

機関番号：13901

研究種目：若手研究(B)

研究期間：2009～2010

課題番号：21700157

研究課題名(和文) 読みやすさを考慮したリアルタイム字幕生成のための話し言葉処理手法の開発

研究課題名(英文) Development of spoken language processing techniques for real-time captioning on the basis of readability

研究代表者：

大野 誠寛 (OHNO TOMOHIRO)

名古屋大学・国際開発研究科・助教

研究者番号：20402472

研究成果の概要(和文)：本研究では、講演や解説などに対して読みやすい字幕をリアルタイムに生成するため、その要素技術として、次の3つの話し言葉処理手法、(1)節の始境界検出に基づいて高精度化した、話し言葉の漸進的係り受け解析手法、(2)話し言葉を読みやすいテキストにするための構文構造に基づく話し言葉の整形手法(主に、読点挿入手法)、(3)字幕テキストを読みやすく表示するための構文構造に基づく改行挿入手法、をそれぞれ開発した。

研究成果の概要(英文)：In this research, I developed the following three spoken language processing techniques as elemental techniques for generating readable captions to support the real-time understanding of spoken monologues such as lectures and commentaries: (1) Incremental dependency parsing technique of spoken language based on clause-start identification (2) Text formatting technique (mainly, comma insertion technique) of spoken language based on dependency structure to translate spoken language to readable texts (3) Linefeed insertion technique based on dependency structure to display caption texts in an easy-to-read manner.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	2,100,000	630,000	2,730,000
2010年度	1,200,000	360,000	1,560,000
総計	3,300,000	990,000	4,290,000

研究分野：音声言語処理, 自然言語処理

科研費の分科・細目：情報学・知能情報学

キーワード：音声言語処理, 係り受け解析, 整形処理, 改行処理

1. 研究開始当初の背景

専門家による講演や解説などから得られる情報は人間の貴重な知的資源と考えられ、これらの情報を聴覚障害者や難聴者にも平等に提供する手段として、リアルタイム字幕生成システムの開発が望まれている。

一方、講演文は、話し言葉であり、また、一文が長い傾向にあるため、たとえ音声認識率が100%であっても、書き起こしテキストをそのまま表示しただけでは読みにくい字幕となる。書き起こしテキストを読みやすく整形し、見やすく改行して表示することが求められている。

このような高度な言語処理を伴うリアルタイム字幕生成の実現には、構文情報が必要不可欠であると考えられる。本研究の応募者は、これまで、漸進的係り受け解析手法を開発し、音声に追従した係り受け構造の出力を実現しており、この解析技術を読みやすい字幕生成のための整形処理と改行処理に応用できると考え、本研究を着想するに至った。

2. 研究の目的

本研究では、講演や解説などに対して読みやすい字幕をリアルタイムに生成するため、

その要素技術として、次の3つの話し言葉処理手法を開発することを目的とした。

- (1) これまでに開発した従来手法を高精度化した漸進的係り受け解析手法
- (2) (1)によって得られる構文構造に基づいた、読みやすい言葉への整形処理手法
- (3) (1)によって整形されたテキストを読みやすく表示するための改行処理手法

上述した本研究の各研究項目(1)～(3)の具体的な達成目標をそれぞれ以下に示す。

- (1) 漸進的係り受け解析手法： 応募者は既に、形態素列のパターンのみから検出可能な節の終境界により挟まれた単位(以下、節境界単位)の内部で係り受けが閉じていることを仮定し、この単位ごとに解析することにより、音声に追従した係り受け構造の出力を実現している。しかし、この従来手法は、埋め込み節が存在する場合に生じる、節境界単位で閉じていない係り受けを解析することができず、係り受け正解率は76.2%に留まっていた。節境界単位で閉じていない係り受けは書き起こしテキストを読みにくくする原因の一つと考えられるため、本研究では、節境界単位で閉じていない係り受けを解析できるように従来手法を改良し、同程度の解析時間、解析の漸進性を保ちつつ、係り受け正解率を80.0%以上に改善することを目指す。
- (2) 読みやすい言葉への整形処理手法： 本研究では、節境界単位で閉じていない係り受けによって生じる読みにくい発話を、文節を移動することによって、読みやすく整形することを試みる。そのために、まず、講演の書き起こしテキストに対して人手で整形と改行を施し、20万形態素規模の字幕コーパスを構築する。このコーパスの分析結果と(1)で得られる係り受け解析結果に基づいて、上記の整形処理を実現する。被験者実験により、書き起こしテキストと比較して読みやすいと判断されることを目指す。
- (3) テキストを読みやすく表示するための改行処理手法： 本研究では、一般講演におけるリアルタイム字幕生成を目指しているため、講演者の横で字幕のみが表示される画面(1行20文字、複数行の字幕が1行ずつスクロール表示される画面)を想定し、上記の字幕コーパスを構築する。構築した字幕コーパスの改行位置について係り受け情報を用いて分析し、字幕出力の同時性も考慮した改行処理プログラムを実装する。字幕出力の同時性を評価し、少なくともテキスト処理部分においては数秒程度の遅れに抑えることを

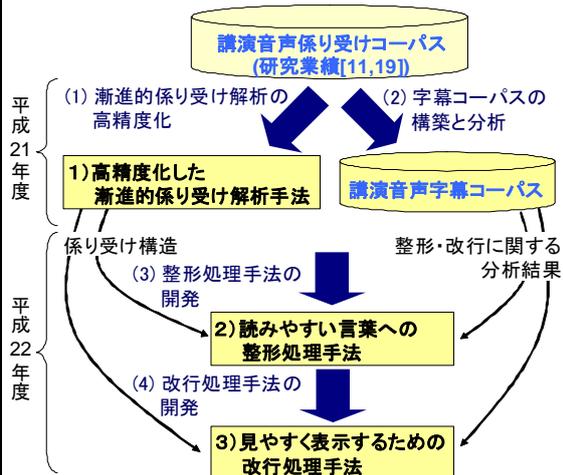


図1 本研究の全体計画

目指す。字幕コーパスとの改行位置の一致率が、人間同士の一致率と同程度になることを目指す。

3. 研究の方法

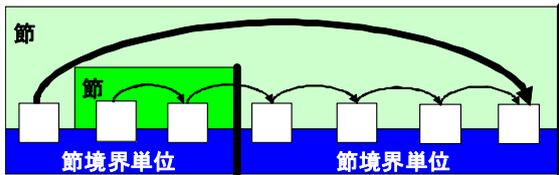
本研究では、講演文のリアルタイム字幕生成の開発を目指し、平成21年度から平成22年度までの2年間で、上述した(1)～(3)の話し言葉処理手法の開発を行った。図1に本研究の全体計画を示す。本研究では、まず、平成21年度に、(1)漸進的係り受け解析の高精度化と(2)字幕コーパスの構築と分析を並行して行った。次に、平成22年度では、初年度に開発した漸進的係り受け解析器から得られる係り受け構造と字幕コーパスの分析結果に基づいて、(3)整形処理手法の開発と(4)改行処理手法の開発を順に実施した。

以下では、年度ごとに実施した研究開発について述べる。

初年度(平成21年度)は、(1)漸進的係り受け解析の高精度化と(2)字幕コーパスの構築と分析を並行して行った。

(1) 漸進的係り受け解析の高精度化では、以下の①～②を順に実施した。

- ① 漸進的係り受け解析の高精度化： 本研究の実施以前に応募者は、機械学習を用いて埋め込み節の始境界を事前に検出することにより、節境界単位で閉じていない係り受け(図2の太線矢印)に対する解析精度を向上させた係り受け解析手法を開発しつつあった。本研究では、この手法とこれまでに開発している漸進的係り受け解析手法を統合し、節境界単位で閉じていない係り受けをも解析可能な漸進的係り受け解析手法を開発した。
- ② 漸進的係り受け解析器の評価： 本研究の



非常に 試合に 参加し 日本の 青年 表情が 明るいと
ている たちの いうのに

図2 節境界単位で閉じていない係り受けの例

実施以前に応募者は、名古屋大学統合音響情報研究拠点で構築された CIAIR 同時通訳データ (Matsubara et al. 2002) の日本語講演音声の書き起こしデータに対して、形態素・節境界・係り受けの情報を人手で付与したコーパス (以下、講演音声係り受けコーパス) を構築していた。このコーパスを用いて実験し、高精度化した漸進的係り受け解析器の有効性を評価した。

(2) 字幕コーパスの構築と分析では、以下の①～③を順に実施した。

- ① 字幕コーパスの構築： 整形処理手法と改行処理手法を開発するための分析用・機械学習用データとして、講演音声係り受けコーパスに対して読みやすくなるように整形や改行を人手で施し、字幕コーパスを構築した。なお、本研究では、一般講演におけるリアルタイム字幕生成を目指しているため、講演者の横で字幕のみが表示される画面 (1行20文字、複数行の字幕が1行ずつスクロール表示される画面) を想定した (図3参照)。改行位置の付与では、1画面に字幕のみが複数行表示される場合の適切な改行位置に関する知見はなく、人による違いも想定されるため、同一データに対して複数人による付与作業を実施した。
- ② 字幕コーパス (整形箇所) の分析： 構築した字幕コーパスを用いて、節境界単位で閉じていない係り受けと読みにくさとの関係を明らかにした。
- ③ 字幕コーパス (改行位置) の分析： 適切な改行位置について、構築した字幕コーパスを用いてその特長を明らかにした。人による改行位置の違いについても分析を試みた。

平成22年度は、(3)整形処理手法の開発と(4)改行処理手法の開発を順に実施した。

(3) 整形処理手法の開発では、以下の①～②を順に実施した。

- ① 整形処理手法の開発： (2)-②の分析結果と漸進的係り受け解析の結果を用いて、

書き起こしテキストをそのまま表示

日本に帰ってきてすぐ私も家のテレビにかじりついてワールドカップの予選を見たわけですが非常に試合に参加している日本の青年たちの表情が明るいというのにもうとても強い印象を受けましてまたすごく気持ちが良い思いがしました



読みやすく整形し、見やすく改行して表示

日本に帰ってきてすぐ
私も家のテレビにかじりついて
ワールドカップの予選を見たわけですが
試合に参加している日本の青年たちの表情が
非常に明るいというのに
とても強い印象を受けまして
すごく気持ちが良い思いがしました

図3 1行20文字以内の字幕表示例
(字幕のみを表示する画面を想定)

節境界単位で閉じていない係り受け (図2太線矢印) で生じる不自然な構文構造を解消し読みやすくする整形処理手法を開発することを試みた。また、新聞記事の読点情報に基づき、適切な位置に読点を挿入する手法を開発し、読みやすいテキスト生成のための基礎技術を開発した。

- ② 整形処理手法の評価： 講演音声係り受けコーパスの書き起こしテキストに対して開発した整形処理手法を適用し、構築した字幕コーパスとの一致率を定量的に評価した。特に、節境界単位で閉じていない係り受けによって読みにくくなっている箇所の整形について考察した。また、読点挿入手法に関しては、新聞記事に対して読点挿入実験を実施し、本手法の有効性を定量的に評価した。

(4) 改行処理手法の開発では、以下の①～②を順に実施した。

- ① 改行処理手法の開発： 本研究の実施以前に応募者は、講演文に対してルールベースの改行挿入手法を開発しているが、ルールベースの手法では、話し言葉の多様性・個性から、ルールの網羅性、保守性に問題があることが分かった。そこで、(2)-③における大量データの分析結果に基づいて、係り受け構造や節境界情報、字幕出力の同時性、1行あたりの文字数を考慮した統計的改行処理手法を開発した。
- ② 改行処理手法の評価： (3)-②の評価と同様に、改行処理後のテキストと字幕コーパスとの改行位置の一致率による定量的評価を行った。また、被験者に実際に

字幕を提示して行うアンケートによる主観評価を実施した。同一データに対して複数人が付与した改行位置を考慮した評価法を検討した。また、被験者実験では、改行前と改行後、字幕コーパスの3つの字幕を比較評価した。

なお、上記の(1)～(4)の研究を効率的に進めるため、以下に示す項目を実施した。

- ① 字幕コーパス構築の効率化： 講演音声及び対話音声の係り受けコーパスを構築した経験を活かし、事前に少人数で字幕コーパスを少量作成した上でマニュアルを作成し外注することにより、大量のデータを効率的に作成した。なお、作業者の多大な負担を軽減するため、作業支援ソフトウェアも事前に開発した。また、随時、発注業者とのコミュニケーションをとり、マニュアルの確認・修正、作業支援ソフトウェアの改良を行った。
- ② 整形処理手法開発の効率化： 講演音声の書き起こしを読みにくくする原因のうち、節境界単位で閉じていない係り受けによるもの以外については、名古屋大学大学院情報科学研究科の大学院生1名の協力を得て、先行研究に基づいた整形処理プログラムを実装することにより、開発時間の短縮を図った。
- ③ 改行処理手法開発の効率化： 機械学習における素性選択のためのデータ分析を効率的に行うために、講演文に対するルールベースの改行挿入手法においてルール化した項目に着目し、字幕コーパスを用いた定量的な分析を実施した。

4. 研究成果

本研究では主に、(1)漸進的係り受け解析の高精度化、(2)字幕コーパスの構築と分析、(3)整形処理手法の開発と、(4)改行処理手法の開発、を行い、以下に示す成果を得ることができた。

- (1) 漸進的係り受け解析の高精度化： 節境界単位で閉じていない係り受けをも解析可能な係り受け解析手法を開発した。名古屋大学統合音響情報研究拠点で構築されたCIAIR同時通訳データの日本語講演音声の書き起こしデータに対して、形態素・節境界・係り受けの情報を人手で付与したコーパス（以下、講演音声係り受けコーパス）を用いて実験し、高精度化した漸進的係り受け解析器の有効性を確認した。
- (2) 字幕コーパスの構築と分析： 整形処理手法と改行処理手法を開発するための分析用・機械学習用データとして、講演音声係り受けコーパスに対して読みや

すくなるように整形や改行を人手で施し、字幕コーパスを構築した。構築した字幕コーパスを用いて、節境界単位で閉じていない係り受けと読みにくさとの関係を明らかにした。また、適切な改行位置について、構築した字幕コーパスを用いてその特長を明らかにした。さらに、人による改行位置の違いについても明らかにした。

- (3) 整形処理手法の開発： 節境界単位で閉じていない係り受けによって生じる読みにくい発話を、文節を移動することによって、読みやすく整形することを試みた。さらに、適切な位置に読点を挿入する手法を開発し、読みやすいテキスト生成のための基礎技術を開発した。新聞記事に対して読点挿入実験を実施し、本手法は適合率 84.1%、再現率 69.1%、F 値 75.9%を達成しており、人間による読点挿入とほぼ同程度の精度を実現することができた。
- (4) 改行処理手法の開発： 係り受け構造や節境界情報、字幕出力の同時性、1行あたりの文字数を考慮した統計的改行処理手法を開発した。字幕出力の同時性を評価し、少なくともテキスト処理部分においては数秒程度の遅れに抑えることを達成した。また、字幕コーパスとの改行位置の一致率を評価し、本手法は、人間による改行挿入と同程度の精度を達成していることを確認した。さらに、主観的評価を実施し、本手法により生成した字幕テキストが、単純なベースラインと比べ、読みやすくなっていることを確認した。

上述の研究成果の国内外における位置づけは以下の通りである。

- ① リアルタイム字幕生成における位置づけ： リアルタイム字幕生成を実現するために、音声認識器の改良が行われている(Saraclar et al. 2002; 西光他 2005)。本研究は、これらの音声認識器によって書き起こされるテキストに対する漸進的な言語処理技術を開発した。
- ② 話し言葉から書き言葉への文体整形における位置づけ： 冗長表現の換言・削除(山本&安達 2005)や助詞の挿入など(下岡他 1995)を行っている研究がある。本研究は、これらの研究で対象としていない話し言葉で生じる不自然に複雑化した構文構造の単純化を試みた。
- ③ 字幕制作の改行処理における位置づけ： 日本のテレビ番組におけるクローズドキャプションの改行位置を形態素列のパターンにより決定する研究(門馬他 2001)があるが、2行に詰め込むことが重視され

意味的なまとまりはあまり考慮されておらず、見やすい改行となっていない。本研究では、意味的なまとまりを考慮し、係り受け構造や節境界情報を用いた改行処理を試みた。

また、上述の研究結果の国内外におけるインパクト・意義は以下の通りである。

- ① 2007年6月1日に閣議決定された『インバージョン 25』では、「一般会話レベルの多言語翻訳の実現」が達成目標の一つとして設定され、そのサブゴールとして「2010年頃までの日常会話の構文解析の実現」が挙げられている。本研究で開発する漸進的係り受け解析手法は、この目標の達成に貢献するものであり、社会的意義が大きいと言える。
- ② 2008年3月にまとめられた『中途失聴・難聴者等聴覚障害者のコミュニケーション支援に関する現状把握調査・研究事業報告書(平成19年度障害者保健福祉推進事業(厚生労働省補助事業))』では、聴覚障害者に対するアンケート結果から、「表記上の見やすさ、読みやすさ(可読性)を実現する技術」がパソコン要約筆記に求められている実態が報告されている。本研究で開発される整形・改行処理手法は、まさにこの要望に応えるものであり、パソコン要約筆記支援ツールとしての応用も期待できる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計3件)

- ① Masaki Murata, Tomohiro Ohno, Shigeki Matsubara : Linefeed Insertion Rules for Improving Readability of Lecture Transcript, International Journal of Knowledge and Web Intelligence, 査読有, Vol.1, No. 3/4, pp. 227-242, Jul. 2010.
- ② 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, 査読有, No. 9, pp. 1621-1631, Sep. 2009.
- ③ Masashi Ito, Tomohiro Ohno, Shigeki Matsubara : Text-Style Conversion of Speech Transcript into Web Document for Lecture Archive, Journal of Advanced Computational Intelligence and Intelligent Informatics, 査読有, Vol.13, No. 4, pp. 499-505, Jul. 2009.

[学会発表] (計6件)

- ① Masaki Murata, Tomohiro Ohno, Shigeki Matsubara : Automatic Comma Insertion for Japanese Text Generation, the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP2010), 査読有, アメリカ・ボストン・マサチューセッツ工科大学, 2010年10月10日
- ② 村田 匡輝, 大野 誠寛, 松原 茂樹 : 読点の用法的分類に基づく自動読点挿入, 情報処理学会 音声言語情報処理研究会, 東京工業大学, 2010年5月27日.
- ③ 村田 匡輝, 大野 誠寛, 松原 茂樹 : 日本語テキストにおける読点位置の検出, 言語処理学会第16回年次大会, 東京大学, 2010年3月10日.
- ④ Tomohiro Ohno, Masaki Murata, Shigeki Matsubara : Linefeed Insertion into Japanese Spoken Monologue for Captioning, Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP2009), 査読有, シンガポール・サンテック市・Suntec Singapore International Convention & Exhibition Centre, 2009年8月4日.
- ⑤ Masaki Murata, Tomohiro Ohno, Shigeki Matsubara : Automatic Linefeed Insertion for Improving Readability of Lecture Transcript, the 2nd International Symposium on Intelligent Interactive Multimedia Systems and Services (IIMSS-2009), 査読有, イタリア・Mogliano Veneto・VILLA BRAIDA ホテル, 2009年7月17日.
- ⑥ 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文への漸進的改行挿入, 情報処理学会 情報処理学会 音声言語情報処理研究会, 東京工業大学, 2009年5月21日.

[その他]

ホームページ等

<http://slp.itc.nagoya-u.ac.jp/~ohno/kaken/readable-caption/>

6. 研究組織

(1) 研究代表者

大野 誠寛 (OHNO TOMOHIRO)

名古屋大学・大学院国際開発研究科・助教
研究者番号 : 20402472

(2) 研究分担者 なし

(3)連携研究者 なし

(4)研究協力者

村田 匡輝 (MURATA MASAKI)

名古屋大学・大学院情報科学研究科・博士
課程後期課程