

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

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

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るため、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。その際、テキストが読みやすくなる位置に改行が挿入されている必要がある。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、行長等の情報に基づき、統計的手法によって改行位置を決定する。日本語講演データを使用した実験によって本手法の有効性を確認した。

Linefeed Insertion Considering Readability of Lecture Transcript

MASAKI MURATA[†] and TOMOHIRO OHNO[‡] and SHIGEKI MATSUBARA[§]

[†]Graduate School of Information Science, Nagoya University

[‡]Graduate School of International Development, Nagoya University

[§]Information Technology Center, Nagoya University

The development of a captioning system that supports the real-time understanding of spoken monologue such as lectures and commentaries is required. In monologues, since a sentence tends to be long, each sentence is often displayed in multi lines on the screen, it is necessary to insert linefeeds into a text so that the text becomes easy to read. This paper proposes a technique for inserting linefeeds into a Japanese spoken monologue as an elemental technique to generate the readable captions. Our method appropriately inserts linefeeds into a sentence by machine learning, based on the information such as dependencies, clause boundaries and line length. An experiment using Japanese speech data has shown the effectiveness of our technique.

1 はじめに

リアルタイム字幕生成とは、講演などの音声ドキュメント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ための技術である。講演では文が長くなる傾向にあるため、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。このため、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しい。

音声からの字幕生成に関する研究はすでにいくつ

が存在している [1, 2, 3, 4, 5]。このうち、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 [5]。ただし、この研究では、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、我々の研究が想定する複数行の字幕のスクロールによる表示環境とは、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

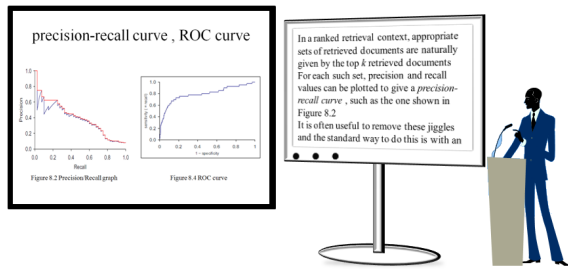


図 1: 講演音声の字幕提示環境

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、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。比較のために設定した 4 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、以下の通りである。2 章では、本研究のめざす字幕表示の概要、および改行位置の分析について述べる。3 章では、本研究が提案する改行挿入手法について述べる。4 章では、本手法による改行挿入実験について報告する。5 章では、本研究のまとめと今後の課題について述べる。

2 講演音声の字幕生成

2.1 講演音声への改行挿入

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 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

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字幕テキスト

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。

本研究では、字幕生成における読みやすい改行挿入として、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。
- 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。

なお、文節は、日本語における意味的なまとまりの基本単位であるため、本手法では、文節境界を改行位置の候補とする。また、本論文の以下では、改行が挿入される文節境界を改行点という。

2.2 改行点の分析

本研究では、統計的手法により改行点を決定する。そのために、音声言語コーパスを用いて分析を与えた。分析には、名古屋大学同時通訳データベース [6] の日本語講演音声データを用いた。データには、形態素、文節境界、係り受け構造、節境界等の構文的情報、ならびに、改行点が、人手により付与されている。分析に使用したデータの規模を表 1 に示す。

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。

なお、改行は、行あたりの文字数が最大 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、

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

節境界	改行挿入率 (%)
主題八	50.8
談話標識	12.0
引用節	22.1
連体節	23.3
テ節	90.2
補足節	68.0
並列節ガ	100.0
並列節ケレドモ	100.0
条件節ト	93.5
連体節トイウ	27.3

急ぐ / 連体節 / 人の 為に エスカレータの 右側を
開ける / 連体節 / 法則 これは 首都圏ルールです
が / 並列節ガ / 関西では 全く 逆ですよ

↓ 改行挿入

急ぐ人の為に
エスカレータの右側を開ける法則
これは首都圏ルールですが
関西では全く逆ですよ

図 4: 節境界と改行点の関係

ディスプレイ上での文字の可読性を考慮して設定した。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。

2.2.1 節境界と改行点

節 (clause) は意味的なまとまりを構成する言語単位の 1 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 箇所に改行が挿入されており、挿入率は 51.1%であった¹。文節境界に対する挿入率よりも高いことから、節境界に改行が挿入されやすいことがわかった。

分析データに出現した 42 種類の節境界について、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 4 に節境界と改行点の関係を示す。図 4 において、改行挿入率が低い節境界「連体節」には改行が挿入されていない。一方、改行挿入率が 100%である節境界「並列節ガ」には改行が挿入されている。

2.2.2 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 2 つの文節で意味的なまとまりを形成するため、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。分析データを調査したところ、係り受け関係にある隣接する文節間の境界 1459 箇所に対して、改行が挿入されたのは 192 箇所であり、挿入率は 13.2%であった。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%である。

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。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。図 5 に係り受け構造と改行点の関係を示す。図 5 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、改行が挿入されているのは、係り受け関係にない隣接文節（「記者が」と「私の」）の境界である。また、改行を挿入した直前の文節「記者が」以外の文節は隣接する文節と係り受け関係にあるため、改行挿入の結果生成された各行は、行内で係り受けが閉じている。

¹本研究では、節境界の種類として節境界解析ツール CBAP[7] で定義されたものを用いた。

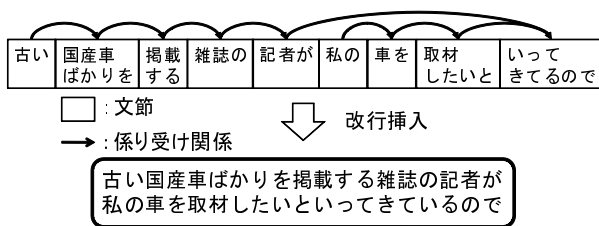


図 5: 係り受け構造と改行点の関係

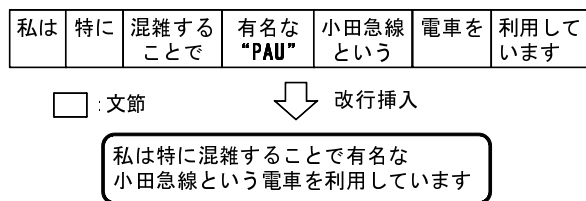


図 6: ポーズと改行点の関係

2.2.3 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、行の文字数が6文字以下となっていた行は全体の7.59%に過ぎなかった。これは、行頭から行末までの文字列がある程度の長さを持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。

2.2.4 ポーズと改行点

ポーズは構文的区切りと一致するという知見が得られている[8]。すなわち、文節境界にポーズが存在する場合に改行が挿入されやすくなる可能性がある。本研究では、200msec.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おり、改行挿入率は62.97%であった。これは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わかった。図6にポーズと改行点の関係を示す。図6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。

2.2.5 行頭の形態素

形態素によっては、行の先頭に出現すると読みにくい字幕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%以下だった形態素には、以下のものがあつた。括弧中の数字は、行頭での出現頻度と分析データ全体での出現頻度を示している。

- 基本形
思う (2/70), 問題 (0/42), する (3/33), なる (2/32), 必要 (1/21)
- 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)

以上から、これらの形態素を第一形態素に持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。

3 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。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、 $k \neq n_j$ ($1 \leq k \leq n_j$) 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。

$$\begin{aligned}
 P(R|B) & \quad (1) \\
 &= P(r_1^1 = 0, \dots, r_{n_1}^1 = 1, \dots, r_1^m = 0, \dots, r_{n_m}^m = 1|B) \\
 &\cong P(r_1^1 = 0|B) \times P(r_2^1 = 0|r_1^1 = 0, B) \times \dots \\
 &\quad \times P(r_{n_1}^1 = 1|r_{n_1-1}^1 = 0, \dots, r_1^1 = 0, B) \times \dots \\
 &\quad \times P(r_1^m = 0|r_{n_{m-1}}^{m-1} = 1, B) \times \dots \\
 &\quad \times P(r_m^m = 1|r_{n_{m-1}}^{m-1} = 0, \dots, r_1^m = 0, r_{n_{m-1}}^{m-1} = 1, B)
 \end{aligned}$$

ここで、 $P(r_k^j = 1 | r_{k-1}^j = 0, \dots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B)$ 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、 $P(r_k^j = 0 | r_{k-1}^j = 0, \dots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B)$ 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。

3.2 最大エントロピー法における素性

本研究では、 $P(r_k^j = 1 | r_{k-1}^j = 0, \dots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B)$ ならびに $P(r_k^j = 0 | r_{k-1}^j = 0, \dots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B)$ を最大エントロピー法により推定する際、2.2 節の分析に基づき、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+1}^j の第一形態素の表層文字が「思う、問題、する、なる、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-ナイ形容詞語幹、名詞-非自立-副詞可能」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 [6] 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挿入処理を実行した。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 [10] 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、さらには、すべての改行点が一致した文数の割合（以下、文正解率）により行った。再現率、精度はそれぞれ、

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

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

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。

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）。
2. すべての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）。
3. 係り受け関係がない隣接文節間を改行点とする。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（係り受け関係に基づく改行）。
4. ポーズが存在するすべての文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（ポーズに基づく改行）。

それから二番目に
先程伊藤さんからもお話をいたしましたように
今年は終戦五十年ということで
特別の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っております

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
どういう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っております

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っております

図 7: 正解データの例

実験では、一行の最長文字数を 20 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で作成した。正解データの例を図 7 に示す。評価データ全体で改行点は 5,497ヶ所存在した。

4.2 実験結果

ベースラインならびに提案手法に対する実験結果を表 3 に示す。

ベースライン 1 では、最右の文節境界に改行を挿入するという単純な手法のため、再現率、精度ともベースライン手法の中で最も低くなっている。

ベースライン 2 は、精度が低くなっている。事前分析でも述べた通り、節境界には改行が挿入されやすい種類と改行が挿入されにくい種類が存在する。全ての節境界に改行を挿入する方法では、改行が挿入されにくい節境界にも改行が挿入されてしまうため、余分な改行が多く挿入される。

ベースライン 3 では高い再現率が達成されている。これは、隣接する文節が互いに係り受け関係にある場合に、その境界に改行が入りにくいことが理由と

表 3: 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

なっている。しかし、係り受け関係にない全ての隣接文節間に改行を挿入するため、他の手法に比べて改行挿入数が多くなり、精度が低くなっている。

ベースライン 4 は再現率、精度とも高い値にはなっていない。ポーズが存在する文節境界と、改行点が必ずしも一致しないことが原因であると考えられる。

一方、提案手法の F 値（精度と再現率の調和平均）は 81.43 であった。すべてのベースラインと比べ高い性能を示しており、本手法の有効性を確認した。

4.3 考察

4.3.1 改行挿入誤りの原因

提案手法による改行挿入誤りの原因を調査した。改行挿入誤り 1,119 箇所中、原因として多く見られたのは、節境界「連体節」への改行であった。「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。このような改行挿入結果の例を以下に示す。

- 以上がこの第一期と私が勝手に呼んでる時期でございます

この例では「呼んでる」の直後に存在する節境界「連体節」に改行が挿入されており、「呼んでる時期でございます」というまとまりが捉えられていない。これは、改行挿入誤りの 10.19% を占めている。

また、正解データでは 1 行であるが、実験結果では 2 行に分割している改行挿入が 291 箇所存在した。そのような改行挿入結果の例を以下に示す。

- どこまで詳しくお話しできるか不安ですが
堅いお話しからやわらかいお話し
織り交ぜてお話ししていこうと思っております

正解データでは「どこまで詳しくお話しできるか不安ですが」が 1 行になっている。1 行にまとめることができる場合でも、提案手法では、改行が挿入され

表 4: 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

再現率	精度	F 値	文正解率
89.82% (459/511)	89.82% (459/511)	89.82	64.84% (83/128)

る傾向にあると言える。

4.3.2 人手による改行挿入の一致率

字幕を読みやすくするための改行点は人によって異なるため、複数の人間が同じテキストに対して改行を挿入した場合でも、異なる位置に改行を挿入する可能性がある。このため、人間同士による改行性能のずれを測定した数値は、自動改行挿入の上限と考えられる。そこで、実験データ 1 講演分 (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により改行挿入を行い、その改行点の正解データに対する再現率、精度、及び文正解率を測定した。結果を表 4 に示す。提案手法は F 値において上限値の 90.65%(81.43/89.82) を達成しており、自動的な改行挿入手法として有効であると言える。

4.3.3 テキストの自動解析に基づく改行挿入性能

実験では、人手で付与した言語情報に基づいて改行点を決定した。しかし、実際に字幕システムを運用する場合は、自動的に言語情報を付与したデータを用いて改行点を決定する必要がある。実際の状況でも本手法によって読みやすい字幕を生成可能であることを確認するため、言語情報を自動的に付与したデータに対して改行挿入実験を行った。ベースラインと提案手法との間で評価の基準を揃えるために、節境界情報、及び係り受け情報を自動で付与した²。実験結果を表 5 に示す。

表 5 の結果を表 3 と比較すると、基本的な言語情報しか利用していないベースラインと比べて、多くの言語情報を利用している提案手法の方が性能の低下が大きくなっているものの、提案手法は各ベースラインの手法より F 値において 10%以上高い数値を示した。

4.3.4 チャンキングによる改行挿入手法との比較

話し言葉を適当な単位に分割する方法として、西光らは、段階的チャンキングによる手法を提案している [11]。この手法では、隣接文節間の係り受け関係の検出が音声認識誤りに頑健であることに着目し、

²係り受け解析には CaboCha[9] を、節境界検出には節境界解析ツール CBAP[7] をそれぞれ用いた。

表 5: 自動的に情報を付与したデータに対する実験結果

	再現率	精度	F 値
提案手法	77.37% (4,253/5,497)	75.04% (4,253/5,668)	76.18
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	69.51% (3,821/5,497)	48.63% (3,821/7,857)	57.23
ベースライン 3	84.01% (4,618/5,497)	52.03% (4,618/8,876)	64.26
ベースライン 4	69.84% (3,893/5,497)	55.60% (3,893/6,905)	61.91

表 6: 西光らの手法による実験結果

再現率	精度	F 値
73.08% (4,017/5,497)	58.28% (4,017/6,893)	64.84

文の主題や述語、格要素などに相当する単位である「構成要素」を抽出する。隣接する構成要素間にポーズやフィラーが存在すれば、さらにそこで分割し、複数の構成要素からなる単位としてフレーズを抽出する。構成要素に基づいて節境界の推定を行うことにより、音声認識誤りに対して頑健な節境界解析が可能であることを示している。

構成要素及びフレーズへのチャンキングを字幕生成における改行挿入に応用することが提案されている。1 行の文字数が最長文字数になるまで構成要素の結合を繰り返し、最長文字数を超える直前で改行する。また、フレーズとフレーズの間で必ず改行している。これにより、均質的で意味的なまとまりとなる字幕を生成できるとしている。

提案手法と、西光らの段階的チャンキングに基づく手法の改行挿入性能を比較するために、西光らの手法を実装し、同様の実験を行った。構成要素の結合は 20 文字を越える直前までとした。西光らの手法による実験結果を表 6 に示す。表 3 に示した提案手法による実験結果と比べて、西光らの手法は再現率は約 10%、精度は約 22%低くなっており、特に、精度において大きく差が見られた。これは、西光らの手法において、フレーズ境界、すなわちポーズ・フィラーが存在する構成要素境界には必ず改行が挿入されるためであると考えられる。構成要素境界に存在したポーズ・フィラーの数は 4,851 個であったが、正解データでは、そのうち 64.05%にしか改行が挿入されていなかった。フレーズの境界が改行点と単純には一致しないことがわかる。

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 8: 西光らの手法による改行挿入結果

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 9: 提案手法による改行挿入結果

また、フレーズ境界以外での改行挿入については、構成要素へのチャンキングは隣接文節間の係り受けのみに着目しているため、単純に構成要素を結合する手法では不自然な位置に改行が挿入されることが多かった。一方、提案手法は多くの言語情報を用いた統計的手法であるため、より決め細やかな改行が実現されている。図 8 に示した例では、「日米の」や「より」の直後に改行が挿入されている。しかし、図 9 に示す提案手法による改行挿入結果では、「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捉えた改行挿入が実現できている。

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%、精度で 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、正解データを用いた評価について述べたが、今後は、実環境下での字幕提示による被験者実験を通して、本手法の実用性について検討することを予定している。

参考文献

- [1] 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 IC-SLP, No.Mon2A2O-1, pp.273-276, Pittsburgh, Pennsylvania, USA, Sept. 2006.
- [2] J. Xue, R. Hu and Y. Zhao, "New Improvements in Decoding Speed and Latency for Automatic Captioning," Proc. 9th IC-SLP, No.Wed1CaP-8, pp.1630-1633, Pittsburgh, Pennsylvania, USA, Sept. 2006.
- [3] W. Daelemans, A. Hothker and E.T.K. Sang, "Automatic Sentence Simplification for Subtitling in Dutch and English," Proc. 4th LREC, pp.1045-1048, Lisbon, Portugal, May. 2004.
- [4] T. Holter, E. Harborg, M.H. Johnsen and T. Svendsen, "Asr-based Subtitling of Live TV-Programs for the Hearing impaired," Proc. 6th ICSLP, pp.570-573, Beijing, China, Oct. 2000.
- [5] 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, Vol.J84-D-II, No.6, pp.888-897, Jun. 2001.
- [6] S. Matsubara, A. Takagi, N. Kawaguchi and Y. Inagaki, "Bilingual Spoken Monologue Corpus for Simultaneous Machine Interpretation Research," Proc. 3rd LREC, pp.153-159, Las Palmas, Canary Islands, Spain, May 2002.
- [7] 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, Jul. 2004.
- [8] 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.
- [9] <http://chasen.org/taku/software/cabocha/>.
- [10] 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].
- [11] 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," 情報処理学会研究報告, Vol.2006, No.40, pp.19-24, May 2006.