

平成 22 年 4 月 1 日現在

研究種目：基盤研究 (C)

研究期間：2007～2010

課題番号：19520342

研究課題名 (和文)

大規模な電子資料の利用による日本語文法の未開拓の基礎的諸問題の原理・実証的考察

研究課題名 (英文)

Re-examination of fundamental problems of Japanese grammar based on large corpora

研究代表者

田野村 忠温 (TANOMURA TADAHARU)

大阪大学・大学院文学研究科・教授

研究者番号：40207204

研究代表者の専門分野：言語学・日本語学

科研費の分科・細目：言語学・言語学

キーワード：日本語文法、電子資料、コーパス

## 1. 研究計画の概要

本研究は、日本語の大規模な電子資料（コーパス）を用いて日本語文法の未開拓の基礎的な諸問題を考察することにより、(1) 電子資料に基づく日本語研究の基盤の形成と発展に寄与するとともに、(2) 電子資料利用の目的である日本語研究そのものに対して実質的な貢献をもたらすことを目的としている。

## 2. 研究の進捗状況

国立国語研究所で構築中の『現代日本語書き言葉均衡コーパス』、自作 Web コーパス、国会会議録データ、Yahoo!知恵袋ベータ版データといった、最近利用可能になった各種の電子資料を用いて、コンピュータ述語文や複合辞を始めとする日本語文法の基礎的な諸問題の考察を進めている。

(1) まず、電子資料に基づく日本語研究の基盤の形成と発展への寄与という目的については、下記のような取り組みを行った。

①日本語 KWIC 索引生成ソフトウェアの開発・公開

②巨大な Web コーパスの作成

③サーチエンジンの比較と評価

④Web 上で日本語の用例を自由に検索できるサイトの作成・公開

⑤入門書『コーパス日本語学ガイドブック』（共著）の刊行

(2) 次に、コーパスを用いた日本語研究そのものの発展という目的については、下記のような研究を行った。

①コンピュータの異形態の使用実態の詳細な調査・分析

②複合辞概念の再考

③個別の複合辞の用法の詳細な分析

④現代日本語文法の通時変化の分析

⑤日本語コーパスからのコロケーション情報の抽出手法の開発とコロケーション概念の再検討

## 3. 現在までの達成度

①当初の計画以上に進展している。

複数の種類の大規模な Web コーパスを作成し、それらを積極的に活用したことにより、文法事象の分析が当初の計画以上に促進された。

## 4. 今後の研究の推進方策

これまでの研究手法をいっそう発展させ、また、それを新たな文法事象に適用することにより、さらなる研究成果の獲得を目指す。

## 5. 代表的な研究成果

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 13 件)

①田野村忠温「サ変動詞の活用のゆれについて・続——大規模な電子資料の利用による分析の精密化——」『日本語科学』第 25 号、91～103 頁、2009 年、査読あり

②田野村忠温「複合辞の本性的について——その構成と単位性——」児玉一宏・小山哲春編『言葉と認知のメカニズム 山梨正明教授還暦記念論文集』、ひつじ書房、489～497 頁、2008 年、査読なし

〔学会発表〕（計2件）

① Tanomura, Tadaharu “A corpus-based analysis of some time-related aspects of contemporary Japanese,” *CLAVIER 09*, University of Modena and Reggio Emilia, 5 November 2009

② Tanomura, Tadaharu “Retrieving collocational information from Japanese corpora: An attempt towards the creation of a dictionary of collocations,” *QUALICO 2009*, Graz University, 18 September 2009

〔図書〕（計1件）

① 田野村忠温・服部匡・杉本武・石井正彦『コーパス日本語学ガイドブック』（特定領域研究「日本語コーパス」日本語学班、199頁、2007年）

〔その他〕

① 自作ソフトウェア公開ページ  
<http://www.let.osaka-u.ac.jp/~tanomura/index.cgi?menu=resources>

② 日本語用例検索システム  
<http://www.tokuteicorpus.jp/team/jpling/kwic/>