

機関番号：12102

研究種目：特定領域研究

研究期間：2006～2010

課題番号：18061001

研究課題名（和文） 代表性を有する書き言葉コーパスを活用した日本語教育研究

研究課題名（英文） Research on Japanese pedagogy using the "Balanced Corpus of Contemporary Written Japanese (BCCWJ)"

研究代表者

砂川 有里子 (SUNAKAWA YURIKO)

筑波大学・人文社会系・教授

研究者番号：40179289

研究成果の概要（和文）：

代表性を有する大規模書き言葉コーパスを日本語教育に活用する方法を開発するという目的のもとに以下の三つの課題を達成した。課題 1：日本語教材コーパスを作成し、中国で開発された日本語教科書との比較などの基礎研究を行った。課題 2：書き言葉均衡コーパスを活用した日本語教材作成を行うための基礎資料として、コロケーションリスト、文法原型シラバス、語彙原型シラバスなどを作成した。課題 3：日本語教育に BCCWJ を活用するためのコロケーション抽出ソフトウェアの開発などを行った。

研究成果の概要（英文）：

The purpose of this project is to investigate methods to utilize the Balanced Corpus of Contemporary Written Japanese (BCCWJ) for Japanese language education. We have completed the following three tasks: (1) construction and analysis of a corpus of Japanese teaching materials; (2) compilation of a Japanese collocation list and construction of a database for basic grammar and basic vocabulary for Japanese education; and (3) development of tools to utilize the BCCWJ for Japanese language education.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2006年度	5,800,000	0	5,800,000
2007年度	4,800,000	0	4,800,000
2008年度	4,800,000	0	4,800,000
2009年度	4,900,000	0	4,900,000
2010年度	3,900,000	0	3,900,000
総計	24,200,000	0	24,200,000

研究分野：人文学

科研費の分科・細目：言語学・日本語教育

キーワード：日本語教育・コーパス・日本語教材・教科書コーパス・コーパス活用ツール・コロケーション・文法シラバス・語彙シラバス

## 1. 研究開始当初の背景

言語コーパスは日本語研究のみならず、辞書の編纂、教育用の語彙シラバスや文法シラバスの作成、語学教材の開発など、国語教育や日本語教育に役立てられる有用な言語資源である。しかしながら、研究開始当初は、日本語でのコーパス整備状況は十分ではなく、特に日本語の多様な全体像を把握することのできる均衡の取れた言語コーパスは皆無の状態であった。一方、日本語教育の世界ではコーパス言語学に基づいた研究成果が上がり始めてはいたが、その実状は、新聞、青空文庫、あるいはウェブなどのごく限られたジャンルの、しかも言語研究のために構築されたわけではないデータを使っただけの研究しか行えない状況であった。以上のような状況で、言語研究のために設計され、しかも均衡の取れた現代日本語コーパスの構築は、日本語研究と日本語教育のどちらの分野においても緊急に達成しなければならない重要な課題であった。

## 2. 研究の目的

日本語教育班の研究目的は、「現代日本語書き言葉均衡コーパスを日本語教育に活用する方法の開発」である。従来教師の経験と勘にもとづいて作成されていた日本語教科書、教材、シラバスについて、客観的な言語データにもとづいて考えるべき部分を見極め、コーパスを日本語教育に活用する方法について検討することをめざし、具体的には、次の3つの課題に取り組んだ。

課題1：日本語教材コーパスの作成と分析

課題2：書き言葉均衡コーパスを活用した日本語教材作成法の開発

課題3：日本語教育のためのコーパス活用ツールの開発

## 3. 研究の方法

### (1) 課題1：日本語教材コーパスの作成と分析

日本語教育のための教材分析、シラバス作成、教材作成、教授法開発などに利用できる基礎的なデータとして日本語教材コーパスを作成する。また、このコーパスを活用し、日本語教育用の語彙の研究、教育用語彙表の試作、中国で作成された教科書コーパスとの比較、などを行うことにより、日本語教育に資する基礎的な情報を提供する。

### (2) 課題2：書き言葉均衡コーパスを活用した日本語教材作成法の開発

日本語教育、特に中級以上の日本語教育において重要なのは、使える表現を増やすことである。そのためには、使用上の特徴が顕著な表現を抽出するとともに、日本語教育におけ

る重要表現の典型的使用パターンと類似表現との使い分けを目に見える形で記述することが必要である。そこで、既存のコーパスとBCCWJ領域内公開データを利用して、使用上明確な特徴を有するコロケーションのリストを作成する。また、既存の文法項目・語彙項目のリストをもとに、BCCWJ領域内公開データを活用した調査を実施し、「具体的な学習場面においてこの項目を教えるかどうか、教えるとしたらどの段階で教えるか」を判断するための情報を各項目に付した「文法原型シラバス」と「語彙原型シラバス」を作成する。

### (3) 課題3：日本語教育のためのコーパス活用ツールの開発

日本語コーパスを教育用に提供するためには、エンドユーザーの検索ニーズやスキルに応じたインタフェースの開発が重要である。そこで、データ班による日本語コーパス整備の進捗を受けて、以下の作業を行う。

- ・語彙プロファイリング・ソフトとして Sketch Engine への日本語データの実装。
- ・日本語教育の観点からのインタフェースの改善に関する調査。
- ・BCCWJ を分析対象とするプロトタイプのコロケーション抽出ソフトウェアの開発。
- ・その他、日本語教育に役立つソフトウェアの開発。

## 4. 研究成果

### 課題1：日本語教材コーパスの作成と分析

- ・初級・中級向け日本語総合教科書および読解教材48種を入力し、テキストの内容を「課題」「解説」「例文」「練習問題」などに種別してタグ付けを行った。また、著作権処理を実施し、公開に向けての準備を行った。
- ・中国で作成された教科書コーパスを利用してシラバス研究、教材分析を行った。
- ・日本語教材コーパスを用いた語彙表を試作し、BCCWJの語彙情報との比較を行った。
- ・BCCWJの評価と試用を目標に、日本語教材分析と作成への多種のコーパス（BCCWJを含む）の比較利用とその可能性の検証を行った。

### 課題2 書き言葉均衡コーパスを活用した

#### 日本語教材作成法の開発

課題2は、研究テーマを「日本語教育のためのコロケーション研究」と「文法項目・語彙項目に関する「原型シラバス」の作成」の2つに分けて実施した。

#### <日本語教育のためのコロケーション研究>

- ・BCCWJ(2009)とGoogle刊行のN-gramファイルを利用し、コロケーションと拡大コロケーション(lexical bundle)の抽出を行い、これまでに作成したリストの精緻化を行った。

それをもとに品詞ごとのコロケーションリストと拡大コロケーションのリストを作成した。

- ・BCCWJ(2008)とBCCWJ(2009)の結果を比較対照させることにより、コーパスサイズとコロケーション認定の統計値の妥当性との関連について検討した。

<文法項目・語彙項目に関する「原型シラバス」の作成>

- ・言語政策班が作成した検定教科書コーパスを用いた文型調査を行い、年少者に対する日本語教育の文型指導に資する調査報告を行った。
- ・BCCWJ領域内公開データを用いた機能語調査を基に複合辞の類義語研究を行った。
- ・BCCWJ領域内公開データ2009年度版を使い、『日本語能力試験出題基準』1・2級の機能語を中心とする439項目についてジャンル別の頻度調査を行った。また、その調査を基に、「BCCWJによる機能語データベース(スタンドアロン版)」を作成した。
- ・初級文型の「原型シラバス」作成のためのケーススタディを行い、発表した。具体的には「イ形容詞」「義務の表現」「原因・理由の表現」「否定疑問(のではないか)」「伝聞表現」について分析することによって、「原型シラバス」に盛り込むべき内容について提案した。
- ・橋本(2008)で提案した話題別語彙リストの改訂作業を行った。対数尤度比を用いてBCCWJに出現する語を100の話題に分類し、その情報を話題別語彙リストに付加して改訂した。
- ・KYコーパス・親密度・BCCWJの三者をつきあわせることにより「使用」という観点から見た場合の実質語の難易度に関する分析を行った。その結果、具体物を表す名詞の使用は、「難易度」ではなく「親密度」から説明することができ、抽象概念を表す名詞の使用は、主に「難易度」で説明することが可能であることが明らかになった。
- ・BCCWJ領域内公開データ2009年度版を使って語彙頻度分析調査に基づく日本語学習語彙表を作成した。

課題3：日本語教育のためのコーパス活用ツールの開発

- ・BCCWJ領域内公開データ2009年度版のサブコーパス10種より無作為抽出された2000語×100ファイル、計20万語の語彙頻度・分析統計を算出した。
- ・BCCWJを用いた語彙統計情報データベースの公開に向け、パッケージ化とドキュメント作成の作業を進めた。
- ・語彙情報データベースのWebサービス化に関する実験とツールの試作を行った。

- ・現在利用できる形態素・係り受け解析ツールを用いて解析したコーパスから語彙パターンを柔軟に検索するための検索式(言語)の使用の策定と検索ツールの試作を行った。

以上の研究成果の一部は『特定領域「日本語コーパス」研究成果報告書』のDVD-ROMに収録されている。このDVD-ROMにはプロジェクトの過程で作成した研究成果、報告書・論文、およびツール・データを収録している。その内容を表1に示し、さらにその概要を表の下に提示する。

表1 日本語教育班：成果報告書

#### DVD-ROMのコンテンツ

表1日本語教育班成果物公開DVD	
作成者	コンテンツ
	平成18年度～22年度の研究成果報告書
小林・小西	BCCWJを用いた初級シラバスの再評価
滝沢	言語研究のためのテキスト処理入門
滝沢	BCCWJ2009Ngrams
妙川・清水・奥川・千葉	BCCWJによる機能語データベース(スタンドアロン版)(Ver. 0.9.1b)
投野・本田	BCCWJサブコーパス10種の評価用語彙頻度・分布統計
千葉	BCCWJによる語彙情報データベース(Ver. 0.6)

1. 平成18年度～22年度の研究成果報告書  
紙媒体で刊行された報告書のpdf版。各年度の進捗状況報告、発表論文などを収録する。
2. BCCWJを用いた初級シラバスの再評価  
BCCWJを用いた初級シラバスの再評価について入門者向けに解説したもの。
3. 言語研究のためのテキスト処理入門  
Windowsのパソコン上で、一般的なソフトと一般に公開されている電子テキストを用いて言語研究に有用な情報を抽出するためのテキスト処理方法について解説したもの。
4. BCCWJ2009Ngrams  
BCCWJ 2009の中のLB(流通書籍)、OM(国会会議録)、OW(白書)、OY(Yahoo!ブログ)、PB(生産書籍)を対象に作成した有意義なコロケーション・拡大コロケーションを認定するためのファイル。
5. BCCWJによる機能語データベース(スタンドアロン版) Ver. 0.9.1b  
BCCWJ領域内公開データ2009年度版における『日本語能力試験出題基準』1・2級の機能語の全用例とジャンル別出現頻度をまとめたデータベース。
6. BCCWJサブコーパス10種の評価用語彙頻度・分布統計

BCCWJ 領域内公開データ 2008 年度版・2009 年度版の基本語彙統計資料の 1 つとして、サブコーパス 10 種から無作為抽出されたテキスト各 20 万語の語彙頻度・分布統計をまとめたもの。

#### 7. BCCJ による語彙情報データベース

任意の日本語テキスト（コーパス）の語彙を BCCWJ と比較するためのツール。本バージョンでは「形態素（語彙素）」「2 グラムの語彙素の連鎖」「隣接する 2 形態素によるコロケーション」の 3 つの単位で分析することができる。

また、以上の研究成果は、毎年開催される特定領域の全体会議、期間中 21 回開催された日本語教育班会議、4 回開催された複合辞研究会のほか、国内および海外での学会発表、パネルセッション、国際フォーラム、国内各地で 4 回開催されたワークショップなどにおいて逐次発表を行った。

#### 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 78 件）

- ① 滝沢直宏，周辺部を記述するための大規模コーパスの利用：その方法と留意点，英語語法文法研究，査読有，17 巻，2010，pp. 23-37
- ② Yukio Tono，"Critical Review of the Theory of Lexicographical Functions" LEXICON，査読有，40，2010，pp. 1-26
- ③ 清水由貴子，「A テモ B テモ」文の分析，日本語文法，査読有，Vol. 1, No. 1, 2010，105-121
- ④ 砂川有里子，コーパスを活用した日本語教育研究，人工知能学会誌，査読有，Vol. 24, No. 5，2009，pp. 656-664
- ⑤ 砂川有里子，清水由貴子，奥川育子，コーパスを用いた類義語研究-「それなのに」「そのくせ」「それでいて」の記述-，日中言語研究と日本語教育，査読有，第 2 巻，2009，pp. 21-33
- ⑥ 橋本直幸・山内博之，日本語教育のための語彙リストの作成，日本語学，査読無，27 巻 10 号，2008，50-58
- ⑦ 小林ミナ，コーパスと日本語教育，国文学解釈と鑑賞，査読有，74 巻 1 号，2008，88-94
- ⑧ 投野由紀夫，NICT JLE vs. JEFLL - n-gram を用いた語彙・品詞使用の発達 -，英語コーパス研究，査読有，15，2008，pp. 119-133
- ⑨ 千葉庄寿，フィンランド語記述文法とコーパスデータの役割，英語コーパス研究，査読有，15，2008，pp. 17-32

- ⑩ 橋本直幸，日本語教育における「話題」の扱い，人文学報，査読有，398，2008，pp. 58-76
- ⑪ 小西円，実態調査からみた「義務の表現」のバリエーションとその出現傾向，日本語教育，査読有，138，2008，pp. 73-82
- ⑫ 砂川有里子，分裂文の文法と機能，日本語文法，査読有，7(2)，2007，pp. 20-36
- ⑬ 滝沢直宏，コーパスを用いた英語研究の方法，日本語学会第 134 回大会予稿集，査読有，2007，pp. 18-23
- ⑭ 曹大峰，多言語コーパスと日本語研究，日本語科学，査読有，22，2007，pp. 59-78
- ⑮ 投野由紀夫，コーパスの医学英語教育への貢献：PERC Corpus プロジェクトを中心に，Journal of Medical English Education，査読有，6-2，2007，pp. 167-178
- ⑯ 滝沢直宏，コーパス利用のためのコンピュータ・リテラシー，日本語教育，査読有，130 号，2006，22-31
- ⑰ 山内博之，初級日本語テキストの形態素解析-「みんなの日本語」をデータとして-，実践国文学，査読有，70，2006，pp. 1-16
- ⑱ 山内博之，第二言語習得過程における言語転位の研究-日本語学習者による「の」の過剰使用を対象に-，第二言語としての日本語の習得研究，査読有，9，2006，pp. 104-114
- ⑲ 投野由起夫，Learner's Dictionary 概観，日本語学，査読有，309，2006，pp. 6-20，

〔学会発表〕（計 74 件）

- ① 砂川有里子，井上優，小林ミナ，橋本直幸，奥川育子，パネルセッション，日本語教育におけるコーパスの活用，世界日本語教育大会，2010. 7. 31，国立政治大学：台湾
- ② Yukio Tono，Corpus-based dynamic wordlists for English language learning and teaching: a critical appraisal of the English Profile Wordlists. TALC Organizing Committee-9th Teaching and Language Corpora (TALC) Conference，2010. 7. 1，Masaryk University, Brno, Czech Republic.
- ③ 砂川有里子，コーパスを活用した類義表現研究-「それなのに」「そのくせ」「それでいて」-，日本語教育連絡会議，2009. 8. 29 ウィーン大学：オーストリア
- ④ 本田ゆかり，大規模コーパスを用いた日本語学習語彙表作成の試み，JSAA-ICJLE2009 日本語教育学会国際研究大会，2009. 7. 15，ニューサウスウェールズ大学：オーストラリア
- ⑤ 山内博之，語彙リストを利用した日本語教育，地域の日本語教育促進に寄与する日本語教員養成特別セミナー，2009. 2. 21，群馬県庁

- ⑥砂川有里子, 日本語教育におけるコーパス活用の可能性, 日本語教育世界大会, 2008. 7. 13, 釜山大学: 韓国
- ⑦千葉庄寿, コリゲーションの抽出における形態統語情報の役割, 言語処理学会第12回年次大会, 2008. 3. 20, 東京大学
- ⑧投野由紀夫, 'A Bird's Eye View of Japanese EFL Learner's Lexico-grammatical Development through the Analysis of Large L2 Writing Corpora', Courpus & Second Language Writing, Symposium on Second Language Writing 2007. 9. 12, 名古屋学院大学
- ⑨滝沢直宏, 他言語コーパス研究の現在: 英語研究への示唆, 英語コーパス学会, 2007. 9. 3, 上智大学
- ⑩砂川有里子, 日本語読解学習支援環境の再構築, CASTELJ in Hawaii, 2007. 8. 4, ハワイ大学: アメリカ
- ⑪滝沢直宏, コーパスを用いた英語研究の方法, 日本言語学会, 2007. 6. 17, 麗澤大学
- ⑫小林ミナ, 記述文法と規範文法のインターフェイスとしての「教育文法」, 早稲田大学日本語教育学会, 2006. 9. 9, 早稲田大学
- ⑬滝沢直宏, コーパスから有意義な情報を得るために必要なこと, 2006年度第1回特別講演会, 2006. 5. 6, (財)交流協会・台北事務所: 台湾

[図書] (計7件)

- ①砂川有里子他編, ころしお出版, 日本語教育研究への招待, 2010, 300ページ
- ②投野由紀夫(監修) & Christian James(英文校閲), 東京書籍, フェイバリット英単語・熟語<テーマ別>コーパス1800, 2010, 272ページ
- ③山内博之編, ひつじ書房, プロフィクションシーから見た日本語教育文法, 2009, 187ページ
- ④山内博之編, 日本語教育スタンダード試案 語彙, ひつじ書房, 2008, 117ページ
- ⑤投野由紀夫, 小学館, 日本人中高生1万人の英語コーパス JEFLL Corpus - 中高生が書く英文の実態とその分析 -, 2007, 191ページ
- ⑥Tono, Yukio, Peter Lang Pub Inc "Spoken Corpora in Applied Linguistics.", 2007, 264ページ
- ⑦滝沢直宏, 小学館, コーパスで一目瞭然, 2006, 207ページ

[その他]

ホームページ等

[http://www.tokuteicorpus.jp/g\\_teaching/index.php](http://www.tokuteicorpus.jp/g_teaching/index.php)

## 6. 研究組織

### (1) 研究代表者

砂川 有里子 (SUNAKAWA YURIKO)  
筑波大学・人文社会系・教授  
研究者番号: 40179289

### (2) 研究分担者

井上 優 (INOUE MASARU)  
麗澤大学・外国語学部・教授  
研究者番号: 30213177

小林 ミナ (KOBAYASHI MINA)  
早稲田大学・大学院日本語教育研究科・教授  
研究者番号: 70252286

滝沢 直宏 (TAKIZAWA NAOHIRO)  
名古屋大学・大学院国際開発研究科・教授  
研究者番号: 60252285

投野 由紀夫 (TOUNO YUKIO)  
東京外国語大学・大学院総合国際学研究院・准教授  
研究者番号: 10211393

山内 博之 (YAMAUCHI HIROYUKI)  
実践女子大学・文学部・教授  
研究者番号: 20252942

### (3) 連携研究者

千葉 庄寿 (CHIBA SHOUJYU)  
麗澤大学・外国語学部・准教授  
研究者番号: 70337723

橋本 直幸 (HASHIMOTO NAOYUKI)  
首都大学東京・大学院人文科学研究科・助教  
研究者番号: 30438113

### (4) 研究協力者

小西 円 (KONISHI MADOKA)  
国立国語研究所・日本語教育研究・情報センター・プロジェクト奨励研究員  
研究者番号: 60460052

曹 大峰 (CAO DAFENG)  
北京日本学研究中心・教授

清水 由貴子 (SHIMIZU YUKIKO)  
日本大学・本部教育推進課・講師

本田 ゆかり (HONDA YUKARI)  
東京外国語大学大学院地域文化研究科・大学院生

奥川 育子 (OKUGAWA IKUKO)  
筑波大学・人文社会科学部研究科・大学院生