講演テキストにおける読みやすさを考慮した改行挿入とその評価

村田 匡輝† 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

1 はじめに

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いる [1]．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 [2, 3]．特に，講演では文が長くなる傾向にあり，講演音声をテキストで提示する場合に，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，形態素列のパタンにより改行位置を決定する手法を提案している [4]．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替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．

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%，精度で 80.24%を達成した．改行挿入結果の被験者による主観的評価により，本手法の有効性を確認した．

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．図 1 に，想定する字幕提示環境を示す．

テレビ番組のクローズドキャプションの場合，通常，画面下部に 2 行程度の字幕が表示され，発声の進行に合わせて表示が切り替わる．一方，本研究では，テキストが行単位で入れ替わり，スクロールしながら常に数行表示される字幕提示システムの利用を前提とする．

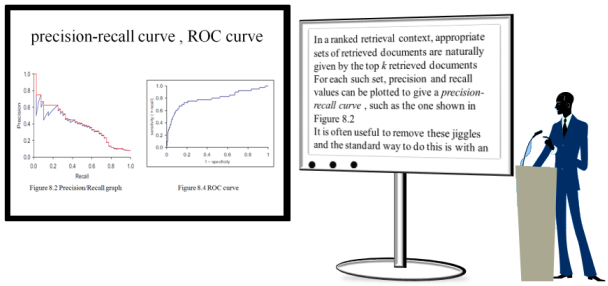

図 1: 講演音声の字幕提示環境

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 [5] 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，

（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．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．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．

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．

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 2: 講演音声の書き起こしテキスト

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

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 改行挿入手法

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．

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 行あたりの文字数が最長文字数を超えないという条件の下，1 文中に挿入されうる改行点の全ての組み合わせの中から，最適な組み合わ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．

3.1 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．

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

ここで，$P(r_k^j = 1|r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B)$ 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，$P(r_k^j = 0|r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B)$ 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．最尤の改行結果は，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．

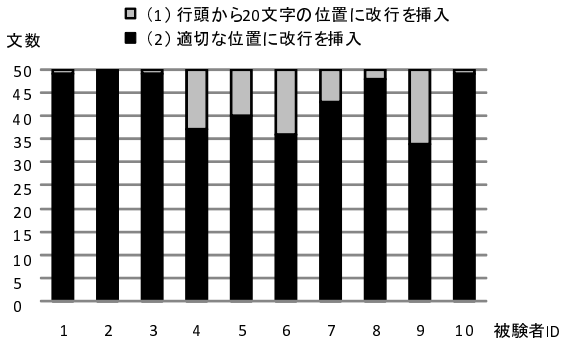

図 4: 講演テキストへの改行挿入の効果

3.2 最大エントロピー法で用いた素性

本研究では，$P(r_k^j = 1|r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B)$ ならびに $P(r_k^j = 0|r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B)$ を最大エントロピー法により推定する．そのための有効な素性に関する分析結果 [6] に基づき，$P(r_k^j = 1|r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B)$ ならびに $P(r_k^j = 0|r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B)$ を推定する際の素性を以下のように決定した．

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 の直後の節境界のラベル（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大表示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大表示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

- 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

文節の第一形態素

b_k^j の直後の文節の第一形態素の基本形が「する，なる，思う，問題，必要」のいずれか，もしくはその品詞が「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

4 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．

4.1 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 [5] 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 講演は最大エントロピー法に用いる素性決定のための分析に用い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 [7] 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精度により行った．再現率，精度はそれぞれ，

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

表 1: 実験結果

	再現率	精度	F 値
提案手法	82.66% (4,544/5,497)	80.24% (4,544/5,663)	81.43
ベースライン 1	27.47% (1,510/5,497)	34.51% (1,510/4,376)	30.59
ベースライン 2	69.35% (3,812/5,497)	48.66% (3,812/7,834)	57.19
ベースライン 3	89.49% (4,919/5,497)	53.73% (4,919/9,155)	67.14
ベースライン 4	69.84% (3,893/5,497)	55.60% (3,893/6,905)	61.91

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

を測定した．

比較のために，以下の 4 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（節境界に基づく改行）．

ベースライン 3：係り受け関係にない隣接文節間を改行点とする（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（ポーズに基づく改行）．

実験では，一行の最長文字数を 20 文字とした．正解の改行データは，人手で改行を付与することにより作成した．評価データ全体で改行点は 5,497 箇所存在した．

4.2 実験結果

提案手法ならびに各ベースラインの精度と再現率を表 1 に示す．提案手法は，再現率で 82.66%，精度で 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．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，係り受け関係に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，精度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，精度ともに，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．

4.3 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

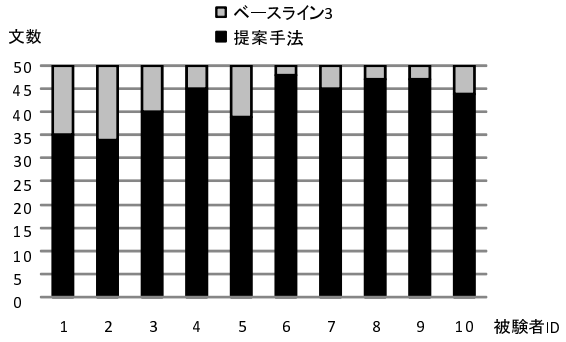

図 5: 被験者による主観的評価の結果

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して
捕まったものです

図 6: 平仮名が連続する場合の例

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した

図 7: 極端に長さが違う行の出現

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 人の被験者が行った．

結果を図 5 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わかった．

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 6，7 にそれぞれ示す．図 6 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 7 では，2 行目の行長が 1 行目や 3 行目と比べて極端に短かくなっており，いずれも読みにくいテキストとなっている．

5 おわりに

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の提案，及びその評価を行った．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%，精度で 80.24%を示した．被験者によるテキストの主観的評価により，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．特に，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，適切な改行位置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，漸進的な改行挿入手法を実現することは今後の課題である．

謝辞 本研究は，一部，科学研究費補助金（基盤研究 (B)）(No. 20300058)，ならびに，財団法人旭硝子財団研究助成により実施したものである．

参考文献

[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，電子情報通信学会誌，Vol.91, No.12, pp.1024-1029 (2008).

[2] 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呈示システムの課題 ―話し言葉の性質が字幕の読みに与える影響―，電子情報通信学会論文誌，Vol.J90-D, No.3, pp.808-814 (2007).

[3] 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 (2008).

[4] 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).

[5] S. Matsubara, A. Takagi, N. Kawaguchi and Y. Inagaki: Bilingual Spoken Monologue Corpus for Simultaneous Machine Interpretation Research, Proc. 3rd LREC, pp.153-159 (2002).

[6] 村田匡輝，大野誠寛，松原茂樹：講演テキストにおける読みやすさを考慮した改行位置同定，情報処理学会研究報告，Vol.NL-188, pp.37-44 (2008).

[7] 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].