

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年 6月 7日現在

機関番号：32608

研究種目：基盤研究（C）

研究期間：2009年度～2011年度

課題番号：21520480

研究課題名（和文） 二葉亭四迷「浮雲」の電子データ化およびそれに基づく各種索引作成と研究

研究課題名（英文）Electronization of Futabatei Shimei's "Ukigumo" Text and Compilation of some indexes by its Data.

研究代表者 半澤 幹一 (Hanzawa kanichi)
共立女子大学・文芸学部・教授

研究者番号：50146528

研究成果の概要（和文）：

この研究のおもな成果は、二葉亭四迷の「浮雲」という明治時代の小説作品の電子テキストデータと、それに基づき作成した、電子版の語彙索引と漢字索引のデータである。近代日本の文学および言語の資料としてきわめて重要な「浮雲」の、当時の本文を極力再現した電子テキストデータと、専門的な研究にとっては必要不可欠な索引は、この研究によって初めて作成されたものであり、今後、その利用者に裨益するところ大であろう。

研究成果の概要（英文）：

The main fruits of this research are a electric text data of Futabatei Shimei's "Ukigumo", and its indexes of vocabulary and kanji. "Ukigumo" is not only a novel in Meiji period, but also a very important material for studies of modern Japanese literature and language. These fruits are its first electric text data and indexes. Therefore hereafter these will be extremely useful tools.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	1,800,000	540,000	2,340,000
2010年度	700,000	210,000	910,000
2011年度	800,000	240,000	1,040,000
年度			
年度			
総計	3,300,000	990,000	4,290,000

研究分野：人文学

科研費の分科・細目：言語学・日本語学

キーワード：二葉亭四迷 浮雲 電子テキスト 索引 漢字 語彙

1. 研究開始当初の背景

今回、研究対象とした、二葉亭四迷の「浮雲」という作品は、日本の近代小説の先駆けであり、いわゆる言文一致体の最初の文章として記念碑的な存在であることは周知のとおりである。それゆえ、近代文学研究においても、近代日本語研究においても、きわめて重要な資料として、さまざまな研究が行われ、その基盤たる個人全集も複数の出版社から

出されている。

にもかかわらず、これまでは「浮雲」の索引はもとより、原本に忠実な本文の電子テキストデータもないという状況にあった。小説をはじめ、近代文学資料の語彙索引は冊子体である程度は出版されているし、電子テキストデータも国立国語研究所から大規模なものが公開されている。しかし、「浮雲」については、国立国会図書館の電子ライブラリー

に収められているものの、画像データであるため、検索にはなじみにくく、インターネットの青空文庫のテキストは、新潮文庫を底本としているため、原文の表記とは異なっていて、学術的な利用には耐えがたい。

当然あるべきと考えられる「浮雲」の索引や電子テキストがこれまで作成されなかった要因としては二つ考えられる。一つは底本の問題、もう一つは技術的な問題である。

「浮雲」は第一篇が明治 20 年、第二篇が明治 21 年に、それぞれ単行本として金港堂書店から出版されたが、第三篇は文芸雑誌『都の花』に明治 24 年に 4 回連載されて終了した（一般には中絶とされる）。その後、全篇を一冊にまとめたものが明治 24 年に「再版」と称して出版されている。従来は、原文に即した研究の場合は、近代文学館が復刻した単行本（第一編・第二篇）と復刻版の『都の花』掲載分（第三篇）を基本資料としてきた。初出というだけならばそれでも適切なのであるが、利用が不便であるだけでなく、1 冊本の再版本の所在確認が困難だったため、両者の比較検討もなされなかった。以上のような事情により、索引なり電子テキストなりを作成するにあたって必須の前提となる底本の問題が解決されないままであった。

技術的な問題としては、単語分節と索引作成の両方に障害があった。「浮雲」は口語と文語、近代語と古代語が入り混じった文章であり、かつ誤植も少なからず認められることから、一律の一般的な基準では単語の分節や同定の判断に窮するものが多く、必然的に索引化するにもその統合・整理も従来の方法では対応しえなかった。さらに、電子化するにあたって、これまでのパソコンの日本語ソフトの技術レベルでは、文字や符号などのバラエティーを忠実には到底、再現しえなかった。

研究代表者自身は、20 年以上も前に、数人で、初出資料を用いて、「浮雲」の手書きによる語彙索引カードの作成に着手し、一応その作業は完了していたものの、もろもろの事情により、その整理に至らぬままとなっていた。そして、その間、余人による、どのような形であれ、ついに「浮雲」の語彙索引が公開されることなく、この研究当初の時点を迎えることになった。

しかし、冒頭で述べたように、この状態を放置しておくことは、近代の文学研究・日本語史研究の今後にとって、致命的な欠如となりかねない。幸いに、従来の問題が解決できる状況が整いつつある。すなわち、底本については国文学資料館から再版本がリプリント版として出版され、広く入手しやすくなったこと、技術についてはパソコンやソフトの急速な進歩と普及により、以前とは比較にならないほどの便宜が可能になったこと、である。さらには、研究分担者（服部隆）が全集本に

よるものであるが、「浮雲」の電子テキストデータを私に作成してあり、土台となりえるものがあつた。

このようなことから、件んの研究の実現が十分に可能であると判断された。

2. 研究の目的

1. をふまえて、研究の目的は次の二つである。

第一に、「浮雲」本文を電子データ化することである。その際、底本には「浮雲」の再版本を使用する。また、その再現にあたっては、版面・文字・符号、さらに誤植までも、極力忠実に反映するようにする。そして、その成果物は DVD に収めて頒布し、研究に自由に利用できるものとする。

第二に、第一の「浮雲」の電子データ化したテキストに基づいて、索引作成を作成することである。その際、可能な限りの「各種」をめざすが、最低限、索引として利便性もつとも高い語彙索引と、底本ならではの漢字索引は完成させる。作成にあたっては、どちらも従来の紙媒体では不可能だった、さまざまな項目による相互検索機能を装備する。合わせて、本文データとともに DVD に収めることにより、その機能を強化する。

第一目的も第二目的も、その達成のための作業の方針・設計には、従来とは異なるものが必要となるため、それらの作業の検討・実行自体が新たな研究となりえた。

3. 研究の方法

第一目的の「浮雲」本文の電子データ化にあたっては、具体的に次のような方法をとった。

(1) 研究分担者（服部隆）の作成した全集本のデータを元にして、底本本文の入力を、篇ごとに、業務委託して行った。入力にあたっては、底本に忠実であることを原則し、ワープロソフト（ワード）で再現できないもの、あるいは誤植とみなされるものは、すべて赤字表示し、必要に応じてコメントを付した。

(2) 入力したデータを、研究代表者および研究分担者が分担して校正した。

(3) 公開用にページごとの版面作成とページ数付けなどを業務委託し、研究代表者が最終的な全体の統一とコメントの整理を行った。

(4) 以上の電子データ化と並行して、本文情報を充実させるために、初出本本文との異同に関するデータおよび再版本の別本の画像データも作成した。

第二目的の索引作成にあたっては、具体的に次のような方法をとった。まず、語彙索引については、

(1) 本文の分節方法について、サンプルを用いて、利用可能な、さまざまな形態素解析

ソフトを適用してみて、もっとも適切な方法を検討した。その結果、「浮雲」に対しては、従来のソフトにはどれも一長一短があり、また費用対効率を考えると、文節単位に分節し、そのうえで単語単位に分節する方法をとることにした。

(2) 本文入力の完了を待って、順次、電子化されたテキストに、文節の切れ目を示す記号を入れていった。

(3) (2) の作業ののち、その記号ごとに切り分けて、テキストファイルをエクセルデータ化した。

(4) (3) のデータを元に、さらに単語に分ける記号の入力を行った。その際、単語単位は、語構成要素に分解し、文字通りの形態素単位にすることを原則とした。そのうえで、それぞれにいくつかのタグ(表記形・語構成・所在箇所・品詞(活用形)・文体・備考)を表示した。(2) から(4) の作業には、研究代表者と研究分担者(服部隆)があたった。

(5) (4) の完了ののち、五十音順にソートし、校正・全体の統一などを行った。これは研究分担者(服部隆)があたった。

漢字索引については、

(1) 語彙索引データをもとに、漢字だけを残して、それを JIS コード配列順にソートし、エクセルデータ化した。この作業には研究分担者(服部隆)があたった。

(2) (1) のデータにより、その全体の統一と情報の整理を行った。これは研究代表者が担当した。

4. 研究成果

以下の研究成果をまとめ、その電子データファイルを1枚のDVDに収めた。

凡例／新編浮雲(再版)本文／新編浮雲(再版)語彙索引／新編浮雲(再版)漢字索引／新編浮雲初版・再版の異同一覧／遠藤好英氏架蔵再版本(PDF)

これらの成果の詳細は実物を閲覧することによってしか確認しえないので、本報告書には、これらのデータの研究上の意義(位置付け・インパクト・展望)について、以下に挙げるにとどめる。

(1) 凡例では、とくに語彙索引・漢字索引における単位認定や語・文字の同定などについて、資料に即した、新たな方針と方法を明らかにした。

語彙索引の作成においては、従来の形態素解析ソフトではなじみがたい時代の資料であるため、最終的にはもっとも素朴な手作業によることになった。それは一方ではソフト開発上の限界の問題、もう一方では資料言語の時代性の問題であって、方法の便宜よりも言

語実態に即すことを優先した結果であった。手作業によった分だけ、時間も労力も予定以上に要することにはなったものの、それを通して、「浮雲」の性質が浮き彫りにされることになり、1. には、それをふまえた手続きが示されている。これは、とりわけ日本語研究にとっては、非常に重要な知見を示しているといえる。

また、漢字索引については、当時の活版印刷の状況もあって、現代とは異なる処理が求められたが、利便性あるいは汎用性を考えれば、常用漢字表やJISコードを基準として、それ以外はテキストデータ自体を参照する形で利用するのが妥当であることを説明した。今後、電子版による漢字索引を作成する場合の参考となることが期待される。

(2) 「浮雲」本文の、再版本による初めての電子データ化自体の意義は改めて言うまでもない。それもさることながら、従来は無視されてきた、漢字の字体、変体仮名、各種符号などを、電子媒体で、しかも検索可能な状態で、極力再現できる方法を試みた点にも、意義がある。

(3) 語彙索引自体が「浮雲」については初めてであり、その電子版により、利用の便宜や方法は飛躍的に広がることが予想される。とくには、語構成要素ごとの見出し語と語構成欄の設定により、どちらからでも検索可能になり、その他のタグも含め、単語の使用実態を精細かつ多面的に調査できる。また紙媒体ではなしえなかった全データ表示により、さまざまな加工・整理が可能になったことも大きい。

(4) (3) 同様、漢字索引もまた「浮雲」では初めての成果であり、近代の文学資料では漢字索引そのものがきわめて少ないという意味でも、当時の漢字の使用状況を知るうえで、利用価値は非常に高いと考えられる。しかも、語彙索引同様、すべてのデータを表示し、かつ語彙情報も付加しあるので、これのみによる漢字の調査・研究が可能であるという点も、従来の漢字索引にはない特長といえる。

(5) 従来の研究の多くが拠っていた初出本のおもな異同を一覧できるようにしたデータであり、これもまた全体的なものとしては初めて公にされるものである。詳細な検討は今後のことになるが、作品テキストにおける著者自身の意図と編集・出版との関係など、新たな問題の提起が期待される。

(6) 再版本の別本も、これまでに公開されることのなかった成果であり、同じく再版とは称しながらも、活字の種類や大きさなど、

微妙な違いも認められ、当時の出版状況を知るうえで好個のデータである。

全体として、近代当初の日本文学・日本語・出版などに関する今後の研究にとって、「浮雲」の本文および索引の資料それ自体の価値はもとより、電子データとして、広く公開することによって、目的に応じて、個々の研究における改変・加工が可能となることの価値は非常に大きいと考えられる。

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 1 件)

服部 隆 「嗟峨の屋おもろ「薄命のすゞ子」の「である」体——節 (Clause) を用いた文体分析の試み (4)」『上智大学国文科紀要』29、pp113-143、2011、査読無し

6. 研究組織

(1) 研究代表者

半澤 幹一 (Hanzawa Kanichi)
共立女子大学・文芸学部・教授
研究者番号：5014628

(2) 研究分担者

服部 隆 (Hattori Takashi)
上智大学・文学部・教授
研究者番号：10289598

倉田 静佳 (Kurata Shizuka)
東京国際大学・人間社会学部・准教授
研究者番号：50585982
(2010 年度～)

鈴木 直枝 (Suzuki Naoe)
東北生活文化大学・家政学部・准教授
研究者番号：40306064
(2009 年度～2010 年度前期)

新井 菜穂子 (Aarai Nahoko)
国際日本文化研究センター・文化資料研究
企画室・研究員
研究者番号：30442536
(2009 年度)

研究協力者

遠藤 好英 (Endo Yoshihide)