

研究種目：特定領域研究

研究期間：2006～2010

課題番号：18061004

研究課題名（和文）

コーパスを用いた日本語研究の精密化と新しい研究領域・手法の開発

研究課題名（英文）

Corpus-linguistic study of Japanese: development of new research areas and methods

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

研究分野：人文学

科研費の分科・細目：言語学・日本語学

キーワード：日本語、日本語学、コーパス

1. 研究計画の概要

本研究課題は、今後の日本語研究にとって不可欠の存在となるはずのコーパスについて、具体的な事例研究を通してその利用価値を明らかにし、日本語の新しい研究領域・手法を開発するとともに、学界に対してコーパスを用いた日本語研究の啓蒙・普及を図ることを目的とする。

また、本特定領域研究において構築の進められている「現代日本語書き言葉均衡コーパス」を日本語研究に適用し、そこで得られた知見をコーパスの構築にフィードバックすることをも目的とする。

2. 研究の進捗状況

前項記載の本研究課題の目的とするところを3つに分け、その区分に従って研究の進捗状況を述べる。

(1) コーパスを用いた日本語研究の精密化と新しい研究領域・手法の開発

研究課題名そのものでもあるこの目的には、従来行われてきた種類の日本語研究の精密化と、新しい研究領域・手法の開発という2つの面がある。

前者、すなわち、日本語研究の精密化に関しては、コンピュータの異形態の分布、副詞と述語の共起関係、複合辞の意味・用法などのオーソドックスな日本語研究のテーマを、コーパスから得られるデータに基づいて分析している。

後者、すなわち、新しい研究領域・手法の開発には特に力を入れてさまざまな試みを行っている。まず、因子分析による副詞と述語の共起関係の分析、「国会会議録」データ

に基づく現代日本語の通時変化の分析、コロケーション情報の抽出法の開発、複合名詞の生産性の分析、コーパスの種類による言語事象の差異の比較、コーパスに基づく研究と内省から得られる情報との比較、新聞記事に見られる語彙使用の季節変動の分析などを行っている。

また、新聞コラムに基づく通時コーパス、画像付きコーパス、巨大なWebコーパスといった、従来日本語研究において試みられることのなかった性質のコーパスの試作とさまざまな視点からの試用を行っている。

さらに、Web上の言語資源へのアクセスに用いられるサーチエンジンの特性の比較・評価を行い、サーチエンジン利用に関わる重大な問題点も明らかにした。

(2) 学界に対するコーパス利用の啓蒙・普及

『コーパス日本語学ガイドブック』（CD-ROM付き）を刊行し、希望する日本語研究者への無料配布を行っている。同書は、コーパスの利用法の紹介、KWICソフト、正規表現の解説、関連研究文献目録などを収めており、コーパス初心者が日本語研究にコーパスを生かすための手引きとなることを目指したものである。

また、青空文庫に収められた約3,500の文学作品から日本語表現の用例を自由に検索してその結果をKWIC形式で表示し、あるいは、エクセルファイルとしても保存することのできる日本語用例検索システムを開発し、Web上で公開した。そのURLは次の通りである。

<http://www.tokuteicorpus.jp/team/jpling/kwic/>
この日本語用例検索システムは、すでに国内

外の多数の利用者によって日々頻繁に利用されている。

(3) コーパス構築へのフィードバック

本領域で構築が進められている「現代日本語書き言葉均衡コーパス」の未完成版が領域内で公開されるごとに試用・検証を行い、コーパスの内容やあり方に関するさまざまな指摘・提言を行い、コーパスの構築に貢献している。

また、開発中の関連ツール類についても、新しいバージョンが公開されるごとに試用のうえ問題点の指摘や改善の提案を行っている。

3. 現在までの達成度

①当初の計画以上に進展している。

(理由)

未開拓の研究領域ということを手探りの状態で研究を開始したが、さまざまな研究の試行を重ねる中で、前項に具体的に記した通り、これまでにすでに広範な研究成果を収めることができた。

4. 今後の研究の推進方策

現在の研究活動をさらに進展・拡大させる。予定している新規の研究内容は多岐にわたるが、そのうち研究計画全体として取り組む予定の重要なものとしては、例えばコーパスのジャンルの特性に関する研究がある。また、従来は考えられなかった巨大な規模のWebコーパスの作成と利用の試みに最近着手したところである。

また、本領域で構築中の「現代日本語書き言葉均衡コーパス」を研究に適用し、他の資料との併用・比較を通して当コーパスの特性と日本語研究上の価値を明らかにする。

最終年度には、コーパスを用いた日本語研究の体系化、組織化を図り、新たな「コーパス日本語学」の可能性と方法論を提示したいと考えている。

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25件)

- ①田野村忠温「サ変動詞の活用のゆれについて・統一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、頁数未定、査読あり、2009年刊行予定)
- ②杉本武「コーパスからみた類義語動詞:『ねじる』と『ひねる』」(『文藝・言語研究 言語篇』55、筑波大学、109～122頁、査読あり、2009年)
- ③服部匡「因子分析を用いた程度副詞と述語等の共起関係分析の試み―新聞コーパス

のデータから―」(『同志社女子大学総合文化研究所紀要』第24号、98～109頁、査読なし、2007年)

- ④石井正彦「日本語研究における探索的データ解析の有用性」(『日本語の教育から研究へ』、くろしお出版、227～237頁、査読なし、2006年)

[学会発表] (計2件)

- ①石井正彦「外来語の20世紀」(日本語学会、関西大学、2007年5月26日)

[図書] (計2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本研究課題のWebサイト
<http://www.tokuteicorpus.jp/team/jpling/>
- ②日本語用例検索システム
<http://www.tokuteicorpus.jp/team/jpling/kwic/>