読みやすいテキスト提示のための
日本語文の整形に関する研究

村田 匡輝

概要

現代社会では，大量の情報が生産，蓄積されており，我々はそれらの情報の中から自分に必要なものを選択的に収集し，活用している．テキストは代表的な情報収集源であり，その媒体として書籍や新聞，雑誌など，紙に印刷されたテキスト，ならびに，Web ニュース，ブログ，メール，また，テレビ番組に付与される字幕など，ディスプレイやスクリーンに提示されるテキストがある．効率的な情報収集を可能にするために，テキストは読みやすいものであることが望ましい．テキストは一般に複数の文によって構成され，その文の内容や見た目がテキストの読みやすさに影響を与える．ディスプレイなどに提示されるテキストを構成する文は，印刷されたテキストのものと異なり，変更することができるため，文を適切に整形することによって読みやすいテキスト提示を行うことが可能となる．

文の整形とは，文の見た目を変更し体裁を整えることをいう．しかし，読みやすいテキストを提示するためには文の見た目以外に文の内容も重要となるため，文の内容と見た目の両方を対象として整形する必要がある．文の内容は文字列によって示されるため，元の文が伝えている情報そのものが変わることがないように，文の内容が大きく変更されない範囲で文字列を修正することが望まれる．また，文の見た目を決定する要素としては，スペースや改行などの記号や，文字のフォント等が挙げられる．記号を使用して文を適切な位置に配置したり，太字やフォントサイズなど，文字の書式を設定することが，テキストを読みやすく提示する上で有効となる．

文字列を整形するための方法として，文を構成する文字列に対する削除，換言，補完等の処理を行うことが挙げられる．これらの処理を施すことによって，文中の誤りの修正や読み手にとって理解が困難な箇所の修正を行うことで，より読みやすいテキストへと変換することが可能となる．一方，配置・書体の整形とは，スペースやタブによるインデントや改行等によって適切な配置を行ったり，文字の装飾やフォントサイズを変更することによって，文の見た目を整形する方法である．文の構造を考慮した配置や，重要箇所を適切な書体で提示することは読みやすいテキストを提示する上で効果的である．

読みやすいテキストを提示する上で，文字列や配置・書体の整形が有効であるが，この際，整形対象の文があらかじめ定まっている場合には，文を対象とした整形処理を行えばよい．しかし，提示するテキストが動的に生成される場合は，順次生成される文字列に対して整形を行う必要がある．そのようなテキストとして，音声や映像の内容理解を促進するために，その発話内容を文字化した字幕テキストが挙げられる．この場合，読みやすいテキストを提示するためには，テキストを提示すべきタイミングまでに文の整形処理を完了させる必要がある．そのため，順次生成される文字列に対して，リアルタイム処理を行うことが重要となる．

以上のことから，文整形においては，文字列の整形，配置・書体の整形という 2 種類の処理が必要となり，さらに，文字列の提示タイミングが重要となる場面においては，提示タイミングを考慮した整形を行う必要があるといえる．

本論文では，読みやすいテキストを提示するための日本語文の自動整形手法を提案する．本研究では，文字列の整形手法として日本語テキストへの読点挿入手法を，配置・書体の整形手法として講演テキストへの改行挿入手法を，テキストの提示タイミングを考慮した整形手法として講演テキストへの逐次的な改行挿入手法を実現した．

本論文は全 5 章から構成される．第 1 章は本論文の序論であり，これまでの文整形に関する研究動向を示すとともに，本論文の位置づけとアプローチについて述べる．

第 2 章では，文字列を整形するための手法として，日本語文に読点を自動挿入する手法を提案する．読点には様々な用法が存在し，その用法によって文中での挿入位置が異なる．用法ごとに有効な特徴を取り出し，素性として用いることで，精度の高い読点挿入を実現する．本研究ではまず，読点に関する文献を調査し，読点の用法を 9 種類に分類した．また，新聞記事テキストを用いて読点の出現傾向をその用法ごとに分析し，読点挿入に有効となる情報を定めた．本手法では，それらの情報を素性とした統計的アプローチによって，入力文中の各形態素境界に対して，その位置が読点位置であるか否かを同定する．評価実験により，読点の用法ごとに定めた各素性，及び，本手法の有効性を確認した．

第 3 章では，配置を整形するための手法として，講演テキストへの改行挿入手法を提案する．講演内容を字幕として提示する字幕生成システムにおいて．読みやすい字幕を提示するためには，発話内容を正しく文字化するだけでなく，字幕の文字列をどのように配置するかということも重要となる．そこで，文中の適切な箇所に改

行を挿入し，意味的なまとまりから構成される行を提示することが考えられる．本研究ではまず，講演音声の書き起こしデータに対して適切な改行位置の付与により改行コーパスを構築し，それを用いた改行挿入位置の言語的な分析を行った．次に，分析から改行挿入に有効な素性を定め，それらの素性を用いた統計的方法によって 1 文中の適切な改行挿入位置を同定する手法を実現した．評価実験により，本手法の有効性を確認した．

第 4 章では，テキストの提示タイミングを考慮した方法として，リアルタイム字幕生成のための講演テキストへの逐次的な改行挿入手法を提案する．講演の進行と同時的に読みやすい字幕を提示するためには，改行挿入によって分割された行が，音声とできる限り追従して提示されることが望ましい．話者の発話と同時的に改行を挿入するために，文よりも短い単位での改行挿入を実現する．まず，字幕提示のリアルタイム性を最も重視する方法として文節単位での改行挿入手法について述べる．本手法では，改行挿入に有効である係り受けやポーズなどの素性のうち，改行同定処理の段階で利用可能な情報を利用して改行挿入を行う．評価実験を行い，文単位での改行挿入手法との比較によって本手法を評価した．次に，字幕提示のリアルタイム性と改行位置の適格さの両方を考慮する方法として節単位での改行挿入について論じる．被験者評価によって文節単位での手法と比較し，同定処理のタイミングと精度が評価に及ぼす影響について考察する．

最後に，第 5 章で本論文をまとめ，今後の研究課題，及び，将来の展望について述べる．

目 次

図一覧

表一覧

第1章　まえがき

1.1　読みやすいテキスト提示のための文整形

現代社会では，大量の情報が生産，蓄積されており，我々はそれらの情報の中から自分に必要なものを選択的に収集し，活用している．テキストは代表的な情報収集源であり，その媒体として書籍や新聞，雑誌など，紙に印刷されたテキスト，ならびに，Webニュース，ブログ，メール，また，テレビ番組に付与される字幕など，ディスプレイやスクリーンに提示されるテキストがある．効率的な情報収集を可能にするために，テキストは読みやすいものであることが望ましい．テキストは一般に複数の文によって構成され，その文の内容や見た目がテキストの読みやすさに影響を与える．ディスプレイなどに提示されるテキストを構成する文は，印刷されたテキストのものと異なり，変更することができるため，文を適切に整形することによって読みやすいテキスト提示を行うことが可能となる．

文の整形とは，文の見た目を変更し体裁を整えることをいう．しかし，読みやすいテキストを提示するためには文の見た目以外に文の内容も重要となるため，文の内容と見た目の両方を対象として整形する必要がある．文の内容は文字列によって示されるため，元の文が伝えている情報そのものが変わることがないように，文の内容が大きく変更されない範囲で文字列を修正することが望まれる．また，文の見た目を決定する要素としては，スペースや改行などの記号や，文字のフォント等が挙げられる．記号を使用して文を適切な位置に配置したり，太字やフォントサイズなど，文字の書式を設定することが，テキストを読みやすく提示する上で有効となる．このように，文字列の整形，配置・書体の整形を自動的に行うことによって，読みやすいテキストを提示することが可能となる．

文字列を整形するための方法として，文を構成する文字列に対する削除，換言，補完等の処理が挙げられる．これらの処理を施すことによって，助詞の誤りや単語の誤用などの文中の誤りの修正が可能である．誤りの存在はテキストの内容を理解する上で大きな障害となるため，読みやすいテキストを提示するためには排除する必

要がある．また，誤り以外に対しても，読み手にとって理解が困難な箇所を修正したり，冗長な箇所，重要ではない箇所を削除することによって，より読みやすいテキストへと変換することが可能となる．

一方，配置・書体の整形とは，スペースやタブによるインデントや改行等によって適切な配置を行ったり，文字の装飾やフォントサイズを変更することによって，文の見た目を整形する方法である．例えば，受信したメールを，PC 上で閲覧する場合と携帯などの小型端末上で閲覧する場合，同じ内容のメールであっても読みやすさが異なる．これは，改行位置の変化やテキストの一覧性の違い，余白の量の違いなど，文の配置が異なっているためである．さらに，同一の閲覧環境においても，単に文字列が記述されたテキストよりも，論文などのように，段落区切り，見出し等が付与された構造を持つテキストの方が読みやすい場合がある [70]．適切な配置を行うことによって，読みやすいテキストへ変換することが望まれる．また，文中の重要箇所を明示的に提示することによって，文の内容が把握しやすくなるため [71]，重要箇所を検出し，太字や大きめのフォントで提示することが効果的である．

読みやすいテキストを提示する上で，文字列や配置・書体の整形が有効であるが，この際，整形対象の文があらかじめ定まっている場合には，文を対象とした整形処理を行えばよい．しかし，提示するテキストが動的に生成される場合は，順次生成される文字列に対して整形を行う必要がある．そのようなテキストとして，音声や映像の内容理解を促進するために，その発話内容を文字化した字幕テキストが挙げられる．近年では，聴覚障害者の情報保障を目的に，テレビ番組へのリアルタイム字幕付与が進んでいる [43]．字幕表示と音声とのずれが大きくなると内容理解の妨げになるため [53]，字幕を音声の進行に合わせた適切なタイミングで提示することが望まれる．

ここで，音声や映像の進行に合わせて読みやすいテキストを提示するためには，テキストを提示すべきタイミングまでに文の整形処理を完了させる必要がある．長い文を整形する場合，文が生成されてから整形を行うと，テキストの提示が遅れる可能性がある．そのため，順次生成される文字列に対して，テキストの提示タイミングを考慮したリアルタイム処理を行うことが重要となる．

以上のことから，文整形においては，

- 文字列の整形
- 配置・書体の整形

という 2 種類の処理が必要となり，さらに，文字列の提示タイミングが重要となる場面においては，

- 提示タイミングを考慮した整形

を行う必要があるといえる．

これらの問題を解決するための文整形手法の開発がこれまでに多く行われている．次節以降では，まず，これまでの文整形手法の研究動向を概観する．次に，本論文の目的と内容について述べる．

1.2 文整形に関する研究動向

本節では，文の自動整形技術に関する研究動向を,「文字列の整形」,「配置・書体の整形」,「提示タイミングを考慮した整形」に分けて概観する．

1.2.1 文字列の整形

文字列の自動整形は，文に含まれる誤りの修正，語彙・文構造を対象とした整形，要約の生成の 3 つに分類できる．

文中の誤りの修正による整形

文に含まれる誤りとして，誤字や脱字などの表記的な誤り，同音異義語に起因する語彙的な誤り，助詞の誤用による文法的な誤りなどがある．誤りを含む文によって構成されるテキストは読みにくい．このため，誤りを修正することによって正しい文に書き換える整形技術が開発されており，そのような研究として以下のものが挙げられる．

- 単語 N-gram を使用した平仮名列中の誤りの検出と訂正と行う手法が開発されている [67]．対象とされる誤りは，ある 1 文字が欠落した削除誤り（例：するどうか→するかどうか），ある 1 文字が挿入された挿入誤り（例：するかがどうか→するかどうか），ある 1 文字が他の 1 文字と入れ替わった置換誤り（例：するかとうか→するかどうか）である．ある平仮名列の出現頻度を N-gram を用いて調べ，頻度が閾値以下である場合に，誤りが存在すると判定する．誤り

の対象を平仮名列中のものに限定することによって，大規模な単語 N-gram の使用を可能にしている．

- スキップマルコフ連鎖モデルを用いた削除，挿入，置換誤りの検出と訂正手法の開発が行われている [40]．隣り合う文字のみでなく，離れた位置に存在する文字との関係を考慮することにより，置換と削除による誤りや置換と挿入による誤りなどの混合的な誤りの検出と訂正が可能となっている．

- 複合語における同音異義語誤り（例：自然化学→自然科学）の検出と訂正手法の開発が行われている [51]．この手法では，あらかじめ，複合語中の同音異義語に対して，その同音異義語に隣接しうる単語のカテゴリ（組織，地域，自然など）の集合を定めておく．ある単語の隣接単語がその集合に含まれない場合，その単語を誤りとして検出し，同じ読みを持つ単語のうち，隣接単語を集合に含む語に訂正する．

- 識別的系列変換による助詞の削除，挿入，置換誤りの訂正手法が開発されている [44]．この手法では，助詞誤りを含む文を誤りを含まない文に翻訳するというアプローチが採用されている．この場合，翻訳のモデルを学習するために，誤りを含む文とそれを訂正した修正文のペアが必要となる．しかし，このペアを大量に収集することは困難である．この問題に対処するため，小規模なデータを使用した誤り訂正の枠組を提案している．

その他に，文の推敲作業を支援するシステムが開発されている [83]．外国語の初学者が書いたテキストには誤りが含まれやすいものの，書き手自身が誤り箇所を判断することは難しい．このシステムでは，外国語の初学者のテキスト推敲支援のために，テキスト中の誤り箇所を自動で検出し，修正候補を提示する．

一方，近年では，情報保障や情報の蓄積の観点から，音声を文字化することが多く行われている．音声では，言い淀みや言い間違い，フィラー，また，助詞の脱落などが頻出するため，音声をそのまま書き起こしたテキストは読みにくいものとなる．そのために，言い淀みや言い間違いの修正とフィラーの削除を行う手法が開発されている．具体的には，言語情報と韻律情報を組み合わせて用いる方法 [22]，統計的機械翻訳アプローチに基づき，言語情報を使用して，言い淀み等を含むテキストを含まないテキストに翻訳する方法 [10, 24] が提案されている．さらに，言い淀みの修正やフィラーの削除に加え，助詞の補完を行う手法も提案されている [65]．

語彙・文構造を対象とした整形

たとえ誤りが含まれていなくても，難解な語が使用されている文，あるいは，係り受け構造が複雑な文を含んだテキストなどは読みにくい．そのような文を修正することによってテキストが読みやすくなる．語彙や文構造の変換による整形手法として，以下が提案されている．

- 日本語の初学者による文の内容理解を支援するために，文中に含まれる専門用語などの難解な単語を検出し，平易な単語に言い換える技術が開発されている [8, 15]．この技術を利用すると，日本語の初学者の人々にとって読んで理解しやすいテキストを提示することが可能となる．

- 外来語であるカタカナ語をより理解しやすい語に言い換えるための知識の自動獲得が行われている [100]．カタカナ語は英単語に基づく外来語であるため，広く浸透していないことも多い．カタカナ語を，同一の意味を持つ別の語に言い換えることによって，文の可読性を向上させることができる．

- 人間が文を理解するときには，係り受け解析作業を行っていると考えられる．ある文節の係り先が遠くの文節となる場合には，その係り先が出現するまでその文節を記憶しておく必要があるため，文の理解が困難になる場合がある．そのため，長い修飾節を文の前方に移動することにより，理解しやすい文に変換する手法が開発されている [99]．

また，音声の書き起こしテキストには，話し言葉特有の語彙や表現が含まれ，読みにくい原因になる．そのために，敬体（「です」「ます」調）から常体（「だ」「である」調）への文体の変換，話し言葉特有の表現から書き言葉で使用される表現への書き換え（「行ってるんですが」→「行っているのですが」,「色んな」→「色々な」など）を行うことによって，読みやすい書き起こしテキストに変換する手法が開発されている [12, 17, 29]．

要約の生成

テキストに記述されている内容の概要を理解するために，要約を生成して提示することは有用である．新聞や論文，近年では，Web ページやメールテキスト，音声の書き起こしテキストなどを対象とした要約生成手法が，以下に示す通り開発されている．

- 元のテキストから重要ではないと思われる文を削除し，残った文を並べて要約を作成する重要文抽出という手法が開発されてきた [1, 16, 19, 33, 81]．抽出した文同士が自然に繋がるように，文末表現の言い換えや接続詞の挿入が行われる．

- 文内の表現を削除したり，より短い表現に換言する文圧縮の手法が開発されている [5, 11, 35, 98]．文圧縮では元のテキストの内容保持を重要視しており，情報を損わない範囲で削除，及び，換言が行われる．文圧縮技術は，電光掲示板やニュース字幕など，限られた表示スペースにテキストを表示するために使用される [97]．

上記の方法以外に，元のテキストから文を抽出するのではなく，表現を別の表現に言い換えることによって，元のテキストの内容を改めて表現し，人間が作成するような流暢な要約を作成する試みが行われている [48, 80]．これらの手法では，2 文を 1 文にまとめるような言い換え処理も行われるため，文を対象とした整形の範疇を超えているものの，要約の作成には有効な方法といえる．

以上のように，文字列の自動整形に関する研究が多く行われおり，これまでの研究では，主に，語や文構造を対象に修正を与える手法が開発されてきた．

1.2.2　配置・書体の整形

文字列の配置や書体を決定する方法として，TEX に代表されるテキスト組版技術が挙げられる．単語間のスペース幅を調整することにより段落中の行の右端を揃える，タイトルや見出しをつける，フォントのスタイルや大きさを変更する，などの方法によって，入力された文字を読みやすいテキストとして提示する．TEX は当初，英語を対象に開発が行われ，その後，日本語を処理できるように改良された pTeX などが開発されている．また，TEX 以外にも，日本語テキストを美しく提示するための文書自動整形システムが開発されている [95]．その他，PowerPoint や Keynote など，読みやすい発表資料を作成するためのアプリケーションが開発されている．これらの技術は，人間がテキストを作成する際に，読みやすいテキストを作成するための支援を行うものである．

一方，配置・書体を自動的に整形するための研究も行われている．読みやすいテキストを提示するために，文中の重要箇所や文の言語的な構造を捉えて文字列の配

置・書体を決定することが重要である．テキストを前から順に読む場合は重要な箇所や構造を判断できるが，大まかに見る場合や飛ばし読みを行っている場合は，その判断が難しくなる．

文字列の重要箇所を示す手法やシステムとして，以下のものが開発されている．

- テキストの可読性を向上させるために文字列を彩色するシステム [46]
- 重要語句を説明している箇所の自動検出手法 [59]
- 授業における学習者支援のための重要語句の強調表示手法 [79, 82]

一方，文の構造を捉えた整形としては，文書から発表スライドを自動生成する研究がいくつか行われている [32, 63, 96]．発表スライドでは，文中の主題や並列構造，重要箇所が箇条書きや見出しとして記述されており，内容の理解が行いやすいテキストであるといえる．

近年では，PC に接続したディスプレイの他，スマートフォンやタブレットなどの小型端末でテキストを閲覧する機会も増加しており，テキストの表示媒体が多様化している．表示媒体の違いによって，テキストの一覧性が異なったり，スクロール表示が必要になるなどの現象が発生することから，表示媒体ごとに適した文字列の配置，書体の設定を行い，提示することが有用である．そのために，テキストが表示される画面サイズに従って文字列の配置を変更する手法が開発されている [75]．

以上のように，配置・書体を適切に整形することによって，読んで理解しやすいテキストへと変換するための研究がいくつか行われている．最近では，ニュース番組や講演の場で，音声を書き起して字幕として提示する機会が増えている．字幕テキストに対しても，文字列を適切に配置し読みやすいテキスト提示を行うことによって，音声の内容理解を支援することが可能となると考えられる．

1.2.3 提示タイミングを考慮した整形

テキストの適切な提示タイミングについて，音声とその文字化テキストを対象とした検討が行われている．被験者評価によって，聴覚障害者のための情報保障を目的として提示されるニュース番組の字幕や講義のキーワードの適切な提示タイミングの分析が行われており [52, 53, 89]，分析結果から，テキストは音声との遅延が少なくなるように提示することが望ましいという知見が得られている．この結果を踏ま

え，テレビ番組に対して聴覚障害者向けの字幕を自動で付与することを目的に，音声と対応する文字列を自動的に検出する手法が開発されている [90].

ここで，音声と同時にテキストを提示する場合にも，文の整形を行い，読みやすいテキストを提示することが必要となると考えられる．しかし，これまでに，文字列の整形を行いながら同時的に文字列を提示するための手法の開発は，ほとんど行われていない.

1.3 本論文の目的

本論文では，読みやすいテキストを提示するための日本語文の自動整形手法を提案する．本研究では，(1) 文字列の整形，(2) 配置・書体の整形， (3) 提示タイミングを考慮した整形，のそれぞれについて，以下の整形手法を実現する.

(1) 読点挿入による文字列の整形

読点は，節や文の主題，並列された要素などの区切りを明示したり，係り受け構造を明確にする記号であり，その挿入位置は，文の読みやすさや読み手による文の解釈に影響を与える [39, 50]．そのため，読みやすいテキストを提示するためには，読点は適切な位置に挿入されている必要がある．しかし，機械翻訳や自動要約，音声筆記などによって生成される文，日本語の初学者が書いた文では，誤った位置に読点が挿入されている可能性がある．誤った読点挿入がなされた文では，文構造や語の区切りが分かりにくくなったり，読み手が文意を誤って解釈してしまう場合がある．そこで，日本語文の適切な位置に，読点を自動挿入できることが望ましい．読点を適切に挿入するとは，文に記述されている内容が読み手にスムーズに伝わるように，文を読みやすく整形することにほかならない．この手法は，文字列の一部である読点を修正していることから，語彙を対象とした整形に相当する.

(2) 改行挿入による配置の整形

改行は文を分割して配置するために利用される記号である．提示したときに複数行にわたるような長い文では，改行位置を考慮せずに画面の幅に合わせて提示すると，文が単語の途中で区切られてしまい読みづらくなる場合がある．音声の進行に合わせて読むことが強いられる字幕やあまり時間をかけずに多くの量を読む必要があるメールテキストなどは，行単位で読んで理解しやす

くなるような位置に改行が挿入されていることが望まれる．実際に，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 [42]，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ものの [3, 9, 14, 27, 31, 34] ，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重要となることが知られている [74, 77]．改行挿入によって適切な行に分割して文を配置することは，この問題を解決するための有効な方法であると考えられる．1 行の長さはテキストが提示される画面の幅よりも短くする必要があるため，行長の制限を満たす中で適切な位置に改行を挿入し，配置することによって，読みやすいテキストを提示する．これは改行で分割された行を配置していることから，配置の整形の一手法である．

(3) 提示タイミングを考慮した逐次的な改行挿入による配置の整形
音声の進行に合わせてテキストを提示すべき場面においては，テキストが音声入力にできる限り追従して提示されることが望ましく，そのためには，音声が発生されてからテキストを提示するまでの時間を少なくすることが要求される．実際，字幕生成のための音声認識に関する研究 [3, 27, 34] において，遅延時間の短縮が重要な課題の 1 つになっている．そのため，文の整形においても，入力に対する出力の同時性を考慮する必要がある．改行挿入によって文を意味的なまとまりに分割することは，読みやすいテキスト提示を行う上で有効であるものの，文が長くなる場合には，文を単位とした改行挿入は，音声の進行と同期したリアルタイムでの処理に適さない．そのため，文よりも細かい単位ごとに改行位置を決定することが望まれる．テキストの提示タイミングを考慮した文への逐次的な改行挿入によって，音声との遅延時間が少なくなるように読みやすいテキストを提示する．

読点は文中の区切りを示す記号である．読点にはいくつかの用法が存在し，その用法によって挿入位置が異なるものの，その挿入位置は，文中の意味の切れ目となっている．また，改行は，テキスト内容の理解を促進するために，各行が意味的なまとまりを構成するような位置に挿入されることが望ましい．以上のことから，読点，改行が挿入される位置は，意味的なまとまりの境界と関係性があると考えられる．そのため，本研究では，読点挿入，改行挿入による文の整形を実現するために，意味的なまとまりに着目する．

文中のまとまりを検出する手法として，形態素，文節，節などの言語的なまとまりを検出するための技術が開発されている [56, 57, 88, 91]．これらのまとまりを検出する目的は，機械的な処理を行うための言語情報を獲得することである．また，テキストからの知識獲得を目的として，固有表現や機能表現といったまとまりの抽出が行われている [72, 73, 76]．しかし，テキストの読みやすさは人の主観的要因に基づいて定まると考えられるため，これらの研究で検出対象とされている言語的なまとまりは，人間がテキストを閲覧する際の読みやすいまとまりとは必ずしも一致しない．

1.4 本論文の内容

本論文ではまず第一に，文字列を整形するための手法として，日本語文に読点を自動挿入する手法を提案する．読点は文中の区切りを示す記号であり，その挿入位置は，文の読みやすさや読み手による文の解釈に影響を与える．意味的なまとまりへの分割に基づいて，読点位置を決定することによって，読み手が文意を解釈しやすい文を生成することが可能となる．読点の挿入位置は，ある程度の規則性が存在するものの，必ずしも明確に定まるわけではない．そこで本研究では，統計的手法に基づく日本語文への自動読点挿入手法を実現した．読点には様々な用法が存在し，その用法によって文中での挿入位置が異なる．読点に関する文献の調査に基づいて読点の用法を分類し，その分類ごとに読点の出現傾向を分析した．本手法では，その分析結果に基づいて定めた素性を用いて学習した確率モデルによって，入力文中の各形態素境界に対して，その位置が読点位置であるか否かを同定する．評価実験により，読点の用法ごとに定めた各素性，及び，提案手法の有効性を確認した．

第二に，配置を整形するための手法として，講演テキストへの改行挿入手法を提案する．講演の場で，話者の発話内容を字幕として提示する字幕生成システムの開発が行われている．読みやすい字幕を生成するためには，発話内容を正しく文字化するだけでなく，字幕の文字列をどのように配置するかも重要となる．そこで，文中の適切な箇所に改行を挿入して提示することによって，読みやすい字幕を生成できることに着目し，講演の書き起こしテキストへの改行挿入手法を提案する．本手法では，確率モデルを用いて1文中の最適な改行位置を同定する．実験によって提案手法の有効性を明らかにした．また，被験者評価により，実際に読みやすい字幕が生成できていることを確認した．

第三に，提示タイミングを考慮した整形手法として，リアルタイム字幕生成システムにおいて，発話と同時的に読みやすい字幕を生成するための逐次的な改行挿入手法を提案する．講演の進行と同時的に字幕生成を行うためには，字幕が音声とできる限り追従して提示されることが望ましい．本研究では，講演音声が同時的に文字化され，その進行に応じて改行位置を同定し，改行位置が定まるごとにその行を提示するシステムを想定する．この場合，音声が入力されてから利用可能な情報をできる限り考慮しつつ，遅延時間を小さくするために改行同定処理を細かい単位ごとに行う必要がある．まず，字幕提示のリアルタイム性を最も重視する方法として文節単位での改行挿入手法を実現した．評価実験により，文単位の改行挿入手法と比較した本手法の有効性を確認した．次に，字幕提示のリアルタイム性と改行位置の適格さの両方を考慮する方法として節単位での改行挿入手法を実現し，被験者実験により，文節単位の改行挿入手法との比較を行った．

1.5 本論文の構成

本論文の構成は以下の通りである．

第 2 章では，高品質なテキストを提示するための要素技術として，日本語文に自動で読点を挿入する手法を提案する．まず，読点に関する文献の調査に基づく読点の用法の分類について論じる．次に，読点挿入に用いる素性を決定するための読点の出現傾向の分析について述べる．分析結果に基づいて定めた素性を用いた読点挿入手法について説明する．最後に，評価実験の知見を考察し，読点の用法ごとに定めた素性の妥当性，及び，本手法の有効性を示す．

第 3 章では，読みやすい字幕テキストを提示するために，講演テキストの適切な位置に改行を挿入する手法を提案する．まず，講演音声の書き起こしデータに対して適切な改行位置の付与により作成した改行コーパスについて述べ，それを用いた改行挿入位置の言語的な分析について論じる．次に，分析結果から得られた情報を素性として用いる統計的な改行挿入手法について述べる．最後に，実験結果を考察し，本手法の有効性を示す．

第 4 章では，リアルタイムに字幕テキストを提示するための，講演テキストへの逐次的な改行挿入手法を提案する．まず，文節単位での改行挿入手法について説明する．評価実験により，文節単位での改行挿入手法の有効性を示す．次に，節単位での改行挿入手法を説明する．文節単位での手法との比較評価によって，同定処理

のタイミングと精度が評価に及ぼす影響について考察する．

最後に，第5章では，本論文のまとめと残された課題，将来の展望について述べる．

第2章　日本語テキストへの読点挿入

2.1　はじめに

日本語テキスト生成は，機械翻訳や自動要約，音声筆記などの性能を決める重要な技術である．生成されたテキストが高い品質を備えているためには，読点が適切な位置に挿入されている必要がある．これまでに，日本語テキストに読点を挿入する技術は，機械翻訳における文生成を目的としたものを中心に，いくつか開発されている [58, 69]．これらの研究では，読点挿入ルールを人手で作成し，それを日本語テキストに適用することによって読点挿入を実現している．読点の挿入位置は，ある程度の規則性が存在するものの，必ずしも明確に定まるわけではなく，文法的には読点が打たれやすい位置であっても，周辺の読点位置や文の長さなどとの関係により，読点が打たれないこともある．すなわち，読点が打たれる位置であるか否かは，複数の要因のバランスのもとに決まると考えられ，人手で作成したルールによって精度よく読点を挿入することは難しい．

本章では，日本語文のチャンキングに基づいて，統計的アプローチにより読点を挿入する手法を提案する．読点には，文構造を明確にする，並列された語の区切りを示すなど，いくつかの用法が存在し，その用法によって文中での挿入位置が異なる．そのため，本研究では，用法ごとに読点の挿入位置を分析し，その出現傾向を捉える特徴素を設定し利用した．

日本語の読点と役割が類似する記号として，外国語におけるコンマがある．これまでに，英語や中国語などを対象に，統計的手法に基づきコンマを挿入する技術が開発されている [6, 7, 23]．しかし，英語におけるコンマの位置は文法的に定まることが多く，その運用は日本語の読点よりも厳密であるなど [84]，日本語の読点と外国語のコンマとでは，運用上の違いが小さくない．そのため，これらの手法で用いられた素性をそのまま日本語に使用することは難しい．また，音声筆記や音声翻訳への応用を目指し，音声言語に対する読点やコンマの挿入技術が提案されているものの [2, 4, 13, 18, 37, 64]，これらは抑揚や強勢，休止など，音声から獲得した情報

表 2.1: 読点の用法の分類

#	用法
1	節を区切る読点
2	係り受け関係を明確にする読点
3	難読・誤読を避ける読点
4	主題を示す読点
5	先頭の接続詞・副詞を区切る読点
6	並列する単語・句を明確にする読点
7	時間を表わす副詞を区切る読点
8	直前の語句を強調するための読点
9	引用を示す読点

を活用している．一方，本研究は，テキストから入手可能な情報のみを用いた読点挿入を実現する．

2.2　読点の用法と分析

日本語テキストにおける読点の自動挿入とは，あらゆる形態素境界に対して，それが読点位置であるか否かを判定することであり，境界周辺の形態素及びその品詞を素性とした機械学習によって判定することは一つの方法である．しかし，このような単純な方法で高い読点挿入性能を達成することは難しいと予想される．というのも，日本語の読点には様々な用法が存在し，用法によってその入り方が異なるためである．これに対して，読点をその用法に応じて分類し，それぞれの読点に対してその特徴を取り出し，素性として用いることが考えられる．

そこでまず，読点に関する文献 [41, 66, 86] に基づき，読点の用法を整理し分類した．結果を表 2.1 に示す．文献によっては，用法が必ずしも網羅的に示されているわけでなく，また，著者によって用法の捉え方や分類の粒度に若干の違いが存在するものの，文献で言及されている読点がいずれかの用法で説明できるように整理すると，おおよそ表 2.1 の 9 つに分類できる．このうち，分類 8 の「直前の語句を強調するための読点」については，執筆者の意向に依存するものであるため，本研究では対象とせず，それ以外の 8 つの用法について，その特徴を明らかにするための分析を与えた．

分析には，京都テキストコーパス 4.0（以下，京都コーパス）[54] の 1 月 1 日から

表 2.2: 分析データの規模

文数	11,821
形態素数	332,885
文節数	115,421
文字数	503,969
読点数	16,596

11 日の全記事を用いた．本研究では，読点を挿入する対象として新聞記事テキストを使用する．新聞記事における読点の運用が必ずしも唯一かつ絶対的なわけではないが，読者の数や執筆者の文章作成力の観点からみて，最も一般的でかつ均一的な運用が施されているテキストの一つである．この意味で，新聞記事と同様に読点を挿入可能な日本語テキスト生成技術が実現されれば，それによって出力されたテキストは，少なくとも読み手が容認できるものであるといえる．京都コーパスのデータフォーマットを IPA 品詞体系 [38] に変換し，使用した．分析データの規模を表 2.2 に示す．コーパス中のテキストには，形態素，文節境界，係り受け構造の情報が，人手により付与されている．また，節境界情報を節境界解析ツール CBAP[91] を用いて機械的に付与した．

形態素境界 321,064 箇所に対する読点挿入率は 5.17%（16,596/321,064）である．また，文節境界への読点挿入率は 15.79%（16,356/103,600）であり，文節境界ではない形態素境界への読点挿入率は 0.11%（240/217,464）であった．用法の性質上，分類 1〜7 の読点は文節境界に挿入され，分類 9 の「引用を示す読点」は文節境界ではない形態素境界に挿入される．上記のそれぞれの読点挿入率を基準に，読点の挿入されやすさを分析した．なお，本章の以下では，言及している例文中の読点に下線を引いて示す．

2.2.1 節境界を明確にする読点

節とは，述語を中心としたまとまりであり [87]，文の構成単位の 1 つである．複文や重文の場合，文は複数の節から構成されるため，節間に読点を挿入することにより文の構造が明確になる．例として，以下の文

- 国連による対イラク制裁解除に向け、関係の深い仏に一層の協力を求めるのが狙いとみられる。

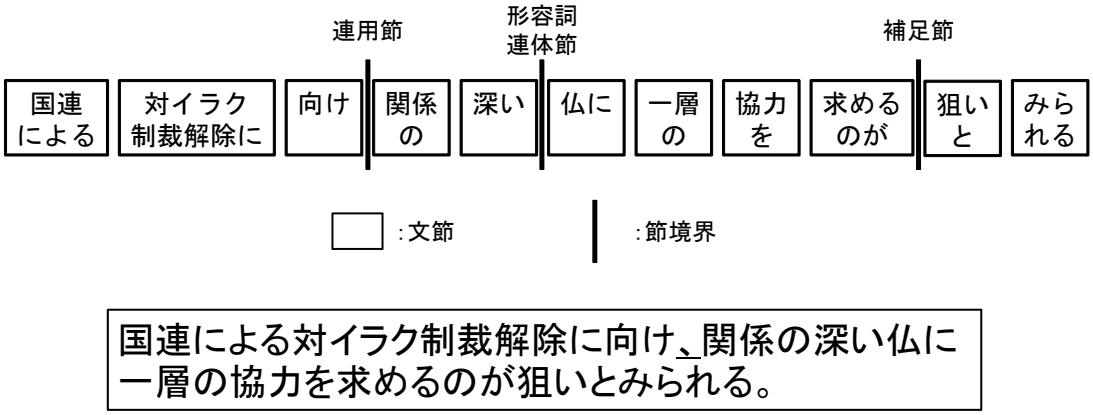


図 2.1: 節境界を明確にする読点

の節境界を図 2.1 に示す．この図では，文中に存在する節境界の位置を縦線によって示している．図 2.1 に示すように,「国連による対イラク制裁解除に向け」の直後に存在する節境界に読点が挿入されている．分析データでは，文末を除く節境界 30,845 箇所のうち 9,223 箇所に読点が挿入されていた．節境界に対する挿入率は 29.90%であり，文節境界に対する挿入率よりも高い．

しかし，単に節境界といってもその役割は様々であり，種類によって読点の挿入されやすさは異なると考えられる．そこで，分析データに出現した 115 種類の節境界について，種類ごとに読点挿入率を調査した．節境界の種類として，節境界解析ツール CBAP[91] で定義されたものを用いた．これらの中には，主題ハや談話標識など，本来の節の定義から逸脱したものも含まれるが，構文的に大きな切れ目になると考え，分析対象に含めた．出現数にして上位 10 種類の節境界とその読点挿入率を表 2.3 に示す．節境界「連用節」や「並列節ガ」,「並列節デ」,「条件節ト」の読点挿入率は 80%を越えているのに対して,「連体節」は 1%にも満たなかった．これらは，節境界の種類によって読点の挿入されやすさが異なることを示している．

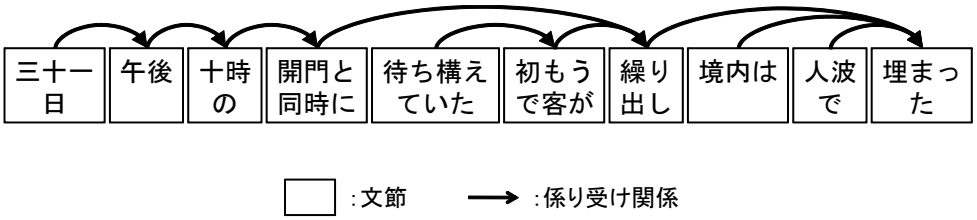
2.2.2　係り受け関係を明確にする読点

読点には係り受け関係を明確にする働きがある．以下の文

- 三十一日午後十時の開門と同時に、待ち構えていた初もうで客が繰り出し、境内は人波で埋まった。

表 2.3: 節境界への読点挿入率

節境界	読点挿入率 (%)	
主題ハ	16.79	(1,472/8,765)
連体節	0.86	(54/6,247)
連用節	82.91	(2,814/3,394)
引用節	4.65	(86/1,850)
テ節	24.21	(437/1,805)
補足節	17.24	(256/1,485)
談話標識	58.22	(659/1,132)
並列節ガ	93.66	(1,005/1,073)
並列節デ	83.44	(630/755)
条件節ト	81.66	(432/529)



三十一日午後十時の開門と同時に、待ち構えていた初もうで客が繰り出し、境内は人波で埋まった。

図 2.2: 係り受け関係を明確にする読点

の係り受け関係を図 2.2 に示す．この図では，矢印の始点が係り文節を，終点が受け文節を示している．図 2.2 では，文節「開門と同時に」は文節「繰り出し」に係る．読点がないと，読み手は文節「開門と同時に」が直後の文節「待ち構えていた」に係ると誤解する可能性があり，読点の挿入にはそれを避ける効果がある．分析データを調査したところ，係り受け関係にある隣接文節間 64,964 箇所に対して，読点が挿入されたのは 2,303 箇所で，挿入率は 3.55%に過ぎなかった．一方，係り受け関係にない隣接文節間への挿入率は 36.37%（14,053/38,636）であった．

さらに，ある文節の係り先が，直後の文節が属する節の節末文節よりも遠い場合，係り受け構造がより複雑になるため，その文節の直後には，読点がより挿入されやすくなると考えられる．例えば，以下の例

- 九三、九四年も、総数がやや減少したとはいえ、同じ傾向が続く。

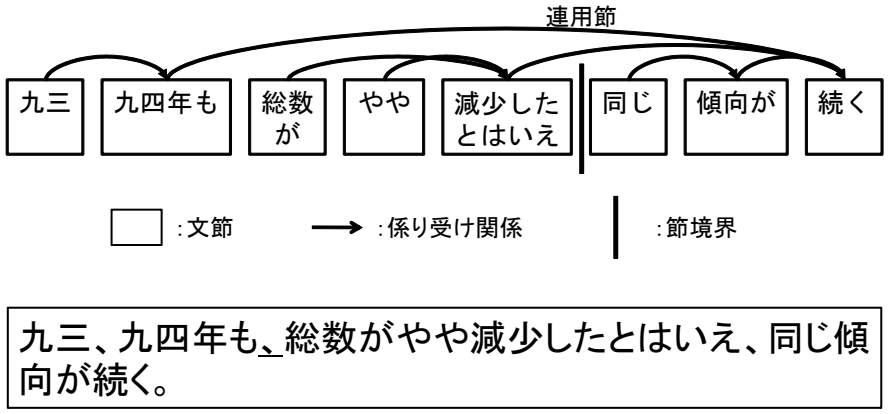


図 2.3: 節末文節よりも遠くの文節に係る文節の直後への読点挿入

では，図 2.3 に示すように，文節「九四年も」が，直後の文節が属する連用節の節末文節である「減少したとはいえ」よりも遠くの文節「続く」に係っており，その係り受け関係を明確にするため読点が挿入されている．分析データでは，そのような場合の読点挿入率は 54.14%（7,881/14,557）であり，読点が挿入されやすい．

また，読点によって挟まれた文字列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文字列とは，文字列外の文節に係る文節が，文字列末の文節以外に存在しない文字列のことをいう．例えば図 2.3 では，読点によって挟まれた文字列「総数がやや減少したとはいえ」内で係り受けが閉じている．読点に挟まれた文字列 16,356 個のうち，12,496 個（76.40%）で係り受けが閉じていた．この結果も，係り受け距離が遠くなる文節の直後には読点が挿入されやすい傾向を反映している．

2.2.3　難読・誤読を避ける読点

日本語は分かち書きがされない言語であるが，ひらがな，カタカナ，漢字と 3 種類の文字が存在しているため，一般には語の境界を識別しやすい．しかし，同じ種類の文字が続けて出現した場合には，読み手が誤読をしたり，読みづらく感じることがあり，読点を打つことによってそれを避けることができる．以下の例では,「優勝」と「賞金」の間の読点が，誤読を避けるために挿入されている．

- プレーオフの末、エルキントンが優勝、賞金１８万ドルを獲得した。

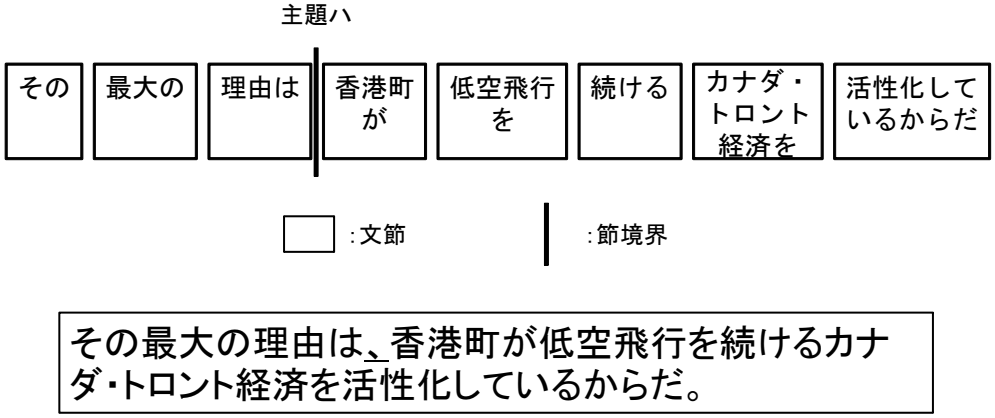


図 2.4: 節境界「主題ハ」への読点挿入

文節にまたがって漢字が出現するような文節境界 2,403 箇所のうち 90.93%（2,185/2,403）に，また，カタカナの場合は 98.15%（212/216）に読点が挿入されていた．すなわち，上記の場合には，そのほとんどの文節境界に読点が打たれる傾向にある．

2.2.4 主題を示す読点

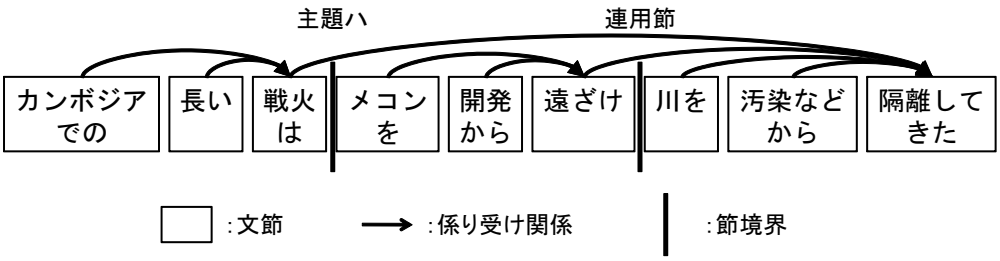
文の主題を明確に示すために読点が挿入されることがある．例えば，

- その最大の理由は、香港町が低空飛行を続けるカナダ・トロント経済を活性化しているからだ。

という文では，図 2.4 に示すように，文の主題「その最大の理由は」を明確にするために，その直後に読点が挿入されている．主題は節境界解析ツール CBAP[91] によってラベル付けされる「主題ハ」に着目することによって検出することができる．節境界「主題ハ」に挿入されている読点は，読点全体の 8.87%（1,472/16,596）を占めている．しかし，節境界「主題ハ」への読点挿入率は 16.79%（1,472/8,765）であり，文節境界に対する読点挿入率と大差はない．すなわち，単に主題であるという理由だけで読点が挿入されやすくなるわけではない．

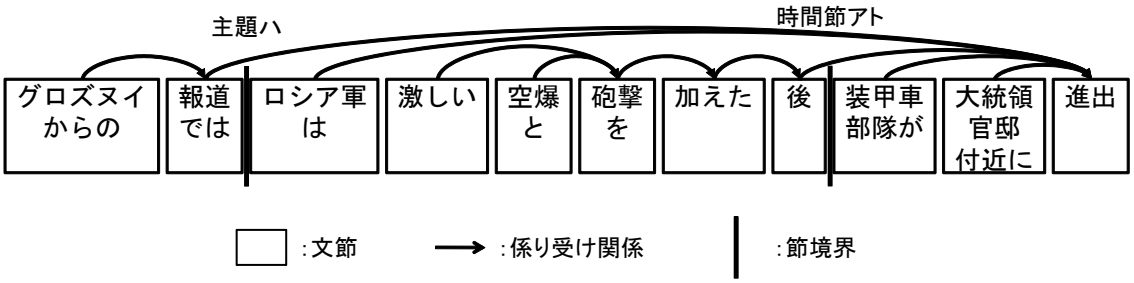
例えば，以下の例の文節「戦火は」のように，節境界「主題ハ」の直前の文節が，隣接する文節に係らない場合 (図 2.5 参照)，読点挿入率は 20.67%と高くなる．

- カンボジアでの長い戦火は、メコンを開発から遠ざけ、川を汚染などから隔離してきた。



カンボジアでの長い戦火は、メコンを開発から遠ざけ、
川を汚染などから隔離してきた。

図 2.5: 隣接する文節に係らない文節の直後に存在する節境界「主題ハ」への読点挿入



グロズヌイからの報道では、ロシア軍は激しい空爆と砲撃
を加えた後、装甲軍部隊が大統領官邸付近に進出。

図 2.6: 直前の文字列が「では」である節境界「主題ハ」への読点挿入

また，図 2.6 の「報道では」のように，節境界「主題ハ」の直前の文字列が「では」であった場合，読点挿入率は 35.61%（256/719）となる．

- グロズヌイからの報道では、ロシア軍は激しい空爆と砲撃を加えた後、装甲軍部隊が大統領官邸付近に進出。

1 文中に動詞が複数存在する場合，節境界「主題ハ」に読点を挿入することによって，節境界「主題ハ」の直前に存在する文節が後方の動詞に係ることを示し，文の主題であるということをより明確にすることが可能である．節境界「主題ハ」の直前の文節と係り先が同一である動詞が存在する場合の読点挿入率は 30.37%(1,802/5,933) であり，節境界「主題ハ」全体に対する挿入率よりも高い．

表 2.4: 主題を示す語句の文字数と読点挿入率

文字数	読点挿入率 (%)
1-5	10.58
6-10	16.10
11-15	22.41
16-20	31.61

主題を示す語句の文字数も，節境界「主題ハ」への読点挿入と関連がある.「主題を示す語句」とは，直後に節境界「主題ハ」が存在する文節と，係り受け関係を係りから受けにたどってその文節に到達可能な全ての文節からなる文字列を意味する. 例えば図 2.5 では,「カンボジアでの長い戦火は」が主題を示す語句である．表 2.4 に主題を示す語句の文字数と読点挿入率との関係を示す．主題を示す語句の文字数が短い場合は読点が挿入されにくいものの，それが長くなるにつれ，読点が挿入されやすくなる傾向にある.

2.2.5 先頭の接続詞・副詞を区切る読点

以下の例の「しかし」のように，文頭に出現する接続詞や副詞の直後には，前置きの語を区切るという目的で読点が挿入される.

- しかし、旧民社党は大半の議員が新進党に参加し、さきがけとの連携も流動的で連携相手は不確定だ。

分析データ中で，最終形態素が接続詞である文頭の文節の直後には，71.65%（498/695）の確率で読点が挿入されていた．また，最終形態素が副詞である場合では，挿入率は 30.97%（140/452）であった．いずれも，文節境界に対する読点挿入率よりも高い.

2.2.6 並列する単語・句の間に打たれる読点

読点には，対等の関係で並列された同じ種類の語や句を区切るという働きがある. 以下に例を示す.

- 共働き、独身など、人の出入りが少ないマンションに被害が集中しているという。

この例では，並列された名詞である「共働き」と「独身」を区切るためにその間に読点が挿入されている．最終形態素が名詞である文節が連続する場合，その文節境界への読点挿入率は 52.50%（4,728/9,006）であった．

また，語の並列以外に句が並列される場合もある．以下の例

- メニューは前夜、首相が何を食べたかを調べて同じ献立を避けたり、和食と洋食のバランスを考えたりして決める。

では，並列されている句「同じ献立を避けたり」と「和食と洋食のバランスを考えたり」の並列を明確にするために，文節「避けたり」の直後に読点が打たれている．並列する二つの句間への読点挿入率は 85.75%（391/456）であり，句が並列される場合の多くにおいて読点が挿入される．

2.2.7　時間を表わす語句の後に打たれる読点

読点の働きの一つとして,「最近」,「先月末」など，時間を表わす語句の直後に挿入することによって，文中の時間を明確に示す働きがある．以下の例では,「最近」の直後に読点が挿入されている．

- ヒトの鋤鼻器官は痕跡だけと考えられていたが、最近、フェロモンの受容器官として働いていることが分かってきた。

本研究では，品詞が名詞-副詞可能と名詞-接辞-副詞可能のいずれかであるとき，その語句を時間を表す語句とした．最終形態素が名詞-副詞可能，及び，名詞-接尾-副詞可能である文節の直後への読点挿入率は 39.97%（1,161/2,905）と高い挿入率を示した．

2.2.8　引用を示す読点

格助詞「と」が，会話文や引用文を受ける場合（引用の「と」），それを示すために格助詞「と」の直前に読点が挿入される．例を以下に示す．

- 実業家として名をはせてはいても、アメリカズカップで頂点に立てるほどの力はない、と思われていたからだ。

この例では,「実業家として名をはせてはいても、アメリカズカップで頂点に立てるほどの力はない」という文が引用されており，それを示すために，文節「ないと」中に存在する格助詞「と」の直前に読点が挿入されている．格助詞「と」が文節の最終形態素である場合,「と」の直前の形態素境界への読点挿入率は 3.38%（131/3,876）であり，文節境界ではない形態素境界への読点挿入率よりも高い．一方，格助詞「と」が名詞と名詞を接続する場合，それは引用の「と」ではないため読点が挿入されにくく（分析データ中に 10 回），また，引用文にかぎ括弧（「」）がついている場合は，すでに引用文であることが示されているため，読点は不要である（分析データ中に該当なし）．すなわち，格助詞「と」の前後の形態素情報が重要となることがわかった．

2.2.9 読点によって挟まれた文字列の文字数

前節までで読点の用法ごとに分析を与えたが，あらゆる用法に共通する特徴として，読点の挿入位置と直前の読点からの文字列の文字数との間の関連が挙げられる．読点が短い間隔で多く打たれすぎても，また，打たれない状態が長く続きすぎても，文の可読性は低下する．

読点によって挟まれた文字列の文字数と，その出現頻度を調査した．結果を図 2.7 に示す．1 文字となる頻度は極めて少ない．また，12 文字を超えるとグラフはおおよそ右肩下がりになっており，長くなると出現頻度が減少することを示している．

2.3 統計的な読点挿入手法

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，入力文中の各形態素境界に対して，その位置が読点位置であるか否かを同定する．入力文に対する適切な読点位置を同定するために，一文において考えうる読点位置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 個の形態素からなる入力文を $M = m_1 \cdots m_n$ とするとき，読点挿入結果を $R = r_1 \cdots r_n$ と記す．ここで，r_i は，形態素 m_i の直後が読点位置であるか $(r_i = 1)$ 否か $(r_i = 0)$ のいずれかの値をとる．なお，文末は句点位置となるが，便宜上 $r_n = 1$ としている．入力文を読点によって l 個に分割した j 個目の形態素列を

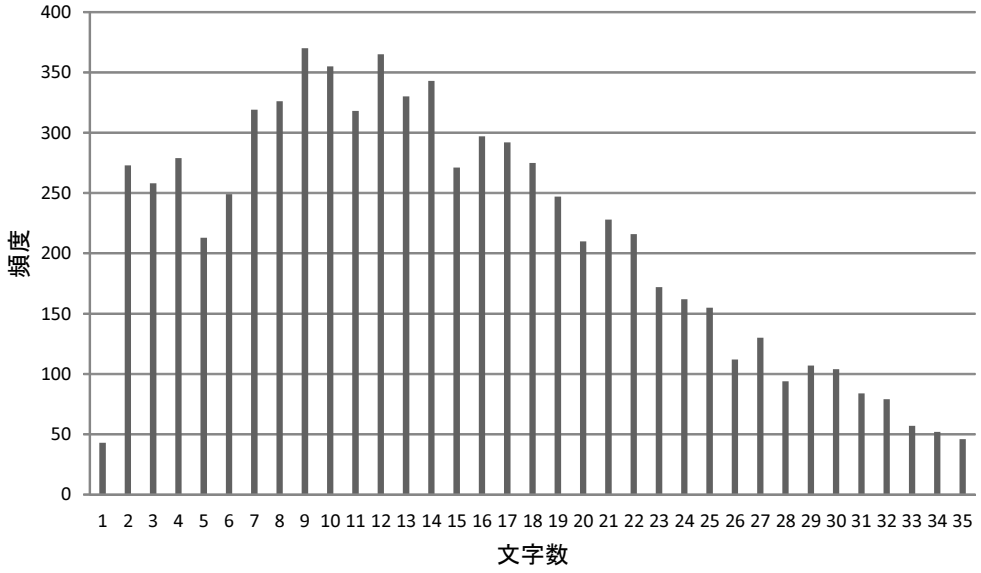


図 2.7: 読点で挟まれた文字列の文字数ごとの頻度

$L_j = m_1^j \cdots m_{n_j}^j (1 \leq j \leq l)$ 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．

2.3.1　読点挿入のための確率モデル

本手法では，入力文の形態素列を M とするとき，$P(R|M)$を最大にする読点挿入結果 R を求める．各形態素境界が読点位置であるか否かは，直前の読点位置を除く，他の読点位置とは独立であると仮定すると，$P(R|M)$ は次のように計算できる．

$$
\begin{aligned}
& P(R|M) \qquad (2.1)\\
= \;& P(r_1^1 = 0, \cdots, r_{n_1-1}^1 = 0, r_{n_1}^1 = 1, \cdots,\\
& \qquad r_1^l = 0, \cdots, r_{n_l-1}^l = 0, r_{n_l}^l = 1 | M)\\
\cong \;& P(r_1^1 = 0|M) \times \cdots\\
& \times P(r_{n_1-1}^1 = 0 | r_{n_1-2}^1 = 0, \cdots, r_1^1 = 0, M)\\
& \times P(r_{n_1}^1 = 1 | r_{n_1-1}^1 = 0, \cdots, r_1^1 = 0, M) \times \cdots\\
& \times P(r_1^l = 0 | r_{n_{l-1}}^{l-1} = 1, M) \times \cdots\\
& \times P(r_{n_l-1}^l = 0 | r_{n_l-2}^l = 0, \cdots, r_1^l = 0, r_{n_{l-1}}^{l-1} = 1, M)\\
& \times P(r_{n_l}^l = 1 | r_{n_l-1}^l = 0, \cdots, r_1^l = 0, r_{n_{l-1}}^{l-1} = 1, M)
\end{aligned}
$$

ここで，$P(r_k^j = 1|r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, M)$ $(1 \leq j \leq l, 1 \leq k \leq n_j)$ は，1 文の形態素列 M が与えられ，$j-1$ 個目の読点位置が同定されているときに，形態素 m_k^j の直後に読点が挿入される確率を表す．同様に，$P(r_k^j = 0|r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, M)$ は，形態素 m_k^j の直後に読点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．

最尤の読点挿入結果は，式 (2.1) の確率を最大とする読点挿入結果であるとして動的計画法を用いて計算する．

2.3.2 最大エントロピー法で用いた素性

本研究では，1 文の形態素列 M が与えられ，直前の読点位置が同定されているときに，形態素 m_i の直後に読点が挿入される確率，ならびに，挿入されない確率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，表 2.5, 2.6 に示す素性を用いた．形態素情報と文節情報は基本的な言語情報として使用する素性であり，それ以外は 2.2 節の分析に基づき定めた素性である．ただし，2.2.7 節の用法に対応する素性は形態素情報の品詞で表現されているため，個別に設定することはしていない．また，読点によって挟まれた文字列の文字数の分類は，学習データを用いたクローズドテストを試行し，いくつか設定した分類のうち，最も高い性能を示したものを採用した．文節情報に関する素性において，主辞は，各文節内で，品詞が記号，助詞，助動詞，名詞-接尾となるものを除き，最も文末に近い形態素を，語形は，各文節内で，記号を除き最も文末に近い形態素をそれぞれ意味する．

2.4 実験

本手法の有効性を評価するため，読点の挿入実験を実施した．

2.4.1 実験概要

実験には京都コーパス [54] に収録されている新聞記事テキストを用いた．テストデータには 1 月 14 日から 17 日の全記事を，学習データには分析データと同一のテキストを使用した．テストデータの規模を表 2.7 に示す．また，テストデータにおける読点の用法の分布を調査した結果を表 2.8 に示す．この調査では，テストデータから 400 文をランダムに抽出し，そのすべての読点（800 個）について，それが

表 2.5: 最大エントロピー法で用いた素性 (1)

形態素情報	$m_{i-1}, m_i, m_{i+1}, m_{i+2}$ の各形態素について，以下の素性を用いる • 品詞 • 活用形
文節情報	• m_i が文節の最終形態素であるか否か
	m_i が文節の最終形態素である場合，以下の素性を用いる • m_i が属する文節の主辞の品詞 • m_i が属する文節の主辞の活用形 • m_i が属する文節の語形の品詞
	• “m_i が文節の最終形態素，かつ，m_i が属する文節の語形の品詞が助詞” である場合，語形の表層文字
節境界を明確にする読点	• m_i が文節の最終形態素である場合，m_i が節の最終形態素であるか否か
	• m_i が節の最終形態素である場合，節のラベル
係り受け関係を明確にする読点	m_i が文節の最終形態素である場合，以下の素性を用いる • m_i が属する文節が，直後の文節に係るか否か • m_i が属する文節が，直前の文節から係られるか否か • m_i が属する文節が，直後の文節が属する節の節末文節より後方に係るか否か • 直前の読点から m_i が属する文節までの文字列内で係り受けが閉じているか否か
難読・誤読を避ける読点	m_i が文節の最終形態素である場合，以下の素性を用いる • “m_i が漢字，かつ，m_{i+1} が漢字” であるか否か • “m_i がカタカナ，かつ，m_{i+1} がカタカナ” であるか否か
主題を示す読点	“m_i が節の最終形態素，かつ，その節ラベルが主題ハ” である場合，以下の素性を用いる • m_i が属する文節が，直後の文節に係るか否か • m_{i-1} が格助詞「で」であるか否か • “m_i が属する文節と係り先が同一，かつ，主辞の品詞が動詞” である文節が存在するか否か • m_i を最終形態素とする主題を示す語句の文字数

どの用法で打たれたものかを特定した．ただし，1 つの読点が複数の用法を担う場合があり，用法の重複を許して調査したため，表 2.8 の分布では，用法ごとにそれを担う読点の割合を示している．

実験では，読点を除いた文を入力とし，形態素，係り受けは京都コーパスのデー

表 2.6: 最大エントロピー法で用いた素性 (2)

先頭の接続詞・副詞を区切る読点	m_i が文頭文節の最終形態素である場合，以下の素性を用いる • m_i の品詞が接続詞であるか否か • m_i の品詞が副詞であるか否か
並列する単語・句の間に打たれる読点	m_i が文節の最終形態素である場合，以下の素性を用いる • “m_i の品詞が名詞，かつ，直後の文節の最終形態素の品詞が名詞” であるか否か • 以下の条件をともに満たすか否か - m_i が属する文節の主辞の品詞が動詞である - 次の 2 つの文節（m_i が属する文節と，主辞の品詞が動詞であり m_i が属する文節より後方に存在する文節）がともに文末の動詞に係る
引用を示す読点	• “m_{i+1} が格助詞「と」，かつ，m_{i+1} が文節の最終形態素” であるか否か
	“m_{i+1} が格助詞「と」，かつ，m_{i+1} が文節の最終形態素” である場合，以下の素性を用いる • m_i の品詞 • m_{i+2} の品詞
	• “m_{i+1} が格助詞「と」，かつ，m_{i+2} が接続助詞「の」，かつ，m_{i+2} が文節の最終形態素”であるか否か
	• “m_{i+1} が格助詞「と」，かつ，m_{i+2} が接続助詞「の」，かつ，m_{i+2} が文節の最終形態素” である場合，m_i の品詞
読点によって挟まれた文字列の文字数	• 直前の読点から m_i までの形態素列の文字数が以下の 4 分類のいずれであるか （1 文字，2 文字以上 4 文字以下，5 文字以上 22 文字以下，23 文字以上）

タを，及び，節境界は解析ツール CBAP[91] で機械的に付与したものを使用した．また，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，京都コーパスにおける読点位置を正解の読点位置とし，正解に対する再現率，及び，適合率により行った．再現率，適合率はそれぞれ，

$$\text{再現率} = \frac{\text{正しく挿入された読点数}}{\text{正解の読点数}}$$

表 2.7: テストデータの規模

文数	4,659
形態素数	131,810
文節数	45,727
文字数	198,899
読点数	6,640

表 2.8: 読点の用法の分布

用法	割合	
節を区切る読点	42.13%	(337/800)
係り受け関係を明確にする読点	7.75%	(62/800)
難読・誤読を避ける読点	2.88%	(23/800)
主題を示す読点	10.50%	(84/800)
先頭の接続詞・副詞を区切る読点	6.13%	(49/800)
並列する単語・句を明確にする読点	19.25%	(154/800)
時間を表わす副詞を区切る読点	8.13%	(65/800)
直前の語句を強調するための読点	6.25%	(50/800)
引用を示す読点	1.50%	(12/800)

$$\text{適合率} = \frac{\text{正しく挿入された読点数}}{\text{挿入された読点数}}$$

を測定した.

読点の用法ごとの分析に基づいて決定した素性の有効性を評価するため，表 2.5 と表 2.6 に示した素性のうち，形態素情報と文節情報のみを用いて読点を挿入する手法をベースラインとして設定し，性能を比較した.

さらに，読点の用法に基づいて定めた各素性の有効性を個別に評価するため，表 2.5 と表 2.6 に示した素性から，各用法に対応する素性を取り除き読点挿入を実施し，全ての素性を使用した場合と性能を比較した.

2.4.2　実験結果

提案手法ならびにベースライン手法の再現率，適合率，及び，それらの調和平均である F 値を表 2.9 に示す．提案手法は，再現率で 70.66%，適合率で 84.65%を達成

表 2.9: 実験結果

	再現率	適合率	F 値
提案手法	70.66% (4,692/6,640)	84.65% (4,692/5,543)	77.02
ベースライン	63.96% (4,247/6,640)	80.59% (4,247/5,270)	71.32

表 2.10: 各用法に対応する素性を除いた場合の読点挿入結果

用法	F 値の差
節境界を明確にする読点	-0.55
係り受け関係を明確にする読点	-3.50
難読・誤読を避ける読点	-0.20
主題を示す読点	-0.19
先頭の接続詞・副詞を区切る読点	-0.01
並列する単語・句の間に打たれる読点	-0.18
引用を示す読点	-0.10
読点によって挟まれた文字列の文字数	-0.05

した．いずれもベースライン手法と比較して高い性能を示しており，提案手法の有効性を確認した．

提案手法とベースライン手法による読点挿入結果のテキスト例を図 2.8 に示す．ベースライン手法によるテキストでは，先頭の接続詞「しかし」の直後や節の境界である「停車したままとなったのをはじめ」の直後，主題を示す語句「その前久と義景では」の直後に読点が挿入されていない．一方，提案手法ではそのような位置に正しく読点が挿入されている．

正解の読点位置のうち，提案手法，ベースライン手法のいずれにおいても読点が挿入された箇所は 3,962 箇所であった．提案手法は，ベースライン手法が正解の位置に挿入した読点の 93.29%（3,962/4,247）をカバーしており，さらに，ベースライン手法が挿入できなかった正解位置の 30.51%（730/2,393）に正しく読点を挿入できている．

表 2.10 に，各用法に対応する素性を除いた場合の読点挿入結果を示す．この表の 2 列目は，全ての素性を使用した場合と各素性を除いた場合との F 値の差を示す．いずれの場合においても F 値が低下しており，各素性の有効性が確認された．

例文 (1)

提案手法:

しかし、激しい雪の影響で復旧作業は進まず、直江津発上野行きの特急「あさま３８号」が妙高高原駅で停車したままとなったのをはじめ、計七本が長野、新潟両県内の駅で立ち往生。

ベースライン:

しかし激しい雪の影響で復旧作業は進まず、直江津発上野行きの特急「あさま３８号」が妙高高原駅で停車したままとなったのをはじめ計七本が長野、新潟両県内の駅で立ち往生。

例文 (2)

提案手法:

その前久と義景では、義景が三つ年上であることから、十中八九、義景に嫁いだのは、娘でなく妹であろう。

ベースライン:

その前久と義景では義景が三つ年上であることから十中八九、義景に嫁いだのは娘でなく妹であろう。

図 2.8: 提案手法とベースライン手法による読点挿入結果の比較

2.5　考察

2.5.1　読点挿入誤りの原因

正解の読点位置のうち，読点が挿入されなかった箇所は 1,948 箇所であった．そのうち，898 箇所は節境界であり，節境界「主題ハ」がその 52.07%（464/891）を占めていた．そもそも節境界「主題ハ」は，読点挿入率がそれほど高いわけでなく判定は容易ではないが，出現数が多く誤りの影響は大きい．表 2.11 に節境界「主題ハ」に対する読点挿入の再現率，及び，適合率を示す．テストデータ中で,「主題ハ」に挿入されている読点は 611 箇所存在したが，そのうち正しく挿入できたのは 147 箇所に過ぎなかった．表 2.5 に示した通り，本手法では節境界「主題ハ」に関する素性を 4 種類導入しているが，それらが十分に機能しなかった可能性がある．素性の更なる検討が今後必要である．

表 2.11: 節境界「主題ハ」に対する読点挿入結果

再現率	適合率	F 値
24.06% (147/611)	63.09% (147/233)	34.84

なお，読点が挿入できなった節境界のうち,「主題ハ」に次いで多かったのは節境界「テ節」であり，114 箇所であった.

節境界以外では，並列された名詞間に読点が挿入されなかった箇所が 68 箇所存在した．以下に挿入結果の例を示す.

- ボウルに豚の背脂ニンニク、ショウガ、ネギのみじん切りを入れ、彩りの赤ピーマンも加えます。

この例では,「背脂」と「ニンニク」の間の読点を挿入できなかった．ただし,「ニンニク」,「ショウガ」,「ネギ」の間には正しく読点が挿入されている．これは名詞が並列されていることの他に，カタカナが文節にまたがって出現していることが要因として考えられる.

一方，読点が挿入された箇所のうち，正解の読点位置と異なるものは 851 箇所存在した．そのうち，文節にまたがって漢字が出現する位置に誤って読点を挿入した箇所が 70 箇所存在した．例えば，以下の例では，文節にまたがって漢字が出現する文節「二千」と「近い」の間に誤って読点が挿入された.

- 専門家集団がいて、常時二千、近いプロジェクトを走らせている。

2.5.2 人間による読点挿入の一致率

実験では，正解データとの比較によって読点挿入の結果を評価した．しかし，実験結果の値がどの程度の値を示せば十分なのかは定かではない．そこで，人間による読点挿入作業を行い，その結果を一つの指標とし，提案手法の読点挿入性能を評価した．京都コーパスの 1 月 14 日の記事中の 250 文 (2,709 文節，7,580 形態素) に対して，作業者 4 名が読点挿入を行った.

正解データに対する作業者，及び，提案手法の再現率，適合率とその F 値を表 2.12 に示す．作業者による差は大きく，読点挿入作業は人間でも揺れが生じ，単純でな

表 2.12: 人間による読点挿入との比較

	再現率	適合率	F 値
作業者 A	77.34% (297/384)	81.37% (297/365)	79.30
作業者 B	87.24% (335/384)	58.06% (335/577)	69.72
作業者 C	41.92% (161/384)	89.44% (161/180)	57.09
作業者 D	65.36% (251/384)	69.15% (251/363)	67.20
提案手法	71.61% (275/384)	85.08% (275/331)	77.77

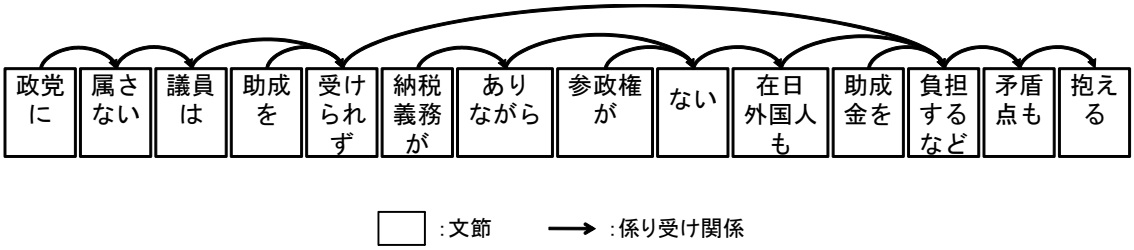
いタスクであることが分かる．提案手法は F 値において，それが最も高かった作業者 A の 98.07%（77.77/79.30）を達成している．また提案手法は，適合率では作業者 A を上回っており，高い読点挿入性能を備えていることがわかる．

2.5.3　不自然な読点挿入

2.4 節の実験の読点挿入誤りには，正解の読点位置とは異なるが許容できるものと，明らかに不自然で許容できないものの 2 種類存在する．後者の場合，読み手が文の意味を取り違える，あるいは，係り受け構造を誤って認識するなど，誤りが与える影響は大きい．そこで，提案手法によって明らかに不自然な位置に挿入された読点について調査した．2.5.2 節で用いたデータと同一の 250 文に対する読点挿入結果のうち，提案手法が正解と異なる位置に挿入した 56 個の読点について，それぞれ許容できるか否かを判定した．判定は，2.5.2 節の作業者とは異なる 3 名の被験者による協議のもと決定した．

調査の結果，挿入誤りのうち，許容できないと判定されたものは 8.92%（5/56）に過ぎず，提案手法は，たとえ正解と異なったとしても，ある程度自然な位置に読点を挿入できているといえる．以下では，許容できないと判定された 5 箇所の読点とその判定理由を示す．該当する読点に下線を記す．

1. 同省の特殊法人は計十三、あり、所管省庁別では運輸省に次いで多い。



政党に属さない議員は助成を受けられず、納税義務がありながら、参政権がない在日外国人も助成金を負担するなど矛盾点も抱える。

図 2.9: 不自然な読点挿入 4. の係り受け構造

2. 残る四党のうち政党助成制度に反対している共産党を除いて、自由連合、民改連、スポーツ平和の三党も締め切り、期限の十七日には届け出るとみられる。

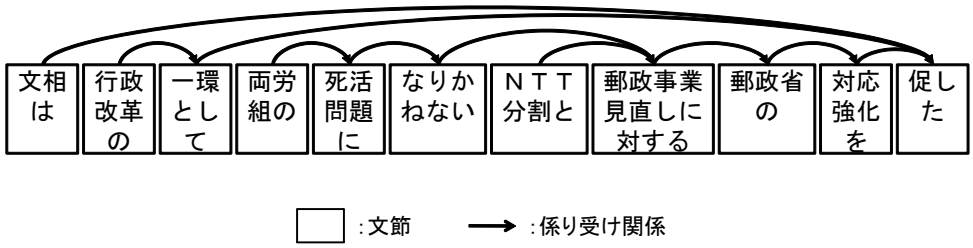
3. 九二年と比べ、一ポイント減で下向き、傾向、という違いがある。

上記の三つの例ではそれぞれ,「十三あり」,「締め切り期限」,「下向き傾向」がそれぞれ強いまとまりを形成しているにも関わらず，それらの間に読点が挿入されている.

4. 政党に属さない議員は助成を受けられず、納税義務がありながら、参政権がない在日外国人も助成金を負担するなど矛盾点も抱える。

5. 文相は行政改革の一環として、両労組の死活問題になりかねないＮＴＴ分割と、郵政事業見直しに対する郵政省の対応強化を促した。

上記二つの例に対する係り受け構造を図 2.9 と図 2.10 にそれぞれ示す．上記一つ目の例では，文節「ありながら」の直後に読点が挿入されることにより,「納税義務がありながら参政権がない在日外国人」という係り受け関係が,「政党に属さない議員は納税義務がありながら」と誤って解釈される可能性がある．また，二つ目の例では，本来,「両労組の死活問題になりかねない」は「ＮＴＴ分割」と「郵政事業見直し」の双方に係る．しかし,「ＮＴＴ分割」の直後に読点が挿入されることにより,「ＮＴＴ分割」のみに係り,「ＮＴＴ分割」と「郵政省の対応強化」が並列した構造であるとして誤って解釈される恐れがある.



文相は行政改革の一環として、両労組の死活問題になりかねないＮＴＴ分割と、郵政事業見直しに対する郵政省の対応強化を促した。

図 2.10: 不自然な読点挿入 5. の係り受け構造

2.5.4　テキストの自動解析に基づく読点挿入性能

2.4 節の実験では人手で解析されたテキストを用いて本手法の性能を評価したが，実際に日本語テキストに読点を自動挿入する場合には，テキストを機械的に解析した結果を用いることになる．そこで，形態素解析，文節まとめ上げ，係り受け解析，節境界解析が機械的に与えられたテキストに基づいて 2.4 節と同様の実験を行い，提案手法とベースライン手法との性能比較を行った．なお，形態素解析には ChaSen [88] を，文節まとめ上げ，係り受け解析には CaboCha [56] を，節境界解析には CBAP [91] をそれぞれ用いた．CaboCha は，京都コーパス 1 月 1 日から 11 日の記事データから読点を取り除いたデータを用いて文節区切り，係り受け解析のモデルを学習した．

結果を表 2.13 に示す．人手で解析されたテキストを使用した実験と比較して，提案手法は適合率で 3.35%，再現率では 9.79%低下した．読点挿入において文節区切りの情報は重要であるが，自動解析では文節区切りを誤ることがあり，読点挿入性能が低下したと考えられる．しかしながら，提案手法は，ベースライン手法の性能を大きく上回っており，また，2.5.2 節の作業者 4 人の F 値の平均値（68.33）と同程度の値を達成しており，自動読点挿入手法としての利用可能性を確認した．

2.5.5　関連研究との性能比較

既存の読点挿入手法と比較することによって本手法の性能を評価した．比較手法として，日本語を対象としており，本研究と同様に統計的アプローチを用いている

表 2.13: 自動解析に基づく読点挿入

	再現率	適合率	F 値
提案手法	60.87% (4,051/6,640)	81.30% (4,051/4,998)	69.62
ベースライン	54.28% (3,604/6,640)	77.86% (3,604/4,629)	63.97

表 2.14: 秋田らの手法による実験結果

再現率	適合率	F 値
61.48% (4,082/6,640)	77.86% (4,082/5,243)	68.70

秋田らの手法 [36] を採用した．秋田らは講演音声を対象に，読点の揺れの分析と自動挿入の検討を与えている．日本語話し言葉コーパス（CSJ）に速記者 3 名が句読点を挿入したテキストを用いて，条件付き確率場に基づく句読点挿入のための識別器を構成している．素性として，単語（出現形），文節境界，直後の文節への係り受け情報，品詞（大分類）を使用し，前後 3 単語分が識別器に入力される．

秋田らの手法を実装し，2.4 節の実験と同一のデータを用いて比較実験を行った．秋田らの手法に基づく実験結果を表 2.14 に示す．提案手法（表 2.9）と比較して，再現率で 9.18%，適合率で 6.79%下回っており，提案手法の優位性を確認した．

秋田らは話し言葉を対象として研究を行っている．話し言葉を構文解析する際には解析誤りが多く含まれるため，秋田らの手法では，使用する素性をあらかじめ限定している．また，使用されている素性も単純なものにとどまっている．このことが，秋田らの手法で再現率が下回った原因として考えられる．それに対し，本手法では，読点の用法を分類し，その分類ごとの詳細な分析に基づいて素性を細かく定めており，それが精度の高い読点挿入を可能にしている．

2.6 おわりに

本章では，文字列を整形して読みやすいテキスト提示を行うための方法として，日本語テキストにおける読点の自動挿入手法を提案した．日本語テキストのチャンキングに基づき適切な読点挿入を行うことによって，高い品質を備えたテキストへと整形することが可能となる．本手法では，読点の用法に注目し， 形態素や係り

受け，節境界等の情報に基づき，統計的手法によって一文中の適切な読点挿入位置を同定する．京都コーパスを用いた読点の挿入実験では再現率で 70.66%，適合率で 84.65%であり，本手法の有効性を確認した．

第3章　講演テキストへの改行挿入

3.1　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．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，改行挿入によって字幕を適切にチャンキングして提示することが望まれる．

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 [94]．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．

本章では，読みやすい字幕を提示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のチャンキングに基づく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，[26, 62, 91]）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．

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

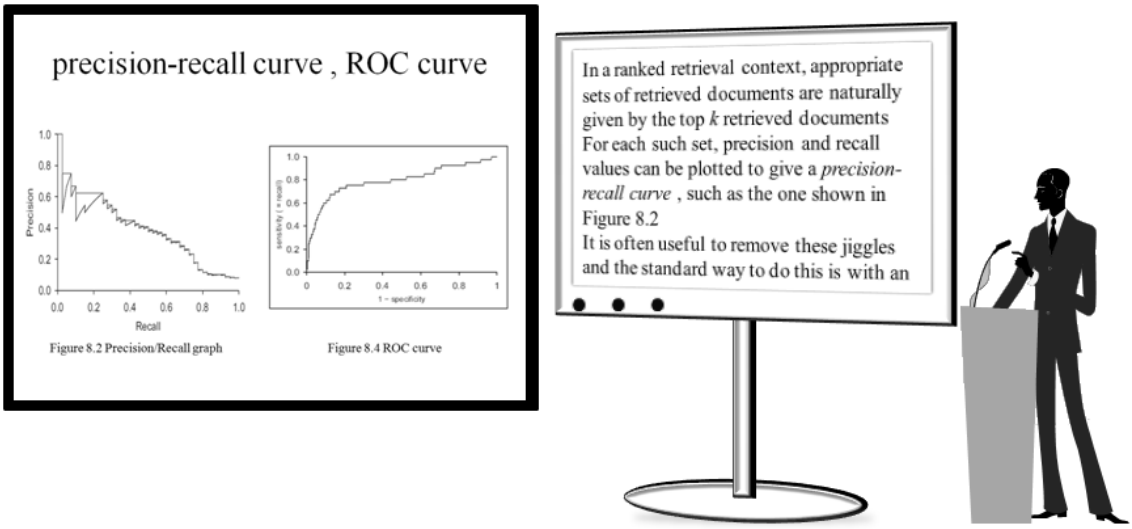


図 3.1: 講演音声の字幕提示環境

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．

3.2　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替わり，スクロールしながら常に数行表示される．図 3.1 に，想定する字幕提示環境を示す．

図 3.2 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 3.3 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．同時通訳データベース [25] 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

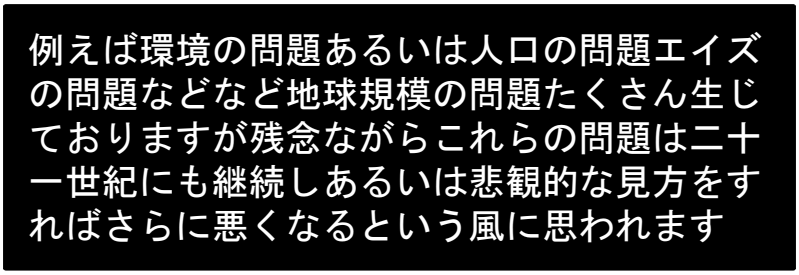


図 3.2: 講演音声の書き起こしテキスト

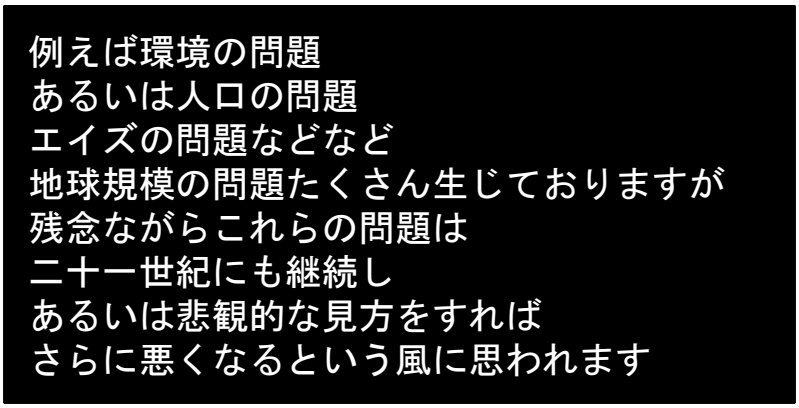


図 3.3: 適切な位置に改行が挿入されたテキスト

を用意した．50 文の平均文字数は 71.0 文字である．なお，同時通訳データベース [25] 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．図 3.2 は（1）のテキストに，図 3.3 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 3.4 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．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．

そこで本章では，講演テキストを読みやすくチャンキングするための改行挿入手法を実現する．改行挿入の手法を実現するにあたり，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，講演音声を文字化する

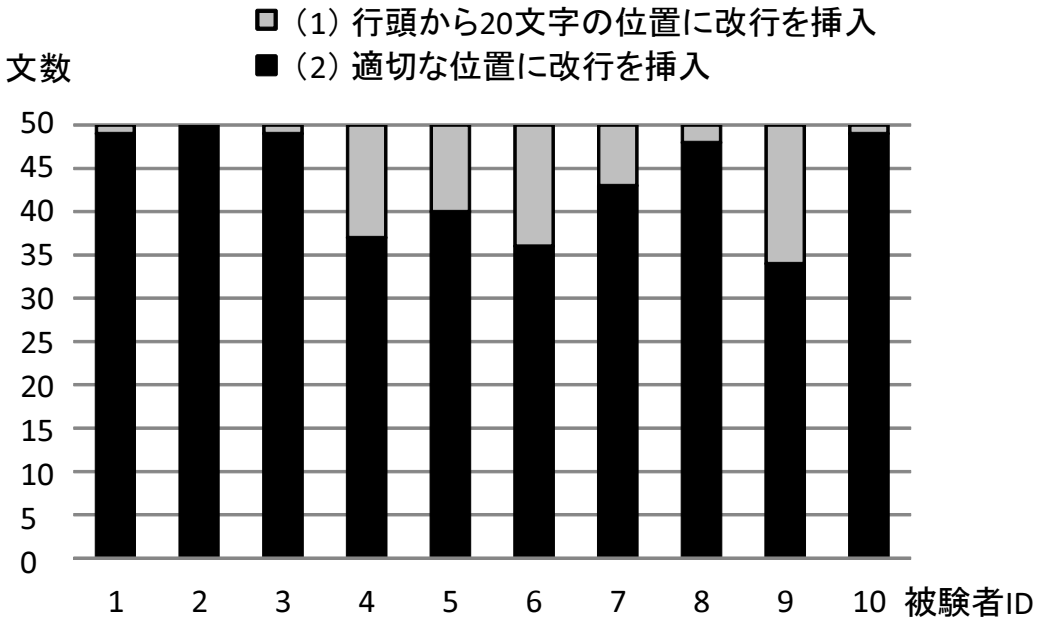


図 3.4: 講演テキストへの改行挿入の効果

のに方法がいくつか考えられるためである [78]．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，音声を直接認識する方式，復唱音声を認識する方式などがあり，さら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記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．

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．

表 3.1: 分析データのサイズ

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

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**という．

3.3 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，同時通訳データベース [25] 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，ならびに，改行点が，人手により付与されている．改行点は 3.2 節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 3.1 に示す．

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，代表的なプレゼンテーションソフトウェアである PowerPoint における文字の可読性を考慮して決定した．文節境界（すなわち，改行点候補）2,670 箇所に対して，833 箇所に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．

3.3.1 節境界と改行点

節は構文的かつ意味的なまとまりを形成する言語単位の 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．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．

表 3.2: 節境界への改行挿入率

節境界	改行挿入率 (%)
主題ハ	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)

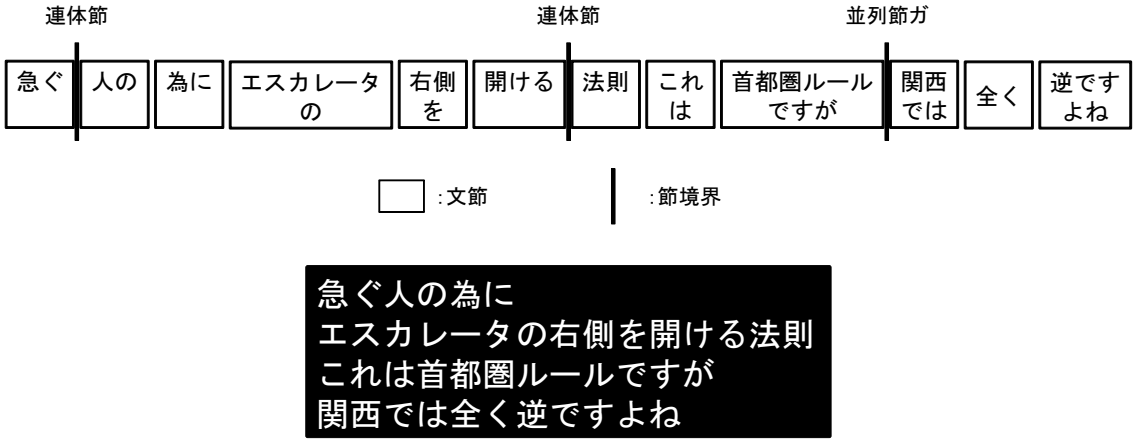


図 3.5: 節境界の種類と改行点の関係

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．節境界の種類として，節境界解析ツール CBAP[91] で定義されたものを用いた．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 3.2 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 3.5 に節境界への改行挿入の例を示す．この例では，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．

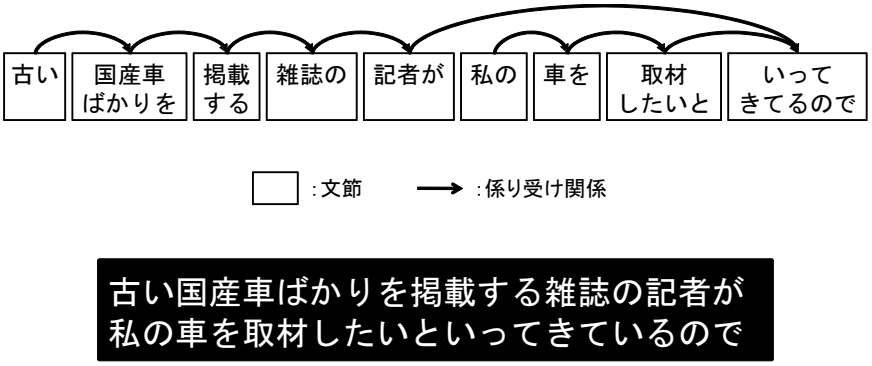


図 3.6: 隣接文節間の係り受け関係と改行点の関係

3.3.2 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 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．図 3.6 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 3.6 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 3.7 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 3.8 にそのような改行挿入の例を示す．図 3.8 の例では，文節「しなくなったんですけれども」までで 1 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「しなくなったんですけれども」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

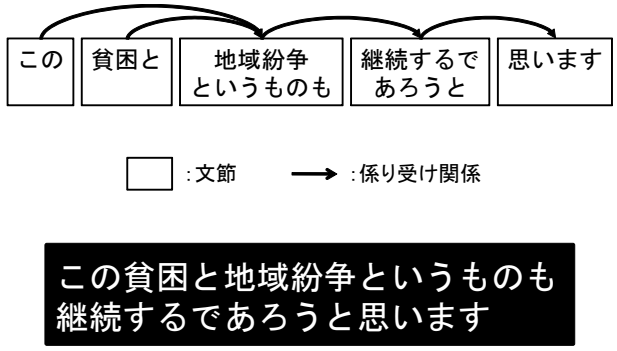


図 3.7: 係り受け関係にある隣接文節間への改行挿入例

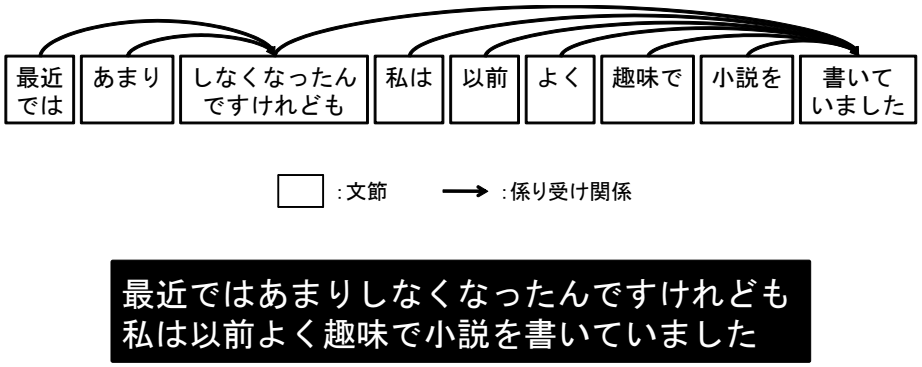


図 3.8: 行内の係り受け構造と改行点の関係

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．

3.3.3　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（本分析では 7 文字以上）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．

3.3.4　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得られている [68]．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 3.9 にお

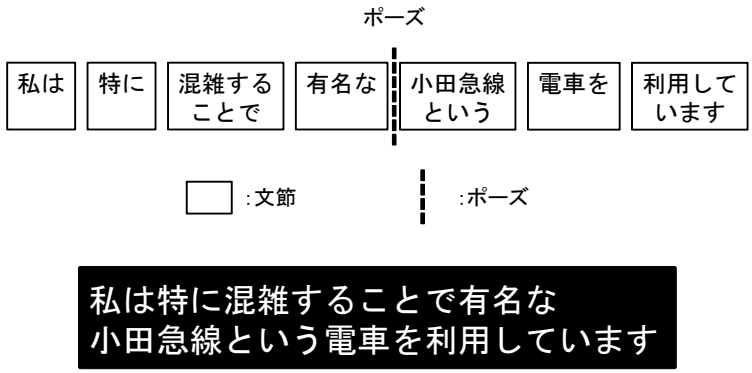


図 3.9: ポーズと改行点の関係

表 3.3: 行頭での出現率が低い形態素

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

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，0.2 秒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わかった．

3.3.5 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 3.3 に示す．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．形態素体系は日本語辞書 IPADIC の品詞体系 [38] に準拠した．表 3.3 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．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．

3.4　統計的な改行挿入手法

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 $L_j = b^j_1 \cdots b^j_{n_j} (1 \leq j \leq m)$ とした場合，$1 \leq k < n_j$ のとき $r^j_k = 0$，$k = n_j$ のとき $r^j_k = 1$ となる．

3.4.1　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．

$$
\begin{aligned}
& P(R|B) \qquad (3.1)\\
= \;& P(r^1_1 = 0, \cdots, r^1_{n_1-1} = 0, r^1_{n_1} = 1, \cdots,\\
& \qquad r^m_1 = 0, \cdots, r^m_{n_m-1} = 0, r^m_{n_m} = 1|B)\\
\cong \;& P(r^1_1 = 0|B) \times \cdots\\
& \times P(r^1_{n_1-1} = 0|r^1_{n_1-2} = 0, \cdots, r^1_1 = 0, B)\\
& \times P(r^1_{n_1} = 1|r^1_{n_1-1} = 0, \cdots, r^1_1 = 0, B) \times \cdots\\
& \times P(r^m_1 = 0|r^{m-1}_{n_{m-1}} = 1, B) \times \cdots\\
& \times P(r^m_{n_m-1} = 0|r^m_{n_m-2} = 0, \cdots, r^m_1 = 0, r^{m-1}_{n_{m-1}} = 1, B)\\
& \times P(r^m_{n_m} = 1|r^m_{n_m-1} = 0, \cdots, r^m_1 = 0, r^{m-1}_{n_{m-1}} = 1, B)
\end{aligned}
$$

ここで，$P\left(r_k^j = 1 | r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B\right)$ 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，$P\left(r_k^j = 0 | r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B\right)$ 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（3.1）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．

3.4.2 最大エントロピー法で用いた素性

本研究では，$P\left(r_k^j = 1 | r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B\right)$ ならびに $P\left(r_k^j = 0 | r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B\right)$ を最大エントロピー法により推定する際，3.3 節の分析に基づき，表 3.4 に示す素性を用いた．なお，行長に関する素性では，3.3.3 節の分析において，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．

3.5 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．

3.5.1 実験概要

実験データとして，同時通訳データベース [25] 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．

表 3.4: 最大エントロピー法で用いた素性

形態素情報	b_k^j の主辞の品詞
	b_k^j の主辞の活用形
	b_k^j の語形の品詞
節境界情報	b_k^j が節の最終文節であるか否か
	b_k^j が節の最終文節である場合，節のラベル
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（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表層文字が「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-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

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 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．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）．

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
それに関することを
若干話させて頂きたいと思います

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います

そして最後に
二十一世紀の日本外交なんて言ってしまって
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と
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います

図 3.10: 正解データの例

- **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．ただし，最長文字数内に係り受け関係に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，一行の最長文字数を 20 文字とした．正解の改行データは，3.2 節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 3.10 に示す．評価データ全体で改行点は 5,497 箇所存在した．

表 3.5: 実験結果

	再現率	適合率	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

3.5.2　実験結果

提案手法ならびに各ベースラインの適合率と再現率を表 3.5 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．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，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．

3.5.3　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

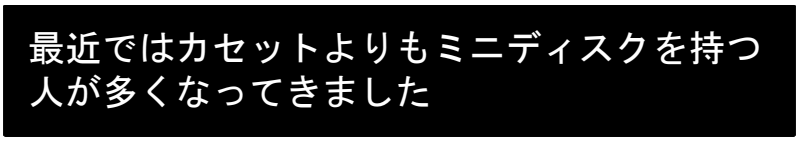


図 3.11: 連体節に関する改行挿入誤り

図 3.12: 主題ハに関する改行挿入誤り

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．3.3.1 節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．図 3.11 に例を示す．図 3.11 の例では,「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」というまとまりが捉えられていない．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「主題ハ」であった．例えば，図 3.12 に示す提案手法による改行挿入結果では,「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「主題ハ」は他の種類の節境界に比べ，出現数が多い．また，表 3.2 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,「主題ハ」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．

3.6 考察

本節では，提案手法の有効性について，より詳細に検討するために，3.5.2 節の実験結果について考察する．

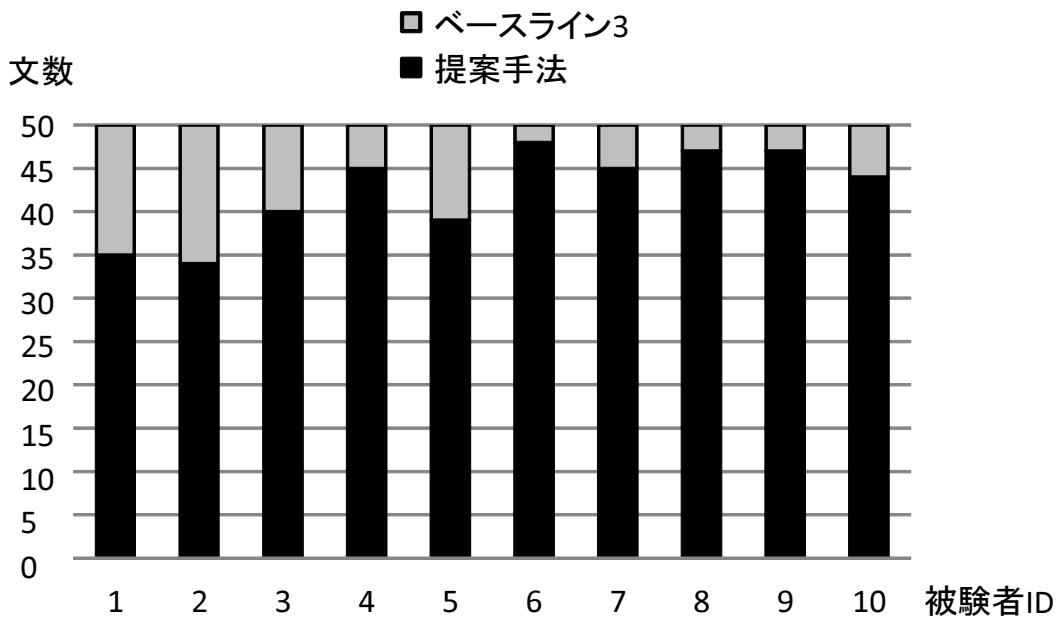


図 3.13: 被験者による主観的評価の結果

3.6.1　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チャンキングし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，3.5 節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 3.13 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．

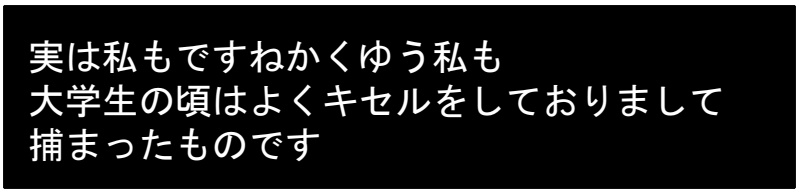


図 3.14: 平仮名が連続する場合の例

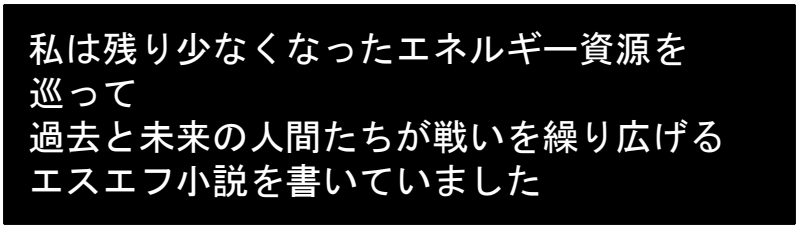


図 3.15: 極端に長さが違う行の出現

- 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 3.14，3.15 にそれぞれ示す．図 3.14 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 3.15 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．

3.6.2 人間による改行挿入の一致率

3.5.2 節では，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 3.6 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．

表 3.6: 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	適合率	F 値
89.82% (459/511)	89.82% (459/511)	89.82

表 3.7: 自動的に言語解析されたデータに対する実験結果

	再現率	適合率	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

3.6.3　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，節境界解析ならびに係り受け解析を機械的に実行した結果に基づいて，3.5.2 節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[56] を，節境界検出には節境界解析ツール CBAP[91] をそれぞれ用いた．

実験結果を表 3.7 に示す．人手で付与した情報を利用した実験結果（表 3.5）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．

3.6.4　チャンクの連結に基づく改行挿入との比較

西光らは，講演などの話し言葉を適当な単位に分割するための段階的チャンキング方式を提案している [60]．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

表 3.8: 西光らの手法による実験結果

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

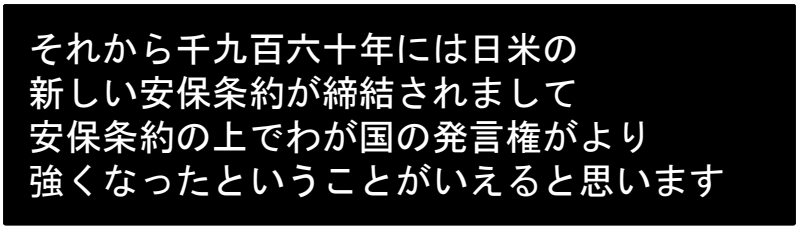


図 3.16: 西光らの手法による改行挿入の例

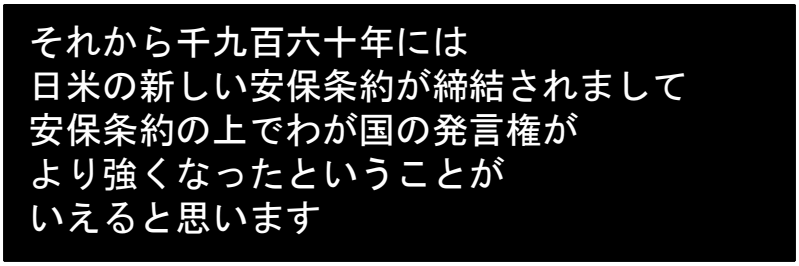


図 3.17: 提案手法による改行挿入の例

目し，そのような係り受け関係，さらには，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 [61].

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さらに考察を与えるために，西光らの手法に基づく改行挿入手法 [61] を実装し，3.5.2 節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 3.8 に示す．提案手法による実験結果（表 3.5）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

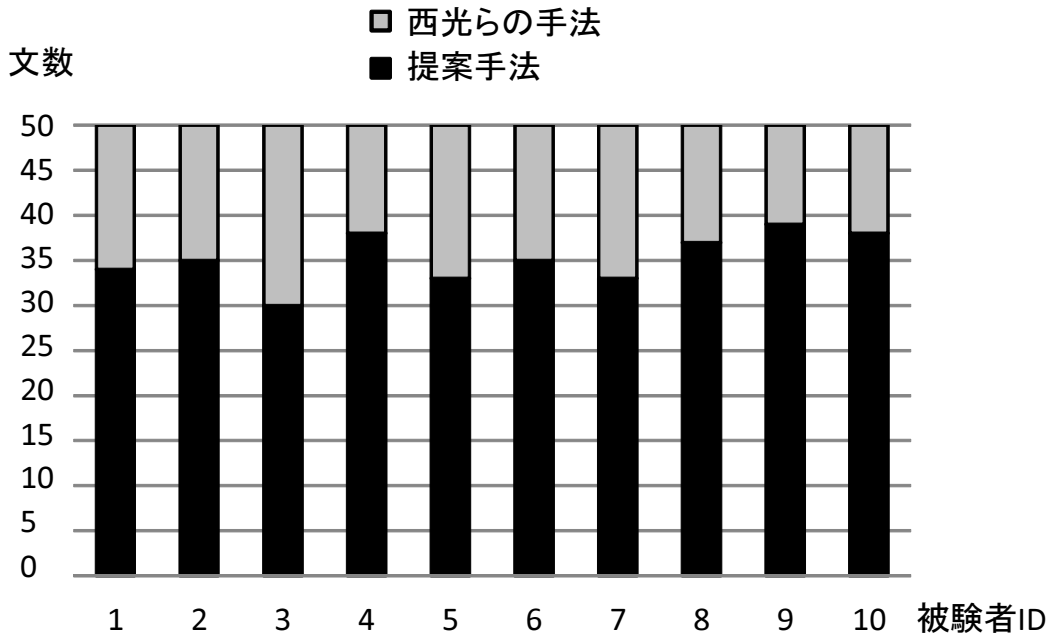


図 3.18: 西光らの手法と提案手法の主観的評価の結果

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わかる．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．

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．西光らの手法による改行挿入の例（図 3.16）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 3.17）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．

主観的評価による比較を行った．3.6.1 節と同様の手順で，同時通訳データベース [25] 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 3.18 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．

3.7 おわりに

本章では，文字列を適切に配置して読みやすい字幕テキストを提示するための方法として，日本語講演データへの改行挿入手法を提案した．この手法は，聴覚障害者，高齢者，外国人等による音声理解の支援に利用できる．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．

第4章　講演テキストへの逐次的な改行挿入

4.1　はじめに

第3章で提案した手法や文献[94]の手法など，字幕テキストへの自動改行挿入に関する研究がいくつか行われており，文中の適切な改行位置を精度よく同定することができる．しかし，これらの方法は，字幕テキストがあらかじめ与えられていることを前提としており，講演の進行と同期したリアルタイムでの字幕生成には必ずしも適さない．なぜなら，聴衆にとっては，字幕が音声入力にできる限り追従して提示されることが望ましく，そのためには，遅延時間，すなわち，音声が発声されてから字幕を表示するまでの時間を少なくすることが要求されるためである．実際，字幕生成のための音声認識に関する研究[3, 27, 34]において，遅延時間の短縮が重要な課題の1つになっており，改行挿入位置の決定においても，入力に対する出力の同時性を考慮する必要がある．

本章では，音声に対してリアルタイムに読みやすい字幕を提示するために，字幕テキストのチャンキングに基づく逐次的な改行挿入手法について述べる．本研究では，自動音声認識やパソコン要約筆記などにより音声が同時的に文字化され，その進行に応じて逐次，改行位置を同定し，改行位置が決まるごとにチャンキングされた行を提示するシステムを想定する．この場合，改行位置の同定をどのようなタイミングで実行するかが問題となる．というのも，精度よく同定するためには，ある程度の長さの音声が入力されてから利用可能な情報をできる限り考慮することが望ましく，一方で，少ない遅延時間で字幕を提示するには，細かい単位ごとに改行位置同定処理を実行し，字幕テキストを行へチャンキングすることが望まれるためである．

本章ではまず，字幕提示のリアルタイム性を最も重視した方式として，文節単位で同定処理を実行する手法を提案する．改行位置は文節境界にあることを前提とするとき，字幕提示の遅延時間が最も少なくなる方式である．本手法では，文節が入

力されるごとに，改行挿入位置を統計的に同定する．第 3 章で，統計的なアプローチによる改行挿入において，係り受けやポーズ，行長などの素性の使用が効果的であることを明らかになっており，本手法では，それらの素性のうち，同定処理の時点で取得可能な情報を利用する．日本語講演データを用いて実験を実施し，本手法の改行挿入性能を，文単位での改行挿入手法と比較することにより評価した．

本章では次に，字幕提示のリアルタイム性と改行位置の適格さの双方を考慮する手法として，文よりも短く，文節よりも長い言語単位での改行挿入手法を実現し，その改行挿入性能について考察する．そのような言語単位として節を採用し，文節ごとの手法との間で比較評価する．評価は，被験者実験により総合的に実施し，同定処理のタイミングと精度が評価に及ぼす影響について議論する．

4.2　リアルタイム字幕生成のための改行挿入

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．本研究では，ディスプレイ上で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ことを前提とする．

図 4.1 に示すように，

> それから南北の格差北側の先進工業国と南側の開発途上国との間の格差を縮めるというのは貧困の国があればですねそれらの国々は政治的にも不安定になりますし工業国からの輸入品も買えないということになるんでそういう南北の格差の是正というのに国連は努力してきました

という 1 文の音声に対して，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 4.2 に示すように，読みやすい位置に改行を挿入して整形した文を提示することが重要である．

また，読みやすい字幕を提示する上で，字幕出力の遅延時間を考慮する必要がある．たとえ，図 4.2 に示すような字幕を表示できたとしても，1 文の発話が終了するまで何も表示されないのであれば，読み手が，話者の音声や身振り，またスライドと同期して内容を理解することは難しくなる．また，ディスプレイが変化しない時間が長くなると，読み手が安定したペースで字幕テキストを読むことが困難になる

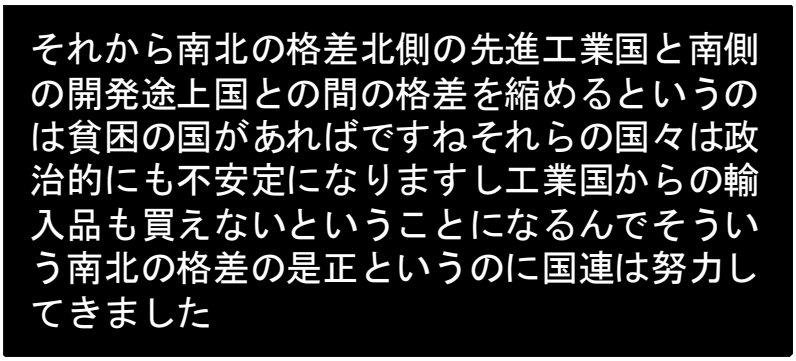


図 4.1: 講演の書き起こし

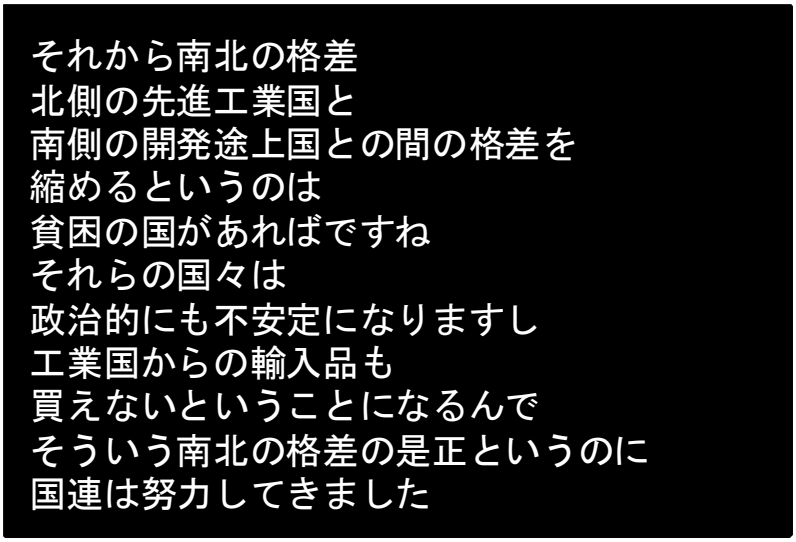


図 4.2: 適切な位置に改行が挿入された書き起こし

とともに，システムが正しく動作しているのかについて読み手に不安を与えることにもなる．図 4.3 のように，改行位置を早い段階で同定し，字幕を即座に出力することにより，字幕出力の遅延時間を短縮することが重要となる．

なお本研究では，字幕生成における改行挿入に関して，以下の前提を設けた．

- ディスプレイの大きさを考慮して行の最長文字数を 20 に設定し，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．
- 改行位置が同定され次第，その改行位置までを，行ごとに字幕出力する．

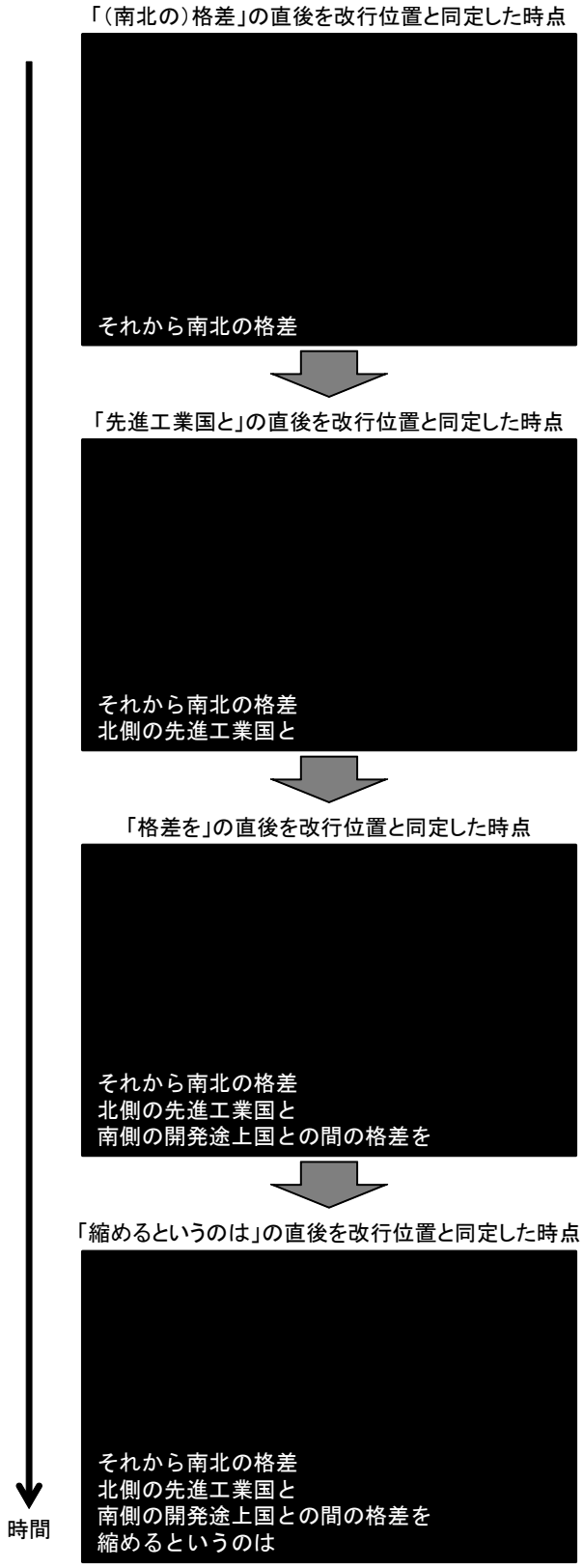


図 4.3: テキストの出力タイミング

4.3 逐次的な改行挿入手法

本節では，音声が入力されてからできる限り早期に字幕テキストの「行」を確定し，それを提示するための方法として，文節が入力されるごとに，入力された文節とその直前の文節の境界に改行を挿入するか否かを決定する手法について述べる．文節境界は前後 2 形態素の情報を用いることにより高精度に検出できる [92]．この手法は，改行位置は文節境界にのみ存在するという前提のもとでは，字幕テキスト提示の遅延時間は最も小さくなる．

4.3.1 文節ごとの改行挿入

講演の最初から数えて $i+1$ 番目の文節 b_{i+1} が入力されたときに，その直前の文節 b_i との境界に，改行を挿入するか否かを判定する．そのために，文節列 $b_1, \cdots, b_{i+1}$，及び，それまでの同定処理の結果 $r_1, \cdots, r_{i-1}$ に対し，文節 b_i と b_{i+1} の間に改行が挿入される確率（以下，改行挿入確率）：

$$P(r_i = 1 | r_1, \cdots, r_{i-1}, b_1, \cdots, b_{i+1})$$

を求める．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．

この改行挿入確率を最大エントロピー法により推定し，確率値が 0.5 より大きい場合に改行を挿入すると判定する．ただし，改行を挿入しないと 1 行の文字数が最長文字数を超える場合は，推定した確率値に関わらず，その文節境界に改行を挿入する．

4.3.2 改行挿入判定に用いる素性

文節 b_i と b_{i+1} の間の改行挿入確率の推定に利用する素性を表 4.1 に示す．本手法では，第 3 章で述べた文ごとの改行挿入手法（1 文が入力された後に，文全体に対する最尤の改行位置を求める改行挿入手法）で用いられた素性（最右列）のうち，文節が入力されるごとに獲得可能な素性をできる限り利用することを基本的な考え方として素性を設定した．

形態素情報や行長，ポーズ情報に関する素性は，音声認識ツールを利用することにより，音声入力に追従して獲得することができる [14, 55]．また，節境界情報に関

表 4.1: 最大エントロピー法で用いた素性

素性		本手法	文ごとの手法
形態素情報	b_i の主辞の品詞	○	○
	b_i の主辞の活用形	○	○
	b_i の語形の品詞	○	○
	b_{i+1} 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	○	○
節境界情報	b_i が節の最終文節であるか否か	○	○
	b_i が節の最終文節である場合，節のラベル	○	○
係り受け情報	b_i が直後の文節に係るか否か	○	○
	b_i が直前の文節から係られるか否か	○	○
	b_i が連体節の節末文節から係られるか否か	○	○
	b_i が節末文節に係るか否か	×	○
	b_i 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	×	○
	行頭から b_i までの間で係り受けが閉じているか否か	×	○
	b_i 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_i と同じ係り先をもつ文節があるか否か	×	○
行長	直前の改行点から b_i までの文字数が，3 つの区分（2 文字以下，3 文字以上 6 文字以下，7 文字以上）のうち，いずれであるか	○	○
ポーズ情報	b_i の直後のポーズ時間が，4 つの区分（0.2 秒未満，0.2 秒以上 1.0 秒未満，1.0 秒以上 3.0 秒未満，3.0 秒以上）のうち，いずれであるか	○	○

する素性（節境界の位置と種類）は，節境界解析ツール CBAP[91] を用いることにより，局所的な形態素列のみを手がかりとして特定することができる．

一方，係り受け関係は，係り文節とその受け文節との間の関係であり，通常，受け

文節が入力されるまで，係り受け情報を獲得することはできない．本研究では，文節が入力されるたびにその直前に改行が挿入されるか否かを判定するため，判定において係り受け関係を利用できるのは，入力文節が直前の文節の係り先となるような場合に限られる．

そこで本研究では，入力文節とその直前の文節とが係り受け関係にあるか否かを係り受け情報として用いる．すなわち，文節 b_{i+1} が入力されると，それが，その直前の文節 b_i の係り先であるか否かを判定し，その結果 $dep(b_i, b_{i+1})$ を係り受け情報として追加する．ここで，$dep(b_x, b_y)$ は，文節 b_x が文節 b_y に係るか否かのいずれかの値をとる．文節 b_i が入力された時点での改行挿入判定では，以下の係り受け情報を利用することができる．

$$dep(b_1, b_2), dep(b_2, b_3), \cdots, dep(b_{i-1}, b_i)$$

文節 b_i が文節 b_{i+1} に係るか否かの判定では，その係り受け確率を最大エントロピー法を用いて推定し，確率値が 0.5 以上であるとき，文節 b_i は文節 b_{i+1} に係るとする．係り受け確率の推定に用いる素性は，節境界に基づく係り受け解析手法 [30] における素性に，表 4.1 のポーズ情報を追加したものである．

なお，本手法では，係り受け情報の「b_i が連体節の節末文節から係られるか否か」の素性は，正確には,「b_i が直前の文節 b_{i-1} から係られ，かつ，その文節 b_{i-1} が連体節の節末文節であるか否か」の素性を用いた．これは，本研究では，ある文節とその直前の文節とが係り受け関係にあるか否かを係り受け情報として用いているためである．また，第 3 章の文ごとの改行挿入では，ポーズ情報として「b_i の直後にポーズがあるか否か」を用いたが，本手法では，ポーズ時間も改行挿入において重要な情報となると考え，このような 4 分類を用いている．

4.4 評価実験

本手法は，できる限り早く字幕テキストを提示するために，文の入力途中の段階で改行位置を決定する．このため，1 文全体を考慮して改行位置を決定する方法に比べ，改行の精度は低下することになる．この低下の程度が大きければ，たとえ，早期に字幕を提示できたとしても，その有用性は損なわれることになる．そこで，文ごとの改行挿入と比較したときの改行点の同定性能を実験的に評価した．

4.4.1　実験概要

実験データとして，同時通訳データベース [25] 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，文節境界情報，係り受け情報，節境界情報，改行情報が人手で付与されている．また，各文節の発話終了時間が連続音声認識エンジン Julius[55] により自動的に付与されている．

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，素性決定のための事前分析において参照したため，評価の対象とせず，残りの 14 講演（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 1,000 に設定し，それ以外はデフォルトのまま使用した．また，1 行の最長文字数を 20 文字とし，1 画面 10 行とした．

実験では，文単位の手法に対する本手法の精度の低下を評価するために，再現率と適合率，及び，その調和平均である F 値を用いた．再現率と適合率はそれぞれ，

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

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

を測定した．

また，本手法で用いた素性の妥当性を確認するために，4 つの比較手法を設定した．これらの手法はすべて，文節が入力されるごとに，その直前の文節境界に改行を挿入するか否かを判定する手法であり，字幕出力の遅延時間は本手法とほぼ同程度になると見込まれる．本手法との違いは，比較手法がそれぞれ単純な 1 つの特徴のみで改行挿入の判定を行っている点である．

- **比較手法 1**（文字数に基づく改行）：
  最長文字数を超えない最右の文節境界を改行点とする．

- **比較手法 2**（節境界に基づく改行）：
  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．

表 4.2: 再現率と適合率（ベースラインとの比較）

	再現率	適合率	F 値
提案手法	76.27% (5,489/7,197)	70.00% (5,489/7,841)	73.00%
文単位の手法	81.21% (5,845/7,197)	79.47% (5,845/7,355)	80.33%
比較手法 1	25.86% (1,861/7,197)	33.90% (1,861/5,490)	29.34%
比較手法 2	75.52% (5,435/7,197)	57.27% (5,435/9,490)	65.14%
比較手法 3	84.08% (6,051/7,197)	52.94% (6,051/11,429)	64.97%
比較手法 4	76.92% (5,536/7,197)	64.34% (5,536/8,604)	70.07%

- **比較手法 3**（係り受け関係に基づく改行）：
  係り受け関係に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に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．係り受け解析には，本手法における係り受け解析（4.3.2 節参照）と同様の方法を用いた．

- **比較手法 4**（ポーズに基づく改行）：
  0.2 秒以上の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．

4.4.2 実験結果

本手法ならびに文単位の改行挿入手法，各比較手法の適合率と再現率，F 値を表 4.2 に示す．文単位の手法では，文末に位置する文節境界の判定結果はすべて正解であるとして評価した．本手法は，文単位の手法の F 値の 90.88% (=73.00% / 80.33%) を達成しており，文節単位での改行位置の同定処理を精度よく実現できることを確認した．

本手法は，4 つの比較手法と比べ，最も高い F 値を達成している．比較手法 3 の再現率が高いが，これは，係り受け関係にないあらゆる文節間に改行を挿入するた

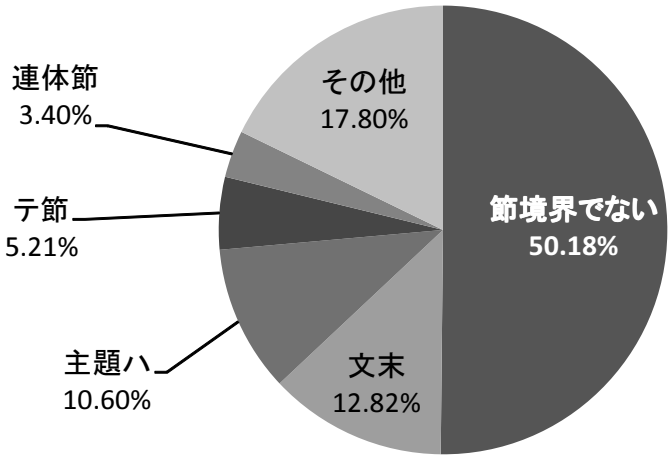


図 4.4: 本手法で改行を挿入できなかった文節境界の内訳

めであり，その分，適合率が低い．また，比較手法の中では，手法 4 の F 値が最も高かった．実験データでは，あらゆる文境界に改行が挿入されており，改行挿入手法としては文境界が検出されればそこを改行位置とすればよい．しかし一般に，言語情報のみを用いて文末を同定することは容易ではなく，比較手法 1〜3 が文末に正しく改行を挿入することは難しい．一方，実験データでは，文境界には 0.2 秒以上のポーズが必ず含まれていたため，比較手法 4 はすべての文末に改行を挿入することができ，それが再現性の高い改行挿入の実現につながった．

4.4.3　改行挿入誤りの分析

再現性に関する誤り

本手法の正解データに対する再現性について分析する．正解データの改行位置に対して，本手法では挿入できていないものが 1,708 箇所存在した．図 4.4 に，本手法で改行を挿入できなかった文節境界がどのような節境界になっているかで分類した内訳を示す．最も多かったのは，節境界ではない文節境界であり，857 箇所であった．節境界ではない文節境界に対する本手法の再現率は，57.99% (1,183/2,040) であり，文節境界全体に対する再現率より低い．節境界ではない文節境界の場合，その場所に改行を挿入するか否かは，前後の文節境界への改行挿入のされやすさに依存して決定されることが多い．本手法では，文節境界ごとに改行を挿入するか否か

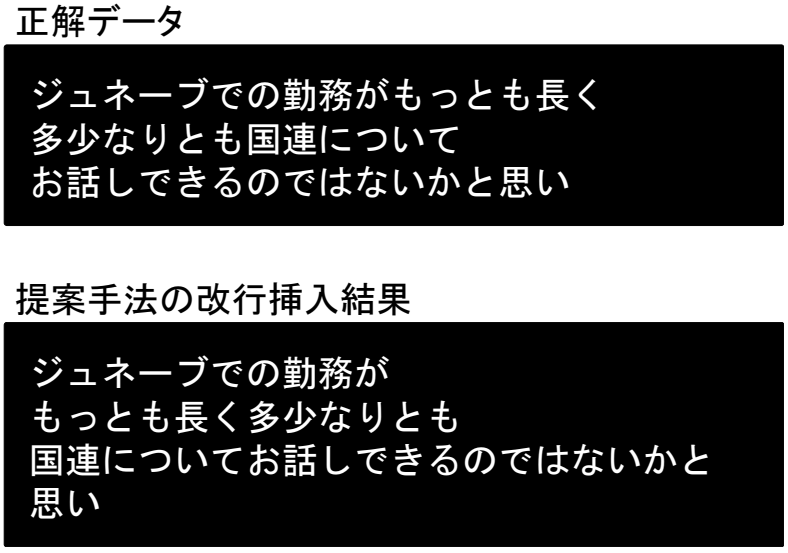


図 4.5: 節境界ではない文節境界への改行結果

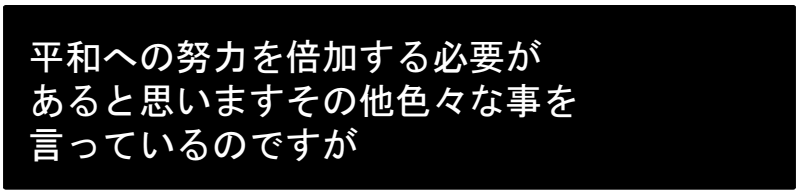


図 4.6: 文境界に改行が挿入されなかった例

を判定するため，後方の文節境界への改行挿入のされやすさを考慮することができず，正解データ通りに改行を挿入できない場合があった．このような例を図 4.5 に示す．正解データでは,「国連について」の直後に改行が挿入されているが，本手法では改行できていない．正しく改行を挿入するためには,「お話しできるのではないかと思い」が入力されるまで待ち，他の文節境界における改行のされやすさを考慮して，周辺テキストとのバランスから適切な改行位置を決定する必要がある．実際，文ごとの改行挿入手法では，正解データ通りの改行挿入を実現できている．

図 4.4 で，2 番目に多かったのは，文境界である文節境界であり，219 箇所であった．本手法では，入力音声の文境界が未知であることを前提としているため，文境界に対しても他の文節境界と同様に改行挿入判定を行う．文境界における改行挿入の再現率は 87.12%(1,481/1,700) であり，文節境界全体における再現率より高い性能を示している．しかし，図 4.6 に示すように，文境界 (「思います」) に改行が挿入できないことは字幕の読みやすさに大きな影響を与えるため，より確実に改行を挿

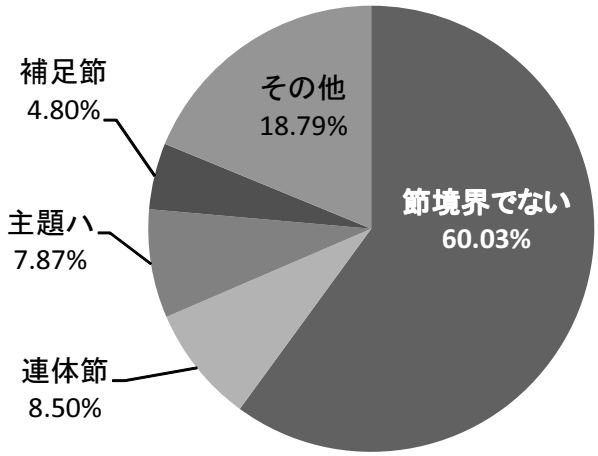


図 4.7: 本手法で余分に改行を挿入した文節境界の内訳

入する必要がある．

適合性に関する誤り

本手法の正解データに対する適合性について分析する．本手法によって 7,841 個の改行が挿入され，そのうち 2,352 個の改行が正解と一致しておらず，余分に挿入されていた．図 4.7 に，図 4.4 と同様に分類した，本手法で余分に改行を挿入した文節境界の内訳を示す．再現性の分析結果と同様に，最も多かったのは節境界ではない文節境界であり，1,412 箇所であった．節境界ではない文節境界に対する本手法の適合率は 45.59% (1,183/2,595) であり，文節境界全体に対する適合率より低い．例えば，図 4.5 の例において，正解データでは,「ジュネーブでの勤務がもっとも長く」は同一行となっているが，本手法では「勤務が」の後に改行が挿入されていた．再現性での考察と同様に，正解データ通りに改行を挿入するためには，周辺テキストとのバランスから適切な改行位置を決定することが必要であると考えられる．

2 番目に多かったのは，節境界「連体節」である文節境界であった（200 箇所）．節境界「連体節」に対する本手法の適合率は，31.51% (92/292) であり，文節境界全体に対する適合率と比べ極めて低い．連体節の最終文節は，基本形の述語であることが多く，文末文節と区別が付きにくいため，文末文節と同様に直後に改行が挿入される傾向にあったと思われる．一方で，正解データでは，意味的なまとまりを捉

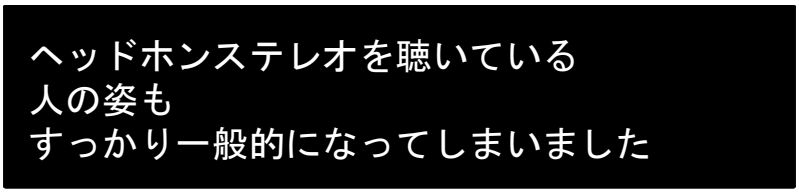


図 4.8: 文境界に改行が挿入されなかった例

えた改行が行われており，連体節が係る文節の後に改行が挿入されているため，正解データと一致しなかったと考えられる．本手法が余分に改行を挿入した例を図 4.8 に示す.「聴いている」と「人の」の間に余分に改行が挿入されており,「ヘッドホンステレオを聴いている人の姿も」というまとまりが捉えられていない.

4.5 節ごとの改行挿入との比較

4.3 節では，リアルタイムに字幕を提示することを目的に，文節が入力されるたびにその直前に改行が挿入されるか否かを判定する手法について述べた．一方で，4.4 節で示したように，1 文に対して改行点を同定する方法を採用すれば，より質の高い改行挿入を実現できる．以上を考慮すると，リアルタイム字幕生成において，同時性がある程度低下することを許容し，文節よりも長い改行処理の単位を採用することにより，改行位置の適格さを高めることが考えられる.

文節より長く文より短い言語単位として節を採用することは有力である．節は述部を中心としたまとまりである [87]．その境界は，構文的かつ意味的な切れ目となるため [91]，改行位置になりやすい．また，節で係り受けがまとまりやすく，節ごとの決定的な係り受け解析を高精度で実行できることが知られている [49]．加えて，節境界解析を用いれば節の切れ目の局所的な情報を使って精度よく節境界を検出でき [91]，同時的な処理との相性もよい.

本節では，節ごとの改行挿入手法について述べ，比較実験により，この手法の改行挿入性能について考察する.

4.5.1 節ごとの改行挿入手法

本研究では，節ごとの改行挿入手法を，以下の通りモデル化する．すなわち，1 講演分の文節列が 1 文節ずつ入力されることを想定し，節境界が検出されるごとに,

節内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．入力された節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もと，その節の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．

なお，節境界の検出では，節境界解析ツール CBAP[91] を用いて，文節が入力されるごとに，その直前の文節との文節境界に節境界があるか否かを随時判定する．

節ごとの改行挿入のための確率モデル

講演の最初から数えて l 番目の節が同定されたときの，改行挿入のための確率モデルについて説明する．本手法では，l 番目の節の文節列を $B_l = b_1 \cdots b_n$ とするとき，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もと，$P(R_l|\mathbf{R}_1^{l-1}, \mathbf{B}_1^l)$ を最大にする改行挿入結果 $R_l = r_1 \cdots r_n$ を求める．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．また，$\mathbf{R}_1^{l-1}$ で $R_1 \cdots R_{l-1}$ を，$\mathbf{B}_1^l$ で $B_1 \cdots B_l$ を表すこととする．

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，$P(R_l|\mathbf{R}_1^{l-1}, \mathbf{B}_1^l)$ は式 (4.1) のように計算できる．式 (4.1) は，B_l が字幕において，講演の最初から数えて x 行目の p $(1 \leq p < n_x)$ 番目の文節から y $(x \leq y)$ 行目の q $(1 \leq q < n_y)$ 番目の文節として表示されるという改行結果 R_l の場合の計算式を示す．なお，講演の最初から数えて j 行目の字幕として出力される文節列を $L_j = b_1^j \cdots b_{n_j}^j$ と記す．このとき，r_k^j は，$1 \leq k < n_j$ のとき $r_k^j = 0$，$k = n_j$ のとき $r_k^j = 1$ となる．

$$
\begin{aligned}
& P(R_l|\mathbf{R}_1^{l-1}, \mathbf{B}_1^l) \qquad (4.1)\\
= \quad & P(r_p^x = 0, \cdots, r_{n_x-1}^x = 0, r_{n_x}^x = 1, \cdots, \\
& \qquad r_1^y = 0, \cdots, r_q^y = 0|\mathbf{R}_1^{l-1}, \mathbf{B}_1^l) \\
\cong \quad & P(r_p^x{=}0|r_{p-1}^x{=}0,\cdots,r_1^x{=}0,r_{n_{(x-1)}}^{x-1}{=}1,\mathbf{B}_1^l)\times\cdots \\
& \times P(r_{n_x-1}^x{=}0|r_{n_x-2}^x{=}0,\cdots,r_1^x{=}0,r_{n_{(x-1)}}^{x-1}{=}1,\mathbf{B}_1^l) \\
& \times P(r_{n_x}^x{=}1|r_{n_x-1}^x{=}0,\cdots,r_1^x{=}0,r_{n_{(x-1)}}^{x-1}{=}1,\mathbf{B}_1^l)\times\cdots \\
& \times P(r_1^y{=}0|r_{n_{(y-1)}}^{y-1}{=}1,\mathbf{B}_1^l)\times\cdots \\
& \times P(r_q^y{=}0|r_{q-1}^y{=}0,\cdots,r_1^y{=}0,r_{n_{(y-1)}}^{y-1}{=}1,\mathbf{B}_1^l)
\end{aligned}
$$

表 4.3: 節ごとの手法で新たに用いた素性

b_i が節内文節である場合，b_i が節末文節に係るか否か
b_i が節内文節である場合，b_i 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
行頭から b_i までの間で係り受けが閉じているか否か

ここで，$P(r_k^j = 1 | r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{(j-1)}}^{j-1} = 1, \mathbf{B}_1^l)$ は，文節列 $B_1 \cdots B_l$ 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，$P(r_k^j = 0 | r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{(j-1)}}^{j-1} = 1, \mathbf{B}_1^l)$ は，文節 b_k^j の直後に改行が挿入されない確率を表す．これらの確率は最大エントロピー法により推定する．最尤の改行結果は，式 (4.1) 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．

改行挿入確率の推定に用いる素性

節ごとの手法は，文節ごとの手法と比べて，改行挿入処理の開始を節境界が検出されるまで遅らせる分だけ，係り受けに関する情報をより多く利用することができる．そこで，節ごとの手法では，文節ごとの手法で利用した素性（表 4.1 の右から 2 列目）の他に，新たに係り受けに関する素性を追加し，それらを合わせて利用する．節ごとの手法で新たに追加した素性を表 4.3 に示す．

節ごとの改行挿入手法では，節境界に基づく係り受け解析 [30] により，係り受け情報を獲得する．この手法では，節境界が検出されるごとに，節内部の係り受け構造と節末文節の係り先を同定することができる．ただし，節末文節の場合には，節内文節の場合と異なり，係り先がまだ入力されていない可能性が高いため，本手法では，同時性を損なわない範囲で可能な限りの係り受け情報を獲得することを試みることとし，直後の文節に係るか否かの判定のみを行うこととする．なお，この判定は，文節ごとの改行挿入手法における係り受け解析と同様の方法で行う．

4.5.2 比較実験

提案手法（文節ごとの手法）と節ごとの手法を比較評価するため，両手法で改行挿入実験を実施し，リアルタイムという観点も含めた主観的評価を実施した．

実験概要

両手法による改行挿入実験は，4.4.1 節と同じ実験環境を用いて実施した．

評価は，改行処理の同時性と品質の両面から総合的に行う必要があり，本研究では，健聴者 10 名（大学院生）を被験者とする主観的評価により行った．各被験者には，ランダムに抽出した 1 分間の講演に対して両手法がそれぞれ生成した，2 種類の字幕映像 10 セットを講演者の音声とともに提示し，読みやすい方を選択するように指示した．なお，事前に，別の 1 セットを提示し，簡単な予行練習を行った．また，字幕の表示には，プロジェクタ用スクリーン（高さ 1525mm ×幅 2033mm）を使用し，評価者は，スクリーンから 5m の範囲で着席し，字幕を視聴した．

各セットの 2 種類の字幕映像は，改行点と各行の出力タイミングのみが異なっている．被験者に対してどちらの字幕を先に提示するかは各セットでランダムに変更した．

なお，各文節の出力時間は，その文節を含む行の行末の改行位置が確定した時間とした．ただし，本研究では書き起こしデータを用いて実験を行っているため，出力時間に書き起こし時間は含まれていない．また，テキストが行単位で入れ替わりスクロールすることを想定しているため，節ごとの手法において，入力された節に対する改行挿入判定結果が複数行になる場合は，複数行が一度に表示されることを避けるために，2 行目以降を順次，1 秒ずつ遅らせて表示した．

実験結果

主観的評価の結果を図 4.9 に示す．グラフは，各セットにおいて，両手法による各字幕映像が何人の被験者により選択されたかを表す．どのセットにおいても半数以上の被験者が文節ごとの手法による字幕映像のほうが読みやすいと判断している．また，文節ごとの手法を読みやすいと判断した被験者は 1 セットあたり平均 6.3 人であった．対応のある場合の平均値の差の検定（t 検定）を行った結果，文節ごとの手法を読みやすいと選択した被験者数は，節ごとの手法と比べて，有意に上回っており（有意水準 5%），文節ごとの手法がより読みやすい字幕を生成していることがわかった．

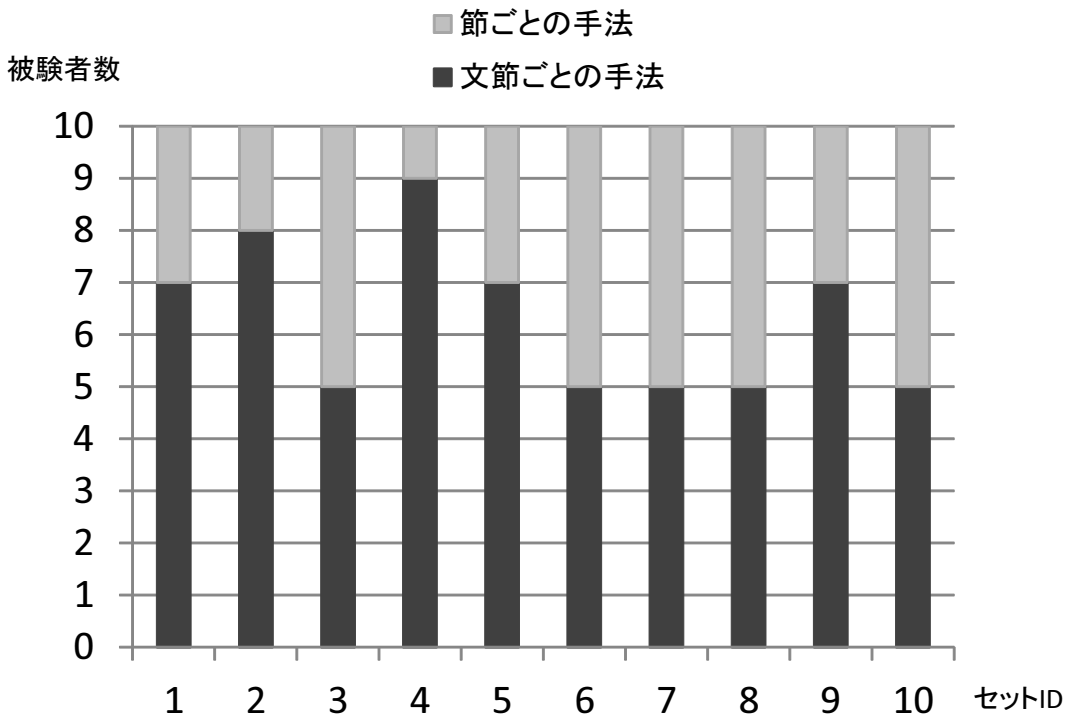


図 4.9: 主観的評価の結果

4.5.3 考察

精度と遅延時間

読みやすい字幕を生成するために，テキストが適切な位置で改行され，かつ，短い遅延時間で出力されることが求められる．しかし，一般に，改行位置同定の精度とタイミングはトレードオフの関係にあり，この意味で，提案手法における節ごとの手法と文節ごとの手法では，それぞれに一長一短があることが予想される．そこで本節では，両手法におけるこれら 2 つの観点の実現の程度を比較分析することにより，実験結果を考察する．

まず，各手法の改行挿入位置の適切さを評価するため，4.4.1 節と同様の再現率と適合率を測定した．両手法の適合率と再現率を表 4.4 に示す．再現率，適合率，これらの調和平均である F 値においていずれも，節ごとの改行挿入手法が，文節ごとの改行挿入手法よりも高いという結果になった．これは，改行処理の単位の長さが長くなるほど，最大エントロピー法の素性として利用可能な情報が多くなるためで

表 4.4: 再現率と適合率

	再現率	適合率	F 値
文節単位の手法	76.27% (5,489/7,197)	70.00% (5,489/7,841)	73.00%
節単位の手法	79.35% (5,711/7,197)	74.90% (5,711/7,625)	77.06%

あると考えられる．

次に，文節ごとに，入力タイミングと出力タイミングの差を遅延時間として測定し，各手法の同時性を評価した．ここで，各文節の入力タイミングは各文節の発話終了時間とした．各文節の遅延時間の累積割合を各手法ごとに図 4.10 に示す．横軸は遅延時間を，縦軸はその遅延時間未満で出力される文節の全文節数に対する割合を示している．文節ごとの手法では，全体の 81.19%の文節が約 3 秒未満の遅延時間で出力されたのに対し，節ごとの手法では，56.49%にとどまった．平均遅延時間（= 遅延時間の総和/総文節数）は，文節ごとの手法が 2.41 秒，節ごとの手法が 3.67 秒であった．改行処理の単位の長さが短くなるほど，遅延時間が減少している．

なお，字幕の平均表示速度（1 分間あたりの平均表示文字数）は，文節ごとの手法が 223.9wpm，節ごとの手法が 221.9wpm であった．両手法間で字幕の出力タイミングは大きく異なるものの，表示する文字数には違いはなく，表示速度の差はほとんど生じない．

以上から，文節ごとの手法は，節ごとの手法と比べて，正解改行位置に対する F 値の若干の低下（5.27% $(= (77.06 - 73.00)/77.06)$）が観られたものの，字幕出力の遅延時間は大幅に短縮（34.33% $(= (3.67 - 2.41)/3.67)$）されていることがわかる．その結果として，主観的評価において，文節ごとの手法が，より読みやすい字幕として判断されたものと考えられる．

精度が主観的評価に与える影響

前節では，字幕出力の遅延時間が短いことが，主観的評価において高い評価につながることを考察した．本節では，改行位置の適切さが主観的評価に与える影響を分析する．

表 4.5 に，各セットの 2 種類の字幕の正解改行位置に対する F 値を示す．なお，4 列目は両手法の F 値の差を示す．例えば，主観的評価（図 4.9）において，節ごとの

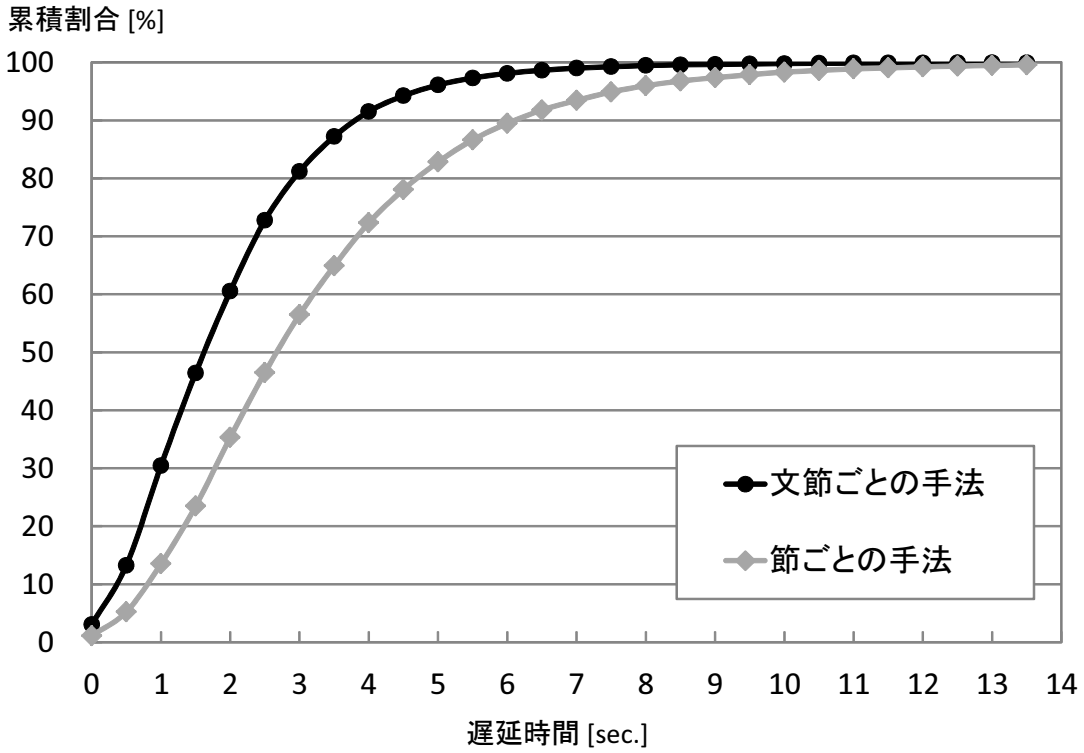


図 4.10: 遅延時間と累積割合

手法を選択した被験者数が最も少ない（1 人）セット (ID=4) では，正解改行位置に対する F 値の差は-8.21%，一方，最も多い（5 人）セット (ID=10) では，正解改行位置に対する F 値の差は 18.75%であるなど，節ごとの手法の F 値が文節ごとの手法と比べて高くなるほど，節ごとの手法による字幕を選択する被験者数が増加する傾向があった．各セットにおける，両手法の F 値の差と各手法を選択した被験者数の間の相関係数は 0.62 であり，中程度の相関がみられ，改行位置の適切さも字幕の読みやすさに影響しないわけではないことを確認した．短い遅延時間を維持しつつ，改行位置の適切さを向上させることが字幕の読みやすさの改善につながると考えられる．

文ごとの改行挿入手法との主観的評価の比較

本章では，文ごとの改行挿入は，講演のリアルタイム字幕生成には適さないという前提のもと，文より短い単位での改行挿入手法を提案した．文ごとの改行挿入では，一文の入力が終わるまで改行挿入処理を実行できず，また，講演では一文が長

表 4.5: 各セットの F 値（文節ごとの手法と節ごとの手法）

セット ID	節ごとの手法	文節ごとの手法	差
1	81.25%	78.79%	2.46%
2	61.90%	62.22%	-0.32%
3	68.57%	70.59%	-2.02%
4	64.52%	72.73%	-8.21%
5	72.73%	65.12%	7.61%
6	90.32%	87.50%	2.82%
7	73.33%	57.14%	16.19%
8	82.76%	75.86%	6.90%
9	71.79%	81.08%	-9.29%
10	81.25%	62.50%	18.75%

くなる傾向にあるため，文の発声が始まってからその字幕表示が開始されるまでの時間が著しく長くなると考えられるためである．

しかし，文ごとの改行挿入の精度が最も高く，また，4.5.3 節で考察したように，改行位置の適切さも字幕の読みやすさに影響を及ぼすことがあり，このことは本章で設けた前提の妥当性について検証の必要性を示唆している．

そこで本研究では，文ごとの改行挿入手法と文より短い単位ごとの改行挿入手法（文節ごとの改行挿入手法，及び，節ごとの改行挿入手法）を比較評価するため，被験者実験を実施した．実験の要領は，基本的に 4.5.2 節と同様である．すなわち，以下の組合せで 2 種類の字幕映像を提示し，被験者は読みやすいと思うどちらか一方を選択する．

(a) 「文ごとの手法」と「文節ごとの手法」（5 セット）

(b) 「文ごとの手法」と「節ごとの手法」（5 セット）

ただし，これら 10 セットは，実験データからランダム抽出し新たに作成したものであり，互いに異なる 1 分間となっている．加えて，以下の点が 4.5.2 節とは異なる．

- 被験者数：健聴者 21 名（大学院生）
- スクリーンの大きさ：高さ 1829mm ×幅 2439mm

図 4.11, 4.12 に被験者実験の結果を示す．図 4.11 が「文節ごとの手法」との比較を，図 4.12 が「節ごとの手法」との比較を示している．グラフの見方は図 4.9 と同

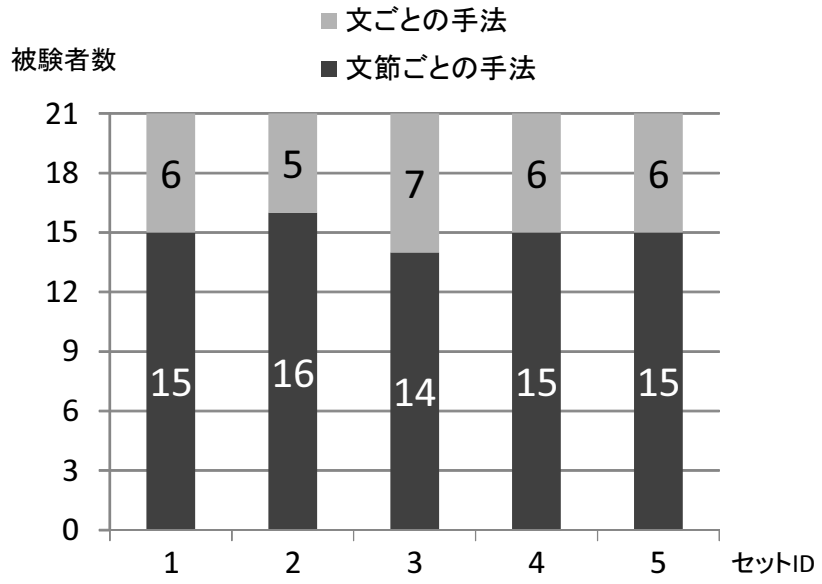


図 4.11: 主観的評価の結果 (a)

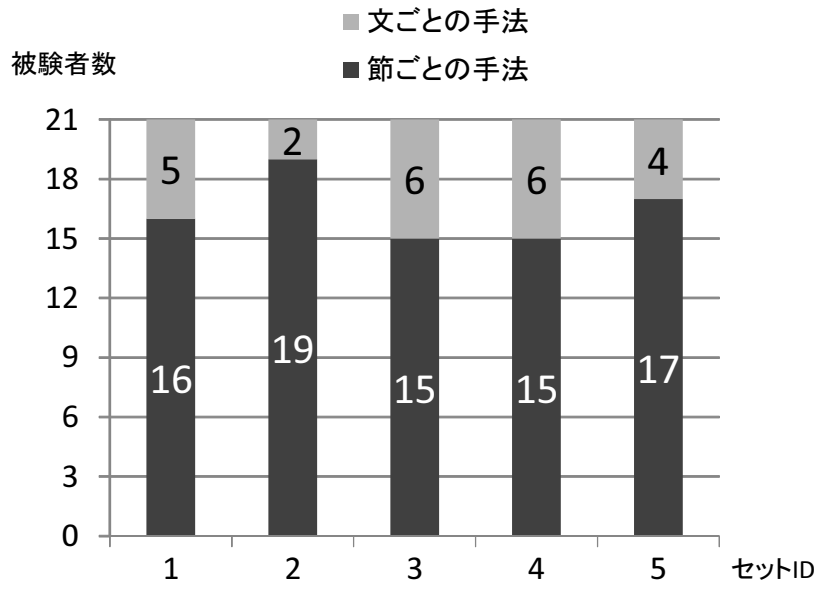


図 4.12: 主観的評価の結果 (b)

様である．いずれのセットも 2/3 以上の被験者が文より短い単位ごとの手法の方を読みやすいと判断しており，本研究で設けた前提を支持する結果となった．

なお，平均して 25.24% (= 53/210) の被験者が文ごとの手法による字幕を読みや

すいと判断した．その原因の 1 つとして，文ごとの手法では文境界が正しく検出されているとし，文境界には改行を必ず挿入できるのに対し，文より短い単位での手法は，文境界の情報を使用しておらず，必ずしも文境界に改行を挿入できるとは限らないことが挙げられる．字幕には句点がなく，文境界に改行が挿入されないという誤りが，読みやすさを著しく損なう場合があり，それが被験者の評価に影響したと考えられる．このようにやや公平性に欠ける実験環境下でも，文より短い単位での改行挿入手法は，文ごとの手法と比べて高い評価を得ており，講演のリアルタイム字幕生成により適した方法であることを確認した．

4.6　おわりに

本章では，提示タイミングを考慮した整形手法として，リアルタイム字幕生成のための逐次的な改行挿入手法を提案した．本手法では，音声の入力に対して早い段階で改行位置を決定し，行にチャンキングした字幕を提示する．まず，改行処理の同時性を優先した方式として，文節ごとの改行処理手法を提案した．実験の結果，F 値にして，文単位の手法の 90.88%の改行挿入精度を達成した．次に，改行処理の同時性と精度の双方を重視する方式として，節ごとの改行挿入手法を示し，文節ごとの手法とで比較評価した．被験者による主観的評価により，文節ごとの手法による字幕の方が評価が高いことを確認した．

第5章　あとがき

5.1　本論文のまとめ

本論文では，読みやすいテキスト提示を行うための，意味的なまとまりへのチャンキングに基づく日本語文の自動整形手法を提案した．

第1章ではまず，文の自動整形問題を,「文字列の整形」,「配置・書体の整形」,「提示タイミングを考慮した整形」の3つの問題に分類し，それぞれの分類ごとに研究動向を概観した．文字列の整形のために，文法的誤りの修正による整形，語彙・文構造に着目した整形，要約の作成が行われており，これまでの研究では，主に，語や文構造の修正が対象とされてきたことを述べた．配置・書体の整形では，重要箇所や文構造を捉えて文字列の配置・書体を決定する手法が開発されているが，字幕テキストを対象とした研究があまり行われていないことを示した．提示タイミングを考慮した整形では，音声と字幕を同時的に提示するための研究が行われているものの，テキストを提示すべきタイミングまでに文の整形を行うことを目的とした研究がほとんど行われていないことを述べた．このような現状を踏まえ，本研究では，読みやすい日本語テキストを提示することを目的に，文字列の整形手法として日本語文への読点挿入手法を，配置の整形手法として日本語講演文への改行挿入手法を，提示タイミングを考慮した整形手法として日本語講演文への逐次的な改行挿入手法を実現した．

第2章では，文字列を整形する手法として，日本語文に読点を挿入する手法を提案した．日本語の読点にはいくつかの用法が存在し，その用法ごとに文中での挿入位置が異なる．本研究ではまず，読点に関する文献を調査し，読点の用法を9種類に分類した．次に，分類した用法ごとの読点の出現傾向を新聞記事テキストを用いて分析し，読点挿入に使用する素性を決定した．本手法では，特定の形態素の出現や節境界の種類，文字の種類，読点間の文字列の長さなどを素性とする統計的手法によって読点の挿入位置を決定する．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，一文中の適切な読点挿入位置を同定する．日

本語テキストコーパスを用いた読点挿入実験を行い，本手法で用いた全ての素性が有効に働いていることを確認した．また，人間の読点挿入性能と比較して，本手法が遜色のない読点挿入性能を示していることから，本手法の有効性を確認した．

第3章では，配置の整形手法として，日本語講演文への改行挿入手法を提案した．本研究ではまず，改行位置が考慮されていない字幕テキストと適切に改行が挿入された字幕テキストを被験者評価によって比較し，テキストを読みやすくするための改行挿入の効果を確認した．次に，講演音声の書き起こしテキストに対して，人手で適切な位置に改行を挿入した改行データを作成し，改行挿入位置の分析を行った．本手法では，分析結果に基づき，節境界や係り受け，ポーズ情報などを素性として用いた統計的手法によって，一文中の各文節境界に対して改行を挿入するか否かを同定する．一文中に挿入され得る改行位置の全ての組み合わせから，最適な組み合わせを確率モデルを用いて決定する．日本語講演の書き起こしテキストを用いた改行挿入実験を行い，精度の高い改行挿入が実現できていることを確認した．また，改行挿入テキストの主観的評価によって，実際に読みやすい字幕テキストが生成できていることを示し，本手法の有効性を確認した．

第4章では，提示タイミングを考慮した整形手法として，リアルタイム字幕生成のための逐次的な改行挿入手法を提案した．本研究ではまず，字幕提示のリアルタイム性を最も重視した方法として，文節単位で改行位置を同定する手法を実現した．本手法では，文節が入力されるごとに，係り受けやポーズ，行長などの素性のうち，同定処理の時点で利用可能な情報を用いた統計的手法によって改行位置を同定する．日本語講演の書き起こしテキストを用いた改行挿入実験を実施し，文単位での改行挿入手法との性能比較を行った．次に，字幕提示のリアルタイム性と改行位置の適格さをの両方を考慮する方法として，節単位での改行挿入手法を実現した．文節単位での手法と比較し，改行位置の同定処理のタイミングと精度が字幕の読みやすさに及ぼす影響について考察した．また，被験者評価を行い，文よりも短い単位での改行挿入手法が，リアルタイム字幕生成に適した方法であることを確認した．

5.2　今後の課題と将来への展望

本論文が残した課題と将来への展望を述べる．

読点挿入に関しては以下の研究課題がある．

- 主題を示す読点に関する素性の検討

実験結果の分析から,「主題を示す読点」の挿入性能が低いことがわかった.「主題を示す読点」の出現数は多く，誤りの影響が大きい．今後は，この用法の読点に関する，より有効な素性を発見・利用し，本手法の再現率を向上させることが課題となる．

- 直前の語句を強調するための読点についての検討
  本研究では読点の用法を 9 種類に分類したが，そのうちの「直前の語句を強調するための読点」については，その挿入位置が執筆者の意向に依存するものであるため，対象としなかった．この「直前の語句を強調するための読点」に関して，今後検討を行う必要がある．

- 整形前の文に含まれる読点の利用
  本手法では読点を除いた文に対して読点挿入を行っている．しかし，整形前の文に含まれる読点の中には正しく挿入されている読点も存在している．それらの読点位置の情報を利用し，正しい読点以外を修正することによって文字列を整形する方法を検討する．さらに，どの読点を修正したかという情報を提示できるようになれば，本手法を書き手による文の推敲の支援に用いることも可能となると考えられる．

改行挿入に関しては以下の研究課題がある．

- 読点挿入との併用
  第 2 章で，読点挿入によって読みやすいテキストが生成できることを述べた．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る．適切な改行位置と読点位置は互いに関係するものの，改行位置は行長に影響されるなど，それらの位置が必ず一致するというわけではない．それらを組み合せて挿入する方法について，今後検討していきたい．

逐次的な改行挿入に関しては以下の研究課題がある．

- 音声認識を考慮した改行挿入
  本研究では音声は正しく書き起されていると仮定して，逐次的な改行挿入手法の開発を行った．しかし，実際のリアルタイム字幕生成に応用するためには，音声認識システム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．

さらに，今後の課題に挙げた技術が実現できれば，本論文で提案した文整形技術と合わせて使用することで，リアルタイム字幕生成システムの開発を行うことが考えられる．実際に運用可能な字幕生成システムを開発するためには，音声認識システムを利用して音声を書き起こし，書き起こされたテキストに対して逐次的に読点挿入，改行挿入などの処理を行うことによって整形する必要がある．文を漸進的に係り受け解析する技術が開発されており [30]，この技術を組み込んだ改行挿入，読点挿入を行うことにより，より読みやすい字幕テキストを提示することが可能となると考えられる．また，限られたスペースに字幕を提示するための文圧縮技術が開発されており [47, 75, 97]，この技術を利用することによっても，高品質な字幕テキストを提示できると考えられる．音声認識，漸進的な構文解析，および，文整形を統一的に扱うための枠組の開発が必要となる．

本研究では，意味的なまとまりに基づき，読点や改行によって文を分割する整形手法を提案した．読みやすいテキストを提示するために，文内の意味的なまとまりに加え，文単位の意味的なまとまりを捉えることもまた重要であると考えられる．特に，メールにおいて読みやすいテキストを書くための技術の 1 つとして，文内の適切な位置に改行を，文間の適切な位置に空行を挿入することが挙げられる [45, 85]．文単位の意味的なまとまりに基づき，空行を自動挿入することによって文書を分割することは，読みやすいメールテキストを提示するうえで有効である．現在，メールテキストを読みやすく変換するための改行・空行挿入手法を開発中である [93]．

将来の展望としては，講演を対象とした同時通訳システムの開発が挙げられる．同時通訳システムの開発が実現できれば，英語の講演において，英語を母国語としない人が講演内容を理解するための支援を行うことが可能となる．これまでに開発されている音声翻訳システムの多くは対話文を対象とした翻訳方式を採用している [21, 28]．講演では 1 文が長くなる傾向にあるため，講演音声の翻訳を行うためには，文よりも小さな単位で翻訳を行い，訳文を講演音声と同時的に提示する必要がある．本研究で提案した読点挿入手法，改行挿入手法は文をより短い意味的なまとまりに分割する手法であるため，読点や改行によって挟まれた単位は翻訳単位と相性が良いと考えられる．本手法の適用によって分割される単位に基づいて翻訳単位を定めることで，品質の高い翻訳を行えると期待できる．また，翻訳結果をテキストとして提示することを考えた場合，音声と同時的に，読みやすく提示することが必要となる．本論文で開発したテキスト整形技術を用いることで，翻訳結果を読みやすく

提示することが可能であると考えられる．本手法を拡張することによって同時通訳のための翻訳単位の検出手法を開発し，音声翻訳技術や話し言葉のための構文解析技術，本論文で提案したテキスト整形技術を併用することによって，同時翻訳システムの開発を実現することが可能であると考えられる．

謝辞

本論文をまとめるにあたり，多大な御教示と御尽力をいただきました，名古屋大学教授の石川佳治先生に厚く謝意を申し上げます．また，本論文の詳細について，貴重な御示唆と御指導をいただきました，名古屋大学教授の渡邉豊英先生，間瀬健二先生，並びに，准教授の加藤芳秀先生に深く感謝いたします．

日頃から，研究や論文執筆等において，様々な懇切丁寧な御指導と御鞭撻を賜りました，名古屋大学准教授の松原茂樹先生に厚く感謝いたします．松原茂樹先生には，公私共に様々な御相談に乗って頂き，また，多くの御足労をお掛けました．心より感謝申し上げます．

本研究の初期の段階より幅広い角度から御指導と御議論をいただき，また研究以外の面においても様々な御支援をいただきました名古屋大学助教の大野誠寛先生に感謝いたします．

本研究を進めるにあたり，また，研究室生活を送るにあたり，御指導と御支援をいただきました，名古屋大学准教授の外山勝彦先生，小川泰弘先生，並びに，助教の笠浩一朗先生に感謝いたします．

研究に関する討論をはじめ，様々な助言をいただき，また，普段の研究生活，研究以外の面においても色々とお世話になりました，松原研究室，石川研究室，外山研究室の皆様に心から感謝いたします．

研究活動を行うにあたり出張や事務手続き等，お世話になりました，名古屋大学松原研究室秘書の土井ひとみさんに深く感謝いたします．

最後に，改めまして，本論文をまとめるにあたり御支援をいただいたすべての皆様に心より御礼申し上げます．

発表文献リスト

種別	論文名	関連する章
論文誌	村田 匡輝，大野 誠寛，松原 茂樹．読点の用法的分類に基づく日本語テキストへの自動読点挿入，電子情報通信学会論文誌，Vol. J95-D, No. 9, pp. 1783–1793, 2012.	2 章
国際会議	Masaki Murata, Tomohiro Ohno and Shigeki Matsubara. Automatic Comma Insertion for Japanese Text Generation, In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP-2010)*, pp. 892–901, 2010.	2 章
論文誌	村田 匡輝，大野 誠寛，松原 茂樹．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，電子情報通信学会論文誌，Vol. J92-D, No. 9, pp. 1621–1631, 2009.	3 章
論文誌	Masaki Murata, Tomohiro Ohno, Shigeki Matsubara. Construction of Linefeed Insertion Rules for Lecture Transcript and Their Evaluation, *International Journal of Knowledge and Web Intelligence*, Vol. 1, No. 3/4, pp. 227-242, 2010.	3 章
国際会議	Masaki Murata, Tomohiro Ohno, Shigeki Matsubara. Automatic Linefeed Insertion for Improving Readability of Lecture Transcript, In *Proceedings of the 2nd KES International Symposium on Intelligent Interactive Multimedia Systems and Services (KES-IIMSS-2009)*, pp. 499-509, 2009.	3 章

種別	論文名	関連する章
国際会議	Tomohiro Ohno, Masaki Murata and Shigeki Matsubara. Linefeed Insertion into Japanese Spoken Monologue for Captioning, In *Proceedings of Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing (ACL-IJCNLP-2009)*, pp. 531–539, 2009.	3 章
論文誌	大野 誠寛，村田 匡輝，松原 茂樹．講演のリアルタイム字幕生成のための逐次的な改行挿入，電気学会論文誌，Vol. 133-C, No. 2, 2013.	4 章

参考文献

[1] Chinatsu Aone, Mary Ellen Okurowski, James Gorlinsky, and Bjornar Larsen. A scalable summarization system using robust NLP. In *Proceedings of ACL-97 Workshop on Intelligent Scalable Text Summarization*, pp. 66–73, 1997.

[2] Doug Beeferman, Adam Berger, and John Lafferty. Cyberpunc: A lightweight punctuation annotation system for speech. In *Proceedings of 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 689–692, 1998.

[3] Gilles Boulianne, Jean-Francois Beaumont, Maryse Boisvert, Julie Brousseau, Patrick Cardinal, Claude Chapdelaine, Michel Comeau, Pierre Ouellet, and Frederic Osterrath. Computer-assisted closed-captioning of live TV broadcasts in French. In *Proceedings of 9th International Conference on Spoken Language Processing*, pp. 273–276, 2006.

[4] Heidi Christensen, Yoshihiko Gotoh, and Steve Renals. Punctuation annotation using statistical prosody models. In *Proceedings of ISCA Workshop on Prosody in Speech Recognition and Understanding*, pp. 35–40, 2001.

[5] Seiji Egawa, Yoshihide Kato, and Shigeki Matsubara. Sentence compression by removing recursive structure from parse tree. In *Proceedings of 10th Pacific Rim International Conference on Artificial Intelligence: Trends in Artificial Intelligence*, pp. 115–127, 2008.

[6] Agustin Gravano, Martin Jansche, and Michiel Bacchiani. Restoring punctuation and capitalization in transcribed speech. In *Proceedings of 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 4741–4744, 2009.

[7] Yuqing Guo, Haifeng Wang, and Josef van Genabith. A linguistically inspired statistical model for Chinese punctuation generation. *ACM Transactions on Asian Language Information Processing*, Vol. 9, No. 2, pp. 6:1–6:27, 2010.

[8] Ryuichiro Higashinaka and Katashi Nagao. Interactive paraphrasing based on linguistic annotation. In *Proceedings of 19th International Conference on Computational Linguistics*, pp. 1218–1222, 2002.

[9] Trym Holter, Erik Harborg, Magne Hallstein Johnsen, and Torbjorn Svendsen. ASR-based subtitling of live TV-programs for the hearing impaired. In *Proceedings of 6th International Conference on Spoken Language Processing*, pp. 570–573, 2000.

[10] Matthias Honal and Tanja Schultz. Correction of disfluencies in spontaneous speech using a noisy-channel approach. In *Proceedings of 8th European Conference on Speech Communication and Technology*, pp. 2781–2784, 2003.

[11] Chiori Hori, Sadaoki Furui, Rob Malkin, Hua Yu, and Alex Waibel. Automatic speech summarization applied to english broadcast news speech. In *Proceedings of 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 9–12, 2002.

[12] Takaoki Hori, Daniel Willett, and Yasuhiro Minami. Language model adaptation using wfst-based speaking-style translation. In *Proceedings of 2003 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 228–231, 2003.

[13] Jing Huang and Geoffrey Zweig. Maximum entropy model for punctuation annotation from speech. In *Proceedings of 7th International Conference on Spoken Language Processing*, pp. 917–920, 2002.

[14] Toru Imai, Shoei Sato, Shinichi Homma, Kazuo Onoe, and Akio Kobayashi. Online speech detection and dual-gender speech recognition for captioning broadcast news. *IEICE Transactions on Information and Systems*, Vol. 90, No. 8, pp. 1286–1291, 2007.

[15] Kentaro Inui, Atsushi Fujita, Tetsuro Takahashi, Ryu Iida, and Tomoya Iwakura. Text simplification for reading assistance: a project note. In *Proceedings of 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications*, pp. 9–16, 2003.

[16] Dong Hyun Jang and Sung-Hyon Myaeng. Development of a document summarization system for effective information services. In *Proceedings of RIAO-97 Computer-Assisted Information Retrieval*, pp. 101–112, 1997.

[17] Tatsuya Kawahara, Kazuya Shitaoka, and Hiroaki Nanjo. Automatic transformation of lecture transcription into document style using statistical framework. In *Proceedings of 8th International Conference on Spoken Language Processing*, pp. 2169–2172, 2004.

[18] Ji-hwan Kim and Philip C. Woodland. The use of prosody in a combined system for punctuation generation and speech recognition. In *Proceedings of 7th European Conference on Speech Communication and Technology*, pp. 2757–2760, 2001.

[19] Julian Kupiec, Jan Pedersen, and Francine Chen. A trainable document summarizer. In *Proceedings of 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 68–73, 1995.

[20] Zhang Le. Maximum entropy modeling toolkit for python and c++. `http://homepages.inf.ed.ac.uk/s0450736/maxent_toolkit.html`, 2008. [Online; accessed 1-March-2008].

[21] Fu-Hua Liu, Yuqing Gao, Liang Gu, and Michael Picheny. Noise robustness in speech to speech translation. In *Proceedings of 8th European Conference on Speech Communication and Technology*, pp. 2797–2800, 2003.

[22] Yang Liu, E. Shriberg, A. Stolcke, D. Hillard, M. Ostendorf, and M. Harper. Enriching speech recognition with automatic detection of sentence boundaries and disfluencies. *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 14, No. 5, pp. 1526–1540, 2006.

[23] Wei Lu and Hwee Tou Ng. Better punctuation prediction with dynamic conditional random fields. In *Proceedings of 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 177–186, 2010.

[24] Sameer Maskey, Bowen Zhou, and Yuqing Gao. A phrase-level machine translation approach for disfluency detection using weighted finite state transducers. In *Proceedings of 9th International Conference on Spoken Language Processing*, pp. 749–752, 2006.

[25] Shigeki Matsubara, Akira Takagi, Nobuo Kawaguchi, and Yasuyoshi Inagaki. Bilingual spoken monologue corpus for simultaneous machine interpretation research. In *Proceedings of 3rd Language Resources and Evaluation Conference*, pp. 153–159, 2002.

[26] Antonio Molina and Ferran Pla. Clause detection using HMM. In *Proceedings of 2001 Workshop on Computational Natural Language Learning*, pp. 70–72, 2001.

[27] Cosmin Munteanu, Gerald Penn, and Ron Baecker. Web-based language modelling for automatic lecture transcription. In *Proceedings of 8th Annual Conference of International Speech Communication Association*, pp. 2353–2356, 2007.

[28] Satoshi Nakamura, Konstantin Markov, Hiromi Nakaiwa, Genichiro Kikui, Hisashi Kawai, Takatoshi Jitsuhiro, Jin-Song Zhang, Hirofumi Yamamoto, Eiichiro Sumita, and Seiichi Yamamoto. The ATR multilingual speech-to-speech translation system. *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 14, No. 2, pp. 365–376, 2006.

[29] Graham Neubig, Shinsuke Mori, and Tatsuya Kawahara. A WFST-based log-linear framework for speaking-style transformation. In *Proceedings of 10th Annual Conference of International Speech Communication Association*, pp. 1495–1498, 2009.

[30] Tomohiro Ohno, Shigeki Matsubara, Hideki Kashioka, Takehiko Maruyama, Hideki Tanaka, and Yasuyoshi Inagaki. Dependency parsing of Japanese mono-

logue using clause boundaries. *Language Resources and Evaluation*, Vol. 40, No. 3-4, pp. 263–279, 2007.

[31] Murat Saraclar, Michael Riley, Enrico Bocchieri, and Vincent Goffin. Towards automatic closed captioning: Low latency real time broadcast news transcription. In *Proceedings of 7th International Conference on Spoken Language Processing*, pp. 1741–1744, 2002.

[32] Masao Utiyama and Kôiti Hasida. Automatic slide presentation from semantically annotated documents. In *Proceedings of 1999 ACL Workshop on Coreference and its Applications*, pp. 25–30, 1999.

[33] Hideo Watanabe. A method for abstracting newspaper articles by using surface clues. In *Proceedings of 16th International Conference on Computational Linguistics*, pp. 974–979, 1996.

[34] Jian Xue, Rusheng Hu, and Yunxin Zhao. New improvements in decoding speed and latency for automatic captioning. In *Proceedings of 9th International Conference on Spoken Language Processing*, pp. 1630–1633, 2006.

[35] Jian Zhang, Ho Yin Chan, Pascale Fung, and Lu Cao. A comparative study on speech summarization of broadcast news and lecture speech. In *Proceedings of 8th Annual Conference of International Speech Communication Association*, pp. 2781–2784, 2007.

[36] 秋田祐哉, 河原達也. 講演の書き起こしに対する読点の自動挿入. 日本音響学会秋季研究発表会講演論文集, pp. 79–80, 2010.

[37] 秋田祐哉, 河原達也. 講演に対する読点の複数アノテーションに基づく自動挿入. 情報処理学会研究報告, Vol. 2011, No. 4, pp. 1–6, 2011.

[38] 浅原正幸, 松本裕治. IPADIC ユーザーズマニュアル, 第 2.5.1 版, 2002.

[39] 阿部紘久. 文章力の基本. 日本実業出版社, 2009.

[40] 荒木哲郎, 池原悟, 橋本憲久. スキップマルコフ連鎖モデルを用いた日本文の誤り検出、訂正方法. 電子情報通信学会技術研究報告, Vol. 99, No. 710, pp. 1–8, 2000.

[41] 犬飼隆. 文字・表記探究法. 朝倉書店, 2002.

[42] 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. 電子情報通信学会論文誌, Vol. 91, No. 12, pp. 1024–1029, 2008.

[43] 今井亨. リアルタイム字幕放送のための音声認識. NHK 技研 R&D, No. 131, pp. 4–13, 2012.

[44] 今村賢治, 齋藤邦子, 貞光九月, 西川仁. 小規模誤りデータからの日本語学習者作文の助詞誤り訂正. 自然言語処理, Vol. 19, No. 5, pp. 381–400, 2012.

[45] 上田晶美, 細田咲江. 超速マスター! E メール・履歴書・エントリーシート成功実例集. 高橋書店, 2009.

[46] 内田友幸, 田中英彦. 可読性向上のための文書自動彩色システム. 電子情報通信学会技術研究報告, Vol. 96, No. 602, pp. 25–30, 1997.

[47] 海野裕也, 二宮崇, 宮尾祐介, 辻井潤一. 機械学習を用いた文脈自由規則の書き換えによる文圧縮. 言語処理学会第 12 回年次大会発表論文集, pp. 1091–1094, 2006.

[48] 大塚敬義, 内海彰, 奥村学. 要約文生成における照応処理. 電子情報通信学会技術研究報告, Vol. 101, No. 61, pp. 19–26, 2001.

[49] 大野誠寛, 松原茂樹, 柏岡秀紀, 加藤直人, 稲垣康善. 節境界に基づく独話の漸進的係り受け解析. 電子情報通信学会論文誌, Vol. J90-D, No. 2, pp. 556–566, 2007.

[50] 小笠原信之. 伝わる！文章力が身につく本. 高橋書店, 2011.

[51] 奥雅博. 日本文推敲支援システム REVISE における複合語同音異義語誤りの検出および訂正支援手法. 電子情報通信学会論文誌, Vol. 79, No. 11, pp. 1836–1846, 1996.

[52] 加藤伸子, 河野純大, 若月大輔, 塩野目剛亮, 黒木速人, 村上裕史, 西岡知之, 皆川洋喜, 白澤麻弓, 三好茂樹, 内藤一郎. 講義の情報保障におけるキーワード提示タイミングに関する基礎的検討. 電子情報通信学会技術研究報告, Vol. 108, No. 170, pp. 51–56, 2008.

[53] 金澤章, 磯野春雄. ニュース字幕の提示タイミングずれの主観評価と補正方法. 2001 年映像情報メディア学会年次大会講演予稿集, pp. 89–90, 2001.

[54] 河原大輔, 黒橋禎夫, 橋田浩一. 「関係」タグ付きコーパスの作成. 言語処理学会第 8 回年次大会発表論文集, pp. 495–498, 2002.

[55] 河原達也, 李晃伸. 連続音声認識ソフトウェア Julius. 人工知能学会誌, Vol. 20, No. 1, pp. 41–49, 2005.

[56] 工藤拓, 松本裕治. チャンキングの段階適用による日本語係り受け解析. 情報処理学会論文誌, Vol. 43, No. 6, pp. 1834–1842, 2002.

[57] 工藤拓, 山本薫, 松本裕治. Conditional Random Fields を用いた日本語形態素解析. 情報処理学会研究報告, Vol. 2004, No. 47, pp. 89–96, 2004.

[58] 熊野明, 吉村裕美子, 野上宏康. 自然な日本語生成のための指針. 情報処理学会第 41 回全国大会講演論文集, pp. 165–166, 1990.

[59] 黒橋禎夫, 白木伸征, 長尾眞. 出現密度分布を用いた語の重要説明箇所の特定. 情報処理学会論文誌, Vol. 38, No. 4, pp. 845–854, 1997.

[60] 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. 情報処理学会研究報告, Vol. 2005, No. 127, pp. 247–252, 2005.

[61] 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, 2006.

[62] 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. 情報処理学会論文誌, Vol. 50, No. 2, pp. 544–552, 2009.

[63] 柴田知秀, 黒橋禎夫. 談話構造解析に基づくスライドの自動生成. 自然言語処理, Vol. 13, No. 3, pp. 91–111, 2006.

[64] 清水徹, 中村哲, 河原達也. 音声翻訳単位の推定における句読点情報の効果. 情報処理学会研究報告, Vol. 2008, No. 123, pp. 127–131, 2008.

[65] 下岡和也, 河原達也, 奥乃博. 講演の書き起こしに対する統計的手法を用いた文体の整形. 情報処理学会研究報告, Vol. 2002, No. 44, pp. 81–88, 2002.

[66] 小学館辞典編集部（編）. 句読点、記号・符号活用辞典。小学館, 2007.

[67] 新納浩幸. 平仮名 N-gram による平仮名列の誤り検出とその修正. 情報処理学会論文誌, Vol. 40, No. 6, pp. 2690–2698, 1999.

[68] 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, pp. 343–364. 明治書院, 1988.

[69] 鈴木英二, 島田静雄, 近藤邦雄, 佐藤尚. 日本語文章における句読点自動最適配置. 情報処理学会第 50 回全国大会講演論文集, pp. 185–186, 1995.

[70] 関友作, 赤堀侃司. テキストにおける段落表示が内容理解に与える影響. 日本教育工学雑誌, Vol. 20, No. 2, pp. 97–108, 1996.

[71] 関友作. テキストの内容把握に対する箇条書とキーワード強調の影響. 日本教育工学雑誌, Vol. 21, pp. 17–20, 1997.

[72] 竹元義美, 福島俊一, 山田洋志. 辞書およびパターンマッチルールの増強と品質強化に基づく日本語固有表現抽出. 情報処理学会論文誌, Vol. 42, No. 6, pp. 1580–1591, 2001.

[73] 土屋雅稔, 注連隆夫, 高木俊宏, 内元清貴, 松吉俊, 宇津呂武仁, 佐藤理史, 中川聖一. 機械学習を用いた日本語機能表現のチャンキング. 自然言語処理, Vol. 14, No. 1, pp. 111–138, 2007.

[74] 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.

[75] 富田紘平, 高村大也, 奥村学. 重要文抽出と文圧縮を組み合わせた新たな抽出的要約手法. 情報処理学会研究報告, Vol. 2009, No. 2, pp. 13–20, 2009.

[76] 中野桂吾, 平井有三. 日本語固有表現抽出における文節情報の利用. 情報処理学会論文誌, Vol. 45, No. 3, pp. 934–941, 2004.

[77] 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.

[78] 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, 2008.

[79] 中村亮太, 井上亮文, 市村哲, 岡田謙一, 松下温. 誘目性の高い講義コンテンツを作成する自動編集システム. 情報処理学会論文誌, Vol. 47, No. 1, pp. 172–180, 2006.

[80] 難波英嗣, 奥村学. 書き換えによる抄録の読みやすさの向上. 情報処理学会研究報告, Vol. 99, No. 73, pp. 53–60, 1999.

[81] 野本忠司, 松本裕治. 人間の重要文判定に基づいた自動要約の試み. 電子情報通信学会技術研究報告, Vol. 97, No. 200, pp. 1–6, 1997.

[82] 則本達哉, 小山登, 小林伸行, 椎名広光, 北川文夫. Vod 講義のための字幕強調や短縮表示法. 情報処理学会研究報告, Vol. 2010, No. 6, pp. 1–7, nov 2010.

[83] 林良彦. 技術文章向けの日本文推敲支援システムの実現と評価. 電子情報通信学会論文誌, Vol. J77-D-II, No. 6, pp. 1124–1134, 1994.

[84] 樋口裕一, 大原理志, 山口雅敏. 「樋口式」文章の書き方 100 のルール:簡潔に書ける!相手に伝わる! PHP 研究所, 2008.

[85] 藤田英時. メール文章力の基本 大切だけど、だれも教えてくれない 77 のルール. 日本実業出版社, 2010.

[86] 本多勝一. 日本語の作文技術. 朝日新聞社出版局, 1982.

[87] 益岡隆志, 田窪行則. 基礎日本語文法 ー改訂版ー. くろしお出版, 1992.

[88] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 高岡一馬, 浅原正幸. 形態素解析システム『茶筌』version 2.2.9 使用説明書, 2002.

[89] 丸山一郎, 阿部芳春, 沢村英治, 三橋哲雄, 江原暉将, 白井克彦. ニュース字幕の提示タイミングずれに対する許容特性. 電子情報通信学会技術研究報告, Vol. 99, No. 123, pp. 21–28, 1999.

[90] 丸山一郎, 阿部芳春, 江原暉将, 白井克彦. ワードスポッティングと動的計画法を用いたテレビ番組に対する字幕提示タイミング検出法. 電子情報通信学会論文誌, Vol. 85, No. 2, pp. 184–192, 2002.

[91] 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.

[92] 村田真樹, 内元清貴, 馬青, 井佐原均. 排反な規則を用いた文節まとめあげ. 情報処理学会論文誌, Vol. 41, No. 1, pp. 59–69, 2000.

[93] 村田匡輝, 大野誠寛, 松原茂樹. 改行・空行挿入によるメールテキストの整形. 言語処理学会第 18 回年次大会発表論文集, pp. 783–786, 2012.

[94] 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.

[95] 安原宏, 小山法孝. 自然言語処理を用いた日本語文書自動整形システム. 情報処理学会論文誌, Vol. 36, No. 6, pp. 1449–1455, 1995.

[96] 安村禎明, 武市雅司, 新田克己. 論文からのプレゼンテーション資料の作成支援. 人工知能学会論文誌, Vol. 18, No. 4, pp. 212–220, 2003.

[97] 山本和英, 池田諭史, 大橋一輝. 「新幹線要約」のための文末の整形. 自然言語処理, Vol. 12, No. 6, pp. 85–112, 2005.

[98] 山本和英, 安達康昭. 国会会議録を対象とする話し言葉要約. 自然言語処理, Vol. 12, No. 1, pp. 51–78, 2005.

[99] 横林博, 菅沼明, 谷口倫一郎. 係り受けの複雑さの指標に基づく文の書き換え候補の生成と推敲支援への応用. 情報処理学会論文誌, Vol. 45, No. 5, pp. 1451–1459, 2004.

[100] 吉田辰巳, 遠間雄二, 増山繁, 酒井浩之. 可読性の向上を目的とした片仮名表記外来語の換言知識獲得. 電子情報通信学会論文誌, Vol. J88-D, No. 7, pp. 1237–1245, 2005.