

**科学研究費助成事業 研究成果報告書**

平成 27 年 6 月 14 日現在

機関番号：32682

研究種目：基盤研究(B)

研究期間：2012～2014

課題番号：24320086

研究課題名(和文)和漢の両系統を統合する平安・鎌倉時代語コーパス構築のための語彙論的研究

研究課題名(英文)A lexical study for compilation of the corpus of Heian Kamakura Japanese which combines kanbun kundoku bun texts and wabun texts

## 研究代表者

田中 牧郎(TANAKA, Makiro)

明治大学・国際日本学部・教授

研究者番号：90217076

交付決定額(研究期間全体)：(直接経費) 14,100,000円

研究成果の概要(和文)：平安鎌倉時代の日本語コーパスを構築するためには、和文の資料だけでなく、漢文系の資料をコーパス化し、その両者を統合することが必要である。本研究では、漢文系の資料を電子テキスト化し形態素解析にかける手順を研究し、平安鎌倉時代の日本語コーパスを試作した。このコーパスを用いて次のような調査を行った。(1)漢文訓読文と和文の語彙を比較することで、文体的特徴語を抽出した。(2)漢文訓読文と和文の間にある、文体的対立を伴う類義対を明らかにした。(3)漢文訓読文と和文の平行コーパスにおいて異なる語が対応する部分に、語彙の文体的変異を指摘した。

研究成果の概要(英文)：For compilation of Heian Kamakura Japanese corpus, it is necessary to build corpus of not only Wabun (native Japanese style) texts but also kanbun kundoku bun (classical Chinese style) texts, and to combine these texts. We studied the procedure from digitization to morphological analysis of kanbun kundoku bun texts, and we also produced the model corpus of Heian Kamakura Japanese. Based on the corpus, we investigated following. (1)We extracted stylistic characteristic words by the lexical comparison of wabun texts and kanbun kundoku bun texts. (2) We clarified the pairs of synonyms with stylistic opposition between wabun texts and kanbun kundoku bun texts. (3) We pointed out stylistic lexical variety in the words corresponding with different words with in the parallel corpus of wabun texts and kanbun kundoku bun texts.

研究分野：日本語学

キーワード：日本語史 コーパス 語彙 文体的変異

### 1. 研究開始当初の背景

日本語史研究の進展のためには、多種多様な文献資料に記録されている言語事象を相互に関連づけながら参照できる総合的な資料体(コーパス)を構築していくことが期待される。国立国語研究所では、「通時コーパス」を設計するプロジェクトを行っており、日本語史における時代的变化をどのようにコーパスに反映させるかの研究が、進められている。一方、日本語史における社会的変異(地域・階層・文体など)をコーパスにどう反映させるかの研究は、当該プロジェクトで本格的に取り組む体制がなく、新たな共同研究体制を構築する必要が生じていた。

文献資料で跡付けることのできる日本語の社会的変異のうち、時代を通じて最も顕著なものは、古典中国語の影響下で書き言葉を創成したことに起因する文体的変異である。特に、現代日本語の書き言葉の源流となった和漢混淆文が一般化する鎌倉時代までのそれを正しく把握することは、日本語史研究において極めて重要である。

話し言葉を反映する「和文」の系統の資料については日本語史研究の中心資料として最も活用されており、「通時コーパス」のプロジェクトでも既に研究対象にしているところである。一方、「漢文」の系統の資料については、和文とは別の層の言語を反映することを明らかにした優れた成果がある(漢文訓読文について築島裕(1963)『平安時代の漢文訓読語につきての研究』(東京大学出版会)和化漢文について峰岸明(1986)『平安時代古記録の国語学的研究』(東京大学出版会)など)。しかし、当時の書き言葉としては、この両系統の文章は連続しており、語彙や語法は共通する部分も大きく、目的や場面によって使い分けられる側面も強かったと考えられるが、和漢の両系統の連続性を解明したり、文体的変異の構造を記述したりする研究は十分に進んでおらず、この方面の研究が求められている。

その全体像を把握した上で、細部も的確に記述できるような研究は、文体的変異が反映したコーパスを作ることによって可能になると考えられる。

### 2. 研究の目的

1の「研究開発当初の背景」に記したことをもとに、本研究で明らかにすることを目指した内容を一言で言えば、平安・鎌倉時代の日本語の文体的変異をとらえることのできるコーパスの具体像を提示することである。

それを実現させるために、次の三つのことに回答を与えることを目的とした。

(1) 当時の漢文系の文献のうちコーパス化するべき資料は何であり、それをどのように電子化すればよいか。

(2) 漢文系資料においては、単語に区切りその読みや品詞を認定することが難しい

場合があるが、その困難を克服して、語彙・語法のデータを集積し、和文系資料のそれと関連づけていくにはどうすればよいか。

(3) 和漢の両系統の資料から得られた語彙・語法のデータを分類し分析していくためのデータベースをどのように作成し、これを用いてどのように研究を進めればよいか。

### 3. 研究の方法

2の「研究の目的」に記した3点それぞれについて次のような方法で研究を進めた。

(1)については、主要な漢文系資料から、平安・鎌倉時代の日本語資料として重要な資料をいくつか選定し、その電子化テキストを試作することを通して、その資料性を詳細に分析することと、コーパス化における問題の所在を明らかにする研究を行った。

(2)については、次の2つの研究を行った。第1に、訓点が付されている漢文系資料を書き下し文にしたものをコーパス試作の対象にし、試作の過程で生じる問題を解決していくことで、コーパスの構築手順を確立させた。第2に、形態素解析辞書 UniDic に、漢文系の語彙を登録して、漢文系資料に形態素解析がかけられるようにした。第3に、漢文系資料と和文系統の資料とで、同一内容が書かれている資料をパラレルコーパスにすることと、和漢の語彙を直接対応づける研究を行った。

(3)については、漢文系資料のコーパスから作成した語彙頻度表と、和文系資料のコーパスから作成した語彙頻度表をもとに、頻度比較を行って、それぞれの特徴語を抽出したり、『分類語彙表』番号を利用して特徴語間の文体的類義対を特定したりして、和漢の観点での語彙分類や語彙分析を行った。

### 4. 研究成果

2「研究の目的」、3「研究の方法」に記した3点それぞれに、次のような成果が得られた。

(1)については、宣命、法華百座聞書抄、尾張国郡司百姓等解文、今昔物語集などをコーパス化し、その過程で生じた諸問題を解決する研究に成果が得られた。特に、異体字、宣命書き、返読文字、訓点など、和文系統の資料にはなかった、特殊な表記について、電子化の仕様を定めて、同種の表記を有する資料をコーパス化する手順を構築した。

例えば、異体字に関しては、現在の日本語の標準文字集合である JISX0213 に示す包摂規準に依拠しながらも、この規準では包摂できない文字についても独自に包摂する規準を定めて、包摂の範囲を規定することと、字形や機能の近い文字で代用する規準を定めて、外字をできるだけ少なくする方策を提示することで、汎用性の高い JISX0213 による今昔物語集の電子化を実現した。

また、返読文字については、その実態を調査した上で、形態素解析が行えるような字順に整形するルールを定めて、そのルールにしたがってテキスト変換を行った上で、人手で点検する手順を構築し、やはり、今昔物語集にこれを適用して、形態素解析をかけることのできるコーパスの作成を実現させた。

(2)については、まず、訓点の付いた資料である、尾張国郡司百姓等解文、高山寺本古往来のコーパス試作を行い、コーパス化において読みを明確化し、その根拠もテキスト中に埋め込む方を整備した。

次に、和文資料用に開発された形態素解析辞書「中古和文 UniDic」に、漢文系資料に使われる語彙を登録して、漢文系資料のテキストにも形態素解析をかけられるように、解析辞書の整備を進めた。さらに、今昔物語集(漢文系資料)と宇治拾遺物語(和文系資料)の同文説話について、文と語の対応付けを行う規準を研究し、それを実装するパラレルコーパスを設計し、6話分について実際にそれを構築した。このパラレルコーパスにおける語の対応付け結果を集計し、二つの作品で異なる語が対応することが多いところに、語の文体的対立が明確に観察できることを明らかにした。さらに、その文体的対立関係を体系化することで、語の文体的価値を明確に記述できることの見通しも示した。

(3)については、今昔物語集の中に、漢文系資料(巻12、巻17)と和文系資料(巻27、巻29)とを設定し、相互の語彙頻度を比較して、それぞれの特徴語を抽出した。さらに、特徴語同士を、『分類語彙表』番号を利用して対応づけることを通して、類義語が文体によって対立している対を数十対抽出した。この対は、築島裕(1963)『平安時代の漢文訓読語につきての研究』(東京大学出版会)で指摘される、和文語と漢文訓読語の「二形対立」の対と同種のものであると考えられた。ただし、築島の対が和漢の両極を見た対であるのに対して、ここで抽出した対は、その両極の間に幾重にも層をなす、語彙の層状構造の中に見られる文体的対立であると考えられた。また、対になることが多い感情形容詞と感情動詞の統語的な特徴を比較して、人称や対象語の取り方で、両者が使い分けられていることを確認した。

以上(1)(2)(3)の研究成果を産出するとともに、試作したコーパスのいくつかが完成を見た。それらのコーパスは、国立国語研究所のWebサイトから順次公開し、一般の利用に供する予定である。その主なものは次の通りである。

- ・五国史宣命コーパス
- ・法華百座聞書抄コーパス
- ・尾張国郡司百姓等解文コーパス
- ・「日本語歴史コーパス鎌倉時代編」所収

## の今昔物語集コーパス

以上のように、試作コーパスの整備と、これを活用した語彙研究を行うことで、平安鎌倉時代を対象とした、文体的変異も反映した歴史コーパスを構築する基盤が整備できた。

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8件)

池田幸恵、須永哲矢、「五国史」宣命コーパスの設計とその利用、『訓点語と訓点資料』、査読有、134、2015、80-98

山本真吾、「あきだる(飽足)」の史的展開—中世軍記物における訓点語の受容—、『日本語史の研究と資料』、査読無、2015、116-130

山本真吾、『今昔物語集』話末評語の漢語の性格、『国語国文』、査読有、84-1、2015、1-16

藤原浩史、『枕草子』における程度副詞「いと」の表現価値、『日本語学』、査読無、30-14、2014、162-171

小木曾智信、歴史コーパスにおける形態素解析と辞書整備、『日本語学』、査読無、30-14、2014、83-95

田中牧郎、『日本語歴史コーパス』の構築、『日本語学』、査読無、30-14、2014、56-67

田中牧郎、山元啓史、『今昔物語集』と『宇治拾遺物語』の同文説話における語の対応—語の文体的価値の記述—、『日本語の研究』、査読有、10-1、2014、16-30

須永哲矢、堤智昭、『日本語歴史コーパス』のための書籍活字の電子化—小学館新全集『今昔物語集』を事例として—、『国立国語研究所論集』、査読有、6、2013、163-181

[学会発表](計15件)

須永哲矢、堤智昭、活字資料のコーパス化における外字チェックと処理、人文科学とコンピュータシンポジウム「じんもんこん2014」、2014年12月12日、国立情報学研究所一橋講堂

山本真吾、自言語による漢文文献の訓読、

日本語学会 2014 年度秋季大会、2014 年 10 月 24 日、北海道大学

河瀬彰宏、野田高広、和文体および漢文体をもつ資料の構造化 法華百座聞書抄の事例研究、第 6 回コーパス日本語学ワークショップ、2014 年 9 月 9~10 日、国立国語研究所

山本真吾、『今昔物語集』話末評語の漢語の性格について、漢字漢語研究会、2014 年 2 月 1 日、早稲田大学

田中牧郎、日本語史における漢語の層コーパスによる記述、日本文芸研究会第 66 回研究大会、2014 年 6 月 14 日、東北大学

富士池優美、岩崎瑠莉恵、『今昔物語集』の捨て仮名、第 5 回コーパス日本語学ワークショップ、2014 年 3 月 6~7 日、国立国語研究所

田中牧郎、『今昔物語集』における文体対立語 巻 12 と巻 27 の語彙比較による、第 5 回コーパス日本語学ワークショップ、2014 年 3 月 6~7 日、国立国語研究所

富士池優美、河瀬彰宏、野田高広、岩崎瑠莉恵、『今昔物語集』のテキスト整形、第 4 回コーパス日本語学ワークショップ、2013 年 9 月 4~5 日、国立国語研究所

田中牧郎、文体から見た『今昔物語集』の語彙 『日本語歴史コーパス 平安時代編』と比較して、第 4 回コーパス日本語学ワークショップ、2013 年 9 月 4~5 日、国立国語研究所

池田幸恵、須永哲矢、「五国史」宣命のコーパス化、第 4 回コーパス日本語学ワークショップ、2013 年 9 月 4~5 日、国立国語研究所

田中牧郎、説話の平行コーパスの設計 平安・鎌倉時代の文体変異の研究に向けて、第 3 回コーパス日本語学ワークショップ、2013 年 3 月 1 日、国立国語研究所

Makiro Tanaka and Hilofumi Yamamoto, Emotive Adjectives and Verbs of the Heian Japanese, 2nd Symposium Japanese Association for Digital Humanities (JADH2012), 2012 年 9 月 16 日、東京大学

Makiro Tanaka and Hilofumi Yamamoto, A Corpus Study of Emotive Adjectives and Verbs of the Heian Japanese, 13th

ACSI International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD2012)、2012 年 8 月 6 日、キャンパスプラザ京都

須永哲矢、堤智昭、小学館新全集『今昔物語集』での漢字活字 コーパス化のための調査と処理の検討、NINJAL「通時コーパス」プロジェクト・Oxford VSARPJ プロジェクト 合同シンポジウム「通時コーパスと日本語史研究」、2012 年 7 月 31 日、国立国語研究所

富士池優美、田中牧郎、今昔物語集における返読文字について 形態素解析の前後処理を通して、日本語学会 2012 年度春季大会、2012 年 5 月 20 日、千葉大学

〔図書〕(計 1 件)

近藤泰弘、田中牧郎、小木曾智信、コーパスと日本語史研究、ひつじ書房、2015、ページ数未定

〔産業財産権〕  
出願状況 (計 0 件)

取得状況 (計 0 件)

## 6. 研究組織

### (1) 研究代表者

田中 牧郎 (TANAKA, Makiro)  
明治大学・国際日本学部・教授  
研究者番号: 90217076

### (2) 研究分担者

池田 幸恵 (IKEDA, Yukie)  
長崎大学・多文化社会学部・准教授  
研究者番号: 10315228

小木曾 智信 (OGISO, Toshinobu)  
国立国語研究所・言語資源研究系・准教授  
研究者番号: 20337489

須永 哲矢 (SUNAGA, Tetsuya)  
昭和女子大学・人間文化学部・専任講師  
研究者番号: 10589584

高田 智和 (TAKADA, Tomokazu)  
国立国語研究所・理論構造研究系・准教授  
研究者番号: 90415612

富士池 優美 (FUJIIKE, Yumi)  
中央大学・文学部・特任准教授  
研究者番号: 20510572

藤原 浩史 (FUJIWARA, Hirofumi)  
中央大学・文学部・教授  
研究者番号：00219065

山本 真吾 (YAMAMOTO, Shingo)  
白百合女子大学・文学部・教授  
研究者番号：70210531

(3)研究協力者

河瀬 彰宏 (KAWASE, Akihiro)  
国立国語研究所・コーパス開発センター・  
非常勤研究員

野田 高広 (NODA, Takahiro)  
啓明大学 (韓国)・講師