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

令和　４ 年　６ 月 １３ 日現在

機関番号：３２６０１
研究種目：基盤研究(C)（一般）
研究期間：2019～2021
課題番号：１９Ｋ００６２９

研究課題名（和文）深層学習による平安時代日本語語彙のジェンダー性の研究

研究課題名（英文）Research on gender of Heian period Japanese vocabulary through deep learning

研究代表者
近藤　泰弘（Kondo, Yasuhiro）

青山学院大学・文学部・教授

研究者番号：２０１２６０６４

交付決定額（研究期間全体）：（直接経費）　　2,100,000円

研究成果の概要（和文）：平安文学作品では、明らかに語彙のレベル・文体のレベルでの、言語使用者のジェンダー差があるように描かれている。そして、そのいずれもが、当時の社会のジェンダー的規範によって構築されたものであり、『古今集』はそのプロトタイプを示すひとつの言語資源となっていると考えられるのである。これらの言語のジェンダー差が現実の言語使用を反映していたものかどうかについての明確な判断はしにくいが、非常に幅広いものであり単なる文学的虚構であるとは言えないように思われる。

研究成果の学術的意義や社会的意義
日本語におけるジェンダー的要素の重要性は言うまでもないが、歴史的研究は遅れていたと言える。特に平安時代言語の研究を行う中で、平安時代文学作品の深い理解や、言語構造の解析のためには、従来考えられている以上に、ジェンダーを含むコノテーションの分析が必要である。深層学習等によって分類された語彙を、文学作品の解釈に用いることで、従来以上の深い読解が可能になる。それによって、さまざまな平安時代語研究に有用なデータを提供できるに留まらず、日本文化の深層にあるものを、実証的にコンピュータによる分析で明らかにすることができる。

研究成果の概要（英文）：In Heian literary works, there are clearly depictions of gender differences among language users at the level of vocabulary and style. And all of them were constructed according to the gendered norms of the society of the time, and the "Kokinshu" can be considered as one language resource that shows the prototype of these differences. While it is difficult to make a clear judgment as to whether these gender differences in language were a reflection of real language use, they seem to be very broad and not mere literary fictions.

研究分野：日本語学

キーワード：ジェンダー　コーパス　平安時代　機械学習　深層学習　分散表現

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

日本語の古典語、特に平安時代語において、言葉の男女差があったかどうかということについてはいくつかの議論があるが、文体的な特徴においては男女の差があるものの、男女それぞれの特有語というような語彙的なレベルでは大きな差を見つけることが困難であるというのが現在の平均的な考え方だと思われる。山口仲美「『源氏物語』の女性語」でも、男性語は、訓読的語彙として存在するが、女性語を見いだすことは非常に困難としている。また、文体的な特徴という点では、文学作品の持つ虚構性に注目し、言語ステレオタイプ(役割語)的なジェンダー差が存在しているのではないかという指摘もなされている。この点も問題ではあるが、現代語ですら、言語の男女差が、実体なのか、ステレオタイプなのかの分離はむずかしい。

このように、日本語におけるジェンダー的要素の重要性は言うまでもないが、歴史的研究は遅れていたと言える。特に平安時代言語の研究を行う中で、平安時代文学作品の深い理解や、言語構造の解析のためには、従来考えられている以上に、ジェンダーを含むコノテーションの分析が必要であると考えている。深層学習等によって分類された語彙を、文学作品の解釈や、文法研究の文法化のメタファー要素などに必要なジェンダーコノテーション辞書（データベース）の形に整理する。それによって、さまざまな平安時代語研究に有用なデータを提供できるに留まらず、日本文化の深層にあるものを、実証的にコンピュータによる分析で明らかにすることができる。

２．研究の目的

日本語の古典語、特に平安時代語において、言葉の男女差があったかどうかということについてはいくつかの議論があるが、文体的な特徴においては男女の差があるものの、男女それぞれの特有語というような語彙的なレベルでは大きな差を見つけることが困難であるというのが現在の平均的な考え方だと思われる。

本研究の研究代表者は、これまでも平安時代語のジェンダー研究を行ってきた。主たるものは、N-gram 分析を使うものである。例えば『古今集』のうち、男性歌か、女性歌かが、明確なものを抜き出し、それをすべて平仮名にする。そして、その仮名文字の連鎖（文字列）をすべて列挙する。例えば、2 文字(2gram)では、あ-い、あ-か、あ-ま、等、3 文字（3gram）では、あ-か-ず、あ-か-に、等。そして、これを意味のない文字列まで含めてすべて 20gram まで収集してから、男女ごとにまとめ、その（集合論的）差分をとる。そうすると、次のように男性特有の文字列と、女性特有の文字列を抽出することができる。これらを調査してみると、男性特有文字列に特徴的なものが多い。

今年度の研究では、以上の点についても、改めて『源氏物語』によって、追加的研究を行った。また、深層学習によって単語の分散表現を算出する word2vec は、単語の共起関係だけから、深層学習により単語の意味を数百次元のベクトルとして算出し、ひとつの表現がその他のジェンダー的表現とどのような関連があるかを調査して、平安時代におけるジェンダー性の本質的特徴を発見することを目的とする。

３．研究の方法

近年の自然言語処理の主流は、機械学習の手法によるものである。従来の統語分析の方法によらず、文字や単語の出現の共起関係などだけを元に膨大な関係性を計算し、分類整理することができる。中でも、深層学習は、複雑な関数を自動的に算出するため、従来の手法では計算できないものを見つけ出すことができる。これらの手法の中でも数年前に開発された word2vec と呼ばれるものは、単語の共起関係だけから、深層学習により単語の意味を数百次元のベクトルとして算出する。このベクトルの演算によって、直接に意味を計算・比較できることが非常に大きな特徴である。今回は、まず、『源氏物語』中の単語すべての分散表現を word2vec で作成し、その中で、ある単語とベクトルが近い(具体的にはコサイン類似度で計測)もののリストを作り、ひとつの表現がその他のジェンダー的表現とどのような関連があるかを調査した。また、それだけでなく、特定の品詞について、その多次元ベクトルを 2 次元に圧縮し、散布図に展開することで、多変量解析を用いて、特徴の軸を分析する。また、軸状の分布について、クラスター分析を行い、どのような特徴を持っているかを把握する。

word2vec による分析は、伝統的な構造言語学での、分布仮説によっているが、同じく構造言語学での「意義素」による分析と極めてよく似ている。二項対立によって示された意義素の項目が 100 個以上のペアで存在するイメージである。また、そのベクトルの加算・減算によって意味を計算することも可能になる。これらの特徴を生かして、語の統計的性格を明らかにしていく。

４．研究成果

本研究の当初の目的の平安時代語ジェンダーの体系について明らかになったことは多い。伝統的な自然言語処理の技術を用いた研究としては次の通りである。従来明らかになっていなかった文長の研究では、長いセンテンスは男性話者に非常に偏って出現することがわかった。『源氏物語』の登場人物では、僧都、源氏、薫、柏木、式部丞、などであり、また例外的に女性で長いのは、（桐壺更衣の）母君、右近、弁の尼といった主人公の周辺女房である。紫の上などの主要女性人物は、いずれも文が短い。（紫の上の最長文は 31 短単位である）これは、そのような女性たちが長々しく語るということがジェンダー規範に反するものであったことを意味すると思われる。女性主人公で文が長い例外的人物が、浮舟であることも興味深い(上記の 102 短単位の例を参照)。なぜなら、最初にあげた N-gram 分析による例外となり、『源氏物語』中で、古今集男性特有語をもっとも頻用する女性は、浮舟だからである。作者はあきらかに浮舟にジェンダー規範を破る人物像を与えていると思われる。

このように、平安文学作品では、明らかに語彙のレベル・文体のレベルでの、言語使用者のジェンダー差があるように描かれている。そして、そのいずれもが、当時の社会のジェンダー的規範によって構築されたものであり、『古今集』はそのプロトタイプを示すひとつの言語資源（リソース）となっていると考えられるのである。これらの言語のジェンダー差が現実の言語使用を反映していたものかどうかについての明確な判断はしにくいが、非常に幅広いものであり単なる文学的虚構であるとは言えないように思われる。

深層学習を用いた研究としては、いくつかを紹介してみよう。まず『古今集』の男性特有語である「飽く」という動詞を word2vec のコサイン類似度で近いものを探すと、「悲し」「恋し」「思ほゆ」などが上位に出てくるのだが、実は、これらの語は、いずれも、古今集での男性特有語である。この分散表現の計算には和歌だけでなく、地の文も含めてすべてを使っていることから、古今集における「飽く」という動詞の文体的特徴を、『源氏物語』の本文も同様に持っていることがわかる。また、同じく男性特有語の「知る」を調べると、「許す・育む・心をさなし」など、上位者から見た年少者に係わる表現が多い。「何ごとをも思し知りにたる御齢なれば」(源氏・蛍)のように、「知る」というのは知識獲得ではなく、一人前になるという文脈で使われるが、これが男性特有語となるひとつの理由だと思われる。また、単純に「男」「女」それぞれの語で単語ベクトルが近いものを見ると、「男」では「子・生む」などが近く、「女」では「姫・男・好きがまし・教へ」などが近く、当時の男女観を知ることができる。

また、シク活用の形容詞の平安時代における分布と、現代語における分布を比較することで、平安時代には、感情表現にかなり特化した存在だったシク活用形容詞が、現代語になると、客観的な形容詞も増えることになっている。その中で、女性のかわいらしさを表す「美し」が、現代語になると、客観的なところに移動していく。これらは『古今集』にも見られる。このように『古今集』でのジェンダー差は、男性や女性に構築された行動規範が元になっており、それにふさわしい特有の語彙が選択されるという形である。また、『古今集』は、当時の知識人にとっての言語知識の源泉(リソース)となっていたため、この使い分けは『源氏物語』の作中和歌や、人物描写にも継承されている。

この研究の発展段階としては、ジェンダー的な表現がどのように分布しているかについてのまとめを行い、これまでの研究と統合して、さらに幅広い研究を行っていく予定である。これによって、従来知られていなかった平安時代語のジェンダー体系について、明確な結論が得られると考えている。この問題は、ジェンダーに留まらず、日本語の会話文の原理や、敬語の成り立ちなど、幅広い射程を持っている。次年度ではなるべくこれらの点についても考察を深めていく予定である。また、この結果について、論文化するだけでなく、『平安時代語のジェンダー』という形の単行本化も構想している。

５．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　2件／うち国際共著　0件／うちオープンアクセス　0件）

１．著者名 近藤泰弘	４．巻 99
２．論文標題 明治元訳新約聖書の諸本と文体	５．発行年 2022年
３．雑誌名 国語と国文学	６．最初と最後の頁 3-20
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 近藤泰弘他	４．巻 17
２．論文標題 データから見る日本語と「性差」	５．発行年 2021年
３．雑誌名 日本語の研究	６．最初と最後の頁 49－55
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計3件（うち招待講演　3件／うち国際学会　0件）

１．発表者名 近藤泰弘
２．発表標題 コンピュータを用いた統語法研究の方法
３．学会等名 日本語文法学会第21回大会（招待講演）
４．発表年 2020年

１．発表者名 近藤泰弘
２．発表標題 平安時代語に見られるジェンダー的性質についてー通時コーパスによる分析ー
３．学会等名 日本語学会2020年度秋季大会（招待講演）
４．発表年 2020年

１．発表者名 近藤泰弘
２．発表標題 古典語の係り結びと情報構造
３．学会等名 国立国語研究所シンポジウム「係り結びと格の通方言的・通時的研究」（招待講演）
４．発表年 2020年

〔図書〕　計1件

１．著者名 近藤泰弘・澤田淳	４．発行年 2022年
２．出版社 開拓社	５．総ページ数 434
３．書名 敬語の文法と語用論	

〔産業財産権〕

〔その他〕

Japanese.gr.jp 日本語日本文学研究のために http://japanese.gr.jp

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