

平成19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 14603 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 特定領域研究 4. 研究期間 平成18年度～平成22年度
5. 課題番号 18061005
6. 研究課題名 書き言葉コーパスの自動アノテーションの研究

7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
10211575	松本, 裕治	情報科学研究科	教授

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
20197875	徳永, 健伸	東京工業大学・大学院情報理工学研究科	教授
60272689	乾, 健太郎	情報科学研究科	准教授
00357766	橋田, 浩一	独立行政法人産業技術研究所・情報技術研究部門	副研究部門長
10345382	橋本, 泰一	東京工業大学・統合研究院ソリューション研究機構イノベーションシステム研究センター	特任准教授
80379528	浅原, 正幸	情報科学研究科	助教

9. 研究実績の概要(国立情報学研究所でデータベース化するため、600字～800字で記入。図、グラフ等は記載しないこと。)

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。

コーパスに対する自動アノテーションツールの開発：係り受け解析において最適の係り先を比較によって決定する手法の提案，述語項構造解析のための手法の事態名詞への拡張を行い，システムの開発の過程で作成した共参照および項構造情報が付与されたコーパスを公開した。

コーパス管理ツールの開発：形態素・係り受け解析コーパス管理ツールに対していくつかの機能拡張を行った。特に、係り受け構造の表示機能の充実、および、高頻度の連続・非連続のパターンのマイニング機能などを開発した。汎用アノテーションツールについてデータ構造の設計の詳細化といくつかの機能拡張を行った。さらに、Webベースのコーパス検索ツールの開発を行った。

大規模な固有表現辞書の構築：Wikipediaの構造情報を利用し、自動的に固有表現を抽出する手法を提案し、実際に大規模な固有表現抽出を行った。抽出した大規模固有表現辞書とコーパスからの統計的手法を組み合わせることにより、精度向上を図った。白書コアデータに対する固有表現タグ付け作業を行った。

談話構造アノテーションツールの開発：文関係，共参照，項構造を記述するためセマンティックエディタの拡張を行った。一般化された木構造表示ユーザインタフェースの実装を行い、談話構造や述語に対する統語的構成素の記述や文章の埋め込み構造、意味構造を編集する機能をこのユーザインタフェースにプラグインする形で実装した。

※ 成果の公表を見合わせる必要がある場合は、その理由及び差し控え期間等を記入した調書(A4 判縦長横書 1 枚)を添付すること。

10. キーワード

- | | | |
|-------------|-----------|----------|
| (1) アノテーション | (2) 形態素解析 | (3) 統語解析 |
| (4) 述語項構造解析 | (5) コーパス | (6) 照応解析 |
| (7) 談話解析 | (8) 言語解析 | (裏面に続く) |

11. 研究発表（平成19年度の研究成果）

〔雑誌論文〕 計（ 3 ）件

著者名	論文標題			
Ryu Iida, Kentaro Inui and Yuji Matsumoto	Zero-anaphora resolution by learning rich syntactic pattern features			
雑誌名	査読の有無	巻	発行年	最初と最後の頁
CM Transactions on Asian Language Information Processing	有	6	2007	1-12

著者名	論文標題			
橋本泰一, 吉田恭祐, 野口正樹, 徳永健伸, 田中穂積	関係データベースを用いた構文木付きコーパス検索手法			
雑誌名	査読の有無	巻	発行年	最初と最後の頁
自然言語処理	有	14	2007	3-22

著者名	論文標題			
Mamoru Komachi, Ryu Iida, Kentaro Inui and Yuji Matsumoto	Learning Based Argument Structure Analysis of Event-nouns in Japanese			
雑誌名	査読の有無	巻	発行年	最初と最後の頁
Conference of the Pacific Association for Computational Linguistics	有	10	2007	120-128

〔学会発表〕 計（ 1 ）件

発表者名	発表標題		
Yuji Matsumoto	Corpus Annotation/Management Tools for the Project: Balanced Corpus of Contemporary Written Japanese		
学会等名	発表年月日	発表場所	
Third International Conference on Large-Scale Knowledge Resources	2008年3月3日	東京工業大学	

〔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webページがある場合は、URLを記載すること。

http://www.tokuteicorpus.jp/g_tool/index.php