

様 式 C - 7 - 1

平成 2 4 年度科学研究費助成事業（科学研究費補助金）実績報告書（研究実績報告書）

1. 機関番号

1	4	6	0	3
---	---	---	---	---

 2. 研究機関名 奈良先端科学技術大学院大学
3. 研究種目名 基盤研究(A) 4. 補助事業期間 平成 2 3 年度 ~ 平成 2 5 年度
5. 課題番号

2	3	2	4	0	0	2	0
---	---	---	---	---	---	---	---
6. 研究課題 大域情報を利用した同時処理による自然言語解析手法の研究

7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
1 0 2 1 1 5 7 5	マツモト ユウジ 松本 裕治	情報科学研究科	教授

8. 研究分担者

研究者番号	研究分担者名	所属研究機関名・部局名	職名
9 0 3 1 1 5 8 9	シンボ マサシ 新保 仁	情報科学研究科	准教授
6 0 5 8 1 3 2 9	コマチ マモル 小町 守	情報科学研究科	助教
8 0 6 3 7 3 2 2	ドゥ ケビン Duh Kevin	情報科学研究科	助教

9. 研究実績の概要

係り受け解析に基づく統語解析の高性能化のため、ボトムアップ的な情報とトップダウンの情報の同時利用に関する研究を行い、両者の情報を融合できる新しい手法の提案を行った。また、遷移に基づく決定的な係り受け解析の誤りの伝搬を除去するため、後戻りなしに解析誤りを修正しつつ解析を継続する手法を、グラフデータマイニングのアルゴリズムを利用して実現する方法について研究し、従来手法の欠点を補う手法を提案した。

言語学習者による誤りを含む可能性のある文の誤り検出・修正を可能にする言語解析手法に関する研究を行った。英語の綴り誤り訂正と品詞解析を同時に行う手法を提案し、これらの処理を逐次的に行う従来手法より誤り訂正効率を向上できることを示した。また、学習者の作文に生じる動詞のテンス・アスペクトに関する誤りを修正するため、文書全体の大域的な情報を利用する方法を提案し、個々の文に現れる動詞の誤りを個別に検出・訂正するよりも良好な結果が得られることを示した。

単語を言語解析の基本要素とする従来の文解析法の性能を向上させるため、複数の語がまとまって役割や意味をもついわゆる多単語表現の英語辞書の構築と、それを用いた英語の品詞解析の方法を提案した。本年度は機能語として働く多単語表現辞書、および、英語の共通データである Penn Treebank への多単語表現のアノテーションコーパスの構築を行った。同時に、この辞書とコーパスを利用して、英語の品詞解析システムの試作を行い、解析精度の向上を確認した。また、英語の統語解析精度の向上を目指し、英語の複文パターンの調査を行い、特に、理め込み文に関係する 300 程度のパターンの抽出を行った。

10. キーワード

(1) 自然言語処理	(2) 形態素解析	(3) 統語解析	(4) 機械学習
(5) 言語資源	(6)	(7)	(8)

11. 現在までの達成度

(区分)(2) おおむね順調に進展している。

(理由)

形態素(品詞)解析, および, 統語解析においては, 大域的な情報の利用および複数の処理の同時処理において, 当初の計画以上の成果を得ることができた。
並列構造の解析, および, 述語項構造解析と統語解析との同時処理については, 継続して研究を行なっているが, 研究発表として具体的な成果を得る結果をだすことができなかった。これらについては, 引き続き次年度も研究を継続する予定である。
言語学習者が書いた誤りを含む可能性のある文の解析法の研究への展開は, 当初は具体的な課題を設定していなかったが, 誤り訂正と品詞解析の融合, および, 動詞のテンス・アスペクト誤りへの大域的な情報の利用に関する具体的な成果を挙げることができた。
英語の多単語表現の辞書構築とそれを利用した英語の品詞解析手法, および, 英語の複文構造のパターンの抽出に関しても, 当初計画していた以上の成果を挙げることができた。

12. 今後の研究の推進方策

(今後の推進方策)

今後は, 以下の項目に関して研究を実施する予定である。
形態素解析と係り受け解析の同時処理を対象に, 大規模コーパスから自動的に大域的な情報を抽出し, 解析精度を向上させる手法の開発に取り組む。また, 表記誤りや表記のバリエーションなどの表層的な異表記に対応する処理と形態素解析の同時処理に関する研究を行う。
係り受け解析と並列構造解析の同時処理を実現するために, 前年度に続き, 系列アラインメントに基づく並列構造解析手法と従来の係り受け解析の同時処理を可能にする手法の研究を行う。
述語項構造解析において, 大規模コーパスから獲得した事象の推移に関する知識と文書全体大域的な情報の利用, および, 2つ以上の述語の項解析の同時最適化に関する研究を行う。
上記の様々なタスクの基本情報となる単語間の類似度尺度の計算法の研究を前年度に続いて行う。類似度計算において, 対象単語を含む文を係り受け解析し, 係り受け関係を利用した部分構造の類似度や, 潜在意味クラスを用いることによって類似度の評価尺度を拡張する手法について研究を行う。
各タスクの基礎データの構築のため, 様々な分野の言語データに対してアノテーション作業を行う。また, これまで開発してきたコーパスアノテーションツールを拡張し, 本研究で開発する種々の言語解析ツールを, コーパスアノテーションツールから呼び出せる機能を追加する。これにより, 自動解析ツールを手動によるアノテーション作業に有効に利用する方法を考案し, それを実現する環境の構築へとつなげる。

13.研究発表(平成24年度の研究成果)

〔雑誌論文〕計(0)件 うち査読付論文 計(0)件

著者名	論文標題			
雑誌名	査読の有無	巻	発行年	最初と最後の頁

掲載論文のDOI(デジタルオブジェクト識別子)				

〔学会発表〕計(5)件 うち招待講演 計(1)件

発表者名	発表標題		
Katsuhiko Hayashi	Head-driven Transition-based Parsing with Top-down Prediction		
学会等名	発表年月日	発表場所	
The 50th Annual Meeting of the Association for Computational Linguistics	2012年07月10日	Jeju Island, Korea	

発表者名	発表標題		
Toshikazu Tajiri	Tense and Aspect Error Correction for ESL Learners Using Global Context		
学会等名	発表年月日	発表場所	
The 50th Annual Meeting of the Association for Computational Linguistics	2012年07月11日	Jeju Island, Korea	

発表者名	発表標題		
Akihiro Inokuchi	Mining Rules for Rewriting States in a Transition-based Dependency Parser		
学会等名	発表年月日	発表場所	
The 12th Pacific Rim International Conference on Artificial Intelligence	2012年09月05日	Sarawak, Malaysia	

発表者名	発表標題	
Keisuke Sakaguchi	Joint English Spelling Error Correction and POS Tagging for Language Learners Writing	
学会等名	発表年月日	発表場所
The 2012 International Conference on Computational Linguistics	2012年12月10日	Mumbai, India

発表者名	発表標題	
Yuji Matsumoto	Things between Lexicon and Grammar	
学会等名	発表年月日	発表場所
26th Pacific Asia Conference on Language Information and Computation(招待講演)	2012年11月08日	Bali, Indonesia

(図書) 計(0)件

著者名	出版社		
書名	発行年	総ページ数	

14. 研究成果による産業財産権の出願・取得状況

(出願) 計(0)件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	出願年月日	国内・外国の別

(取得) 計(0)件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	取得年月日	国内・外国の別
				出願年月日	

15.備考

--