

機関番号：14401

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) Study 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・文学研究科・教授

研究者番号：40207204

研究成果の概要(和文)：

日本語研究におけるコーパスの利用は他の主要言語の研究に比べて大きな遅れを取ってきたが、昨今さまざまな要因により状況は変わりつつある。

本研究では、日本語研究におけるコーパスの利用の可能性を多角的に追求した。具体的には、独自の特徴のあるいくつかのコーパスを自作し、日本語の文法・形態・意味に関わる諸問題を分析した。また、研究成果の学界・一般社会への還元として、コーパス関連ソフトウェアの開発・公開や日本語用例検索サイトの製作・公開を行った。

研究成果の概要(英文)：

Japanese has been lagging behind the other major languages of the world in the utilization of electronic corpora in linguistic studies, but recently the situation is changing due to several factors.

The major accomplishments of this research project include: 1) analysis of various phenomena of Japanese grammar, morphology and semantics using corpora, 2) construction of a few corpus including a huge Web corpus, 3) developments and publication of corpus tools and a Web site for searching Japanese expressions.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	1,000,000	300,000	1,300,000
2008年度	800,000	240,000	1,040,000
2009年度	800,000	240,000	1,040,000
2010年度	800,000	240,000	1,040,000
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：人文学

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. 研究開始当初の背景

日本語研究におけるコーパス、すなわち、電子媒体の言語研究資料の利用は、英語を始めとする他の主要言語の研究に比べて大きな遅れを取ってきた。しかし、現在国立国語

研究所で進められている『現代日本語書き言葉均衡コーパス』の構築を始めとするさまざまな要因により、この数年間のあいだに状況は変化しつつある。

2. 研究の目的

本研究は、応募者が従来長年にわたって継続的に行ってきた研究をさらに発展させる形で、日本語研究におけるコーパス（電子媒体の言語資料）の高度な利用の可能性をさまざまな角度から追求し、それを通じて日本語研究の新たな領域・手法を開拓することを主たる目的とした。

コーパスは言語研究の質を高めその領域を拡大する大きな潜在的可能性を秘めているが、コーパス利用の歴史の浅い日本語研究においてはどのようにすればその潜在的可能性を引き出せるかがまだ十分に分かっていない。日本語のコーパスの利用環境が整いつつある今、その有効利用の方途を明らかにすることは今後の日本語研究の発展にとって喫緊の課題である。

3. 研究の方法

従来の日本語研究におけるコーパスの一般的な利用法はそれを単に効率的な用例収集の手段とするという素朴なものであった。しかし、それはコーパスの有益な用途の1つではあるが、すべてではない。大量性と処理の柔軟性という著しい2つの著しい特性を併せ持つコーパスという新しい種類の言語研究資料を活用することにより、従来は望むこともできなかったような日本語研究へのさまざまな道が拓かれる。

本研究では、そうした認識に基づき、大規模なコーパスの有効な利用法を見出すことによって日本語研究の新たな地平を切り拓くと同時に、得られた知見の発表や関連ソフトウェアの開発・公開などにより、学界におけるコーパスの有効利用を普及促進させることを目指した。

4. 研究成果

コーパスを利用した日本語研究の可能性を追求すべく、多数の事例研究に取り組んだ。主な研究テーマとしては、日本語のコピュラ述語文の考察、いわゆる複合辞の本性の考察、現代日本語文法の形態や文法の通時変化の分析、コロケーションの概念の考察とコーパスからのコロケーション情報の抽出などに取り組んだ。

それらの研究においては、従来広く使われてきた種類の文学作品や新聞記事テキスト、現在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』、ヤフー株式会社より日本語研究者に対して提供されている『Yahoo!知恵袋ベータ版』データを利用するだけでなく、国立国会図書館のWebサイトから国会会議録のデータを取得・加工してコーパス化したものや、インターネット上の日本

語文書を大量に収集して作成した巨大な Web コーパス（単行本約 40 万冊相当の規模）など独自の特徴を有する大規模なコーパスも利用した。

また、日本語研究者が言語表現の使用実態の調査にインターネットのサーチエンジンを利用することがあるが、サーチエンジンの表示する検索ヒット件数には論理的整合性と時間的安定性という2つの点において深刻な問題がある。主要なサーチエンジンの比較分析を通してその様相を具体的に明らかにし、サーチエンジンへの安易な依存の危険性を指摘した。

研究成果の学界および一般社会への還元としては、日本語の任意の電子テキストから用例を検索して KWIC 形式で出力するソフトウェアを開発し、その後も改良を続けている（Web サイトで公開している）。それを Web 上に移植し、青空文庫所収の約 3,400 作品から用例を自由に検索して KWIC 形式で出力できるようにした日本語用例検索サイト（<http://www.tokuteicorpus.jp/team/jpling/kwic/>）は、国内外の日本語研究者および一般の人々によって日々頻繁に利用されている。

5. 主な発表論文等

（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16件）

- ① Tanomura, Tadaharu ‘A corpus-based analysis of some time-related aspects of contemporary Japanese,’ 2011刊行予定、頁数未定、査読有
- ② 田野村忠温「日本語コーパスとコロケーション—辞書記述への応用の可能性—」『言語研究』第138号、1～23頁、2010年、査読有
- ③ Tanomura, Tadaharu ‘Retrieving collocational information from Japanese corpora: Its methods and the notion of “circumcollocate”,’ Peter Grzybek, Emmerich Kelih and Ján Mačutek (eds.) *Text and Language: Structures · Functions · Interrelations*, Wien, Austria: Praesens Verlag, pp.213-222, 2010, 査読有
- ④ Tanomura, Tadaharu ‘The concept of “circumcollocate” and its significance for lexicography: A discussion with particular reference to the Japanese language,’ Isabel

Moskowich-Spiegel Fandiño, Begoña Crespo García, Inés Lareo Martín and Paula Lojo Sandino (eds.) *Language Windowing Through Corpora* (Conference proceedings in the electronic format), A Coruña, Spain: Universidade da Coruña, pp. 873-879, 2010, 査読有

- ⑤ 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、91～103頁、2009年、査読有
- ⑥ 田野村忠温 「日本語研究の観点からのサーチエンジンの評価・統一検索ヒット件数の時間変動のその後とWeb文書量の推計の修正—」『計量国語学』第26巻第8号、290～294頁、2009年、査読有
- ⑦ 田野村忠温 「コーパスからのコロケーション情報抽出—分析手法の検討とコロケーション辞典項目の試作—」『阪大日本語研究』21、21～41頁、大阪大学大学院文学研究科日本語学講座、2009年、査読無
- ⑧ 田野村忠温 「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」『人工知能学会誌』第24巻第5号、647～655頁、2009年、査読無
- ⑨ 田野村忠温 「コーパスと文法研究」『国文学 解釈と鑑賞』第74巻第1号、79～87頁、2009年、査読無
- ⑩ 田野村忠温 「日本語研究の観点からのサーチエンジンの比較評価—Yahoo!とGoogleの比較を中心に—」『計量国語学』第26巻第5号、147～157頁、2008年、査読有
- ⑪ 田野村忠温 「大規模な電子資料に見る現代日本語の動態」『待兼山論叢』第42号文化動態論篇、55～76頁、大阪大学大学院文学研究科、2008年、査読無
- ⑫ 田野村忠温 「複合辞の本性について—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、査読無
- [学会発表] (計39件)
- ① 田野村忠温 「電子資料と日本語研究—動向と展望—」コーパス日本語学セミナー、台湾大学 (台湾・台北市)、2010年12月4日
- ② 田野村忠温 「Webコーパスとコロケーション」日本語学会2010年度秋季大会ワークシ

ョップ「コーパス日本語学の新展開—コーパスと方法論の多様化—」、愛知大学、2010年10月23日

- ③ 田野村忠温 「コーパスと日本語文法研究」北京日本学研究中心創立25周年記念国際シンポジウム (パネルディスカッション「コーパスと日本語学及び日本語教育学」)、北京日本学研究中心 (中国・北京市)、2010年10月17日
- ④ 田野村忠温 「電子資料と日本語研究—動向と展望—」コーパス言語学ワークショップ、高麗大学校 (韓国・ソウル市)、2010年10月10日
- ⑤ Tanomura, Tadaharu 'The concept of "circumcollocate" and its significance for lexicography: A discussion with particular reference to the Japanese language,' *CILC10 (Congreso Internacional de Lingüística de Corpus)*, Asociación Española de Lingüística de Corpus, Universidade da Coruña (A Coruña, Spain), 14 May 2010.
- ⑥ 田野村忠温 「コーパスからのコロケーション情報の抽出・統一辞書記述への応用の可能性を前提とした考察—」国立国語研究所共同研究プロジェクト「コーパス日本語学の創成」研究発表会、国立国語研究所、2010年2月1日
- ⑦ Tanomura, Tadaharu 'A corpus-based analysis of some time-related aspects of contemporary Japanese,' *CLAVIER 09 (Corpus Linguistics and Language Variation)*, Università degli Studi di Modena e Reggio Emilia (Modena, Italy), 5 November 2009.
- ⑧ Tanomura, Tadaharu 'Retrieving collocational information from Japanese corpora: An attempt towards the creation of a dictionary of collocations,' *QUALICO 2009 (Quantitative Linguistics Conference)*, International Quantitative Linguistics Association, Universität Graz (Graz, Austria), 18 September 2009.
- ⑨ 田野村忠温 「コーパスと日本語研究—最近 (過去3年) の動向—」日本学総合講座、北京日本学研究中心 (中国・北京市)、2009年4月9日

[その他]

ホームページ等

・個人のホームページ

<http://www.let.osaka-u.ac.jp/~tanomura/>

・共同研究のホームページ

<http://www.tokuteicorpus.jp/team/jpling/>

6. 研究組織

(1) 研究代表者

田野村 忠温 (TANOMURA TADAHARU)

大阪大学・文学研究科・教授

研究者番号：40207204

(2) 研究分担者

なし

(3) 連携研究者

なし