

平成22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 若手研究(B) 4. 研究期間 平成21年度～平成23年度
5. 課題番号

2	1	7	0	0	1	7	0
---	---	---	---	---	---	---	---
6. 研究課題名 系列アラインメントに基づく並列構造解析と統語解析の統合に関する研究
7. 研究代表者

研究者番号							研究代表者名		所属部局名			職名	
3	0	4	6	7	6	9	1	ハラ 原	カズオ 一夫	情報科学研究科			研究員

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

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。

自然言語処理研究において、並列構造解析は困難な課題の一つであり、既存の優れた句構造解析器を使用しても並列構造を含む文の解析誤りが生じやすい。この誤りは、医学/生物学分野の学術論文テキストを対象にすると特に頻出する。というのは、そこで主に記述されるのは生命科学実験の問題設定および実験結果であり、これらは並列構造を用いて記述されやすいからである（典型的には、新規療法/仮説に基づく手法と既存療法/コントロール手法との対比）。

我々はすでに、英語文を対象に、系列アラインメントを応用した手法を用い、並列構造構成要素間の類似性を測ることで、並列構造を構成する単語系列の範囲を同定する機械学習による手法（識別モデル）を開発し、既存解析器を上回る精度を得ることに成功した。しかし、そこで開発した方法は、複数の並列構造が文に含まれる場合に対応できないため、実用的ではない。

この問題を解決するために、今年度の研究では、複数の並列構造が成す階層構造を表現できる（並列構造解析に特化した）文法を用いることとし、その上で、構文解析で用いられるCKYアルゴリズムを応用して、それら並列構造を構成する単語系列の範囲と階層構造を同時に決定する方法を提案した。

10. キーワード

(1) 自然言語処理	(2) 並列句解析	(3) アラインメント
(4)	(5)	(6)
(7)	(8)	(裏面に続く)

11.研究発表（平成22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				
原 一夫, 新保 仁, 松本 裕治	文法制約と系列アラインメントによる並列構造の解析				
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	569-578	

著 者 名	論 文 標 題				
Javier Tejada-Cárcamo, Hiram Calvo, Alexander Gelbukh, and Kazuo Hara	Unsupervised WSD by finding the predominant sense using context as a dynamic thesaurus				
雑 誌 名	査読の有無	巻	発 行 年	最初と最後の頁	
Journal of Computer Science and Technology	有	25	2010	1030-1039	

〔学会発表〕 計（ 4 ）件 うち招待講演 計（ 0 ）件

発 表 者 名	発 表 標 題		
原一夫, 新保仁, 松本裕治	バイオ医療テキストに対する並列構造解析と固有表現抽出の統合解析		
学 会 等 名	発表年月日	発 表 場 所	
第33回日本分子生物学会年会・第83回日本生化学会大会合同大会(BMB2010)	2010年12月10日	神戸ポートアイランド	

発 表 者 名	発 表 標 題		
鈴木郁美, 原一夫, 新保仁, 松本裕治	バイオ医療専門用語のシソーラス拡張のための分布類似度計算法の提案		
学 会 等 名	発表年月日	発 表 場 所	
第33回日本分子生物学会年会・第83回日本生化学会大会合同大会(BMB2010)	2010年12月10日	神戸ポートアイランド	

発 表 者 名	発 表 標 題		
鈴木 郁美, 原 一夫, 新保 仁, 松本 裕治	係り受け木を利用した単語類似度計算方法とそのシソーラス拡張への応用		
学 会 等 名	発表年月日	発 表 場 所	
情報処理学会研究報告, 自然言語処理研究会, 2009-NL-199	2010年11月18日	広島市立大学	

発 表 者 名	発 表 標 題		
Kazuo Hara	Towards automatic biomedical entity annotation by reducing error propagation		
学 会 等 名	発表年月日	発 表 場 所	
The 1st CALBC workshop	2010年6月18日	European Bioinformatics Institute, Hinxton, Cambridge, U.K	

〔図 書〕 計（ 0 ）件

著 者 名	出 版 社		
書 名			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。

--