

機関番号：14401  
 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 the Japanese Language: Towards the Development of New Research Areas and Methods  
 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）：コーパス、すなわち、電子媒体の言語資料を日本語研究に利用するさまざまな可能性に取り組み、その有効性を数々の事例研究に基づいて明らかにした。多岐にわたるその研究成果は国内外で論文、口頭発表、講演などの形で発表した。また、日本語研究におけるコーパス利用の普及のために、『コーパス日本語学ガイドブック』を刊行し、国内外でコーパス日本語研究に関するワークショップやセミナーを開催した。

研究成果の概要（英文）：We explored a variety of possibilities of applying electronic corpora to the study of the Japanese language, demonstrating a number of ways we may advance our understanding of the language with the use of corpora. With a view to promoting effective uses of Japanese corpora, we published “A Handbook of Japanese Corpus Linguistics”, as well as held several workshops and seminars on Japanese corpus linguistics.

#### 交付決定額

（金額単位：円）

|        | 直接経費       | 間接経費 | 合計         |
|--------|------------|------|------------|
| 2006年度 | 10,600,000 | 0    | 10,600,000 |
| 2007年度 | 8,800,000  | 0    | 8,800,000  |
| 2008年度 | 9,200,000  | 0    | 9,200,000  |
| 2009年度 | 7,100,000  | 0    | 7,100,000  |
| 2010年度 | 7,300,000  | 0    | 7,300,000  |
| 総計     | 43,000,000 | 0    | 43,000,000 |

研究分野：人文学

科研費の分科・細目：言語学・日本語学、情報学・知能情報学

キーワード：日本語、コーパス、日本語学

#### 1. 研究開始当初の背景

英語の研究においては、早くからコンピュータを利用した文法研究や辞書編纂が行われており、すでに膨大な実績の蓄積がある。英語以外の言語でも同様の努力が国家レベルで精力的に行われつつある。これに対して、日本語研究におけるコーパスの利用ははなはだしく立ち遅れ、コーパスの価値や可能性が大多数の研究者に十分に知られておらず、そのために研究者も少ないという状況にあった。本研究課題は、そうした状況を背景と

して、コーパス利用の意義を明らかにし、コーパスを利用した日本語研究法の普及、そしてそれによる日本語研究の精密化と深化・発展に貢献することを意図して計画された。

#### 2. 研究の目的

本研究課題は、コーパスを用いた日本語研究の深化発展とコーパス構築へのフィードバックという2つの課題を自らに課した。すなわち、第1に、今後の日本語研究にとって不可欠の存在となることが確実なコーパス

について、具体的な事例研究を通してその利用の価値を明らかにし、日本語の新しい研究領域・手法を開発するとともに、それにより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した。第2に、本特定領域研究の中心的な成果となる、我が国初でありかつ将来日本語研究の標準的資料として広範に利用されるであろう大規模な書き言葉コーパスの構築の進行に伴い、それを日本語研究に適用し、その過程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目的とした。

### 3. 研究の方法

研究内容は多岐にわたるが、本研究課題の中心的な目的であるコーパスを用いた日本語研究の深化発展という課題については、研究課題名にもある通り、主に次の2つの観点から研究に取り組んだ。

#### (1) 日本語研究の精密化

コーパスから得られる豊富な用例を利用することにより、従来少数の用例や内省に依存して行われてきた種類の日本語研究を精密化し、より実証的なものにすることができる。種々のコーパスを利用して、日本語の文法や意味に関わるさまざまな問題を題材としてその可能性を検討した。

#### (2) 新しい研究領域・手法

コーパス固有の特性を生かすことにより、旧来の言語研究資料では実現できなかった様々な種類の研究への道が拓かれる。新しく独自に作成したいくつかのコーパスを利用しつつ、さまざまな観点からその可能性を追求した。

### 4. 研究成果

本研究課題の主要な研究成果を、上記の(1)、(2)の別に従って記す。その範疇に入らない研究成果等については毎年度に刊行した研究成果報告書、特に最終報告書である『コーパス日本語学の新展開』を参照願いたい。

#### (1) 日本語研究の精密化

##### ・複合辞の研究

複合助詞には、「に関する」と「に関しての」のように連体用法において2通りの語形を持つもの、「に対して」と「に対し」、「ために」と「ため」のように連用用法において2通りの語形を持つものがある。このような、各用法で複数の語形を持つ複合助詞「によって」「において」「に関して」「に対して」「に際して」「ために」について、『現代日本語書き言葉均衡コーパス(BCCWJ)』等のコーパスにおいて、いずれの語形が用いられているか調査した。その結果、「によって」は、意味

用法によって異なりはあるものの、連体用法において圧倒的に「による」の形が用いられること、「に際して」は、他の複合助詞と異なり、連体用法において圧倒的に「に際しての」の形が用いられること、「に対して」は、接続助詞用法の場合「に対して」「に対し」のいずれも用いられるのに対して、格助詞用法の場合「に対して」が好まれること、「ために」は、目的用法の場合「ために」が、理由用法の場合「ため」が好まれることなどを明らかにした。さらに、このような語形の偏りは、複合助詞の品詞性（格助詞的か接続助詞的か）の違いによるものであることを示した。

日本語のコピュラの形式と分布の問題は日本語文法の根幹に位置する基礎的な問題でありながら、これまで研究の関心が向けられることすらなかった。2006年にこの問題に関する主として内省に基づく考察の結果を発表したが、コピュラ形式の分布はその一部にゆれの側面があり、内省ではその様相を正確に判断することができない。本研究では、数種類の大規模なコーパスを用いて、連体的な位置におけるコピュラ形式「な」と「の」の使い分けや、引用の「と」の直前におけるコピュラの潜在の様相について調査・分析を行った。

##### ・意味分析

接続助詞のように使われる「代わり（に）」は興味深い意味的二重性を示す。これは単なる意味上の問題ではなく、その背後には文法的な性格の違いがあると考えられるが、ともあれその意味的二重性はその区別が不明瞭なケースがあることもあって問題の本質が正確に捉えにくい。2008年に作成した巨大なWebコーパスを利用し、そこから得られる多様な用例の観察・分析を通して、「代わり」の意味的二重性について立ち入った考察を行った。

「移動先」、「留学先」、「投資先」などのように目標地点や相手方などを表す「漢語動名詞＋‘先’」の種類と用法について、特に新聞記事データベースの実例観察を行った。その結果、「動名詞＋‘先’」の中に、ある種の多義の解釈を許すものがあることが分かった。例えば、「受注先」は「受注者」を指すことと「被受注者（発注者）」を指すこととがあり、「引用先」は「引用主体（の文書）」を指すことと「被引用者（引用の出典）」を指すことがある。多様な「～先」の意味分析を行った上で、多義が生じるメカニズムの解明を行った。

#### (2) 新しい研究領域・手法の開発

##### ・現代日本語の通時変化

国会会議録は、1947年から今日に至る日本

語の話しことばの変化の様相をうかがい知るのに有効に使える膨大な資料である。これを国立国会図書館のWebサイトから取得してコーパスとして用いることにより、過去60年間における日本語の変化を明らかにする試みを行った。

まず、「属する>属す」や「論ずる>論じる」に代表される一字漢語サ変動詞の活用の変化とそのゆれの様相については、1990年前後の新聞記事に基づく調査・分析の結果を過去に発表したことがあったが、国会会議録から得られる用例を分析することにより、特定の一時期だけの日本語の観察からは見えてこない、活用の変化のより詳細な様相と動向を明らかにすることができた。ほかにも異形態や各種の慣用句形の選択に関わる通時変化の様相を観察・分析した。

また、「可能性」のように尺度的な属性を表わす名詞とその属性の値の大きさ・小ささを表わす形容詞とが構成する「可能性が大きい／多い／高い／強い／濃い」のようなコロケーションについて、その頻度の通時的な推移を調査した。その結果、「～性」「～率」「～度」のような名詞では、「高い」が主として用いられる方向へと変化が進行中の可能性があることが分かった。さらに、発話年代と発話者の出生年代という2つの変数に注目し、補助動詞のイルとオルの選択、人を主語とする存在文でのイル(オル)とアルの選択、「～的{ナ/ノ}」の選択、「まする」の使用、副詞と否定辞との共起率などについて分析を行った。その結果、どの現象でも、同じ年代に出生した話者でも発話年代が進むにつれて使用傾向が変化する現象が見られた。これは、発話年代での区分に基づく分析では知りえなかった事実である。

現代語の通時コーパスとして、20世紀後半の新聞コラムのコーパスを作成し、現代語の通時的な研究を試みると同時に、通時コーパスの分析に関する問題点の探索を継続した。新聞のコラムを対象としたのは、テキストタイプが同じである、各年の延べ語数がほぼ等しくなる、話題が適当に分散していて特定の語彙に集中しない、という理由のほかに、各年の書き手が原則として同一個人であり、言語変化に関する書き手の影響を検討するのに好適だからである。作成したコーパスは、毎日新聞のコラム「余録」欄の、1950・60・70・80・90・2000年の各1年分、計6年分を収めたプレイン・コーパスで、データ量(「茶筌」の処理結果)は(固有名詞・数・記号・助詞・助動詞を除く)実質的な単語で各年7～8万語(延べ)である。これにより、20世紀後半で増減する単語の抽出や、単語使用における「著作年代の差」と「書き手の違い」との影響関係の分析などを行った。

科学技術振興機構(JST)が運営する科学

技術文献情報のオンライン・データベースサービス“JDreamII”を現代語の通時コーパスとみなして、専門用語の借用／翻訳の選択がどのようになされるのかを、事例研究により試行的に検討した。“JDreamII”には1976年以降の文献情報データが大量に蓄積されており(約4,800万件)、それを経年的に調べることで、専門家が原語に接触して借用／翻訳を主体的に選択した段階をとりだすことができると思込まれる。“JDreamII”の(医学分野を除く)「科学技術全般ファイル」を使って、コンピュータ関連分野の“ubiquitous”という英単語(原語)が日本語の標題でどのように表現されているかを調べた結果、外国論文の訳題では翻訳(「遍在」等)を選択し、日本語論文の原題では借用(「ユビキタス」)を選択する傾向が確認でき、借用が「自国語に訳出する」というより「自国語で案出する」という位相の中で行われることが示唆された。

#### ・コロケーション・呼応

従来、日本語研究においてコロケーションは基本的なキーワードとして広く認知されてこなかったが、昨今辞書編集や日本語教育への応用の観点から興味を引くようになってきた。日本語コーパスからコロケーション情報を抽出するには、英語の場合とはまた違った手法が必要となる。2008年に作成した巨大なWebコーパスを用いて、日本語コーパスからのコロケーション情報の抽出手法を検討し、それによって得られた情報に基づいて日本語コロケーション辞典の数項目を試作し、また、用言の文法的性格の分析や類義的な複合辞の違いの分析などを行った。また、複合的な性質を持つコロケーションの概念に着目する意義について、事例に即して考察した。

従来「呼応」と呼ばれた現象について、通時的・共時的の両面からコーパスを用いた分析を行った。通時的な面では、「全然」と「全く」の2つの副詞について分析したところ、「全く」に関しては、否定辞とよく共起する方向に変化しており「違う」との共起率も高まってきていることが分かった。また共時的な面では、否定形の述語とよく共起する副詞について分析し、従来知られていなかった傾向をいくつか発見した。

#### ・マルチメディア・コーパス

コーパスを用いた言語研究を、話しことばを基本とする「言語使用」の研究に拡大・発展させていくためには、言語形式が検索できるというだけでなく、実際の発話場面における映像と音声参照して、その使用にかかわる各種の情報をも同時に得ることのできる「マルチメディア・コーパス」が必要にな

る。その試作版として、国立国語研究所「テレビ放送の語彙調査」(1989年4~6月)のNHK総合・教育テレビのデータを収めた「NHKコーパス」と、大阪の6放送局7チャンネルが放送した23種類の対談番組(2009年3~8月、106回分の放送)を収めた「対談番組コーパス」とを作成した。どちらも、特定の単語を指定すれば、それを含む文が発話された際の映像・実音声再生できるようにしたものである。これらを用いて、特定の言語形式・表現と非言語行動との関係(思考動詞「思う」が一人称主語の述語として発話される際の話し手の視線、擬音語・擬態語の発話時の身振りの有無、指示詞(直示用法)の発話時の指差しの共起など)の分析や、(言語使用としての)談話と映像との関係を談話分析の手法を用いて検討することなどを試みた。

#### ・統計的手法

コーパスを用いた発見的研究の新たな可能性を探るため、程度副詞と述語の間の共起関係を、新聞記事のデータに多変量解析(因子分析)の手法を適用することによって分析し、それに基づいてそれぞれの程度副詞の特徴づけを行うことを試みた。それぞれの程度副詞と述語との共起の有無のデータを基に、程度副詞の共起傾向に関わる因子を抽出し、因子得点と述語の意味的(文体的)特性を参照しながら意義付けを試み、内省に基づく従来の記述と対照した。その結果、共起例の有無という単純なデータに基づく方法でもおよそ従来の記述と合致する結果が導かれることが分かり、この方法を他のまだ十分記述されていない現象の分析に利用しうる可能性が示された。

コーパス日本語学において「探索的データ解析」(Exploratory Data Analysis)という統計手法が有効な分析ツールとなることを、これまでの計量的日本語研究の成果・知見を検証・追試することによって確認し、その上で、探索的データ解析が用意する一連の手法のうちのどれが、日本語についてのどのような調査・研究に利用可能であるのかを、独自に用意した中学校歴史教科書や新聞コラムのコーパスを試料として明らかにした。とくに、探索的データ解析の手法のうち、日本語研究において有効と考えられる10の手法(幹葉表示、数値要約と平行箱型図、データのならし、ヒンジ散布度一定化のための再表現、抵抗直線、蛇行箱型図、ルートグラム、二元分析(中央値精練法)、リジット解析、ロジット変換)をとりあげ、具体的な日本語研究への適用事例とともに、その利用法を紹介・解説した。

#### 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50件)

- ① Tanomura, Tadaharu ‘A corpus-based analysis of some time-related aspects of contemporary Japanese,’ 2011, To be published in Italy, 査読有
- ② 服部匡「程度の側面を持つ名詞とそれを量る形容詞類との共起関係—通時的研究—」『言語研究』第140号掲載予定、頁未定、2011年、査読有
- ③ 石井正彦「日本語コーパス言語学の展開」『日本言語文化』第16輯、5~21頁、韓国日本言語文化学会、2010年、査読無
- ④ 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
- ⑤ 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
- ⑥ 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
- ⑦ 服部匡「『全く』と『全然』の使用傾向の変遷—国会会議録のデータより—」『同志社女子大学総合文化研究所紀要』第27巻、162~176頁、2010年、査読無
- ⑧ 石井正彦「テレビ放送のマルチメディア・コーパス—映像・音声を利用した計量的言語使用研究の可能性—」『阪大日本語研究』21、1~20頁、大阪大学大学院文学研究科日本語学講座、2009年、査読無
- ⑨ 石井正彦「借用の『位相』—JST・科学技術文献情報の『ユビキタス』を例に—」『待兼山論叢』第43号日本学篇、73~90頁、大阪大学大学院文学研究科、2009年、査読無
- ⑩ 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』第55巻、109~122頁、筑波大学、

2009年、査読有

- ⑪ 杉本武「格助詞『を』再考」『日語日文学研究』第71輯1巻、3～12頁、韓国日語日文学会、2009年、査読無
- ⑫ 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
- ⑬ 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
- ⑭ 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
- ⑮ 服部匡「『～シテイル』と『～シテオル』一戦後の国会会議録における使用傾向調査」『計量国語学』第27巻第1号、1～17頁、2009年、査読有
- ⑯ 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
- ⑰ 杉本武「複合格助詞『にとつて』の意味と文法機能」藤田保幸・山崎誠編『複合辞研究の現在』、137～167頁、和泉書院、2006年、査読無

[学会発表] (計16件)

- ① 杉本武「コーパスと複合辞の記述」国際研究フォーラム「日本語学習辞書の開発と日本語研究」、筑波大学(茨城県)、2010年12月12日
- ② 石井正彦「マルチメディア・コーパスと言語使用」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 杉本武「コーパスに現れる言語使用と文法記述」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
- ④ 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
- ⑤ 服部匡「国会会議録データと現代語の通時変化」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
- ⑥ 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

- ⑦ 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.
- ⑧ 服部匡「気づかれにくい語結合の傾向とその変化—共時的/通時的コーパスのデータから—」第26回表現学会近畿例会、同志社大学(京都府)、2010年2月27日
- ⑨ 石井正彦「日本語コーパス言語学の新展開」韓国日本言語文化学会、明知大学校(韓国・龍仁市)、2009年11月14日
- ⑩ 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.
- ⑪ 杉本武「コーパスからみたモダリティー『かもしれない』と『可能性がある』—」第3回清華大学・北京大学・筑波大学合同研究会、北京大学(中国・北京市)、2009年9月19日
- ⑫ 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.
- ⑬ 杉本武「格助詞『を』再考」韓国日語日文学会「韓日語の対照研究シンポジウム」、釜山大学校(韓国・釜山市)、2009年5月25日
- ⑭ 石井正彦「テレビ放送のマルチメディア・コーパス」計量国語学会第52回大会、武庫川女子大学(兵庫県)、2008年9月20日
- ⑮ 石井正彦「外来語の20世紀」日本語学会平成19年度春季大会、関西大学(大阪府)、2007年5月26日

[図書] (計3件)

- ① 田野村忠温・服部匡・杉本武・石井正彦『コーパス日本語学の新展開』特定領域研究「日本語コーパス」日本語学班平成18年度～平成22年度研究成果報告書、500頁、

2010年

- ② 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、200頁、2007年
- ③ 石井正彦『現代日本語の複合語形成論』、ひつじ書房、497頁、2007年

[その他]

ホームページ等

<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

服部 匡 (HATTORI TADASU)  
同志社女子大学・表象文化学部・教授  
研究者番号：40228490  
杉本 武 (SUGIMOTO TAKESHI)  
筑波大学・大学院人文社会科学研究科・教授  
研究者番号：70196749  
石井 正彦 (ISHII MASAHIKO)  
大阪大学・文学研究科・教授  
研究者番号：10159676

### (3) 連携研究者

なし