

様 式 C - 7 - 1

平成 2 7 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) (一般) 4. 研究期間 平成 2 6 年度 ~ 平成 2 8 年度
5. 課題番号

2	6	2	4	0	0	3	5
---	---	---	---	---	---	---	---
6. 研究課題名 構文パターン獲得と並列構造解析による統語的依存構造解析の高精度化

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
8 0 6 3 7 3 2 2	ドゥ ケビン Duh Kevin	情報科学研究科	助教
2 0 7 3 4 7 8 4	シンドウ ヒロユキ 進藤 裕之	情報科学研究科	助教

9. 研究実績の概要

並列構造解析を考慮した依存構造解析アルゴリズムを実装し、英語の代表的な統語解析アノテーション付コーパスである Penn Treebank から自動変換によって得られた依存構造木におけるカバレッジの評価を行った。また、部分並列構造を含む文の解析に対応するため、Penn Treebank における並列構造の範囲および並列構造に関連する依存構造のアノテーションの仕様設計を開始した。

英語の構造パターン収集のために作成した例文集に対して、複単語表現および複文構文パターンのアノテーション作業を昨年度より継続し、収集した 700 文に対するアノテーションを完了した。複単語表現および複文構文パターンを辞書管理システムに格納するためのフォーマットの詳細設計を行った。

機能表現として働く複単語表現および句動詞の網羅的な辞書構築を行い、Penn Treebank におけるすべての出現について、正用例かどうかの確認を半自動で行う方法を実行し、複単語機能表現および句動詞のアノテーションがほどこされたコーパスを構築した。大規模なコーパスから階層的な言語表現パターンを抽出する方法について検討し、2 種類の手法を実装し、言語モデルとしての性能評価実験を行った。

10. キーワード

- | | | | |
|------------|------------|------------|-----------|
| (1) 自然言語処理 | (2) 並列構造解析 | (3) 依存構造解析 | (4) 多言語処理 |
| (5) 機械学習 | (6) 構文パターン | (7) | (8) |

11. 現在までの進捗状況

(区分)(2) おおむね順調に進展している。

(理由)

本研究の目標である並列構造解析および複文パターンの構築について基礎的な実験および基盤となる辞書の構築とコーパスへのアノテーション作業に一定の区切りをつけることができた。

12. 今後の研究の推進方策

(今後の推進方策)

Penn Treebank中の並列構造と依存構造のアノテーションの詳細設計を行い、句構造情報を利用した自動アノテーションと人手による修正作業をコーパス全体に対して行う。このコーパスを用いて並列構造の範囲同定および依存構造解析の実験を行い、性能を評価する。特に、並列構造解析と依存構造解析を同時に行う手法について再検討を行い、部分並列構造の解析にも対応させるとともに、その性能評価を行う。

本年度収集した英語の複文構文パターンを辞書として完成させて辞書管理システムに収納するフォーマットを決定し、辞書データとして完結させる。さらに、収集した複文パターンのPenn Treebankにおける出現を網羅的に探索し、一部人手による確認を行いながら半自動でアノテーションを行う方法を検討し、複文構文パターンを網羅的にアノテーションされたコーパスを構築する。

複単語表現および複文構文パターンを利用して依存構造解析の性能を向上させるアルゴリズムの詳細設計と実装を行い、アルゴリズムの性能評価実験を行う。

句動詞以外に、修飾語等を伴うことのできる柔軟な複単語表現の収集を行う。

本研究課題で構築した複単語表現、句動詞、複文構文パターンの辞書、および、並列構造、複単語表現、複文構文パターンのアノテーションを行ったコーパスの情報を研究利用可能な言語資源として無償公開する。

13. 研究発表(平成27年度の研究成果)

(雑誌論文) 計(0)件/うち査読付論文 計(0)件/うち国際共著論文 計(0)件/うちオープンアクセス 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	国際共著	
掲載論文のDOI(デジタルオブジェクト識別子)						
オープンアクセス						

(学会発表) 計(7)件/うち招待講演 計(0)件/うち国際学会 計(5)件

発表者名		発表標題	
Akifumi Yoshimoto, Kazuo Hara, Masashi Shimbo, Yuji Matsumoto		Coordination-aware dependency parsing	
学会等名	発表年月日	発表場所	
The 14th International Conference on Parsing Technologies(国際学会)	2015年07月22日 ~ 2015年07月24日	Bilbao, the Basque Country	

発表者名		発表標題	
Joseph Irwin, Yuji Matsumoto		CKY Parsing with Independence Constraints	
学会等名	発表年月日	発表場所	
The 14th International Conference on Parsing Technologies(国際学会)	2015年07月22日 ~ 2015年07月24日	Bilbao, the Basque Country	

発表者名	発表標題	
Masayuki Komai, Hiroyuki Shindo, Yuji Matsumoto	An Efficient Annotation for Phrasal Verbs using Dependency Information	
学会等名	発表年月日	発表場所
The 29th Pacific Asia Conference on Language, Information and Computation(国際学会)	2015年10月30日 ~ 2015年11月01日	Shanghai, China

発表者名	発表標題	
Xiaoyi Wu, Yuji Matsumoto	An Improved Hierarchical Word Sequence Language Model Using Directional Information	
学会等名	発表年月日	発表場所
The 29th Pacific Asia Conference on Language, Information and Computation(国際学会)	2015年10月30日 ~ 2015年11月01日	Shanghai, China

発表者名	発表標題	
Xiaoyi Wu, Yuji Matsumoto, Kevin Duh, Hiroyuki Shindo	An Improved Hierarchical Word Sequence Language Model Using Word Association	
学会等名	発表年月日	発表場所
Statistical Language and Speech Processing(国際学会)	2015年11月24日 ~ 2015年11月26日	Budapest, Hungary

発表者名	発表標題	
吉本暁文, 新保仁, 原一夫, 松本裕治	並列構造解析に向けた依存構造解析アルゴリズムの拡張	
学会等名	発表年月日	発表場所
情報処理学会 第221回自然言語処理研究会 第106回音声言語情報処理研究会 共催研究会 Vol.2015-NL-221	2015年05月25日 ~ 2015年05月26日	東北大学(宮城県仙台市)

発表者名	発表標題	
加藤明彦, 進藤裕之, 松本裕治	複合機能語を考慮した英語の依存構造コーパスの構築	
学会等名	発表年月日	発表場所
情報処理学会 第223回自然言語処理研究会 Vol.2015-NL-223	2015年09月27日 ~ 2015年09月28日	広島経済大学(広島県広島市)

(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. 科研費を使用して開催した国際研究集会

(国際研究集会) 計(0)件

国際研究集会名	開催年月日	開催場所

16.本研究に関連して実施した国際共同研究の実施状況

(1)国際共同研究: -

17.備考

--