zhao, “New improvements in decoding speed and latency for automatic captioning,” Proc. 9th International Conference on Spoken Language Processing, no.Wed1CaP-8, pp.1630–1633, Pittsburgh, Pennsylvania, USA, Sept. 2006.
- [4] T. Holter, E. Harborg, M.H. Johnsen, and T. Svendsen, “Asr-based subtitling of live TV-programs for the hearing impaired,” Proc. 6th International Conference on Spoken Language Processing, pp.570–573, Beijing, China, Oct. 2000.
- [5] M. Saraclar, M. Riley, E. Bocchieri, and V. Goffin, “Towards automatic closed captioning: Low latency real time broadcast news transcription,” Proc. 7th International Conference on Spoken Language Processing, no.ThB39o.1, pp.1741–1744, Denver, Colorado, USA, Sept. 2002.
- [6] T. Imai, S. Sato, A. Kobayashi, K. Onoe, and S. Homma, “Online speech detection and dual-gender speech recognition for captioning broadcast news,” Proc. 9th International Conference on Spoken Language Processing, no.Wed1CaP-1, pp.1602–1605, Pittsburgh, Pennsylvania, USA, Sept. 2006.
- [7] C. Munteanu, G. Penn, and R. Baecker, “Web-based language modelling for automatic lecture transcription,” Proc. 8th Annual Conference of the International Speech Communication Association, no.ThD.P3a-2, pp.2353–2356, Antwerp, Belgium, Aug. 2007.
- [8] 中野聡子, 牧原 功, 金澤貴之, 中野泰志, 新井哲也, 黒木速人, 井野秀一, 伊福部達, “音声認識技術を用いた聴覚障害者向け字幕表示システムの課題—話し言葉の性質が字幕の読みに与える影響,” 信学論 (D), vol.J90-D, no.3, pp.808–814, March 2007.
- [9] 特定非営利活動法人全国要約筆記問題研究会調査研究委員会, 中途失聴・難聴者等聴覚障害者のコミュニケーションに関する現状把握調査・研究事業報告書, March 2008.
- [10] 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,” 信学論 (D-II), vol.J84-D-II, no.6, pp.888–897, June 2001.
- [11] 西光雅弘, 秋田祐哉, 高梨克也, 尾嶋憲治, 河原達也, “局所的な係り受けの情報を用いた話し言葉の節・文境界の推定,” 情処学論, vol.50, no.2, pp.544–552, Feb. 2009.
- [12] 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, July 2004.
- [13] A. Molina and F. Pla, “Clause detection using HMM,” Proc. 2001 workshop on Computational Natural Language Learning - Volume 7, no.25, pp.70–72, Toulouse, France, July 2001.
- [14] S. Matsubara, A. Takagi, N. Kawaguchi, and Y. Inagaki, “Bilingual spoken monologue corpus for simultaneous machine interpretation research,” Proc. 3rd Language Resources and Evaluation Conference, pp.153–159, Las Palmas, Canary Islands, Spain, May 2002.

- [15] 中野聡子, 金澤貴之, 牧原 功, 黒木速人, 上田一貴, 井野秀一, 伊福部達, “音声認識技術を利用した字幕呈示システムの活用に関する研究—聴覚障害者のニーズに即した呈示方法?” メディア教育研究, vol.5, no.2, pp.63-72, Dec. 2008.
 - [16] 技術評論社編集部, 今すぐ使えるかんたん PowerPoint 2007, 技術評論社, 東京, 2008.
 - [17] 杉藤美代子, “談話におけるポーズとイントネーション?” 日本語と日本語教育 2, 杉藤美代子 (編), pp.343-364, 明治書院, 東京, 1988.
 - [18] 浅原正幸, 松本裕治, “IPADIC ユーザーズマニュアル” version 2.5.1 2002.
 - [19] L. Zhang, “Maximum entropy modeling toolkit for python and c++,” http://homepages.inf.ed.ac.uk/s0450736/maxent_toolkit.html, 2007, [Online; accessed 6-September-2007].
 - [20] T. Kudo and Y. Matsumoto, “Japanese dependency analysis using cascaded chunking,” Proc. 6th Conference on Natural Language Learning, pp.63-69, Taipei, Taiwan, Aug. 2002.
 - [21] 西光雅弘, 高梨克也, 河原達也, “係り受けとポーズ・フィラーの情報を用いた話し言葉の段階的チャンキング” 情処学研報, vol.2005, no.127, pp.247-252, Dec. 2005.
 - [22] 西光雅弘, 河原達也, 高梨克也, “隣接文節間の係り受け情報に着目した話し言葉のチャンキングの評価” 情処学研報, vol.2006, no.40, pp.19-24, May 2006.
- (平成 20 年 12 月 22 日受付, 21 年 4 月 13 日再受付)



松原 茂樹 (正員)

1993 名工大・工・電気情報卒。1998 名大大学院博士課程了。博士(工学)。同年, 同大助手を経て, 2002 名古屋大学情報連携基盤センター助教授。現在, 情報基盤センター准教授。この間, 日本学術振興会特別研究員, ATR 音声言語コミュニケーション研究所客員研究員, 情報通信研究機構研究員。自然言語処理, 音声言語処理, 情報検索, デジタル図書館の研究に従事。情報処理学会, 人工知能学会, 言語処理学会, IEEE, ACM 各会員。



村田 匡輝 (学生員)

2008 名大・工・情報卒。現在, 同大大学院情報科学研究科博士前期課程在学中。自然言語処理に関する研究に従事。



大野 誠寛 (正員)

2003 名大・工・情報卒。2007 同大大学院情報科学研究科博士後期課程了。博士(情報科学)。同年名古屋大学大学院国際開発研究科助教。2006~2007 日本学術振興会特別研究員。自然言語処理, 音声言語処理の研究に従事。情報処理学会, 言語処理学会各会員。