

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 27 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 27 年度は、複単語表現のコーパス作成と、複単語表現を漏れなく同定する手法の研究を行った。コーパス作成では、Wiktionary という知識データベースから網羅的に複単語表現（2 語以上の辞書項目）を抽出し、英語の OntoNotes コーパス（新聞記事データ）に対して正例か負例かのアノテーションを行った。複単語表現は様々な種類があり、Wiktionary で取得できるものは、主に句動詞や機能表現である。一方、Wiktionary に不足している複単語表現としては、not only ... but also ... のような文全体にまたがるパターンや、複合名詞などであることがわかった。これらの複単語表現に関しては、今後データマイニング手法を採用して大規模なテキストコーパスから半自動的に獲得し、複単語表現の辞書へ取り入れる必要がある。複単語表現を漏れなく同定する手法に関しては、畳み込みニューラルネットワークを用いて品詞のタグ付けと複単語表現の同定を同時に行う手法を提案し、既存手法やルールベースの手法に対して精度面で優位であることを実証した。従来とは異なり、本手法では文字や単語に関わる特徴量を明示的に定義する必要がなく、ニューラルネットワークによって特徴量を自動的に獲得できるため、実装も容易で精度も高いことがわかった。今後は、本手法と構文解析手法とを組み合わせることによって、最終的に文から複単語表現および構文木を予測するモデルを構築する必要がある。

## 10. キーワード

(1) 複単語表現	(2) 構文解析	(3)	(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)

当初計画していた二つの項目(コーパス作成と同定手法の考案)について、ほぼ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、一部の複単語表現に関してアノテーションが完了した。複単語表現の同定手法に関しては、ニューラルネットワークを用いた新たな手法を提案し、既存手法に対する優位性を示すことができた。

## 12. 今後の研究の推進方策 等

(今後の推進方策)

今度は、前年度に引き続き複単語表現の辞書作成およびコーパスの構築を進めていく。特に、Wiktionaryでは網羅できない種類の複単語表現に関して、どのようなデータベースまたはコーパスから辞書項目を収集するかが大きな課題である。また、複合語の同定と構文解析を統合して同時に実行できるモデルや手法の開発を推進していく予定である。これは、完全に同時に行う、細かいモデルに分けて最後に統合するなどのいくつかバリエーションが考えられるため、理論的、経験的の両側面から問題に合った方法を考案し評価実験を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 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(6)件/うち招待講演 計(0)件/うち国際学会 計(6)件

発表者名		発表標題	
Hiroki Ouchi, Hiroyuki Shindo, Kevin Duh and Yuji Matsumoto		Joint Case Argument Identification for Japanese Predicate Argument Structure Analysis	
学会等名	発表年月日	発表場所	
Association for Computational Linguistics(国際学会)	2015年07月26日 ~ 2015年07月31日	北京, 中国	

発表者名		発表標題	
Yuichiro Sawai, Hiroyuki Shindo, and Yuji Matsumoto		Semantic Structure Analysis of Noun Phrases using Abstract Meaning Representation	
学会等名	発表年月日	発表場所	
Association for Computational Linguistics(国際学会)	2015年07月26日 ~ 2015年07月31日	北京, 中国	

発表者名	発表標題	
The Tung Nguyen, Graham Neubig, Hiroyuki Shindo, Sakriani Sakti, Tomoki Toda and Satoshi Nakamura	A Latent Variable Model for Joint Pause Prediction and Dependency Parsing	
学会等名	発表年月日	発表場所
Interspeech(国際学会)	2015年09月06日 ~ 2015年09月10日	ドレスデン, ドイツ

発表者名	発表標題	
Masayuki Komai, Hiroyuki Shindo and Yuji Matsumoto	An Efficient Annotation for Phrasal Verbs using Dependency Information	
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)	2015年10月30日 ~ 2015年11月01日	上海, 中国

発表者名	発表標題	
Xiaoyi Wu, Yuji Matsumoto, Kevin Duh and 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
International Conference on Statistical Language and Speech Processing (国際学会)	2015年11月24日 ~ 2015年11月26日	ブタペスト, ハンガリー

発表者名	発表標題 【発表確定】	
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
Language Resources and Evaluation Conference (国際学会)	2016年05月23日 ~ 2016年05月28日	ポルトロス, スロベニア

(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

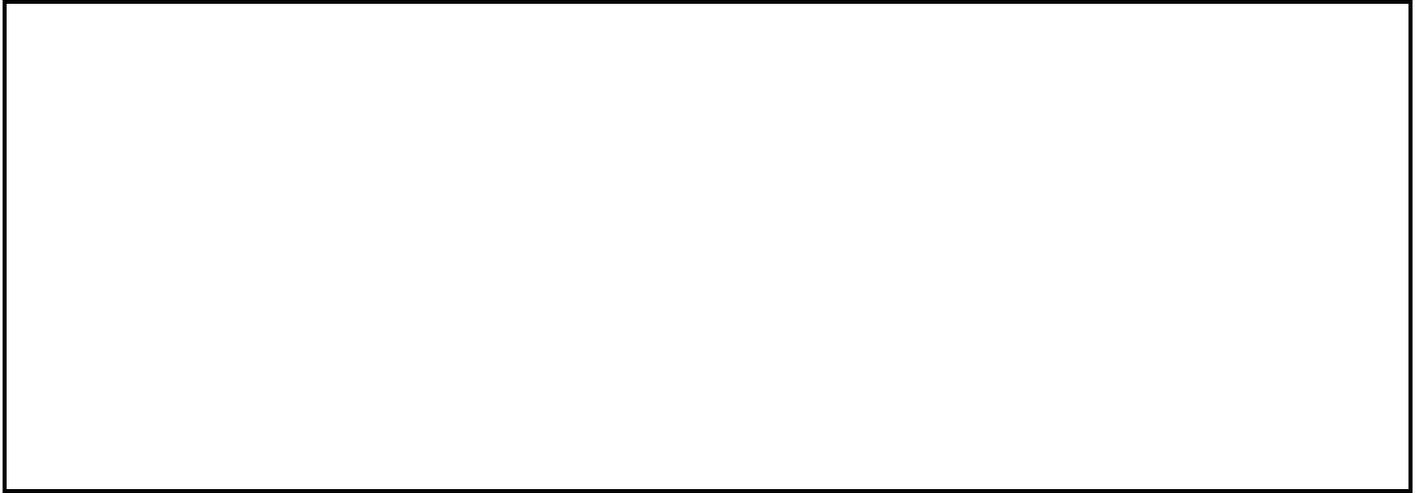
(国際研究集会) 計(0)件

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

A large empty rectangular box with a black border, intended for handwritten notes or additional information.