

令和 6 年 6 月 27 日現在

機関番号：12608

研究種目：基盤研究(C)（一般）

研究期間：2018～2023

課題番号：18K00528

研究課題名（和文）歌ことばの効果的可視化技術と通時的言語変化記述に関する基礎研究

研究課題名（英文）A fundamental study of effective visualization and description of longitudinal changes of classical Japanese poetic vocabulary

研究代表者

山元 啓史（YAMAMOTO, HIROFUMI）

東京工業大学・リベラルアーツ研究教育院・教授

研究者番号：30241756

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）：古今集から新古今集までのテキストを整理し、2つのデータセットを開発した。1つは、八代集の和歌すべてについての単語のデータである。このデータセットは、八代集の各和歌に含まれる単語を1次結合の場合と結合しない場合の分析が行なわれており、さらに各要素には国立国語研究所の分類語彙表準拠のコードを古語に対して付与されている。これは、Hachidaishu vocabulary datasetというデータ名で、Zenodo(4744170)にアップロードされている。これらのデータセットを利用し、国際学会で発表し、JCLSに原著論文を発表することができた。

研究成果の学術的意義や社会的意義

開発した2つのデータセットは、Zenodo にアップロードされ、日本だけでなく世界中の日本文学研究者に提供できた。Zenodoは、CERN（欧州原子核研究機構）によって運営されているオープンリサーチリポジトリで、研究データを簡単に公開および共有できるように設計されており、データの永続的な保存とアクセスを保証するためにDOI（デジタルオブジェクト識別子）が割り当てられているだけでなく、GitHubとの統合により、ソフトウェアリリースの自動アーカイブも可能です。本研究において、研究にしたデータを世界のリポジトリに公開することにより、日本文学を世界の研究として参画することに貢献した。

研究成果の概要（英文）：We organized the texts from KOKINSHU to SHINKOKINSHU developed two datasets. The first is the word data for all of the waka in the Yashiroshu. This dataset analyzes the words contained in each waka of the Yashiroshu with and without first-order combinations, and each element is given a code for archaic words that conforms to the classification lexicon table of the National Institute for Japanese Language and Linguistics. This data is called the Hachidaishu vocabulary dataset and has been uploaded to Zenodo (4744170). Using these datasets, I was able to present at international conferences and publish original papers on JCLS.

研究分野：言語学

キーワード：和歌 歌ことば モデリング 言語変化 八代集 データセット

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 (共通)

## 1. 研究開始当初の背景

本研究の代表者である山元は、2007年に和歌のための形態素解析ツールを開発した。このツールは和歌の文法構造を解析するものであり、その解析対象は、平安時代から鎌倉時代にかけて編纂された八代集（およそ905年から1205年）に限定されていた。本研究の大きな目的の一つは、解析可能対象を八代集から二十一代集にまで拡大することである。もう一つの目的は、これに対応する辞書を開発することである。この辞書には、二十一代集の語彙とその文法情報が含まれる。

しかしながら、従来八代集解析で用いられていた最長一致法では、すべての解析結果を一つ一つ手作業で確認する必要があるため、非常に時間と労力を要する。そこで、本研究では八代集の解析済みデータを用いて、単語の接続（つながり）に関する規則をコンピュータ処理によって自動的に獲得する。この規則を用いて、二十一代集の解析処理を行い、品詞タグづけを自動化する。この作業を通して、二十一代集から獲得された接続規則を分析する。これにより、和歌用語の語彙的結束性や接続規則の理論化を試みる。

## 2. 研究の目的

本研究の目的は、古代語の通時的変遷を俯瞰的に分析するための効果的な可視化システムと語彙データベースの開発にある。古代語は現代語とは異なり、これまで可視化技術を利用した研究がほとんど行われてこなかった。そのため、従来は限られた資料から目視による丹念な分析が主流であった。しかし、本研究では新たなアプローチとして、古代語のデータを動的に表示し、意味的類似度の計算や語と語の関係データの生成を行うシステムを構築することを目指す。

古代語の研究には、大量のデータがなく、限られた範囲で研究を進める必要があるという制約がある。また、現存するテキストの内容や語の意味が限定的であるため、現代語における自然言語処理技術のような高度な分析が難しい。このような背景から、基本的な八代集を対象にし、和歌のデータを活用して動的表示を実現することが本研究の主眼である。具体的には、和歌のデータから単語の意味的類似度を計算し、語と語の関係データを生成することで、これまで静的に切り出してきたデータとの差分を明らかにする。これにより、古代語のデータを動的に表示し、俯瞰的に分析するための基盤を構築することを目指す。また、現代語の分析に利用される自然言語処理技術は目覚ましい発展を遂げており、人工知能技術やニューラルネット、ベイズ統計学、時系列分析などの基礎技術と融合している。これらの技術を古代語研究にも応用し、データサイエンスとしての成果を収めることが期待される。最終的に、本研究は古代語の通時研究における新たな手法として、効果的な可視化システムと語彙データベースの開発を通じて学術的創造性を発揮すること

を目指している。

### 3. 研究の方法

可視化の研究の中心的な課題は、基礎となる類似対・関係対の弁別を合理的に説明できるかどうかである。関係対のデータを古代語から得るツールと類似対・関連対の数理的特徴について比較・分析し、考察を行う。二語の共起による重み (cw) を使えば、そのパターン毎の頻度分布が正規分布になることがわかっている。cw の数理的特徴について比較・分析し、考察を行う。現状では、cw の値によって描画されるエッジの数は、抽出されたテキスト量に依存するため、グラフの見通しがよくないことがある。テキスト量に依存しない cw もしくは cw に加える方法を検討する。

### 4. 研究成果

本研究では、古代語の通時的変遷を俯瞰的に分析するための効果的な可視化システムと語彙データベースの開発を行った。従来、古代語の研究においては、限られた資料から目視によって丹念に分析する方法が主流であり、可視化技術を利用した研究はほとんど行われてこなかった。

本研究の結果、単語の意味的な領域を低頻度、中頻度、高頻度に自動的に分類できる方法は存在せず、特に中頻度の単語の性質（例えば、文脈での使用頻度や意味の変動）がほぼ無視されていたことが明らかになった。用語の意味を明瞭に表すために、単語単位でなく「香り-花」のようなペア・パターンの出現頻度を計算し、共出現ウェイトが正規分布に近似することを利用した。これにより、値の上位を描画して毛玉現象を取り除き、グラフの見通しを良くした。また、値の下位を描画することで、機能語を抽出することができた。

研究に利用した八代集の処理済みテキストは、Hachidaishu vocabulary dataset として Zenodo に公開されており、ID は Zenodo(4744170) である。これらのデータセットは、研究者が古代語の研究を進める際に利用できる重要なリソースである。さらに、これらのデータセットと可視化システムの利用方法について、国際学会で発表し、その意義を説明した。また、JCLS に原著論文として発表することで、古代語研究における新たな手法の重要性を示すことができた。

## 5. 主な発表論文等

〔雑誌論文〕 計11件（うち査読付論文 8件 / うち国際共著 3件 / うちオープンアクセス 10件）

1. 著者名 山元 啓史, ホドシチェク ボル, 陳旭東	4. 巻 2022
2. 論文標題 単語アライメントの誤り対応を用いた歌ことばのコノテーション検出	5. 発行年 2022年
3. 雑誌名 じんもんこん2022論文集	6. 最初と最後の頁 111-118
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 オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Bor Hodoscek, Yamamoto Hilofumi	4. 巻 2022
2. 論文標題 Development of datasets of the Hachidaishu and tools for the understanding of the characteristics and historical evolution of classical Japanese poetic vocabulary	5. 発行年 2022年
3. 雑誌名 Digital Humanities 2022 RESPONDING TO ASIAN DIVERSITY	6. 最初と最後の頁 647
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 オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Hilofumi Yamamoto	4. 巻 version1.01
2. 論文標題 Hachidaishu vocabulary dataset	5. 発行年 2021年
3. 雑誌名 Zenodo	6. 最初と最後の頁 online
掲載論文のDOI (デジタルオブジェクト識別子) 10.5281/zenodo.4744170	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Xudong Chen, Hilofumi Yamamoto, Bor Hodoscek	4. 巻 2021
2. 論文標題 Token-based semantic vector space model for classic poetic Japanese	5. 発行年 2021年
3. 雑誌名 The 11th Conference of Japanese Association for Digital Humanities	6. 最初と最後の頁 77-81
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 オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Hilofumi Yamamoto, Bor Hodoscek	4. 巻 2021
2. 論文標題 Open source datasets of the Hachidaish for the research of classical Japanese poetic vocabulary	5. 発行年 2021年
3. 雑誌名 The 11th Conference of Japanese Association for Digital Humanities	6. 最初と最後の頁 82-87
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 オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Hilofumi Yamamoto	4. 巻 version 1.01
2. 論文標題 Hachidaishu vocabulary dataset	5. 発行年 2021年
3. 雑誌名 Zenodo	6. 最初と最後の頁 online
掲載論文のDOI (デジタルオブジェクト識別子) 10.5281/zenodo.4744170	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Hilofumi Yamamoto, Bor Hodoscek	4. 巻 2020
2. 論文標題 A fundamental study of effective visualization and description of longitudinal changes of classical Japanese poetic vocabulary	5. 発行年 2020年
3. 雑誌名 The Proceedings of the 10th Conference of the Japanese Association of Digital Humanities, JADH2020 "A New Decade in Digital Scholarship: Microcosms and Hubs"	6. 最初と最後の頁 82-86
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 オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Hilofumi Yamamoto, Bor Hodoscek	4. 巻 JADH2019
2. 論文標題 An Analysis of the Differences Between Classical and Contemporary Poetic Vocabulary of the Kokinshu	5. 発行年 2019年
3. 雑誌名 The 9th Conference of Japanese Association for Digital Humanities (JADH2019) "Localization in Global DH"	6. 最初と最後の頁 68-71
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 オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Hilofumi Yamamoto, Griffen Schwiesow	4. 巻 2019-CH-121-3
2. 論文標題 Changes in classical Japanese poetic vocabulary between the Kokinshu; and the Shinkokinshu: ame (rain) and shiguru (drizzling rain)	5. 発行年 2019年
3. 雑誌名 第121回人文科学とコンピュータ研究発表会研究報告	6. 最初と最後の頁 1-2
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 無
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 山元 啓史, ホドシチェク ボル	4. 巻 2018
2. 論文標題 八代集「桜の花」歌における作者の分類	5. 発行年 2018年
3. 雑誌名 じんもんこん2018論文集	6. 最初と最後の頁 175-180
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 オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Hilofumi Yamamoto, Bor Hodoscek	4. 巻 2018
2. 論文標題 A study on the distribution of cooccurrence weight patterns of classical Japanese poetic vocabulary	5. 発行年 2018年
3. 雑誌名 JADH2018 Proceedings of the 8th Conference of Japanese Association for Digital Humanities	6. 最初と最後の頁 179-182
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 オープンアクセスとしている (また、その予定である)	国際共著 該当する

〔学会発表〕 計5件 (うち招待講演 0件 / うち国際学会 4件)

1. 発表者名 陳 旭東(東京工業大学), ホドシチェクボル(大阪大学), 山元 啓史(東京工業大学)
2. 発表標題 単語アライメントの誤り対応を用いた歌ことばのコノテーション検出
3. 学会等名 じんもんこん2022
4. 発表年 2022年

1. 発表者名 Xudong Chen, Hiiofumi Yamamoto, Bor Hodoscek
2. 発表標題 Token-based semantic vector space model for classic poetic Japanese
3. 学会等名 The 11th Conference of Japanese Association for Digital Humanities (国際学会)
4. 発表年 2021年

1. 発表者名 Hiiofumi Yamamoto, Bor Hodoscek
2. 発表標題 Open source datasets of the Hachidaish for the research of classical Japanese poetic vocabulary
3. 学会等名 The 11th Conference of Japanese Association for Digital Humanities (国際学会)
4. 発表年 2021年

1. 発表者名 Hiiofumi Yamamoto
2. 発表標題 Co-occurrence weights and the distribution pattern of classical Japanese vocabulary
3. 学会等名 2020 Korea-Japan Symposium on Digital Humanities (国際学会)
4. 発表年 2020年

1. 発表者名 Bor Hodoscek
2. 発表標題 Visualization of classical Japanese poetic vocabulary
3. 学会等名 2020 Korea-Japan Symposium on Digital Humanities (国際学会)
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

Hachidaishu vocabulary dataset  
<https://zenodo.org/record/4744170#.YKXpl-tUuAk>  
Hachidaishu part of speech dataset  
<https://zenodo.org/record/4835806#.Ynf3D9tUthH>  
Hachidaishu vocabulary dataset  
<https://zenodo.org/record/4744170#.YKXpl-tUuAk>  
歌ことばの効果的可視化技術と通時的言語変化記述に関する基礎研究  
<https://cuckoo.js.ila.titech.ac.jp/~yamagen/waka/kaken2018.html>

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
研究分担者	ホドシチェク ボル (HODOSCEK BOR)  (10748768)	大阪大学・大学院人文学研究科(言語文化学専攻)・准教授  (14401)	

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
---------	---------