

# 語彙プロファイルに見られる八代集における変化

陳 旭東・山元 啓史（東京科学大学 環境・社会理工学院）

ホドシチェク ボル（大阪大学 大学院人文学研究科）

**概要：**本稿では、八代集の語彙変化について、言語の内容を見る視座と言語の様式を見る視座を調整するための計算手法を論じた。方法としては、同義類義の語群、同上位概念の語群、関係なしの語群の3水準で小さいサンプルを用意し、2歌集間のサンプル内の語彙変化量を語形分布の非類似度で計算した。この変化量に基づき、八代集の(1)変化最大の隣接2歌集、(2)時代区分、(3)推移のパターンを検討した。結果として、3水準の分析結果が共通し、調整の有効性はさらなる検証を要する。一方、分析結果は文学史においても合理的に説明しうる内容となった。

**キーワード：**八代集、語彙分析、和歌、言語変化

## Profile-based lexical change in the Hachidaishu

Xudong Chen / Hilofumi Yamamoto (School of Environment and Society, Institute of Science Tokyo)

Bor Hodošček (Graduate School of Humanities, Osaka University)

**Abstract:** This paper discusses computational methods for adjusting the perspectives of examining linguistic content and linguistic style in relation to lexical changes in the Hachidaishū (Eight Imperial Anthologies). As a method, small samples were prepared at three levels: groups of synonymous and near-synonymous words, groups of words sharing the same superordinate concept, and groups of unrelated words. The amount of lexical change within samples between two anthologies was calculated using the dissimilarity of word form distributions. Based on this amount of change, we examined (1) the adjacent pair of anthologies with the maximum change, (2) periodization, and (3) patterns of transition in the Hachidaishū. As a result, the analytical results at the three levels were consistent, though the effectiveness of the adjustment requires further verification. On the other hand, the analytical results proved to be content that could be rationally explained within literary history as well.

**Keywords:** the Hachidaishu, waka, lexical analysis, language change

### 1. はじめに

本研究の目的は、古典日本語の語彙分析において、全語彙によるマクロレベルの調査に代わり、関連語群の部分語彙サンプルを用いた調査手法を八代集の語彙変化に適用することである。この方法に関して関連研究と関連理論を整理し、語彙変化の研究が蓄積している八代集のテキストデータを分析する。

複数の言語変種間の語彙変異・変化は、それぞれの変種の全語彙の差集合で検討することの古典語における有効性が指摘されている [20], [48]。一方、語彙の変化・変異の比較手法には、以下の3点の拡張が可能であると考えられる。

第一に、語の推敲など様式の側面と、テーマ・トピックなど内容の側面が混在し、両者を分離して調べられていない。

第二に、「変種 A の話者はみな X を言い変種 B ではまったく言わない」といった表現 X よりも、「変種 A の話者は Y を多用し変種 B はあまり使わない」といった

数量的差異のある表現 Y が一般的だ [50] が、集合差ではこうした Y が拾えない。

第三に、集合差分では語の消滅・出現のみ把握できるものの、その消長を用いた語彙差（語彙変化）の数量化は十分に検討されていない。

そこで、2点の調整を考える。第一に、語彙差の調査では形式と内容のいずれかを統制し、片方を分析する方法を開発する。第二に、集合の差による分析に代わり、語彙における語の確率分布の差を利用し、語彙変化を測定する方法を検討する。

第一の調整については、日本語学で既に実践例がある。文献 [9], [30] は、同一作品の複数バージョンに注目して、語彙・語の変化を検討した。文献 [18], [19], [44] は、パラレルテキストで語の対応づけを行い、内容の差を捨象したうえで対応づけをもつ言語要素の間の言語変化を調べた<sup>\*1</sup>。これらの研究は、内容の統制としてパラレルコーパス相当の資料を利用しているが、パラレルコーパスは語彙史の資料としてごく限られ、特定資料への依存

<sup>\*1</sup> [19] は語彙に着目していない。

を避ける方法論も求められる。たとえば、本稿の和歌資料では異本こそあるが、同一作品の異なる「変種」は存在しない。また、既存研究の多くは特定変種の特徴的な表現や語彙交替を事例として記述するにとどまり、交替しう語の選好を数量化する、すなわち第二の調整の趣旨は十分に扱われていない。

このように、この第二の調整は古典語の語彙論の応用可能な領域の文体・位相の研究におよぶが、日本語の研究では語彙論での実践が少ない\*2。対して、計量方言学は、交替しうる表現の分布を用いて変種間の違いを定量的に測定し言語変種の分類に貢献する方法論が充実している。そこで、それらを参考としつつ、勅撰和歌集の代表である八代集における変遷について、語彙を用いた計量分析を検討する。

本稿が扱う八代集は、成立年がおおむね 300 年の範囲で均一なため、平安期の貴族社会が好んだ言語の変遷を見る手がかりとなる。表現論と語彙論では、その変遷が多く議論されてきたが、「形式・内容」を切り離して語彙変化を検討する方法論はまだ確立されていない。そこで本稿では、方法論的に形式的変化と内容的変化の 2 視座を調整する手法として、語彙変種計量論 (lexical lectometry) [39] に基づく拡張 (後述第 2.3 小節) を提案する。八代集を広義に時期別の「変種」とみなし、その変化量を測定して従来の研究の知見と照合する。さらに、視座の調整による結果の変化の有無と、変化があった場合の規則性も検討する。

## 2. 先行研究

本稿ではまず八代集の言語変化と転換点を説明し、方法論の前提として、形式的変化と内容的変化の視座調整の理論的背景と日本語語彙論の研究手法に基づく拡張可能性を先行研究に照らして整理する。

### 2.1 八代集の言語変化

八代集は、『古今和歌集』(905 年頃) から『新古今和歌集』(1205 年) までを含む 8 つの勅撰和歌集であり、同一の題材かつ成立年が比較的等間隔なため、時期ごとのスナップとして和歌語彙の通時的分析に適した資料である。和歌の言語研究では、語形の選好の交替とテーマの流行が混在し、両者の境界が曖昧である。文芸研究としては、和歌の詠まれる情景 [49]、流行 [17]、歌語辞典としては「桜」「梅」など歌語間の共出現 [16] が指摘されている。八代集の表現の転換点としては、『拾遺和歌

集』(拾遺) から『後拾遺和歌集』(後拾遺) への移行が、「褻」(私的・日常的) から「晴」(儀礼的・公式的) への転換として言及されている [17]、[49]。

一方、前述脚注でも言及したとおり、語彙を視座に歌語の変化を調査した研究では、語彙そのものではなく、語彙の各種品詞の比率や修飾語対動詞比率などの語彙指標を利用する傾向がみられる [10]、[35]、[47]。本稿の趣旨である、語の言い換えなどの特定の語群内の選好を視座とした数量化は確認されていない。また、計算文学研究として、文芸研究での八代集の転換点と、言語学 (語彙論) としての八代集の転換点を包括的に計算する架橋となる方法論的枠組みが期待される。

### 2.2 語彙変化分析における意味論的統制

前述の語彙論における「内容の統制」について述べる。日本語の語彙の研究では、コーパスの内容面面の干渉を捨象しようとする考えは文献 [30] ですでに実践されている。一方、後述する「語彙の変種計量論 (lexical lectometry)」の枠組み [5]、[7]、[8]、[39] では、「意味論的統制 (semantic control)」として理論と方法を体系的に整理している。その系統を踏まえ、日本語語史・語彙史の分析視座を改めて概観する。

#### 2.2.1 意味変化と命名変化の区分の明確化

言語変異の研究対象について、onomasiological variation と semasiological variation の二分が [7] で提唱されている。onomasiological variation は同じ意味を表す表現の変異 (命名の変化) であり、semasiological variation は同じ表現の指す意味の変異 (意味の変化) である [7]。

前者「意味の変化」には、語史研究による精緻な分析 [46]、計量的手法の開発と応用 [2]、[43] など、多様な分析が展開されている。後者「命名の変化」には、同語の異表記の選好変化として論じられる場合が多い [27]、[41]、[42]。後述にとりあげる。また、類義語同士の交替でも、最終的には意味分担の相違と使い分けに焦点が置かれる [12]、[45] \*3。このように、命名変化には常に意味変化の要素が絡むと考えられる。

一方で、日本語の語彙の集合・体系としての変化では、(1) 語彙集合・部分集合の各要素の消滅・出現と (2) その集合がもつ数量的特徴\*4の 2 点が主な分析対象となりやすい。この場合、語史のように命名変化・意味変化を区別する必要はない。文体や時期、ジェンダなどによる語彙差を論じる際も、数量的指標を利用することが多い

\*2 語彙を用いた歌語の変化については、語彙そのものではなく、各種の品詞分布、修飾語対動詞比率など、語彙から計算された二次的なマクロ指標を用いる傾向がある [10]、[35]、[47]。

\*3 語彙でなく構文・統語変異の場合、方言研究や社会言語学変異理論を視野に入れた特定表現群の交替分析が [29]、[34]、[51] など比較的多い。

\*4 各種品詞の比例、語種率、タイプ/トークン率など。

[21], [22]. 他方, 方言地理学では同義の語形分布を用いる計量方言学の研究があり, 同概念異語形の分布で方言間の距離を計算する方法が多用されている [13]. これらの手法にならない, 語史・語彙史で検討した使い分けうる類語群・関連語群を用いて語彙の違いを測る手法が語彙史でも考えられる. その考えは, 八代集の各時期にみられる語彙表現の差を分析するうえで示唆となる.

### 2.2.2 語彙的変種計量論

計量方言学の手法をより一般化した方法論として「変種計量論 (lectometry)」の枠組みが提案されている. 時期ごとの言語変化は地域方言とは異なり, 狭義の言語「変種」や日本語学でいう「位相」としては扱えないが, 変種計量論ではそのような時期差も「変種」の1つとして考慮される [8].

「変種計量論」は, 複数の言語変数<sup>\*5</sup>を収集し, それらを定量的に分析して言語変種間の距離を測定する方法論群である [37]. 言語表現の要素の1つである語彙を用いた「語彙的変種計量論 (lexical lectometry)」はその下位区分に位置し, 本稿では主にこの語彙的変種計量論の枠組みを採用する.

語彙的変種計量論では, 意味論的統制として語彙変数 (言い換えうる語形の群) を利用している. 1つの語彙変数に属す言い換えの2変種間の使用度数・率の差で2変種の変化量を計算し, 内容による変化量を変化量全体から取り除く (第3.3.1小節にて後述). こうして得られた変化量を複数の語彙変数で統合すれば, 2変種の全体的な語彙差が算出できる<sup>\*6</sup>. この測定手法は, 日本語の語彙論・語彙史ではまだ実践されていない.

### 2.2.3 2種類の語彙変数の認定アプローチ

上記「語彙変数」, または言い換えうる語形の群の認定は完全に客観にはなりえない. 本来, 変異理論の「変数」概念では, 変数内の語形はあらゆる条件で互換でき, 命題として真値が同値である必要がある. しかし, 語彙レベルでその厳格な同義判定を行うのは難しく [26], 結果として語彙変異は音韻などに比べ, 変異理論の分析の周縁に置かれがちであった [5].

変種計量論では, このような語形たちの真理条件レベルでの同値を「形式的同値 (formal equivalence)」とし, それよりルーズな同値である「概念的同値 (conceptual equivalence)」を提起している [5], [8]. 概念的同値とは, 同一のデノテーションを異なる概念として語彙化した2語の間の同義のことである. たとえば, 同じズボンについて, *breeches* とでも *pants* とでも言えるため, 形式的

同値ではない2語は, 概念的同値にあたりうる [5], [8]<sup>\*7</sup>.

その結果, [ *trouser = pants* ] のような形式的同値の語彙変数のみならず, 半ズボンとして [ *breeches = trouser = pants* ] のような概念的同値の語形たちも語彙変数であると捉えられる. さらに, 概念的同値にある変異形の選好は, 異なる変種の話者が行う言語化, とくに認知面の概念のカテゴリ化のエビデンスとして成立するとされている [5].

本稿の文脈では, このような2種類の同値による語彙変数の規定は, 意味統制として2段階での調整を可能にし, 異なる観点で語彙差を観測できるようにしたと考える.

## 2.3 語彙変化の統制の拡張の可能性

上述した語彙変化・変異の統制は語彙変種計量論の枠組みに基づくが, 他方で日本語の語彙論的研究を踏まえれば, 前掲の2段階に加えて統制の水準をさらに拡張しうる可能性がある. 具体的には, 全体的な語彙変化を対象とする代わりに, 部分語彙を取り出して詳しく検討する手法がすでに事例として報告されている.

### 2.3.1 上位概念での語群比較: 意味分野別構造分析法

意味分野別構造分析法 [40] は, 語彙の意味分野を単位に部分語彙を区分し, その比較を行う手法である. たとえば, 親族呼称や色彩表現などの意味分野ごとに2変種を比べ, 片方にのみ含まれる語を意味分野レベルで特定できる. 意味分野 (上位概念) は語彙変数に当たらないが, 意味の近さを粗く統制する点で類似すると考えられる. 一方, 2変種の語を直接比較する代わりに意味分野で整理すれば, 一定の統制をかけつつ詳細な差を捉えやすくなる.

### 2.3.2 同語異表記群の比較: 表記の変遷研究

日本語では, 同語であっても異なる表記が用いられる. 研究 [42] は「カワル・カエル」「アラワレル・アラワス」を例に表記の変遷を示し, 研究 [41] は「ハレル」「オビル」「ツナグ」などの複数表記の合一の傾向を指摘した. 研究 [27] は, 近代の二字漢語で多様な表記が衰退し統一へ向かう要因を論じる. また, 研究 [41], [42] は表記と意味の結び付きに基づくより精緻な分析を提示している. 同語異表記で括った語形の比較は, 意味統制の観点からは語彙変数を最も厳格に設定したと認識できる<sup>\*8</sup>.

<sup>\*7</sup> 踵の上までの長ズボンか, 膝を覆う程度の半ズボンか明確にできない場面が現実には多く存在する. また, 命名の実験 [25] からこの点は確認できる.

<sup>\*8</sup> ただし, 同語であっても, 表記ごとに表意機能が異なり, 使用者に別語意識をもつ場合もある. その使い分けは, 研究 [42], [43] でも議論されている. 同一の語形の多義性は, 後述の展望にて触れる.

<sup>\*5</sup> 言語変数は, 同一の内容を意味する表現群を指す [24], 188.

<sup>\*6</sup> 全体的な語彙差は, 各語彙変数で計算した変化量の平均あるいは重みづけ平均を用いる [37], [39].

### 2.3.3 まとめ：意味統制の水準の多段階化

語彙変化の特定側面を見るための統制は、ここまでに形式的同値から概念的同値の2段階を取り上げたが、日本語語彙研究では、同語異表記と意味分野の調査手法を参考に、以下のように多段階化できる\*9

統制なし → 意味分野統制 → 概念的同値語群統制 → 形式的同値語群統制 → 同語異表記群統制 → (同義トークン異表記統制)\*10

本稿では、データの制約上「ランダム語群 → 同概念語群 → 類義同義語群」に簡略化し、この軸に沿って八代集各時期の語彙を部分語彙に分け、サンプルの分布差を用いた計算で分析する。こうすることで、用語の選択の変化をテーマの選択から段階的に濾過できると想定される。最終的に、方法論的見解として、八代集の語彙変化の見え方がこの統制水準によって異なるか、異なる場合その違いが連続的か離散的かを明らかにする。

## 3. 方法

### 3.1 材料

#### 3.1.1 八代集と八代集語彙データセット

八代集のリストと本稿で用いる略称は、表1に示す。本研究では八代集語彙データセット[11]を使用した。このデータセットは、新編国歌大観CD-ROM版の二十一代集データ[38]を基に、新日本古典文学大系本二十一代集の書籍を参照して正規化と単位分割を行った。作者と歌番号は資料[32]所収の作者タグを利用し、各単語には国立国語研究所の分類語彙表[33]に準じた分類番号を付与した。同表にない固有名詞(人名、地名、表外語彙)には新規番号を与えている。八代集の分類語彙表の詳細は次節で述べる。

### 3.2 分類語彙表

八代集データセットにおける八代集用の語彙分類番号

No.	歌集名	略称	成立時期	和歌数
1	古今集	古今	ca. 905	1100
2	後撰集	後撰	ca. 951	1425
3	拾遺集	拾遺	ca. 1007	1351
4	後拾遺集	後拾遺	1086	1218
5	金葉集	金葉	ca. 1121	665
6	詞花集	詞花	ca. 1151	415
7	千載集	千載	1188	1288
8	新古今集	新古今	1205	1978

\*9 この拡張はあくまでも変種計量論の枠組みを日本語語彙研究に適用する試みであり、その枠組みの本来もつ認知社会言語学的意図とは必ずしも合致しない点に留意されたい。

\*10 括弧内は本稿では扱わない。

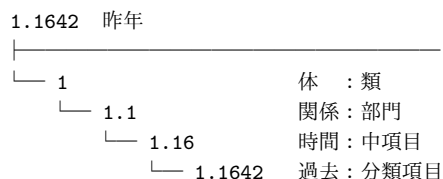


図1 分類語彙表番号の例([3]に基づく)。「昨年」の番号は1.1642である。先頭の1は体言を意味し、その下位に1.1(関係)、1.16(時間)、1.1642(過去)が階層的に分類されている。このように、番号は階層構造を反映しており、語彙の意味的・統語的な位置づけを示している。

表2 プロファイルの例:「葎」のプロファイル;数字は各時期の使用率を意味する。

	後拾遺以前	後拾遺以降
ムグラ	100%	62.5%
ヤヘムグラ	0%	37.5%

の階層を用いることで、「ランダム語群」「同概念語群」「類義同義語群」の語群をサンプリングする。

分類語彙表番号は、国立国語研究所によって2004年に編纂された日本語の大規模なシソーラスであり、階層的な意味カテゴリを伴うエントリが収録されている[3]。図1に示す通り、5桁の数字からなり、各桁が階層的な意味カテゴリを表す。

分類語彙表番号は当初、現代語への付与のみが行われていたが、後に古典語への付与も進められた[3],[31]。本稿で用いる分類は、現行の分類語彙表番号の旧版(1994年フロッピー版)[33]に準じており、現行番号とは一部異なる。しかし、数桁の数字で意味カテゴリを階層表示する仕様は共通している。さらに、和歌に特有で現代語には存在しないカテゴリを追加し、同語の異表記を細かく分類する拡張も行った。

### 3.3 手続き

語彙変化を測定するには、語彙の変種計量論の初期手法であるプロファイル基盤分析[39]が提示するプロファイルの非類似度を用い、各歌集の語彙的選好の差を数値化する。さらにクラスタリング分析と統計モデリングにより、通時的変遷の転換点(差の最も大きい隣接歌集)を調べる。本稿でのプロファイル概念の拡張、その非類似度の計算、そしてサンプリング手法を述べる。

#### 3.3.1 プロファイルと変化量の計算

プロファイル基盤分析における「プロファイル」とは、語彙変数の下位語形と変種とのクロス表で、各変種における同義・類義語形の使用頻度・率を示す(表2)。本稿では、語彙変数をさらに一般化し、3水準の意味の類似度をもつ語群を用いてプロファイルを作成する。

プロファイル非類似度は、その語群に限定した変種間の変化量を示す値である。計算手法は文献 [39] に準拠し、変種をベクトル、語形の相対頻度を各変種ベクトルの次元として扱い、2 者の 1 ノルムの距離を求める。「ムグラ・ヤヘムグラ」では、後拾遺以前 (1.00,0.00) と後拾遺以降 (0.625,0.375) の距離は  $|1.00 - 0.625| + |0.00 - 0.375| = 0.75$  となる。さらに、変化量の有意性は絶対頻度での対数尤度検定で判定し、有意でない場合変化量を 0 と見做す。

2 変種の全般の語彙差は複数のプロファイル非類似度を統合して計算される。複数の類似度を統合する方法としては、平均、もしくは、重みづけ平均があげられる<sup>\*11</sup>。本稿では単純平均を用いる。すなわち、 $n$  個のプロファイルで計算されたプロファイル非類似度  $d_1, d_2, \dots, d_n$  を用い、全体の語彙差を  $D = \frac{1}{n} \sum_{i=1}^n d_i$  とする。

### 3.3.2 3 水準の語群のサンプリング

前述のように、本稿では意味統制の水準を 3 段階に拡張した。これら 3 水準に対応する語群を用いてプロファイルのサンプルを生成し、変化量を計算する。

サンプリングは、第 3.2 小節で説明した分類語彙表番号を基に実施する (表 3)。研究 [39] では、概念的価値に基づく語群は 2~3 個の語形からなる例が多い。また、八代集の語彙中で類義・同義語群のサイズも基本的に 2~5 個に収まる。これらに合わせ、ランダム水準と同概念水準の語群のサンプルも 2~5 個の語形を含めるように設定した。

具体的には、ランダム水準の語群は、対象となる 30 の意味分野<sup>\*12</sup>の全語彙から、サイズが 2~5 の非同概念の語形の群をそれぞれのサイズで 50 回ずつ無作為抽出 (非復元抽出) し、合計で 200 イテレーションを実施した。同概念語群は、対象の 30 の意味分野ごとにサイズ 2~5 個の同概念の語形の群をそれぞれのサイズで 30 回ずつ無作為抽出 (非復元抽出) した。同義類義語群は分類番号を参考に目視で全数抽出した。最終的にランダム群 196 サンプル、同概念語群 928 サンプル、同義類義語群 43 サンプルを得て、これらを変化量の計算と分析に用いる。

### 3.3.3 変化量の分析

(1) 統計モデリング：変化量最大の隣接 2 歌集 それぞれの意味統制の水準で変化量が最大となる隣接 2 歌

集の有無と相違を調べるために、語群サンプルの変化量 (プロファイル非類似度) をハードル対数正規回帰モデル<sup>\*13</sup>で分析する。

モデルでは、固定効果として隣接する勅撰集の組 (phase)<sup>\*14</sup>と意味統制の水準、さらに二者の交互作用 (phase × 統制水準) を含める。サンプルとなる語群のサイズも統制変数として、固定効果に組み込む。ランダム効果項 (ランダム切片) として語群の属する意味分野 (profile) と語群のサンプル ID を設定する。最終的に、各統制水準ごとに phase のレベル間のコントラスト (隣接ペア間の差) を事後分布からサンプリングし、変化が特に大きい phase の有無と 3 水準での違いを検討する。

- (2) クラスタ分析：八代集語彙の分類 上記統計モデリングはサンプルごとの変化量を用いた。次に研究 [39] を踏襲し、サンプルの非類似度の平均を求めて大域的な語彙変化量を算出し分析する。八代集の語彙の分断を分類問題と見做し、クラスタ数を 3 と仮定し、非類似度行列を基に k-means 法でクラスタリングを行う。最後に、3 水準で見る八代集の語彙の分類の相違を確認する。
- (3) 多次元尺度構成法：語彙の揺れのパターンの可視化 最後に、上記の大域的な語彙変化量の行列を基に多次元尺度構成法 [23] を実施し、第 1 主軸に八つの勅撰集の語彙の相対的位置づけを  $x$  軸に、成立年順に  $y$  軸へ等間隔に配置して、可視化する。さらに、上記 2 つの分析と同様に、3 水準での相違を検討する。

## 4. 結果

### 4.1 変化量最大の隣接 2 歌集

隣接 2 歌集は統計モデルでは phase と呼び、八代集

<sup>\*13</sup> 語群サンプルのプロファイル非類似度の分布は対数正規分布に従うと観測されているが、対数正規分布に含みえない 0 値データも含まれている。このような分布にはハードル対数正規分布を適用する [4]。ハードル対数正規モデルは、非類似度が 0 かどうかをベルヌーイ分布でモデル化し、0 を超える場合の非類似度の大きさを対数正規分布でモデル化し、二者のパラメータを同時に推定する手法である。本稿では、主に対数正規分布のパラメータを用いて分析を行う。統計分析にあたり、R 統計言語 (バージョン 4.2.1; [36])、brms (バージョン 2.20.4)、stan(cmdstanr) (バージョン 2.36.0) と cmdstanr (バージョン 0.8.1) を用い、2021 年 macOS 14.3 で実施した。モデルについては、事前分布をデフォルトのままとし、MCMC (Hamiltonian Monte Carlo) を用いた。4 本のチェーンと各チェーン 4000 ステップ (うち 1000 ステップはウォームアップ) でサンプリングを行っている。 $\hat{R}$  はすべて 1.00 程度で、事後分布の有効標本サイズの Bulk 部と Tail 部も十分な値を示し、各パラメータのチェーンの混合状態は良好と判断できる。

<sup>\*14</sup> たとえば、「古今→後撰」「拾遺→後拾遺」「千載→新古今」など、7 つの組がある。

<sup>\*11</sup> 重みづけの方法については研究 [37] がプロファイルの重要度、内的整合性などに基づく計算方法を詳しくとりあげている

<sup>\*12</sup> 各和歌集において異なり語数が 2 以上存在し、かつ語群全体で延べ語数が 1 以上になるように選んだ。この基準は、データセットを 8 つの時代のサブセットに分けて 8 つの時代の変化を調べる際に、時代ごとの異なり語数が 1 か 0 になるような意味分野を避けるために考えた。

表 3 3 水準の語形集合のサンプル例

水準	説明	サンプリング基準	例	サンプル数
ランダム群	意味的統制なし	分類番号分類項目不一致	{ 花橘, 雁 }	196
同上位概念群	広義の意味分野共有 (例:「植物」語彙)	分類番号分類項目一致	{ 花橘, 稲, 桜花 }	928
類義・同義語群	狭義の概念共有 (例:「植物-葎」語彙)	分類番号同語判定+目視選別	{ 葎, 八重葎 }	43

には計 7 つの phase がある。各統制水準を通じ、それらの phase のうち他のすべての phase より変化量が大きい例は存在しなかった。

具体的に、各統制水準での phase の変化量の差は図 2 で確認できる。同図では、ランダム・同概念・同義類義の 3 水準における phase の変化量の比較のうち、95% の信用区間 (CrI) で有意に 0 とならない比較のみを提示している。すべての phase のレベル間コントラクト (計 21 対) のうち、ランダム水準は 5 対、同概念水準は 14 対、同義類義水準は 4 対に差が確認された。

それぞれの水準で共通した変化量の差としては、「拾遺→後拾遺」より「古今→後撰」が小さい (ランダム水準: Median=0.003, 95%CrI[0, 0.06]; 同概念水準: Median=0.002, 95%CrI[0.01, 0.03]; 類義語群: Median=0.006, 95%CrI[0, 0.13]) ことと、「金葉→詞花」より「古今→後撰」が小さい (ランダム群: Median=0.003, 95%CrI[0, 0.06]; 同概念群: Median=0.004, 95%CrI[0.02, 0.05]; 類義語群: Median=0.008, 95%CrI[0.01, 0.17]) こと、2 点あった。phase 間の変化量の差はすべて 0.01 未満だった。それ以外では表 4 において「詞花→千載」に正の効果が確認された (Median=0.212, 95%CrI[0.07, 0.35])\*<sup>15</sup>。また、サンプルの語群サイズには変化量への正の効果が見られた (表 4)。

## 4.2 八代集の分類

非類似度の平均をもとに計算した 2 つずつの和歌集のグローバルな語彙差の行列でクラスタリングした結果を示す (図 3)。

ランダム水準では、「古今・後撰・拾遺」「後拾遺・金葉・詞花・千載」「新古今」の 3 クラスに分類された。同義類義水準は、ランダム水準と同様な分類であった。

中間水準である同概念水準では、「古今・後撰・拾遺」「後拾遺・詞花・千載・新古今」「金葉」の 3 クラスに分類された。他の 2 水準とは、「金葉」「新古今」の扱いに相違があった。

## 4.3 語彙の揺れのパターンの可視化

多次元尺度構成法で語彙差の行列を可視化し、第 1 主

成分軸を  $x$  軸、成立年順を  $y$  軸に配置した (図 図 3)。その結果、3 水準で類似した傾向がみられ、「古今→後撰→拾遺→後拾遺→金葉」は負から正へ移行し、「拾遺→後拾遺」で 0 値を切り、「金葉→詞花→千載→新古今」は負方向と正方向を往復するパターンを示した。

## 5. 考察

### 5.1 3 水準における八代集の語彙変化の同調

統計モデリングの結果 (第 4.1 小節) では、最大の変化を示す隣接 2 歌集は見られなかったが、「古今→後撰」の語彙変化は小さく、「拾遺→後拾遺」「金葉→詞花」「詞

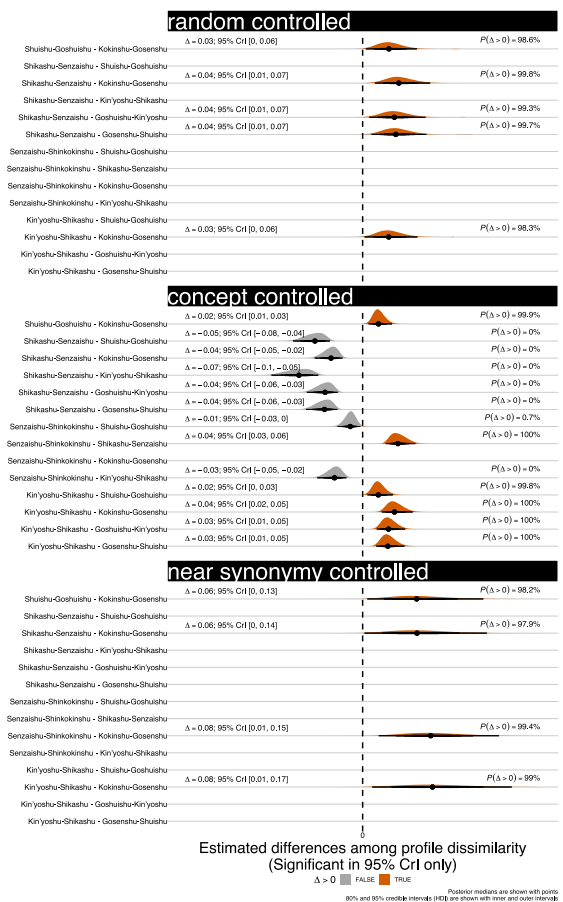


図 2 異なる統制の水準に基づく移行期間の非類似度の有意差 (95% CrI)。それぞれの有意差の事後分布の、中央値と 95% の信用区間 (CrI) を  $\Delta = [\text{<Median>, 95\% CrI [<lower.CrI>, <upper.CrI>]}$  で提示している。差が 0 より大きい確率は  $P(\Delta > 0) = \text{<possibility>}$  で提示している。

\*<sup>15</sup> モデルでは「古今→後撰」を比較の基準として効果を変換していない estimates であることに留意されたい



表 4 モデル推定値. 太字は 95%信頼区間が 0 を含まない係数.

係数	推定値	95% CrI
$\alpha_{\mu}$	-2.472	[-2.596, -2.342]
$\alpha_{hu}$	-5.043	[-5.843, -4.460]
$\beta_{size}$	<b>0.139</b>	[0.122, 0.156]
$\beta_{concept\ controlled}$	-0.055	[-0.261, 0.141]
$\beta_{near\ synonymy\ controlled}$	-0.252	[-0.578, 0.072]
$\beta_{Gosenshu-Shuishu}$	0.020	[-0.119, 0.158]
$\beta_{Shuishu-Goshuishu}$	<b>0.158</b>	[0.019, 0.296]
$\beta_{Goshuishu-Kin'yoshu}$	0.029	[-0.114, 0.169]
$\beta_{Kin'yoshu-Shikashu}$	<b>0.156</b>	[0.013, 0.297]
$\beta_{Shikashu-Senzaishu}$	<b>0.212</b>	[0.073, 0.352]
$\beta_{Senzaishu-Shinkokinshu}$	0.134	[-0.006, 0.275]
$\sigma$	0.702	[0.692, 0.714]
$\sigma_{profile}$	0.358	[0.310, 0.412]
$\sigma_{hu\ profile}$	1.342	[0.950, 1.907]
観測数	8145	
$R^2$	0.222	
周辺 $R^2$	0.094	

花→千載」の語彙変化は大きかった。このことは「後拾遺」を境目とする初期の安定性と後期の変動性を支持した。ただし、いずれも転換点といえるほどの大きな変化ではなく、漸進的とみなすべきである。

クラスタ分析 (第 4.2 小節) では、連続した勅撰集が 1 つのクラスタになりやすい結果から勅撰集の変化の連続性が推測される。同概念水準とその他 2 水準との比較では金葉と新古今のクラスタの扱いに違いがみられたことは、可視化の結果 (第 4.3 小節) の原理と関連していると考えられる。そのため、次にまとめてとりあげる。

可視化の結果 (第 4.3 小節) では、三代集の「古今→後撰→拾遺」が比較的安定していた一方、後拾遺以降は軸の正負が転換した。先行研究が示す「褻」から「晴」への体裁変化 [49] とは整合するが、同義類義水準の語形選択にも転換が見られる点は注目に値する。同義類義水

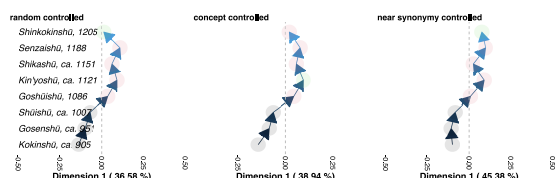


図 3 多次元尺度構成法とクラスタリングによる第 1 主成分の可視化。矢印は成立年順を示す。色はクラスタを示す。クラスタリングにより三代集である古今・後撰・拾遺、および詞花・千載が安定のクラスタを形成する一方、金葉と新古今の位置は水準により変動する。 $x$  軸での変動パターンは、「古今→後撰→拾遺→後拾遺→金葉」は負から正值へと移行し、「拾遺→後拾遺」で 0 値を切る。「金葉→詞花→千載→新古今」は、負方向と正方向の方向転換の繰り返しを示す。

準での変化は「褻→晴」として単純に解釈しがたく、さらなる検討が必要である。また、金葉まで負軸から正軸への連続的推移があったが、金葉から詞花への転換は成立年が近いにもかかわらず軸の正負が初めて反転した。これは金葉の撰者が同時代の歌人を多く採録するのに対し、詞花が後拾遺の歌人の作を多く収録したこと [28] と関連し、詞花の後拾遺寄りへの回帰を示唆すると考えられる。さらに、「金葉→詞花→千載→新古今」は正の値の範囲内で揺れを示し、(1) 同時代重視から旧時代志向への回帰、(2) 再び同時代志向への転換、(3) 最終的に新古今集における古歌を取り入れる本歌取り<sup>\*16</sup>の隆盛、といった新古の選好交替を反映している可能性がある。

全体的には、本稿で設定した 3 水準の結果はいずれも近似しており、中間的水準である同概念水準だけが異なる傾向を示した。文学史の記述とも感覚的に整合するが、本稿の分析手続き (語彙のサンプル調査) と文学史的事実が必ずしも対応しているわけではなく、偶然是排除できない。しかし、言語学的分析手法の計算文学研究への応用可能性 (主に数量的裏づけ) を示す予備的結果といえる。この異なる水準での分析結果の同調については、次節で詳述する。

## 5.2 語彙分析における意味統制の水準

統計モデリングの結果 (第 4.1 小節) によると、統制水準が異なっても大きな変化はなかった。これは、八代集は「部立」などに規定されている共通のテーマをもとに編纂され、テーマ差による変化が本来小さく、内容の統制を行っても結果は変わりにくいかもしれない。別資料での再検証が必要である。

一方、同概念水準でのみ変化量の有効なコントラストが多くみられた理由としては、次のように考えられる。同概念水準のサンプル語群は、部立など共通上位概念内の語形で構成される場合が多いため、テーマ内で語彙選択の変化を計算する傾向がある。このようにテーマ別に絞ることで、全体では目立たない時代差が顕著になりやすい。この論理では、同義類義水準の類語選択は八代集を通じて安定的ともいえる。

中間の同概念統制水準の結果は他の 2 水準とやや異なったため、語群サンプリング時の統制水準をルーズから厳格へ段階的に変えても、語彙変化が一方向に連続的に見えやすく・見えにくくなるわけではなかった。その影響が連続量的に振る舞わないことがわかった。この意味では、語彙変化の調査における意味・内容面の統制は、語彙変化の視座を離散的なものとして切り替える操作に

<sup>\*16</sup> 古歌の一部を新たな歌に取り入れる技法。

近いといえる。

### 5.3 サンプル語群のサイズと変化量の関係

第 4.1 小節 の統計モデリングでは、サイズが 2~5 個のわずかな幅の中でもサンプル語群のサイズ拡大に伴い変化量が大きくなる正の効果が確認された (表 4)。プロファイル基盤分析の変化量はサイズに敏感であることがわかった。この結果は、マクロとミクロの語彙変化量の差が必ずしも意味の統制によって生じるだけでなく、語群のサイズによっても生じる可能性を示唆する。

### 5.4 展望

本稿の手法をめぐって展望を 3 点提案する。

#### 5.4.1 意味統制の水準調節の自動化

本稿では、意味統制の水準に分類番号を用いたが、分類番号のない古典語資料も多い。分類番号に依存せず同様の手法を実現するには、語の階層的分類を自動的に生成する手法が必要である。

#### 5.4.2 意味統制の水準の精緻化

第 2.3 小節 で述べた「統制なし → 意味分野統制 → 概念的同値語群統制 → 形式的同値語群統制 → 同語異表記群統制 → 同義トークン異表記統制」の 5 段階のうち、本稿では 3 段階に簡略化して採用したが、さらに精緻な水準設定も考えられる。

また、先行研究で触れた多義性のある語形では、該当しないトークンを除外する処理が必要となる。たとえば *bank* は「銀行」や「土手」を指し、別の同義語群に割り当てることが望ましい。こうした同一語形の多義性の意味統制については [6] が詳しい。さらに、計算上での変化量の重みづけに語群内の整合性や語群の重要度を考慮する工夫 [37] も、語彙変化の把握をより精緻化しうる。

#### 5.4.3 意味以外の統制軸からみる語彙変化

本稿の検討は、意味の軸で語彙変化を段階的に把握する構想である。一方、語彙研究では、意味以外の観点で語を部分語彙にまとめ、比較する手法が発達している。たとえば「量的構造史モデル」[14], [15] は、使用度数 (高・中・低) と使用範囲 (広・中・狭) で語彙を分類し、時代別の変化を検証する。語彙史の研究では、ある視座で同一とみなす語を統合し、通時の変化を整理することによって「語彙のカテゴリー [1]」と呼ばれている。語彙のカテゴリーは、日本語語彙史の研究領域 (語の形態、語種、位相、語形成など) を多面的に含む [1]。これらの細分領域で分類した部分語彙を比較すれば、意味の軸を超えた多様な語彙変化を観測できる。本稿では意味軸のみを扱ったが、計算の型は他の軸でも応用可能である。

## 6. 結論

本研究は、八代集の語彙変化を分析するため、語彙的変種計量論の枠組みにおけるプロファイル基盤分析 [39] を基に、意味統制の水準を拡張し日本語学の語彙論の観点と切り替える手法を提案し、八代集の語彙に適用した。

具体的には、無関係語群・同概念語群・類義同義語群の低・中・高の意味統制の 3 水準を設け、八代集の各歌集でそれらのサンプル語群を収集し、語群内の語形分布の非類似度を計算した。この変化量の分析から、(1) 変化最大の隣接 2 歌集、(2) 時代区分、(3) 推移のパターンを検討した。

結果として、低・高の水準は類似し、中間水準のみが異なる傾向を示した。このことにより、プロファイルでの意味統制は視座の切り替えに近く、語彙変化の各々の側面を強調する役割を分担していると考えられる。さらに、文学史の事実とも感覚的に整合したことから、プロファイル基盤分析が文学研究の論点の裏づけに活用しうる可能性が示された。

謝辞 本研究は、日本学術振興会外国人特別研究員制度 (課題番号: ) より支援を受けました。

## 参考文献

- [1] 安部清哉: 語彙の特性から見る語彙史研究の諸相—「語彙のカテゴリー」「部分語彙」「反現象」「反作用」「中和」、人文、No. 8, pp. 246–210 (2009).
- [2] 相田太一, 小町 守, 小木曾智信, 高村大也, 持橋大地: 異なる時期での意味の違いを捉える単語分散表現の結合学習, 自然言語処理, Vol. 30, No. 2, pp. 275–303 (オンライン), 10.5715/jnlp.30.275 (2023).
- [3] Asahara, M., Ikegami, N., Suzuki, T., Ichimura, T., Kondo, A., Kato, S. and Yamazaki, M.: CHJ-WLSP: Annotation of ‘Word List by Semantic Principles’ Labels for the Corpus of Historical Japanese, *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages* (Sprugnoli, R. and Passarotti, M., eds.), Marseille, France, European Language Resources Association, pp. 31–37 (2022).
- [4] Chaudhry, S. and Heiss, A.: NGO Crackdowns and Foreign Aid - Are Donors Really Responding? Analyzing the Impact of Global Restrictions on NGOs (2018).
- [5] De Pascale, S.: Token-Based Vector Space Models as Semantic Control in Lexical Sociolectometry, PhD Thesis, KU Leuven, Leuven (2019).
- [6] De Pascale, S. and Zhang, W.: Scoring with Token-based Models, *Cognitive Sociolinguistics Revisited* (Kristiansen, G., Franco, K., De Pascale, S., Rosseel, L. and Zhang, W., eds.), De Gruyter, pp. 186–200 (online), 10.1515/9783110733945-016 (2021).
- [7] Geeraerts, D., Grondelaers, S. and Bakema, P.: *The Structure of Lexical Variation: Meaning, Naming, and Context*, Cognitive Linguistics Research, No. 5, Mouton de Gruyter, Berlin (1994).



- [8] Geeraerts, D., Speelman, D., Heylen, K., Montes, M., De Pascale, S., Franco, K. and Lang, M.: Lexical Variation and the Lexeme-Lecture-Lect Triangle, *Lexical Variation and Change: A Distributional Semantic Approach* (Geeraerts, D., Speelman, D., Heylen, K., Montes, M., De Pascale, S., Franco, K. and Lang, M., eds.), Oxford University Press, pp. 3–31 (online), 10.1093/oso/9780198890676.003.0001 (2023).
- [9] 蓮見陽子: 同一情報に基づく文章表現の異同についての分析, 計量国語学, Vol. 18, No. 3, pp. 136–144 (1991).
- [10] 波多野完治: 和歌と俳句, 文章心理學の問題, 三省堂, 東京, pp. 47–62 (1941).
- [11] Hodošček, B. and Yamamoto, H.: Development of Datasets of the Hachidaishū and Tools for the Understanding of the Characteristics and Historical Evolution of Classical Japanese Poetic Vocabulary, *Digital Humanities 2022 Conference Abstracts*, Tokyo, The University of Tokyo, pp. 647–648 (2022).
- [12] 池上 尚: 嗅覚表現名詞カ・ニホヒの史的変遷 -その交替に着目して-, 早稲田日本語研究, Vol. 21, pp. 75–85 (2012).
- [13] Inoue, F. and Kasai, H.: Dialect Classification by Standard Japanese Forms, *Japanese Quantitative Linguistics*, Vol. 39, pp. 220–235 (1989).
- [14] 伊藤雅光: 語彙の量的構造史モデル, 日本語の研究, Vol. 4, No. 1, pp. 112–129 (オンライン), 10.20666/nihongonokenkyu.4.1\_112 (2008).
- [15] 伊藤雅光: 計量語彙論から見た語彙史, 岩波書店 (2009).
- [16] 片桐洋一: 歌枕歌ことば辞典, 角川書店, 東京 (1983).
- [17] 川村晃生: 撰関期和歌史の研究, 博士論文, 慶應義塾大学, 東京 (1991).
- [18] 北崎勇帆: 洒落本を対象とした東西対照コーパスの設計と構築, 研究報告人文科学とコンピュータ (CH), Vol. 2015-CH-106 (5), 情報処理学会, pp. 1–6 (2015).
- [19] 北崎勇帆: 平家物語の対照コーパスを通して見る中世日本語, 計量国語学, Vol. 34, No. 7, pp. 483–497 (2024).
- [20] 近藤泰弘: 平安時代の漢文訓読語の分類, 訓点語と訓点資料, Vol. 127, pp. 120–131 (2011).
- [21] 近藤明日子: 近代文語文の通時的变化の分析: 語種率・品詞率に着目して, 言語資源活用ワークショップ発表論文集, Vol. 1, pp. 355–364 (2018).
- [22] 近藤明日子: 明治・大正期の口語体会話文の位相差: 語種率・品詞率を観点として, 言語資源活用ワークショップ発表論文集, Vol. 4, pp. 119–127 (2020).
- [23] Kruskal, J. B.: Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis, *Psychometrika*, Vol. 29, No. 1, pp. 1–27 (online), 10.1007/BF02289565 (1964).
- [24] Labov, W.: *Sociolinguistic Patterns*, University of Pennsylvania Press (1972).
- [25] Labov, W.: The Boundaries of Words and Their Meanings, *New Ways of Analyzing Variation in English* (Bailey, C. J. N. and Shuy, R. W., eds.), Georgetown University Press, Washington, pp. 340–373 (1973).
- [26] Lavandera, B. R.: Where Does the Sociolinguistic Variable Stop?, *Language in Society*, Vol. 7, No. 2, pp. 171–182 (online), 10/d82gxm (1978).
- [27] 間瀬洋子: 近代二字漢語における同語異表記の実態と変化: 形態論情報付きコーパスを用いて, 計量国語学 = Mathematical linguistics: 計量国語学会機関誌, Vol. 30, No. 5, pp. 257–274 (2016).
- [28] 松田武夫 (編): 詞華和歌集, 岩波書店, 東京 (1939).
- [29] 松田謙次郎: 岡崎敬語調査に見る「足りない」～「足らない」の変異と変化, 計量国語学, Vol. 32, No. 2, pp. 66–81 (オンライン), 10.24701/mathling.32.2\_66 (2019).
- [30] 宮島達夫: 「共産党宣言」の訳語, 言語の研究 (言語学研究会, 編), むぎ書房, 東京 (1979).
- [31] 宮島達夫, 鈴木 泰, 石井久雄, 安部清哉: 日本古典対照分類語彙表, 笠間書院 (2014).
- [32] 中村康夫, 立川美彦, 杉田まゆ子: 国文学研究資料館データベース古典コレクション『二十一代集』(正保版) CD-ROM, 岩波書店, 東京 (1999).
- [33] 中野 洋, 林 大, 石井久雄, 山崎 誠, 石井正彦, 加藤安彦, 宮島達夫, 鶴岡昭夫: 分類語彙表 フロッピー版, 国立国語研究所言語処理データ集 5, 大日本書店, 東京 (1994).
- [34] 南部智史: 定量的分析に基づく「が／の」交替再考, 言語研究, Vol. 131, pp. 115–149 (オンライン), 10.11435/gengo.131.0\_115 (2007).
- [35] 西端幸雄: 和歌と語彙: 八代集和歌語彙の変遷, 樟蔭国文学, Vol. 29, pp. 83–106 (1992).
- [36] R Core Team: R: A Language and Environment for Statistical Computing (2022).
- [37] Ruette, T., Geeraerts, D., Peirsman, Y. and Speelman, D.: Semantic Weighting Mechanisms in Scalable Lexical Sociolinguistics, *Aggregating Dialectology, Typology, and Register Analysis* (Szmrecsanyi, B. and Wälchli, B., eds.), DE GRUYTER, pp. 205–230 (online), 10.1515/9783110317558.205 (2014).
- [38] 新編国歌大観編集委員会 (編): CDROM 版新編国歌大観, 角川書店, 東京 (1996).
- [39] Speelman, D., Grondelaers, S. and Geeraerts, D.: Profile-Based Linguistic Uniformity as a Generic Method for Comparing Language Varieties, *Computers and the Humanities*, Vol. 37, No. 3, pp. 317–337 (online), 10.1023/a:1025019216574 (2003).
- [40] 田島毓堂: 語彙研究法としての「意味分野別構造分析法」のために—阪倉論文の顕彰と批判, 国語と国文学, Vol. 77, No. 6, pp. 43–56 (2000).
- [41] 高橋雄太: 近代における和語の表記の変遷 -複数表記から単一表記へ-, 国際日本学研究所論集, Vol. 4, pp. 37–48 (2016).
- [42] 高橋雄太: 近代における和語の用字法の変化 -カワル・カエルとアラワレル・アラワスを中心に-, 明治大学人文科学研究所紀要, Vol. 85, pp. 173–208 (2019).
- [43] 高橋雄太, 相田太一, 小木曾智信: 単語分散表現による日本語和語動詞の書き分けの研究, デジタル・ヒューマニティーズ, Vol. 4, No. 1, pp. 1–11 (オンライン), 10.24576/jadh.4.1\_1 (2025).
- [44] 田中牧郎, 山元啓史: 『今昔物語集』と『宇治拾遺物語』同文説話における語の対応 —語の文体的価値の記述—, 日本語の研究, Vol. 10, No. 1, pp. 16–31 (2014).
- [45] 田中牧郎: 「すぐれる」「まさる」の意味変化 —コーパスによる語史研究—, 明治大学国際日本学研究, Vol. 12, No. 1, pp. 31–45 (2020).
- [46] 田野村忠温: 「科学」の語史: 漸次的・段階的変貌と普及の様相, 大阪大学大学院文学研究科紀要, Vol. 56, pp. 123–181 (オンライン), 10.18910/56924 (2016).
- [47] 辻 勝美: 歌語の研究史—現状と展望, 歌ことばの歴史 (小町谷照彦, 三角洋一, 編), 笠間書院, 東京, pp. 217–238 (1998).
- [48] 築島 裕: 訓点資料に見える『和文特有語』について,

文化言語学: その提言と建設 (文化言語学編集委員会, 編), 三省堂, 東京 (1992).

- [49] 上野 理: 後拾遺集前後, 笠間書院, 東京 (1976).
- [50] Wolfram, W.: *Social Varieties of American English, Language in the USA: Themes for the Twenty-first Century* (Finegan, E. and Rickford, J. R., eds.), Cambridge University Press, Cambridge, pp. 58-75 (online), [10.1017/CB09780511809880.006](https://doi.org/10.1017/CB09780511809880.006) (2004).
- [51] 山田彬堯: 江戸期から昭和期にかけての尊敬語のバリエーション: 多項混合効果ロジスティック回帰によるベイズ統計学的アプローチ, 言語文化共同研究プロジェクト, No. 2020, pp. 5-25 (2021).