

科学研究費助成事業 研究成果報告書

平成 27 年 5 月 22 日現在

機関番号：13901

研究種目：基盤研究(B)

研究期間：2010～2014

課題番号：22300051

研究課題名(和文) 同時的な発話理解のための話し言葉処理に関する研究

研究課題名(英文) Spoken language processing for simultaneous speech understanding

研究代表者

松原 茂樹 (Matsubara, Shigeki)

名古屋大学・情報科学研究科・准教授

研究者番号：20303589

交付決定額(研究期間全体)：(直接経費) 13,400,000円

研究成果の概要(和文)：リアルタイム音声言語システムの実現のための基盤技術として、話し言葉の逐次的な解析方式、すなわち、音声の入力と同時進行で解析し、入力途中の段階で随時、解析結果を出力する技術の開発を目的に研究を遂行した。逐次的な解析のための係り受け構造を策定し、それを生成する解析過程の実行可能性を示した。また、開発した技術をリアルタイム字幕生成などの音声言語システムに応用し、実験により話し言葉の逐次解析の有効性を示した。

研究成果の概要(英文)：As one of language technologies for realization of real-time spoken language systems, we developed methods for incremental analysis of spoken languages. The incremental analysis means processing spoken languages and generating the analysis results synchronously with speech inputs. We defined a style of the dependency structure for incremental analysis and demonstrated the feasibility of such the analysis process. We applied the incremental analysis to spoken language systems such as real-time captioning and experimentally indicated the effectiveness of incremental language analysis.

研究分野：自然言語処理

キーワード：音声言語処理 構文解析 意味解析 機械翻訳 対話処理 字幕生成 言語資源 コーパス

1. 研究開始当初の背景

音声は、人間にとって極めて手軽で効率のよいコミュニケーション形態である。これまで、「人と人」「人と機械」のコミュニケーション支援に向け、音声対話、音声翻訳、音声要約などの音声言語処理システムの開発が進められてきた。音声は、時系列上で発生する。発話は、テキスト上に記せば文字列としてまとめて存在するものの、音声としては音素(または音節)の連続として一過的に出現するに過ぎない。音声におけるこの連続性・一過性という性質は、聞き手が発話をその出現と同時に理解することを強いる。このような人間の言語理解における同時性は、我々の直観が強く支持している。

しかし、話し言葉処理の研究ではこれまで、音声におけるこのような性質はほとんど考慮されることはなく、文字化された文を単位とした技術開発が進められてきた。もちろん、人間と計算機とでは、演算や記憶などの物理的な制約が異なるため、計算機が人間と同じ方法論を採る必然性はない。しかし、言語処理プロセスの違いは、音声言語処理システムの性能を左右するものであり、実際、「音声翻訳における同時通訳」「音声対話における相槌・割り込み」「音声要約におけるリアルタイム提示」などの機能は、同時的な発話理解の仕組み無しに実現は難しい。音声の入力と同時進行で文字化するリアルタイム音声認識方式が提案されており、同時的な発話理解を可能にする音声処理技術が整備されつつある現在、一方の言語処理技術の整備が強く望まれる状況にあった。

同時的な発話理解のための言語処理技術として、文脈自由文法を用いた逐次的な構文解析がいくつか提案されているものの、日本語についてはほとんど試みがなかった。また、逐次的な意味解析についても同様の状況にあった。

2. 研究の目的

リアルタイム音声言語処理システムの実現のための基盤技術として、本研究では、話し言葉の逐次的な構文・意味解析方式の開発を目的に研究を遂行した。本研究を進めるにあたり、実現する話し言葉処理の機構を、

- 音声の入力と同時進行的に解析を実行できること

- 入力途中の段階で随時、それまでの解析結果を出力できること

を満たすもの、として定義した。また、このような技術を開発するために、

- 話し言葉の逐次的な構文・意味解析技術を開発すること

- 字幕提示などの音声言語処理システムに開発した解析プログラムを移植し、導入の効果を実証すること

を目標として設定した。なお、話し言葉処理では、文法に逸脱する現象への頑健な対処も重要な課題となるが、これについてはすでに多

くの取り組みがあり、本研究のターゲットからは除外した。

3. 研究の方法

本研究で設定した目標に対して、それぞれアプローチした。具体的には、「逐次的な構文・意味解析手法」及び「解析手法の音声言語処理への応用」を実施した。

(1) 逐次的な構文・意味解析手法の開発

リアルタイム音声言語システムでは、入力音声に追従して結果を出力することが求められる。このようなシステムが、入力文を構成する要素間の依存関係を早期に利用できることを目的に、文の入力と同時進行的に解析を実行する逐次的な依存構造解析が開発されてきた。従来の研究は、一般的な依存構造解析と同様、入力文中の依存関係、すなわち、依存元と依存先の組を同定することとして問題を捉え、その同定処理の実現方法に焦点を当ててきた。ここでは、入力が進むごとに解析の途中結果を更新し、依存元と依存先の組を同定するたびに、その組を解析結果として出力することを想定していた。

しかし、上述の枠組みによる逐次的な依存構造解析は、依存元と依存先のどちらか一方だけでなく双方が入力されるまで、その依存関係を出力できないという問題がある。このことは、入力された構成要素に対して、その解析結果を用いてシステムの上位層が処理を実行できるタイミングは早くても、その要素と依存関係にある要素が入力された後であることを意味し、処理のリアルタイム性を高める上で障害となる。特に、日本語の依存関係、すなわち、文節間の係り受け関係の場合、一般に、依存元の文節は依存先の文節より後方に現れ、かつ、文末の文節が多く文節に対する係り先となりやすいため、システムの上位層の処理タイミングへの影響はより大きくなる。

このため本研究では、リアルタイム音声言語システムにおいて、文節間の依存関係に関する情報をできる限り早期に利用するという観点から、逐次的係り受け解析が生成する係り受け構造について考察した。具体的には、文の入力途中において、係り先がすでに入力されている文節については係り先文節との依存関係を、また、係り先が入力されていない文節については入力済みのどの文節とも依存関係にないことを、係り受け構造が明示することとした。これは、文節間の依存性だけでなく、非依存性に関する情報を係り受け構造が併せて保持することにより、音声言語システムの上位層は、入力された文節列における構文的なまとまりをより考慮した処理の実行が可能となると考えたためである。

また、本研究では、提案する係り受け構造の逐次的生成手法を、係り受け解析の従

来法を拡張することにより実現することとし、その解析精度を評価するために、日本語講演データを用いて係り受け解析実験を実施した。

(2) 逐次解析の音声言語処理への応用

講演などの音声に対して、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。講演では、文が長くなる傾向にあり、1文が字幕上で複数行にまたがって表示されることになる。提示されたテキストが読みやすくなるように、適切な箇所に改行が挿入されていることが重要である。この問題に対して、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いる。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。

そこで本研究では、音声に対してリアルタイムに字幕を提示するための、字幕テキストへの逐次的な改行挿入手法について検討した。その際、自動音声認識やパソコン要約筆記などにより音声と同時に文字化され、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その行を提示するシステムを想定した。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が、それは、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、細かい単位ごとに同定処理を実行し改行位置を決定することが望まれるためである。

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式として、文節単位で同定処理を実行する手法を検討した。改行位置は文節境界にあることを前提とすると、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。統計的なアプローチによる改行挿入では、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がすでに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した。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。

本研究では次に、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した。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した。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考察した。

4. 研究成果

本研究により、同時的な話し言葉処理に関していくつかの知見を得ることができた。音声言語の解析技術の開発、及び、音声言語処理システムへの応用においてそれぞれ以下の成果が得られた。

(1) 逐次的な解析の研究成果

本研究では、文の入力途中の段階で逐次的係り受け解析が生成する係り受け構造を定めた。この構造では、入力済みの係り元と係り先の組だけでなく、係り先が未だ入力されていない文節については、入力済みのどの文節にも係らないことを明示することができる。また、そのような係り受け構造を生成可能な逐次的係り受け解析の実現可能性を示した。評価実験により、そのような係り受け構造の利用可能性を確認した。さらに、読みやすい字幕生成のための改行位置の逐次的な同定に本手法を応用し、その有効性を確認することができた。

(2) 音声言語処理システムへの応用

本研究では、リアルタイム字幕生成のための改行挿入手法として、音声の入力に対して早い段階で改行位置を決定する手法を実現した。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、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それぞれ実現した。評価実験の結果、両手法のいずれにおいても高い改行挿入精度を達成できることを確認することができた。

5. 主な発表論文等

[雑誌論文](計6件)

Yoshihide Kato, Shigeki Matsubara, Identifying Nonlocal Dependencies in Incremental Parsing, IEICE Trans. on Information and Systems, 査読有, E98-D(4), 994-998, 2015.

大野 誠寛, 松原 茂樹, 文節間の依存・非依存を同定する漸進的係り受け解析, 電子情報通信学会論文誌, 査読有, D-98(4), 709-718, 2015.

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のな改行挿入, 電気学会論文誌, 査読有, 133-C(2), 418-426, 2013.

[学会発表](計11件)

加藤 芳秀, 松原 茂樹, 組合せ範疇文法に基づく漸進的な意味解析, 言語処理学会第20回年次大会論文集, 356-359, 2014.

村田 匡輝, 大野 誠寛, 松原 茂樹, 未入力文節との構文的関係を考慮する漸進的な係り受け解析, 言語処理学会第20回年次大会論文集, 193-196, 2014.

Tomohiro Ohno, Shigeki Matsubara, Dependency Structure for Incremental Parsing of Japanese and its Application, Proceedings of 13th Int. Conference on Parsing Technologies, 91-97, 2013.

加藤 芳秀, 松原 茂樹: 漸進的構文解析における長距離依存関係の同定, 言語処理学会第18回年次大会論文集, 963-966, 2012.

Koichiro Ryu, Haibei Yu, Shigeki Matsubara, Corpus-based Analysis of Simultaneous Interpreters' Speech Rates, Proceedings of 9th International Symposium on Natural Language Processing, 110-114, 2012.

6. 研究組織

(1) 研究代表者

松原 茂樹 (MATSUBARA, Shigeki)

名古屋大学・大学院情報科学研究科・准教授

研究者番号: 20303589

(2) 研究分担者

柏岡 秀紀 (KASHIOKA Hideki)

独立行政法人情報通信研究機構・脳情報通信融合研究センター・室長

研究者番号: 10395022

加藤 芳秀 (KATO Yoshihide)

名古屋大学・情報連携統括本部・准教授

研究者番号: 20362220

大野 誠寛 (OHNO Tomohiro)

名古屋大学・情報基盤センター・助教

研究者番号: 20402472