

平成28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26年度～平成28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
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
0 0 7 8 2 5 4 1	ノウジ ヒロシ 能地 宏	情報科学研究科	助教

9. 研究実績の概要

自然言語の統語解析の高性能化を目指し、特に英語の複単語表現と複文パターンの収集とコーパスへのアノテーション、および、英語の並列構造解析のためのリソース構築と解析手法の開発に関する研究を行った。

英語の複単語表現については、これまでの機能表現に加えて固有表現や動詞句に関する表現を収集し、Penn Treebankへの半自動アノテーションによるコーパス構築と、複単語表現を考慮した依存構造解析手法に関する研究を行い、複単語表現の出現を考慮することで依存構造解析の高精度化を達成する手法を提案した。また、修飾語を内部に含むような柔軟な構造をもった複単語表現の網羅的な収集を行い、Penn Treebank上の出現を自動的に同定する手法を提案した。

英語の並列構造解析については、昨年度は従来の依存構造解析アルゴリズムを拡張し、並列構造の範囲を動的に考慮しながら依存構造解析を同時に行う手法を提案したが、本年度は、並列構造の範囲同定を行う手法の高性能化について検討し、系列の類似性を測るニューラルネットワークモデルの設計と実装を行った。

上記の研究成果を統合し、複単語表現と並列構造を同時に考慮しながら英文の解析を行うシステムの構築まで達成することができなかったが、引き続き統合的な解析システムとアノテーションを施したデータの構築を行っていく予定である。

なお、本研究で構築した複単語表現アノテーションコーパス（Ontonotes corpusのPenn Treebank対応の部分に複単語表現のアノテーションとそれを考慮した依存構造木）をLinguistic Data Consortiumを通じて公開した。

10. キーワード

- | | | | |
|------------|----------|-------------|----------|
| (1) 自然言語処理 | (2) 統語解析 | (3) 並列構造 | (4) 複文構造 |
| (5) 依存構造解析 | (6) 機械学習 | (7) アノテーション | (8) コーパス |

11. 現在までの進捗状況

(区分)

(理由)

28年度が最終年度であるため、記入しない。

12. 今後の研究の推進方策

(今後の推進方策)

28年度が最終年度であるため、記入しない。

13. 研究発表 (平成 28 年度の研究成果)

〔雑誌論文〕 計 (1) 件 / うち査読付論文 計 (1) 件 / うち国際共著論文 計 (1) 件 / うちオープンアクセス 計 (0) 件

著 者 名		論 文 標 題				
Hiroki Ouchi, Kevin Duh, Hiroyuki Shindo, and Yuji Matsumoto		Transition-Based Dependency Parsing Exploiting Supertags				
雑 誌 名	査読の有無	巻	発行年	最初と最後の頁	国際共著	
IEEE/ACM Transactions on Audio, Speech, and Language Processing	有	24	2016	2059-2068	該当する	
掲載論文のDOI (デジタルオブジェクト識別子)						
10.1109/TASLP.2016.2598310						
オープンアクセス						
オープンアクセスではない、又はオープンアクセスが困難						

〔学会発表〕 計 (8) 件 / うち招待講演 計 (1) 件 / うち国際学会 計 (4) 件

発 表 者 名		発 表 標 題	
Akihiko Kato, Hiroyuki Shindo, and Yuji Matsumoto		Construction of an English Dependency Corpus incorporating Compound Function Words	
学 会 等 名	発表年月日	発 表 場 所	
the Tenth International Conference on Language Resources and Evaluation (国際学会)	2016年05月26日	Portoroz, Slovenia	

発 表 者 名		発 表 標 題	
Yuji Matsumoto		Parsing Complex Linguistic Constructions	
学 会 等 名	発表年月日	発 表 場 所	
The Eighth International Conference on Knowledge and Systems Engineering (招待講演) (国際学会)	2016年10月07日	Hanoi, Vietnam	

発表者名	発表標 題	
Xiaoyi Wu, Kevin Duh, and Yuji Matsumoto	A Generalized Framework for Hierarchical Word Sequence Language Model	
学 会 等 名	発表年月日	発 表 場 所
The 30th Pacific Asia Conference on Language, Information and Computation (国際学会)	2016年10月29日	Seoul, Korea

発表者名	発表標 題	
Ayaka Morimoto, Akifumi Yoshimoto, Akihiko Kato, Hiroyuki Shindo, and Yuji Matsumoto	Identification of Flexible Multiword Expressions with the Help of Dependency Structure Annotation	
学 会 等 名	発表年月日	発 表 場 所
the Workshop on Grammar and Lexicon: interactions and interfaces (国際学会)	2016年12月11日	Osaka, Japan

発表者名	発表標 題	
加藤明彦・進藤裕之・松本裕治	複単語表現を考慮した英語の依存構造解析モデリング	
学 会 等 名	発表年月日	発 表 場 所
情報処理学会 第229回自然言語処理研究会	2016年12月22日	東京都武蔵野市

発表者名	発表標 題	
加藤明彦, 進藤裕之, 松本裕治	固有表現と複合機能語を考慮したMWEベースの依存構造コーパス構築と解析	
学 会 等 名	発表年月日	発 表 場 所
言語処理学会第23回年次大会	2017年03月14日	茨城県つくば市

発 表 者 名	発 表 標 題	
久保大輝, 田中貴秋, 進藤裕之, 松本裕治, 永田昌明	日本語Universal Dependenciesへの複合辞情報付加の試み	
学 会 等 名	発 表 年 月 日	発 表 場 所
言語処理学会第23回年次大会	2017年03月14日	茨城県つくば市

発 表 者 名	発 表 標 題	
森元彩華, 吉本暁文, 加藤明彦, 進藤裕之, 松本裕治	依存構造情報を用いた柔軟な複単語表現の同定	
学 会 等 名	発 表 年 月 日	発 表 場 所
言語処理学会第23回年次大会	2017年03月16日	茨城県つくば市

〔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

--