

平成21年度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(C)(一般) 4. 研究期間 平成21年度～平成23年度

5. 課題番号

2	1	5	0	0	1	4	1
---	---	---	---	---	---	---	---

6. 研究課題名 リンク解析に基づく自然言語処理・テキストマイニング技術の開発

7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
90311589	フリガナ シンボ マサシ 新保 仁	情報科学研究科	助教

8. 研究分担者(所属研究機関名については、研究代表者の所属研究機関と異なる場合のみ記入すること。)

研究者番号	研究分担者名	所属研究機関名・部局名	職名
	フリガナ		
	フリガナ		
	フリガナ		
	フリガナ		
	フリガナ		

9. 研究実績の概要

下欄には、当該年度に実施した研究の成果について、その具体的内容、意義、重要性等を、交付申請書に記載した「研究の目的」、「研究実施計画」に照らし、600字～800字で、できるだけ分かりやすく記述すること。また、国立情報学研究所でデータベース化するため、図、グラフ等は記載しないこと。

平成21年度は、自然言語文の構文解析タスクの一つである並列句解析のうち、日本語並列句および、ネストした並列句の解析、という、より困難な問題に取り組み、これらの問題におけるアラインメントグラフ上のコスト学習法の実用性を示したことが主な成果である。

具体的には、アラインメントグラフを機械学習法的一种である系列学習パーセプトロンと組み合わせ、日本語文においても並列句同士の類似度が適切に学習できることを実証した。英文の並列句がほぼ間違いなく「and」「or」といった並列句マーカーを伴うのに対し、日本語の並列句は「と」といったマーカーは必ずしも並列句を導くとは限らない。例として、「清水寺と二条城に行った」、「友達と二条城に行った」はいずれも助詞「と」を含むが、前者は並列句を含むが、後者は含まない。このように日本語並列句解析は、「並列句の範囲同定」に加え「並列句が存在するか否かの判定」という処理も必要とされ、問題はより困難である。我々はこのような問題に対しても、グラフ形状を工夫する(バイパスと呼ばれる経路を追加する)ことで対処した。

また、並列句はしばしばネストして文中に出現するが、従来のアラインメントグラフを用いる並列句解析法ではこのような文を扱うことが不可能であった。我々は、このような文についても、アラインメントグラフとともに構文規則に基づく制約を併用し、効率良く学習・解析が可能であることを示した。

10. キーワード

(1) リンク解析	(2) 自然言語処理	(3) テキストマイニング
(4)	(5)	(6)
(7)	(8)	(裏面に続く)

11.研究発表（平成21年度の研究成果）

〔雑誌論文〕 計（ 2 ）件 うち査読付論文 計（ 2 ）件

著 者 名	論 文 標 題				
大熊秀治, 原一夫, 新保仁, 松本裕治	バイパス付きアラインメントグラフを用いた日本語並列句検出と範囲同定				
雑 誌 名	査読の有無	巻	発 行 年	最初と最後の頁	
人工知能学会論文誌	有	25	2009	206-214	

著 者 名	論 文 標 題				
小町守, 工藤拓, 新保仁, 松本裕治	Espresso 型ブートストラッピング法における意味ドリフトのグラフ理論的分析				
雑 誌 名	査読の有無	巻	発 行 年	最初と最後の頁	
人工知能学会論文誌	有	25	2010	233-242	

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	

〔学会発表〕 計（ 3 ）件 うち招待講演 計（ 0 ）件

発 表 者 名	発 表 標 題		
Kazuo Hara, Masashi Shimbo, Hideharu Okuma, and Yuji Matsumoto	Coordinate Structure Analysis with Global Structural Constraints and Alignment-Based Local Features		
学 会 等 名	発表年月日	発 表 場 所	
Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2009)	2009年8月5日	シンガポール	

発 表 者 名	発 表 標 題		
Hideharu Okuma, Kazuo Hara, Masashi Shimbo, and Yuji Matsumoto	Bypassed Alignment Graph for Learning Coordination in Japanese Sentences		
学 会 等 名	発表年月日	発 表 場 所	
Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2009)	2009年8月4日	シンガポール	

発 表 者 名	発 表 標 題		
Ikumi Suzuki, Kazuo Hara, Masashi Shimbo, and Yuji Matsumoto	A Graph-based Approach for Biomedical Thesaurus Expansion		
学 会 等 名	発表年月日	発 表 場 所	
Third ACM International Workshop on Data and Text Mining in Bioinformatics (DTMBIO)	2009年11月6日	香港	

〔図 書〕 計（ 0 ）件

著 者 名	出 版 社		
書 名	発 行 年	総ページ数	

12. 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

13. 備考

※ 研究者又は所属研究機関が作成した研究内容又は研究成果に関するw e b ページがある場合は、U R Lを記載すること。

--