

令和 5 年 6 月 29 日現在

機関番号：32507

研究種目：若手研究

研究期間：2020～2022

課題番号：20K13060

研究課題名（和文）新語彙定着期の言語変化 コーパスに基づく通時的語彙研究の実践

研究課題名（英文）Language change during the new vocabulary establishment period

研究代表者

間淵 洋子（MABUCHI, Yoko）

和洋女子大学・人文学部・准教授

研究者番号：10415614

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：本研究では、言語変化と時代・社会の変化との関わりを解明する研究の一つとして、新しい語彙がどのように定着するかを、日本語の実資料を元に実証的に明らかにすることを旨とした。近年整備が進む日本語の歴史的資料のコーパス（言語研究用のテキストデータベース）を用いて、新語彙を把握し、淘汰・定着の様相を観察・分析して、日本語の語彙変化を捉えた結果、前近代の語彙変化では、近代漢語語彙に見られた「流動性とその固定化」という様相は見られず、新語彙の受容・生成における近代漢語語彙の特殊性が浮き彫りになった。

また、近代漢語語彙の精緻な分析を行うため、明治期の科学・思想分野の論文を対象としたコーパスを新たに作成した。

研究成果の学術的意義や社会的意義

本研究は、これまで一部の資料を対象とした語彙調査や辞書に基づく語彙調査によって、断片的に観察されてきた「日本語語彙の変化」を、現存のコーパスに基づき、大量の歴史資料を総合的に観察することで大局的に捉えようとするものであり、日本語の歴史、日本語の変化を、実証的に、より詳細に記述したものとして、意義のあるものと考えられる。

また、本研究で作成した明治期の啓蒙的文章のコーパスは、日本語の語彙変化において極めて重要な位置づけとなる「近代漢語語彙」の研究の精緻化に寄与する資料として意義深いものと考えられる。

研究成果の概要（英文）：In this study, I aimed to empirically clarify how new vocabulary is established based on actual Japanese materials. Using a corpus of historical Japanese materials, I observed and analyzed the selection and establishment of new vocabulary, and captured changes in Japanese vocabulary. As a result, the lexical changes in the pre-modern era did not show the aspect of "fluidity and its fixation" seen in modern Sino-Japanese vocabulary. In addition, in order to perform a detailed analysis of modern Sino-Japanese vocabulary, a new corpus was created for articles in the fields of science and thought in the Meiji period.

研究分野：日本語学

キーワード：コーパス 新語彙 言語変化 通時研究 日本語史 近代漢語 バリエーション

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

近年、大規模なコーパス(言語データベース)の開発・公開が相次いでいることで、日本語の諸相を計量的・実証的に分析しようとする研究が増えている。書き言葉を対象としたコーパスとしては、現代語を収録する『現代日本語書き言葉均衡コーパス』(国立国語研究所、2011年)、奈良時代から主に大正時代までの日本語を収録する『日本語歴史コーパス』(国立国語研究所、2012年から順次拡張公開中)が代表的なものであり、これらを用いた語彙の変化に関する研究も多く行われてきた。

ただ、それらの多くは、特定の意味分野の語彙、あるいは特定の語について、主に意味に着目しその変遷や様相を捉えるもので、語彙の変化を大局的に把握しようとする研究や、語形・表記・語法変化といった形態的側面の変化を、語彙の定着と関連付けて総合的に捉えようとする研究も十分に行われていなかった。

2. 研究の目的

本研究では、語彙の流動的な時期と、その時期に新たに出現し定着する語彙(=新語彙)を把握すること、また、その定着過程において、どのような語形・表記・語法の変化が見られるかを詳細に観察・分析することを目指す。そのために、まず、既存のコーパスを用いて、語彙の変化を捉えることを本研究の第一の目的とする。また、加えて、語彙変化の大きいことが分かっている近代初期において、新語彙が取り入れられ定着する過程をより詳細に捉えるために、新語彙の創造・受容に大きく寄与したと思われる明治期の啓蒙思想関連資料のコーパスを新たに構築し、これに基づく新語彙の調査・分析を実施することを第2の目的とする。

3. 研究の方法

まず、語彙の変化を捉えるために、コーパスを用いた大規模な実例調査・分析を行う。使用コーパスは、国立国語研究所が開発・公開する形態論情報付きコーパス『現代日本語書き言葉均衡コーパス』(以下BCCWJ)、『日本語歴史コーパス』(以下CSJ)とする。参照として、開発中の『昭和・平成書き言葉コーパス』を適宜用いる。主に、国立国語研究所が提供する各コーパスの語彙表を利用し、語種構成の変化、語彙の変化を把握する。

次に、明治期の啓蒙思想関連資料のコーパスを作成する。底本に『明治文学全集3 明治啓蒙思想集』(大久保利謙編、1967年、筑摩書房)、『明治文学全集91 明治新聞人文学集』(西田長寿編、1979年、筑摩書房)を選定し、漢文を除く本文を全て入力した上で、Webサイト上で駆動する形態素解析器「Web茶まめ」(<https://chamame.ninjal.ac.jp>、国立国語研究所)を用いて形態素解析を実施する。ここで構築した新たなコーパスを用いて、新語彙を特定し、その定着状況と語形・表記・語法等の実態を捉える。

4. 研究成果

4.1 語彙の変化

既存のコーパスを用いた総合的な語彙調査の結果、語種構成では、やはり、平安から鎌倉期にかけて漢語が、江戸から明治にかけて漢語が、大正から現代にかけて外来語が大きく比率を上昇させていることが分かった。これは、従来の辞書等の調査に基づき指摘されてきた語彙における変化に合致する。

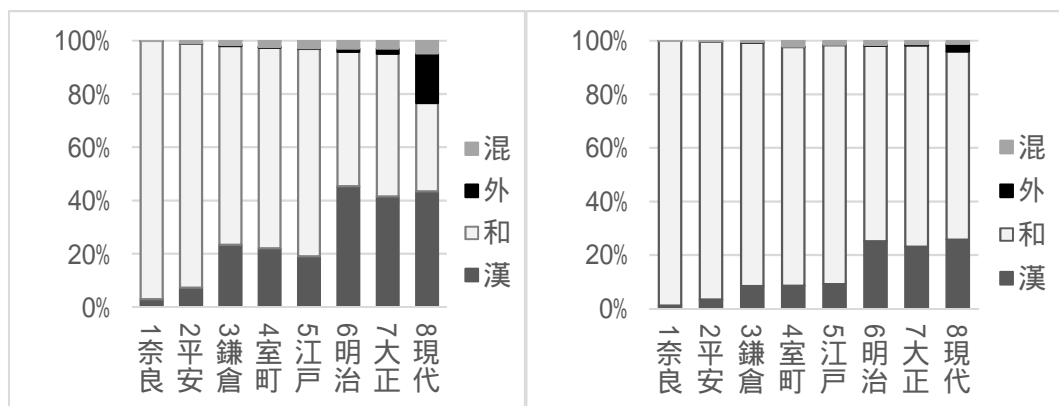


図1: 『CHJ』『BCCWJ』に見る語種構成比率の変遷(左:異なり語数、右:延べ語数)

そこで、時代間で使用度数(100万語あたりの調整頻度に基づく)に大きな差の見られる語群を特定し、主に、前近代を中心に語形・表記・語法等について詳細な調査を行った。

その結果、顕著な減少・衰退の見られる語群として、敬語表現、律令に関わる役職名などを、顕著な増加・興隆が見られる語群としては、人称代名詞、指示詞とその派生語(代名詞、副詞、接続詞等)感動詞、仏教語に由来する漢語等を、特徴的なものとして見出すことができた。

ただし、これらの変化は、時代によりコーパスに収録されている歴史的資料の性質が異なっていること（王朝文学、説話文学、軍記、大衆演劇、大衆文学など）に起因するものも多く、文章の位相差と語彙の関係性をより慎重に検討する必要があることが看取された。

また、これらの変化過程においては、近代漢語語彙に見られた「流動性とその固定化」と言うべき様相が、同様に現れているとは言い難く、新語彙の受容・生成における近代漢語語彙の特殊性が浮き彫りになった。

4.2 明治期啓蒙資料コーパスの構築

近代漢語語彙の創生・受容・定着の背景をより詳細に観察・分析するために、明治期の啓蒙思想関連資料のコーパスを作成した。

収録対象として、『明治文学全集 3 明治啓蒙思想集』（大久保利謙編、1967年、筑摩書房）『明治文学全集 91 明治新聞人文学集』（西田長寿編、1979年、筑摩書房）を選定し、漢文を除くすべての本文を文字入力しコーパステキストのベースを作成した。文章（以下「記事」と呼ぶ）の著者は、犬飼木堂、高田半峰、沼間守一、石河幹明、朝比奈知泉、渡邊巳之次、島田三郎、肥塚篁、矢野龍溪、加藤弘之、森有礼、神田孝平、杉享二、西周、西村茂樹、中村正直、津田真道、箕作麟祥といった思想家・学者・官僚などのいわゆる知識人で、収録記事の多くは、雑誌・新聞等に寄稿された論文で、漢字カタカナ交じりの近代文語文が中心であった。入力したテキストは、Web サイト上で稼働する形態素解析ツール『Web 茶まめ』を用いて解析した。解析辞書には「近代文語 UniDic」を選択し、解析前処理として、「踊り字を展開」「カタカナひらがな反転」を利用した。文字コード UTF-8 の CSV 形式で解析結果をダウンロードし、Microsoft Excel で解析データを整備・集計した。

完成したコーパスの言語量は、延べ語数約 60 万語、異なり語約 2.3 万語となった。

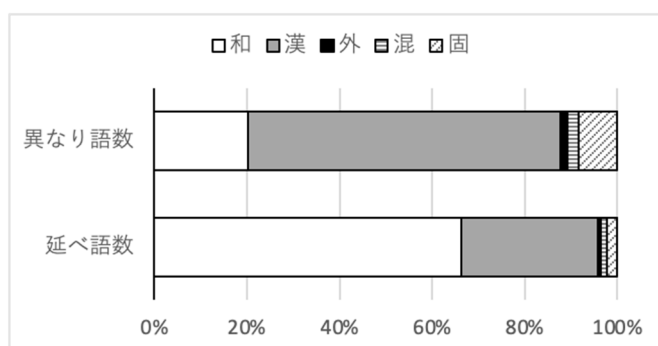


図 2：作成コーパスの語種構成

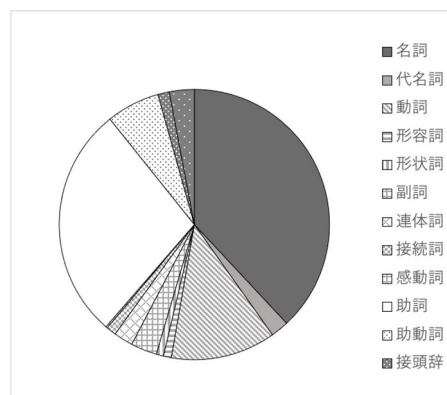


図 3：作成コーパスの品詞構成

データは、延べ語数で 30%程度、異なり語数で 60%以上が漢語であることから、近代漢語語彙の観察に適したコーパスとなっていることが期待される。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 間淵洋子
2. 発表標題 字順転倒漢語の語形交替について
3. 学会等名 研究発表会 「昭和・平成書き言葉コーパスによる近現代日本語の実証的研究」
4. 発表年 2021年

1. 発表者名 間淵洋子・小木曾智信
2. 発表標題 近現代日本語の意味変化分析のための単語データセット構築の試み
3. 学会等名 言語処理学会第27回年次大会 (NLP2021)
4. 発表年 2021年

〔図書〕 計1件

1. 著者名 田中牧郎・橋本行洋・小木曾智信	4. 発行年 2021年
2. 出版社 ひつじ書房	5. 総ページ数 375
3. 書名 コーパスによる日本語史研究 近代編	

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------