様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ３０ 年　 ５ 月 ３１ 日現在

機関番号：３３８０１
研究種目：若手研究(B)
研究期間：2015～2017
課題番号：１５Ｋ１６７６５

研究課題名（和文）コーパス内話者情報を利用した近世後期江戸語・上方語の研究

研究課題名（英文）Study of the Edo and Kamigata dialects in the late Edo period employing speaker information from the Corpus of Historical Japanese

研究代表者
市村　太郎（ICHIMURA, Taro）

常葉大学・教育学部・講師

研究者番号：１０７０１３５２

交付決定額（研究期間全体）：（直接経費）　　2,400,000円

研究成果の概要（和文）：本研究は、国立国語研究所『日本語歴史コーパス』内の洒落本資料について、会話部分の話者情報に関する情報を具体化・精密化し、既存の文書情報・形態論情報と連動させて高度なデータ収集を行うことのできる環境を整え、近世後期における江戸語・上方語の口語資料の言語を、主に位相的・数量的な面から分析し、その使用状況を明らかにすることを目指した。結果、洒落本コーパスの話者情報の整備に関しては、29年度末に国立国語研究所より公開されたデータに反映させることができた。また、研究期間内において、試論的なものや本研究に着想を得たものを発表・論文化した。現在、引き続き本研究の成果を活用した論文執筆を行っている。

研究成果の概要（英文）：This research annotated and refined the speaker information from conversation segments in the corpus of Share-Bon in CHJ. In addition, I prepared an environment to perform advanced data collection in conjunction with text and morphological information. I aimed to analyze the language of colloquial documents from Edo and Kamigata in the late Edo period using this information.

This research achieved the following results. On completion, I cooperated with the National Institute for Japanese Language and released the corpus of Share-Bon in CHJ with detailed, written speaker information. Moreover, I made research presentations and wrote papers in an attempt, and I am still writing papers.

研究分野：日本語学

キーワード：コーパス　江戸語　上方語　近世語　洒落本

１．研究開始当初の背景

近世期は、士農工商に代表される職業の別、男女の別、出身地域の別等、現代に比べ各位相の間により大きな開きがあり、このような位相・地域差は言語にも反映される。近世期の言語状況を正確に分析・記述するためには、発話者の社会的・生得的情報を考慮に入れることが望まれる。

このような話者の社会言語学的な情報に配慮した研究は、これまで主として、敬語表現や職分（たとえば「遊里語」「武家語」など）等の特定の表現について行われてきた。

その一方で、敬語形式以外の、例えば語彙の選択、文法形式、感動詞の使用等、様々な表現が話者の位相に左右されることも、容易に想像される。代表者はこれまで近世～近代語の副詞の研究に従事しきたが、その中で、明治期の「ほんに」は女性や遊里関係者の使用に偏ること（市村 2014）などを指摘したことがある。敬語や特定の階層の言葉以外にも、さまざまな面で位相差・地域差等が見込まれるのである。

しかし、近世期の語彙・文法に関する多くの研究では、用例に「会話」「地の文」の別、話者名等の情報が挙げられはするものの、個別的に分析されるというのが、現状どちらかといえば一般的であろう。従来の「手集計」・索引の利用・限られた電子化資料の文字列検索では、用例の収集・整理自体に膨大な手間がかかるうえ、同一話者全体の使用状況を確かめる等を行うためには、改めて本文を確認して話者に関する記述を収集しなおす必要があり、データ構築が極めて煩雑である。このような状況では、分析者の「勘」に頼って調査を進める他なく、確実な結果が出そうなテーマでもないかぎりは、データ集計のコストをかけることに慎重にならざるを得ないであろう。

しかし近年、国立国語研究所において『日本語歴史コーパス』が開発され、代表者が携わってきた『江戸時代編Ⅰ洒落本』も公開が予定されている状況であった。これにより用例収集自体の煩雑さは克服されることが予測され、上述の問題に関して光明がみえはじめた。ただ、本コーパスを用いても、当初の仕様では話者に関する必要な情報を得ることはできず、社会言語学的観点からの計量分析等はなおも難しい状況であることが予想された。予定されたコーパスの仕様に加え、上述のような近世後期語研究に必要な情報をいかに分析の俎上に載せるかが課題として残っていた。

そのような状況に鑑み、本研究では、上記のように、「近世日本語研究ならでは」ではないにせよ、近世語研究上特に必要とされる話者に関する情報を、既存のコーパスの仕様にプラスアルファする形で表現し、またそれを利用して近世後期語のいくつかの側面について分析を試みることとした。

２．研究の目的

本研究の目的は、国立国語研究所で構築中の『日本語歴史コーパス』内の洒落本資料を対象とするコーパスについて、会話部分の話者情報に関する情報を具体化・精密化することによって、既存のタグ情報・形態論情報と連動した高度なデータ収集を行うことのできる環境を整え、近世後期における江戸語・上方語の口語資料の言語を、主に位相的・数量的な面から分析し、その使用状況の特徴を明らかにすることであった。

期間内では、具体的に下記の２点を目指した。

まず第１に、『日本語歴史コーパス』洒落本編に対する話者情報の記述追加と、そのデータの公開である。

『日本語歴史コーパス江戸時代編』収載の洒落本各作品に関して、作品内の各話者に関する情報を収集し、研究に有用な情報を検討する。その上で、『日本語歴史コーパス江戸時代編Ⅰ洒落本』に話者情報の記述を加える。具体的には、表記を統一した話者・性別・使用言語の地域・身分(職業)・年齢等を調査したうえで、XML タグに記述し、検索ツール上で単語情報と共に表示させる。データは『日本語歴史コーパス江戸時代編Ⅰ洒落本』とともに公開することを目指すこととした。下図に、構想時におけるデータのイメージを示す(図１)。

図１　研究計画時に Himawari 版での「洒落本コーパス」を想定したイメージ図

また、話者情報の収集・記述を行うと同時に、話者情報に基づいた研究を行う上で有用な近世後期口語資料で、『日本語歴史コーパス江戸時代編Ⅰ洒落本』の当初案に含まれないもの数作品を選定し、「日本語歴史コーパス」洒落本編と同じ仕様・同じ形式でデータを構築することを目指した。

第２に、構築したコーパスの話者情報を利用し、近世後期江戸語・上方語の研究を行うことである。

構築したデータを基に、近世後期江戸語・上方語の語彙・文法形式等について、いかなる位相差・地域差があるかを分析することを試みた。具体的には、助詞・助動詞や副詞に関して、使用者の属性による使用状況の特徴等を明らかにする、またはそのような試みを示すことを目指した。

本研究の遂行により、次のような学術的結果と意義が予想された。

『日本語歴史コーパス』では、基本仕様として、品詞や活用など単語（短単位）に関する情報が付与されている。本研究ではこの『日本語歴史コーパス江戸時代編Ⅰ洒落本』のデータに、話者に関する社会言語学的情報を別に調査して連動させ、語を検索と同時にその語を含む会話の話者情報を電子テキスト上で取得可能にすることで、あらゆる語に関する話者の位相を知り、計量的に分析することが可能となる。これは既存の日本語歴史資料を対象としたテキストデータ・コーパスでは行われなかった試みであった。そして話者情報と単語情報との複合的な計量分析を行うことにより、敬語や遊里語等はもちろん、様々文法形式や語彙に関する社会言語学的な状況について明らかにする端緒を得ることが可能となろう。

また、単語情報からだけでなく話者情報からの調査も可能になるため、ある登場人物、あるいはある地域の話者全体、あるいは女性全体といった、従来難しかったある話者や話者集団全体の使用語を対象とした計量的な分析を行うことが可能となり、「作者が登場人物をどう描いたか」というような文学的な分析も含め、多面的な言語の使用状況を明らかにすることが可能となることが予想された。

本研究における研究成果を公開することにより、本研究以後の日本語史研究・方言研究等に対する大きな貢献が見込まれるとともに、長期的視野で見れば、近世文学、江戸文化史、風俗史など、洒落本を研究対象とする同時代の関連諸分野への貢献も予想される。

３．研究の方法

本研究で行ったのは大きく以下の３点である。

①『日本語歴史コーパス江戸時代編Ⅰ洒落本』洒落本編に対する話者情報データの構築

②話者情報の利用に適した作品の捜索とデータ化

③コーパスの話者情報を利用した近世後期上方語・江戸語の研究

このうちデータ構築に関しては、代表者が計画・管理を行い、データの入力作業・確認について、歴史的資料を扱った経験のある大学院生・修了生に作業を依頼した。また、データの分析・研究成果の公開に関しては、代表者が中心となって、代表者が常葉大学教育学部に転出した後も、共同研究プロジェクト「通時コーパスの構築と日本語史研究の新展開」の共同研究員として、国立国語研究所との連携のもと行った。

初年度はまず、①『日本語歴史コーパス江戸時代編Ⅰ洒落本』に対する話者情報のデータの構築に関して、代表者がいかなる情報を付与するのが妥当かを検討し、『日本語歴史コーパス江戸時代編Ⅰ洒落本』の試作版 XML データでの話者情報データ構築作業の試行に着手した。この過程で、話者情報の内「年齢」に関しては、作品本文に記述された年齢を詳細に拾い上げていくのはコストが大きいと判断し、「子供」「老人」の別程度にとどめることとした。

そして、『日本語歴史コーパス江戸時代編Ⅰ洒落本』に選定されることが見込まれた作品について、試作品への話者情報付与の試行を経て、『日本語歴史コーパス江戸時代編Ⅰ洒落本』の XML 構築が完了した作品から、大学院生１名に依頼し、順次話者情報データの構築に着手した。

②話者情報の利用に適した作品の捜索とデータ化に関しては、版本の確保や、画像リンクの利用可能性など、コーパスに追加するに適した作品として『総籬』『仕懸文庫』『辰巳婦言』を選定し、入力仕様を検討するとともに、入力の元となる翻刻資料の選定と準備資料の作成を行い、テキストデータを構築した。入力が終わったデータのうち前者２作品は、市村他（2012）等で示した仕様を改良して XML での構築を行い、同様に話者情報の追加記述を行った。

平成 28 年度以降は、引き続き、以下を行った。

①『日本語歴史コーパス江戸時代編Ⅰ洒落本』に対する話者情報のデータの構築に関しては、『日本語歴史コーパス江戸時代編Ⅰ洒落本』構築の作業の進捗状況に合わせ、データ構築予定の作品について、話者情報のデータ構築作業を行った。

その後、公開に向けてデータの調整・確認作業を行い、平成 29 年度末の 3 月までに『日本語歴史コーパス江戸時代編Ⅰ洒落本』の一部として Web 上で公開した。

②話者情報の利用に適した作品の捜索とデータ化に関しては、『総籬』『仕懸文庫』の 2 作品について、データの調整・確認作業を行い、平成 29 年度末の 3 月に『日本語歴史コーパス江戸時代編Ⅰ洒落本』の一部として Web 上で公開した。

③コーパスの話者情報を利用した近世後期上方語・江戸語の研究に関しては。データ構築が完了したものについて、副詞について地域差やその他の情報に着目して分析するとともに、助詞・助動詞や地域差の調査等を行い、論文や利用法の解説等を執筆した。また、データ構築に関する論文の投稿や研究発表、およびエントリー等を行った。

４．研究成果

本研究の研究成果として、本研究で作成した話者情報を収載した洒落本のコーパスを構築・公開したことが大きい。

まず、試作版として、2015 年 10 月に『ひまわり版「洒落本コーパス」Ver．0.5（『江戸時代編Ｉ洒落本』試行版）』を作成し、公開した。これは、大坂を舞台とする『聖遊郭』（内題「雪月花」・宝暦７・著者不明）、京都を舞台とする『河東方言箱まくら』（文政５・大極堂有長）、江戸を舞台とする『花街鑑』（文政５・鼻山人）を対象とし、文書情報・形態論情報に話者情報を加え、さらに国語研所蔵の版本との画像リンクを行い、コーパス検索ツール「Ｈｉｍａｗａｒｉ」（山口・田中 2005）に対応させたものである。代表者は、当時国語研のコーパス開発担当者として従事していたが、文書情報・形態論情報・話者情報・参照用画像が付与されたコーパスという、当時構想したデータのモデルを示したものであった。

この試行データの構築に大きな問題が生じなかったため、作品数を広げ、本研究で作成した『総籬』『仕懸文庫』も含め、同様の情報をもって、形態論情報の判断基準を改善するなどし、中納言形式で公開したのが、『日本語歴史コーパス江戸時代編Ｉ洒落本』である。京都・大坂・江戸を舞台とする洒落本各 10 作品、計 30 作品を対象とし、約 22 万短単位の規模で、文書情報・形態論情報・話者情報がそろい、また多くの作品につき参照用画像へのリンクも付されている（図２）。

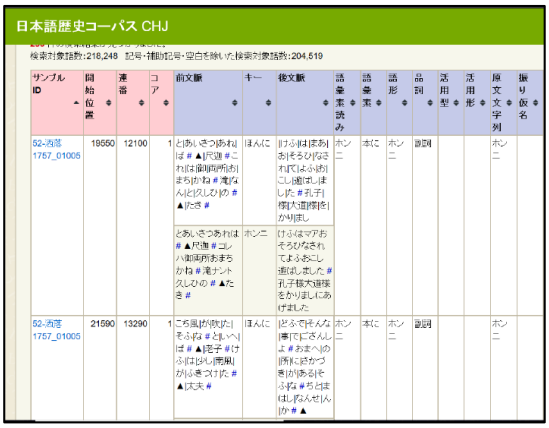
近世後期口語の基礎資料として、また「通時コーパス」における近世期の資料として、大きく活用されることが見込まれる資料となった。

また、このデータ構築に関して、データ構築を進める上での展望を示したもの、留意点や問題点等を示したもの、洒落本テキストの多様性と形態素解析について論じたものなど、各種の学会発表および論文執筆を行った。

データを利用した研究として、副詞を対象とした、代表者の博士論文をはじめとする一連の研究がある。加えて、本研究に着想を得て、近代初期の小説を対象とし、「ほんに」「じつに」「まことに」という 3 種類の副詞に関して話者の男女をもとに集計し、分析した論文を発表した。

また、研究期間内の刊行には至らなかったものの、助動詞を対象として話者の地域別に分類した、コーパス利用に関するテキストや、江戸時代の語彙をテーマとして本研究で構築したデータを利用して語彙の諸側面について論じたものについて執筆し、刊行の準備を進めているところである。

図２『日本語歴史コーパス江戸時代編Ｉ洒落本』検索結果画面例（上は左部、下は右部で、話者情報等の表示箇所）



今後さらに、データの誤りを見直し、コーパスを改善（可能ならば拡張）しながら、次期研究課題においても活用していきたい。

また本研究で得た話者情報付与の手法を他のコーパスに応用するなどし、本研究の成果を発展・継承させていきたいと考えている。

＜引用文献＞

①市村 太郎、副詞「ほんに」をめぐって―「ほん」とその周辺―、日本語の研究、10 巻 2 号、2014、1－16

②市村 太郎、河瀬 彰宏、小木曽 智信、近世口語テキストの構造化とその課題、情報処理学会研究報告、2012-CH-96、2012、1－8、https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=85858&item_no=1&page_id=13&block_id=8

③国立国語研究所、『日本語歴史コーパス 江戸時代編Ｉ洒落本』（短単位データ 0.9・開発担当者として従事）、2018、http://pj.ninjal.ac.jp/corpus_center/

chj/edo.html

④山口 昌也、田中 牧郎、構造化された言語資料に対する全文検索システムの設計と実現、自然言語処理、12 巻 4 号、2005、pp. 55-77

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４件）

①市村 太郎、近代初期における「じつに」「まことに」「ほんとうに」の用法、近代語学会編『近代語研究』（武蔵野書院）、第二十集、2018、pp.61-80

②市村 太郎、村山 実和子、洒落本コーパス構築の試行、国立国語研究所論集、査読有、12 巻、pp.29-45、Doi:10.15084/00000852

③市村 太郎、洒落本における「いっそ」と「いっこう」、近代語学会編『近代語研究』（武蔵野書院）、第十九集、2016、pp.21-41

④市村 太郎、近世・近代における程度副詞・強意副詞の研究、早稲田大学大学院文学研究科博士論文、査読有、2016、pp.1-207

〔学会発表〕（計６件）

①村山 実和子、市村 太郎、小木曽 智信、『日本語歴史コーパス　江戸時代編Ⅰ洒落本』の公開、日本語学会 2018 年度春季大会、明治大学、2018-5-20

②市村 太郎、江戸・上方洒落本における程度副詞の使用状況、「通時コーパス」シンポジウム 2017、国立国語研究所、2017-3-11

③市村 太郎、「江戸時代編」の構築と課題、日本語学会 2016 年度春季大会ワークショップ『日本語歴史コーパス』の拡張とその課題―「通時コーパス」をめざして―、学習院大学、2016-05-15

④市村 太郎、小木曽 智信、文書構造を利用した近世期洒落本の形態素解析、言語処理学会第 22 回年次大会、東北大学、2016-03-08

⑤市村 太郎、村山 実和子、洒落本コーパス試行版の作成と公開、第 328 回日本近代語研究会、国立国語研究所、2015-11-28

⑥市村 太郎、上野 左絵、『日本語歴史コーパス　江戸時代編』の現状と今後、「通時コーパス」国際シンポジウム、国立国語研究所、2015-10-04

〔その他〕
ホームページ等
①国立国語研究所、『日本語歴史コーパス 江戸時代編Ⅰ洒落本』（短単位データ 0.9・開発担当者として従事）、2018、http://pj.ninjal.ac.jp/corpus_center/chj/edo.html

②国立国語研究所、『ひまわり版「洒落本コーパス」（日本語歴史コーパス江戸時代編）』（Ver. 0.5・開発担当者として従事）、2015、http://pj.ninjal.ac.jp/corpus_center/chj/edo.html#share

６．研究組織
(1)研究代表者
市村　太郎（ICHIMURA, Taro）
①2015 年度：大学共同利用機関法人人間文化研究機構国立国語研究所・コーパス開発センター・ プロジェクト非常勤研究員
②2016・2017 年度：常葉大学・教育学部・講師
研究者番号：10701352