

様 式 F - 7 - 1

科学研究費助成事業（学術研究助成基金助成金）実施状況報告書（研究実施状況報告書）（平成 28 年度）

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 若手研究(B) 4. 補助事業期間 平成 27 年度～平成 29 年度

5. 課題番号

1	5	K	1	6	0	5	3
---	---	---	---	---	---	---	---

6. 研究課題名 複合表現を考慮した構文解析手法に関する研究

7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
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

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. 研究実績の概要

平成 28 年度は、27 年度に引き続き複単語表現のコーパス作成と、複単語表現の解析手法に関する研究を行った。まず、コーパス作成に関しては、英語だけでなく日本語の機能的複単語表現の辞書収集とコーパスアノテーションについて検討を行った。英語に関しては、これまでに機能表現となる複単語表現を中心にコーパス構築を行ってきたが、より網羅性を高めるために、動詞および形容詞となりうる複単語表現の辞書収集およびコーパスアノテーションを実施する準備を進めている。更に、これまで無視されてきた、“a great number of ~”のような、複単語表現の中に修飾語（この場合、“great”）が入り込む場合について、コーパスから事例を収集してアノテーションを行う準備を進めている。このような非連続となる複単語表現は、コーパス作成に多大なコストがかかるため、アノテーションを全て人手で行うのではなく、計算機を利用して半自動的にコーパスアノテーションを行う等の工夫が必要になると考えられる。

次に、複単語表現の解析手法に関しては、これまでのように文を入力として複単語表現の位置を同定するだけでなく、複単語表現の同定と構文解析とを同時に行う手法を提案し、それぞれ独立に実施するよりも高精度であることを実証した。この実験には、昨年度までに我々が構築した英語の複単語表現のコーパスを用いている。今後は、上記で述べた非連続パターンとなる複単語表現を許容するモデルを提案し、なるべく少ない計算量で動作するアルゴリズムを開発する必要がある。

10. キーワード

(1) 複単語表現	(2) 構文解析	(3) 言語コーパス	(4)
(5)	(6)	(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)

平成28年度に実施予定であった、コーパス作成と、複単語表現の同定と構文解析を同時に行う手法の提案について、どちらも計画通りに研究が進捗している。コーパス作成に関しては、さらに網羅性を高めるため、動詞や形容詞および非連続パターンに拡大していく準備を進めている。複単語表現の解析手法に関しては、当初期待した通り、構文解析と同時に実行することで精度が向上することを実証することができた。

12. 今後の研究の推進方策 等

(今後の推進方策)

今後は、前年度に引き続き複単語表現のコーパス・辞書作成を中心に進めていく。また、作成したコーパスや辞書は、LDCなどの言語リソースを配布する組織から世界へ公開し、それが広く使われるように整備していく必要があるため、論文だけでなく、成果物を今後も利用される言語リソースにしていくための作業も行っていく予定である。

(次年度使用額が生じた理由と使用計画)

(理由)

(使用計画)

(課題番号: 15K16053)

(注)・印刷に当たっては、A4判(縦長)・両面印刷すること。

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 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 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 (5) 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	
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	
COLING 2016 Workshop: Grammar and Lexicon: Interactions and Interfaces (国際学会)	2016年12月11日 ~ 2016年12月12日	Osaka (Japan)	

発 表 者 名		発 表 標 題	
Taishi Ikeda, Hiroyuki Shindo and Yuji Matsumoto		Japanese Text Normalization with Encoder-Decoder Model	
学 会 等 名	発表年月日	発 表 場 所	
2nd Workshop on Noisy User-generated Text (W-NUT) (国際学会)	2016年12月11日 ~ 2016年12月12日	Osaka (Japan)	

発表者名	発表標題	
Masashi Yoshikawa, Hiroyuki Shindo and Yuji Matsumoto	Joint Transition-based Dependency Parsing and Disfluency Detection for Automatic Speech Recognition Texts	
学会等名	発表年月日	発表場所
Conference on Empirical Methods on Natural Language Processing (国際学会)	2016年11月01日 ~ 2016年11月05日	Austin (USA)

発表者名	発表標題	
Duc Anh Phan, Hiroyuki Shindo and Yuji Matsumoto	Multiple Emotions Detection in Conversation Transcripts	
学会等名	発表年月日	発表場所
Pacific Asia Conference on Language, Information and Computation (国際学会)	2016年10月28日 ~ 2016年10月30日	Seoul (Korea)

発表者名	発表標題	
Ikuya Yamada, Hiroyuki Shindo, Hideaki Takeda and Yoshiyasu Takefuji	Joint Learning of the Embedding of Words and Entities for Named Entity Disambiguation	
学会等名	発表年月日	発表場所
The SIGLL Conference on Computational Natural Language Learning (国際学会)	2016年08月11日 ~ 2016年08月12日	Berlin (Germany)

〔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

--