

文部省科学研究費補助金
1998年度研究成果報告書

特定領域研究

人文科学とコンピュータ

—コンピュータ支援による人文科学研究の推進—



1999年3月

平安時代和歌資料における特殊語彙抽出についての 計量的研究と利用ツールの公開

研究代表者：近藤みゆき（千葉大学文学部・助教授）

代表者連絡先：

〒263-8522 千葉市稲毛区弥生町1-33

Tel.043-290-3641 Fax. 043-290-3641

Email: mkondo@bun.l.chiba-u.ac.jp

1 古今和歌集の歌語と表現のジェンダー性について

1.1 これまでの研究と98年度の研究概要

平安和歌研究の課題の一つに、和歌表現特有の語彙—いわゆる歌語、を対象とした研究がある。平安和歌の歌語は、古今和歌集和歌における語彙を一つの美的規範としながら、三代集時代、院政期、平安末～新古今時代と、画期ごとに、万葉集や漢文訓読語・俗語・物語の用語など外部からの語彙を取り込みつつ、活性化し史的展開を遂げていく。ゆえに古今の歌語がどのように継承されたか、また、どのような歌語が新たに開拓されたかを、表現を開拓した歌人群、史的背景とあわせてあきらかにする事は和歌研究においてきわめて重要な作業といえる。和歌研究の分野では、歌語・表現の研究が盛んであるにもかかわらず、データベース化によった網羅的な表現研究は、未だ行われていない。平安和歌資料を、歌語をはじめとする表現の分析に適した形でデータベース化する方法を検討し、また具体的な分析結果を提示することは、和歌研究に大きな進展をもたらすことになると思われる。

このような観点から、97年度には『古今和歌六帖』を分析の対象とした。同作品は、平安中期の成立、作歌の手引書として作られた類題和歌集で、総歌数は4,494首（桂宮本）を数える。平安中期の特殊歌語の宝庫と言ってよい。古今六帖全歌に形態素解析タグを付してこれをデータベース化し、総語彙を網羅するとともに出現頻度数でソートして、最も特殊な語彙として出現頻度数1の語彙2,295語を得て、内容が万葉語、古歌や伝承歌の用語、古今・後撰の一度語や先行歌人の特異用語、歌謡での用語など、様々な位相の語彙の集合体であること

を明らかにした。またそれら語彙の、同時代私家集における使用数の分布を調べ、六帖語彙が、新しい歌語として、同時代歌人に受容されていた実態を明らかにした。

また同じく97年度には、『相模集』を対象として、データベースを作成した。総歌数は557首であり、平安時代の代表的女流歌人の私家集である。これは「じんもんこんデータベース」の一環として、CD-ROMにより刊行した。詳細は、当該CD-ROMを参照されたい。

こうした六帖特殊歌語の研究とデータベース作成の研究とあわせて、今年度は新たに、和歌表現・和歌語彙におけるジェンダー性の分析というテーマに着手した。和歌という表現形式の特質の一つに、それが詠歌主体の性—男性の歌であるのか、女性の歌であるのか—を大きく反映する文学形態であるということがある。折口信夫以来、多くの論や研究の続いている「女歌」論は、この問題を扱った研究を代表するものだが、そうした文芸論・文学論とあわせて、性差が、語彙や用語にどう現れているのか、具体的・数量的に明らかにしていく研究が、一方で必要であると考えられる。「女歌」論が集中してきた恋歌だけではなく、和歌全般に範囲を広げて、男の歌と女の歌の用語を計量的に比較し、平安和歌における、男性的な言葉の型・女性的な言葉の型を明らかにしたい。

なお、従来の和歌研究におけるジェンダー性についての文献の一部を次に示しておく。

- 折口信夫「巫女としての女性」（折口信夫全集第一巻『万葉集研究』）
- 鈴木日出男「相聞歌の展開」『古代和歌史論』、1990年
- 後藤祥子「女流による男歌」（『平安文学論集』1992年）

1.2 n グラム統計処理と長尾・森プログラム

それでは、王朝和歌という言語表現を対象として、語彙や表現のジェンダー性を認定するには、どのような方法があるだろうか。古代の雅語に対しては、現代の研究者の内省はほとんど通用しないと言って良い。ここでは大量データの計量分析を、内省に匹敵する有力な手段として、王朝和歌の言語表現から、和歌におけるジェンダーを反映した用語・語彙を出来るだけ多く抽出する方法を考えたい。

まず男性歌人・女性歌人それぞれの性別に語彙・句法などを抽出し、それぞれが多く用いている語彙や表現、また一方が他方に比して多く用いている語彙や表現を、網羅的・計量的に抽出することが必要である。そしてその場合の言語処理方法として、ここで注目したいのが、京都大学の長尾真・森信介両氏の開発した日本語語句の自動抽出プログラムである（以下、長尾・森プログラムと称する）。そもそもこのプログラムはnグラム統計（シャノンの情報理論において展開された言語分析のための理論・テキスト中で任意の長さの文字列を抽出し、その出現頻度を求める）を大規模テキストに対して行うためのプログラムとして開発されたものであるが、nグラム統計の性質上、そのテキスト上で、高頻度に出現する文字列や、逆に、きわめてまれに出現する文字列などを求めることができる。両氏の研究では、現代語の大量データの分析結果についての例が報告されているが、本研究ではこれを平安時代和歌資料のデータ分析に応用してみることとする。なお、nグラム統計の理論的側面および当該プログラムの内容については、次の参考文献を参照されたい。

- Claude E. Shannon & Warren Weaver, The Mathematical Theory Of Communication, The University Of Illinois Press, 1949
- 長尾真・森信介、「大規模日本語テキストのnグラム統計の作り方と語句の自動抽出」（自然言語処理 96-1、1993）

この長尾・森プログラムによるnグラムの統計を、言語処理に利用すると、次のような多くのメリットがある。

1. 大規模テキストに対して、高速に分析が可能であること。
2. nグラムを任意の長さで設定できること。
3. 単なるプレーンテキストから、語彙を含むさまざまな文字列を抽出することが可能であること。

そもそも、nグラム統計においては、テキスト中の文字列を、単語であるかどうかにかかわらず特定の長さごとにすべて列挙し、頻度を計る。例えば、『古今集』男性歌人の2グラムであれば、次のような出現頻度になる。

【古今集男性歌人2グラムの文字列例】

157 ひと
150 はな
136 やま
131 おも
97 ころ
92 りけ
92 なく

※頭の数字157、150、136、97、92、96は出現頻度数を示す（以下同じ）

「ひと」が157回、「はな」が150回、また単語ではない「りけ」という文字列も92回出現することがわかる。このように単なる偶然による連鎖である文字列まで網羅することは、文学研究の上では、邪魔になる場合もあるが、逆に、次に述べるように、テキストを単語に分割（形態素解析）しなくても分析できる結果、様々なメリットがあるという側面もあるのである。

1.3 nグラムで和歌データを分析することの利点

このnグラム統計を和歌データの分析に応用すると、和歌データを計算機処理する際の、次のような問題を解決できることになる。

- 形態素タグを付ける作業を必要としない。

解釈の安定しない古典和歌の、すべての歌集（勅撰集・私家集・私撰集）を対象に、和歌一首ごとに統一のとれた形態素タグを付けることつまり単語を認定して文節に切る作業を行なうことは不可能に近い。特に私家集群など本文上の問題もある作品は清濁を付すことさえ決定できない場合も少なくない。古典和歌の語彙分析用のデータベースを作成する際の最大の問題は、解釈を確定し統一タグを付けることがたいへん困難であるという点に尽きる。この方式であれば、全文清濁無しの平仮名書き本文データから単語・擬似単語を抽出出来るので、このタグ付け作業の困難さをすべて解消できるのである。形態

素タグ付きのデータで特徴語句・語彙を厳密に析出する研究は、理想ではあるが、時間と労力は比較にならない。

- 掛詞処理

形態素解析してしまうと、掛詞のように、複数の形態素を兼ね備えた部分の分析が一面的になってしまうが、文字列のままならば、かえってそのような問題はなく、単純でかつ見落としなく、掛詞に関する要素抽出が可能である。

- 文字表記の統一が容易である

n グラムという形で、文字列を抽出する方法は、分析対象となるデータが、漢字・片仮名・平仮名のいずれにせよ、表記体系の統一されたものであればあるほど結果の精度は高くなる。漢字仮名交じり表記が多く、しかも、複数の作品間では表記方針がほとんど不統一な現代文に対して、和歌資料は、写本など、もともと原テキストが平仮名表記を基本としており、平仮名表記で統一をとることが極めて容易である。結果として、簡便にこのプログラムを使用してきて、しかも高い精度の結果を得ることが期待出来る。

- 長い文字列の慣用的表現や類句の抽出が可能である
長い n を設定することが自由であるので、たとえば n = 8 で設定すると、同じ『古今集』男性歌人で次のような文字列を検出することが出来る。

【古今集男性歌人 8 グラムの文字列例】

- 4 かみなづきしぐれ
- 3 やまのさくらばな
- 3 ものにぞありける
- 3 むちはやぶるかみ
- 3 はなすすきほにい
- 3 めものにぞありけ
- 3 なにこそありけれ
- 3 ぞあやまたれける
- 3 あらじとおもへば
- 3 あしひきのやまの
- 2 をみなへしおほか

「かみなづきしぐれ」が 4 回、「やまのさくらばな」「ものにぞありける」「むちはやぶるかみ」「はなすす

きほにい」「めものにぞありけれ」「ぞあやまたれける」「あらじとおもへば」「あしひきのやまの」がそれぞれ 3 回、「をみなへしおほか」が 2 回出現することがわかるが、これらの多くは、語彙ではなく、「特定のことは続き」、とでも称すべき表現である。このような特定のことは続きの表現は、文節を単位とする通常の索引では、「やまのさくらばな」は「やま」「の」「さくらばな」、「ものにぞありける」は「もの」「に」「ぞ」「あり」「ける」のように、最も小さい単語の単位に分類されてしまうので、一括して検出することは非常に難しい。しかしこの方式であれば、それらを容易に、かつ遺漏なく抽出できるのである。

以上のように、この方法によれば、比較的容易に、男性・女性のそれぞれの和歌から、ジェンダー性を反映した語彙やことは続き、言い回しなど、様々なレベルの表現を抽出することができると考えられる。そこで次に「古今和歌集」を対象に検証した結果を報告する。

1.4 古今和歌集の分析

今回は、ジェンダー性をもった和歌語彙を抽出する分析のモデルとして、古今和歌集を対象とした研究を行った。古今和歌集は平安時代に第一番目に成立した勅撰和歌集で、平安文学への影響力はきわめて大きく、特にその歌語・表現は、それに続く王朝和歌の基準となったことが知られている（例えば、小町谷照彦『古今和歌集と歌ことは表現』岩波書店、1994 年など参照）。よって、これを機械可読形式にデータ化し、あわせて詠作者の性別を示すタグを歌頭に付すことにした。なお、上記で繰り返し述べたように、n グラム統計には形態素タグは不要なので、今回もそれらはいっさい付していない。作業内容とタグ・記号の詳細は次の通りである。

古今和歌集のデータ化方針

1. 入力本文：岩波書店 日本文学大系『古今和歌集』（梅沢本）

歌数：1111 首

n グラムによる分析用に、すべて平仮名表記にあためた。（なお、東大国語研究室有志によるテキストデータを参考にした）

2. 性別のタグ付け

作者名に応じて性別のタグを付した。男性歌人は m（male の頭文字）、女性歌人は f（female の頭文字）、

読人不知は n（neutral の頭文字）とした。また次の歌のように、表示自体は「読人不知」でも、左注などで作者が示唆されている場合、男性は m@、女性は f@とし、今回はそれぞれの性に含めて統計した。

【例】古今和歌集・7 番歌

題しらず 読人不知
心ざしふかく染めてしおりければ消えあへぬ雪の花
とみゆらん

ある人の曰く、前太政大臣の歌也
このような場合には m@ とする。

3. 性別の後に、和歌集名の「こきん」と歌番号をつけた。歌末には「／」をつけた。

性別タグ付け本文例

春歌上

- m, こきん 0001 としのうちにはるはきにけりひとと
せをこぞとやいはむことしとやいはむ／
m, こきん 0002 そでひちてむすびしみづのこほれる
をはるたつけふのかぜやとくらむ。／
n, こきん 0003 はるがすみたてるやいづこみよしの
のよしののやまにゆきはふりつつ／
f, こきん 0004 ゆきのうちにはるはきにけりうぐひす
のこほれるなみだいまやとくらむ／
n, こきん 0005 うめーがえにきゐるうぐひすはるか
けてなけどもいまだゆきはふりつつ／
m, こきん 0006 はる、たてばはなとやみらむしらゆ
きのかかれるえだにうぐひすのなく／
m@, こきん 0007 こころざしふかくそめてしをりけ
ればきえあへぬゆきのはなとみゆらむ／
m, こきん 0008 はるのひのひかりにあたるわれなれ
どかしらのゆきとなるぞわびしき／
m, こきん 0009 かすみたちこのめもはるのゆきふれ
ばはななきさともはなぞちりける／
m, こきん 0010 はるやときはなやおそきとききわか
むうぐひすだにもなかずもあるかな／
m, こきん 0011 はるきぬとひとはいへどもうぐひす
のなかぬかざりはあらじとぞおもふ／
m, こきん 0012 たにかぜにとくるこほりのひまごと
にうちいづるなみやはるのはつはな／
m, こきん 0013 はなのかをかぜのたよりにたぐへて
ぞうぐひすさそふしるべにはやる／
m, こきん 0014 うぐひすのたによりいづるこゑなく
ははるくることをたれかしらまし／
m, こきん 0015 はるたてどはなもにほはぬやまざと

- はものうかるねにうぐひすぞなく／
n, こきん 0016 のべちかくいへゐしせればうぐひす
のなくなるこゑはあさなあさなきく／
n, こきん 0017 かすがのはけふはなやきそわかくさ
のつまもこもれりわれもこもれり／
n, こきん 0018 みやまにはまつのゆきだにきえなく
にみやこはのべのわかなつみけり／
n, こきん 0019 かすがののとぶひののもりいでてみ
よいまいくかありてわかなつみてむ／
n, こきん 0020 あづさゆみおしてはるさめけふふり
ぬあすさへふらばわかなつみてむ／

1.5 それぞれの n グラムの値（出現頻度）

上記のデータを対象に、男女の歌人別に、2～15 グラムまでをとった異なり数値の結果が、次の表 1 である。当該表で、たとえば 2－男性－2 3 0 3 とある「2 3 0 3」の数値は、文字列の異なり数をあらわしている。すなわち、2 グラムだと、男性歌人の作品の場合「ひと」「はな」「やま」などをはじめとして、2 3 0 3 通りの文字列が存在したということになる。

表 1

n グラム	男性	女性	読人不知
2	2 3 0 3	1 2 0 1	2 1 6 1
3	1 0 5 1 1	2 3 8 3	8 4 3 1
4	1 5 7 5 7	2 7 1 8	1 1 6 5 8
5	1 7 8 8 9	2 8 0 3	1 2 8 0 9
6	1 8 8 6 5	2 8 3 0	1 3 4 8 5
7	1 9 2 7 4	2 8 3 8	1 3 4 8 5
8	1 9 4 9 8	2 8 4 4	1 3 5 9 2
9	1 9 5 8 1	2 8 4 6	1 3 6 4 0
1 0	1 9 6 1 3	2 8 4 8	1 3 6 6 0
1 1	1 9 6 2 7	2 8 4 9	1 3 6 7 4
1 2	1 9 6 3 4	2 8 5 0	1 3 6 8 6
1 3	<u>1 9 6 4 1</u>	2 8 5 1	1 3 6 9 5
1 4	1 9 6 4 1	2 8 5 2	1 3 1 7 0
1 5	1 9 6 4 1	2 8 5 3	1 3 7 0 2
1 6	1 9 6 4 1	2 8 5 4	1 3 7 0 3
1 7		2 8 5 5	1 3 7 0 4
1 8		2 8 5 6	1 3 7 0 5
1 9		<u>2 8 5 7</u>	1 3 7 0 6
2 0		2 8 5 7	1 3 7 0 7
2 1			1 3 7 0 8

2 2	1 3 7 0 9
2 3	1 3 7 1 0
2 4	1 3 7 1 1
2 5	<u>1 3 7 1 2</u>
2 6	1 3 7 1 2

1.6 n グラムの飽和と考察の範囲

ところで、考察の対象とする n グラムの範囲であるが、1 グラムには、1 音節の単語「き」「め」なども含まれるものの大半は文字の頻度であり利用しにくいので今回は考察の対象とはしない。2 グラムからは「ひと」「はな」「やま」など、歌の表現の一部として意味のある文字列が存在する。そして、だんだんと n の値を増加させてゆくと、男性は 1 3 グラム、女性は 1 9 グラム、読人不知は 2 5 グラムで、すべての文字列の出現頻度が 1 となる（表 1 で下線を施した部分）。そしてこの、すべての文字列の出現頻度が 1 になった段階で、それ以上の n グラムをとっても、次の例のように、頻度統計としては意味をもたないことになる。

【男性歌人 1 3 グラム例】

1 ををみなへしうしろめたくも
1 ををしむべらなりもろともに
1 ををしみしたゆふひものむす
1 ををしまぬひとしなければさ
1 をわれはへにけむはなみつ
1 をわたらむもみぢばはあめと

【男性歌人 1 4 グラムの例】

1 ををみなへしうしろめたくもみ
1 ををしむべらなりもろともにな
1 ををしみしたゆふひものむすば
1 ををしまぬひとしなければさく
1 をわれはへにけむはなみつつひ
1 をわたらむもみぢばはあめとふ

【男性 1 5 グラムの例】

1 ををみなへしうしろめたくもみゆ
1 ををしむべらなりもろともになき
1 ををしみしたゆふひものむすばほ
1 ををしまぬひとしなければさくら
1 をわれはへにけむはなみつつひと
1 をわたらむもみぢばはあめとふる

このように、すべての文字列の出現頻度が 1 になると、それ以降は、1 3 グラム「ををみなへしうしろめたくも」→1 4 グラム「ををみなへしうしろめたくもみゆ」と、実態としては、既に出現している各文字列が 1 字（1 グラム）ずつ増加していくにすぎないのである。このような、文字列出現頻度がすべて 1 になる段階を、ここでは仮に n グラムの飽和と呼ぶことにする。文字列出現頻度を問題とする場合は、飽和値以前が考察対象となるわけで、それは古今集和歌データの場合、男性歌では 1 2 グラム、女性歌では 1 8 グラム、読人不知歌では 2 4 グラムまでということになる。

言い方を変えれば、男性歌では 1 2 グラム、女性歌では 1 8 グラム、読人不知歌では 2 4 グラムまでは、文字列が複数回出現するという結果が出た訳だが、長歌という例外はあるが、一般には和歌は三十一文字一すなわち 3 1 グラムで完結するので、1 2・1 8・2 4 グラムとは、歌一首の三分の一から三分の二に相当することになる。このような長い文字列の一致を見るパターンとは、簡単に言えば類似度の高い歌が抽出されるということになるが、一口に類似度が高いと言っても、そこには異伝歌や類歌の問題もあれば、歌人相互の影響関係として考察するのが適当な例もあるなど、実態は一通りではない。これについては 2 節で述べることにし、次には、ここで求められた n グラムが飽和するまでに範囲を定めて、男性・女性の特有表現を抽出する。

1.7 各グラム別に見た男性特有文字列と女性特有文字列の抽出

さて、以上のような範囲で n グラムを求めるとして、次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を試みる。その手順は次のようになる。

1. 元データの性別タグが m になっているもの（男性歌人の歌）と f になっているもの（女性歌人の歌）をそれぞれ取り出し、そのグループごとに n グラムを計算する。（既に算出して提示したもの）
2. 次に同じ n グラムの男性グループと女性グループの文字列群相互を比較し、その相互の差分を求める。（差分の抽出と操作には UNIX の標準ユーティリティの comm と join を用いた。）男性文字列から、女性文字列と共通するものを引き算したものが、男性特有文字列であり、その逆が女性文字列となる。

男性 — 女性 = 男性特有文字列
女性 — 男性 = 女性特有文字列

以上のような手順によって男性特有文字列と女性特有文字列を求めた。ただし、ここで一つ問題点がある。古今和歌集では、女性歌人の用例数が予想以上に少なかったということである。タグ付け後の性別歌数は次に示すとおりである。

- 男性歌人 — 595 首
- 女性歌人 — 87 首
- 読人不知 — 429 首

古今時代、名のある女流歌人は極めて少なく、総数として100首を割り込んでいる。ここから男女別にnグラムをとり、さらに上記のように特有文字列を求めた結果、女性特有文字列の方には、ほとんどが出現頻度1のものばかりになってしまっていて、その結果はあまり意味のないものになってしまうのである。たとえば、次は女性特有文字列のn=5の例の「あ」で始まる文字列の一部である。

1 あかざりし
1 あかれやは
1 あきのもみ
1 あきのよも
1 あけぬるも
1 あさぢには
1 あさつゆの
1 あさなけに
1 あさなわが
1 あしたゆく
1 あしもやす
1 あぢきなし
1 あはむつき
1 あはれとは
1 あひくるみ
1 あひそふる
1 あひてもの
1 あひにあひ
1 あひみぬも
1 あふたのみ
1 あふといへ
1 あへずたぎ

以上のようにいずれも頻度1であることから、女性特有であることの有意性を論じることはほとんど意味をも

たない。それに対して、男性特有の文字列においては、次に述べるように、多くの有効な結果を得ることが出来きたために、それを用いて論じて行くこととしたい。

1.8 男性特有のことば・表現

男性歌の特有文字列は、

n = 2で1169
n = 3で9162
n = 4で15053
n = 5で17522
n = 6で18624
n = 7で19096
n = 8で19353
n = 9で19452
n = 10で19488

の異なり数をもとめることが出来た。これらの文字列には、単語・歌語・表現として捉えることのできるものも非常に多く含まれている。相当大部になるので、データ処理結果の全体をあげることは出来ないが、単語としてのまとまりの多く見られるn=4、n=5、n=6、n=7から、その一部をあげて次節で述べることにする。

1.9 考察

古今和歌集において、男性だけが用いていることばがここに網羅・抽出できたわけであるが、単語や文の断片といった文字列も含まれてはいるが、まとまりや意味のあるものも相当数存在する。具体例をあげてみよう。ことばの種類からみて、次の2つのものが得られる。

1. 歌語

n = 4：もみぢば・しらゆき・あきのの・しらくも・しらつゆ・たなばた・しらやま・しらたま・よろづよ・やまかぜ・みよしの・みなづき・ふくかぜ・たまのを・かよひぢ・うつせみ・あらたま・わびびと
n = 5：をみなへし・うめのはな・あしひきの・ちはやぶる・たつたがは・かみなづき・あまのかは・わすれぐさ・やまたかみ・ふじばかま・なみだがは・きりぎりす・よしのがは・はなのかけ

2. 特定のことば続きをもった表現。

- n = 4：あふこと・まにまに・もしらず

- n = 5 : ならなくに・やはあらぬ・もあるかな・わがこひは・わがやどの・やわたらむ・にけるかな・よをさむみ・よそにのみ・わがこころ・はるくれば
- n = 6 : おもふこころ・ものをおもふ・をおもふこころ・わがみひとつ
- n = 7 : おもほゆるかな・こひやわたらむ・こひもするかな・こひしきものを・おもふこころは・あやまたれける・わがやどのはな・わがおもふひと

4グラム、5グラムの短い文字列では、いわゆる歌語が多く抽出される。また6グラム、7グラムと文字列が長くなると、特定のことば続きをもった表現で、注目すべきものが抽出されてくる。これらの歌語・表現は、古今集において、女性には全く用例のない、男性側にシフトした用語ということになるが、更にこれらの用語を詠んだ歌そのものを検討すると、内容から、およそ次のような3つの位相に分類できることがわかった。

- a. 文学的に創出されたことばで、特に男性性の反映されたことば

しらゆき（白雪）、たまのを（玉の緒）、わびびと（侘び人）、たなばた（七夕）、しらたま（白玉）、うつせみ（空蟬）、をみなへし（女郎花）、うめのはな（梅の花）、ふじばかま（藤袴）、はなのかげ（花の蔭）、なみだがは（涙川）、わがやどのはな（我が宿の花）

- b. 男性の当時の社会生活を反映したことば

しらくも（白雲）、しらやま（白山）、かよひぢ（通い路）、こひやわたらむ（恋や渡らむ）、こひもするかな（恋もするかな）、こひしきものを（恋しきものを）

- c. 男ことばとして意識されていたことば

まにまに、あしひきの、おもほゆるかな、

いくつか詠歌例をあげながら検証する。まず、b・cから述べておく。b「男性の当時の社会生活を反映したことば」にあげた「しらやま」は、白山（はくさん）のことで、古今集で、はじめてよまれるようになった歌枕である。古今集での用例6首はすべて男性歌人の作例で、作者は躬恒、貫之、藤原兼輔らだが、その詠歌例は

越国へまかりける人に、よみて、遣はしける

よそにのみこひやわたらむ白山の雪みるべくもあらぬ
わが身は（羈旅歌・383・凡河内躬恒）

のように、国司などの任務について実際に越前の国に赴く人に歌われたものや、あるいは、赴任した者がその当地で詠んだものが大半で、当時の受領層歌人たちの社会生活から生まれたことばと言える。名所・歌枕は行動の自由な男性、地方下向を余儀なくされる男性によって、まず開拓されていくというありかたがうかがえる。また、平安時代の恋愛の定型がことばのレベルにあらわれたものと言えるのが「通ひ路」「恋やわたらむ」「恋もするかな」「恋しきものを」などである。男性が女性のもとに通う結婚形態であったことから「通ひ路」が男性の歌語として展開したことは容易に想像されようが、現代から見ると特徴のない「恋もするかな」などの表現が男性のことばに限られているのは注目されよう。

題しらず

五月山こずゑを高みほととぎすなくねそらなる恋もするかな（恋歌二・579・貫之）

「恋もするかな」＝男性表現、という結びつきはかなり強固で、後撰・拾遺～千載集と後続の勅撰集でも、また私家集でも女性の詠歌例は全く検出されない。それはこの語の古今集での用例が、恋一～五までである古今集恋歌のうち、すべて恋一・恋二という恋の初期段階の歌を集めた巻に集中していることとも関連しており、恋愛の最初のアクションは男性からという当時の通念、言いかえれば、女性はあくまで受け身で、求婚は男性の側からという当時の恋愛のパターンが、ことばのレベルにあらわれたものと考えられる。「恋やわたらむ」「恋しきものを」も同様で、すなわち、平安時代に「恋をする」のは男性、ということになるのである。

行動圏も広く、恋愛にも主導権を持つ男性と、反対にほとんど都を出ることもなく、恋愛においても受け身の女性という、両者の社会的役割、生活圏の相違が、ことばに反映されている実態が、明瞭に把握できる。

それに対して、cの「男ことばとして意識されていたことば」としたのは、aにもbにも入らず、用例が男性に集中していることから、平安時代にはいわゆる「男ことば」として認識かつ使用されていたと判断されることばである。「まにまに」を一例にとろう。

山に登りて、帰りまうできて、人びと別れにけるついでによめる

わかれをば山のさくらにまかせてむ止めむとめじは花のまにまに（離別歌・393・幽仙法師）

古今集での用例は、この他、清原深養父、藤原兼輔、菅原道真、小野貞樹にあり、後撰・拾遺でも、読人不知以外は男性の用例で、女性の歌はない。私家集に範囲を広げると、女流歌人の伊勢に

みそぎつつおもふことをぞ祈るりつる八百万よの神のまにまに（伊勢集・82番）

の例などがあるが、禊ぎ・神事を詠んだ特殊な例で、やはり特殊な位相の語であると思われる。女性が使うには、固い、あるいは厳格な語感があつたのだろうか。現代人にはもっとも内省の利かない領域だが、今回の方法ではこのような用語も検出が可能で、データ全体では、このcに該当することばもかなり見いだすことが出来た。

文学研究の立場から、もっとも注目されるのは、aの「文学的に創出されたことばで、特に男性性の反映されたことば」であろう。この範疇に分類されることばは数量的にも多い。更にいくつかのグループに分けることが可能であると思われるが、ここでは、このaに分類されることばには、人事性を付与された景物の多いこと、それも男性の視点で捉えた女性のイメージの付与された景物が大きな割合をしめるていることを特記しておきたい。例歌として「藤袴」をあげておく。

是貞親王家歌合に、よめる

なにひとかきてぬぎかけしふぢばかまくるあきごとのべをにははす（秋歌上・239・敏行朝臣）

藤袴を、よみて、人に遣わしける

やどりせしひとのかたみかふぢばかまわすられがたきかににはひつつ（同・240・貫之）

藤袴を、よめる

ぬししらぬかこそにはへれあきののにたがぬぎかけしふぢばかまども（同・241・素性）

すべて秋歌、すなわち季節の歌として、自然の景物としての藤袴を詠んだ歌であるのだが、詠歌内容は、単純に秋の七草として詠んだだけの万葉歌「萩の花尾花葛花なでしこの花をみなへしまた藤袴あさがほの花」（巻八・1538・山上憶良）とは全く異なっている。三首とも

が、袴を衣服としての袴と、花の香を人の移り香と関連づけて詠んでおり、美しい女性の脱ぎ置いた藤袴にはほかに恋情をかき立てられるという、官能性を内包したものととなっているのである。同様に、男性歌のみで10例をしめる「梅の花」が「よそにのみあはれとぞ見し梅の花あかぬ色香は折りてなりけり」（春歌上・37・素性）のように、限りなく女性をイメージした官能性をたたえていることや、5グラムではトップに立つ「女郎花」が、「誰が秋にあらぬものゆへ女郎花なぞ色にいでてまだきうつろふ」（秋歌上・232・貫之）のように、多情で心変わりしやすい女性のイメージとして詠まれていることは従来から指摘されてきた通りである。従来の研究では、これらを、擬人・比喩というレトリックの問題として扱ってきたのだが、ここでは、実はそれらの大半が、男性の視点で捉えた女性性というジェンダー性を属性とすることを、あらためて問題提起しておきたい。こうした属性を持つ歌語は、今回のデータによれば、花すすき、やまとなでしこ、秋萩、など、更に多くが加わるのであって、擬人という手法が、実は、その後平安時代に展開する女性美の型、それも男性の視点で捉えた女性性の型の形成と定着に大きな役割を果たすものであったことが判明するのである。

一方で、男性が男性を比喩したことばもある。「夏虫」である。「夏虫」は、4グラムで出現頻度2の語彙なのだが、古今と、それ以後の用例を新編国歌大観CD-ROM等で検索し、検討すると次のような事がわかってくる。

三代集での用例（ただし、拾遺集には用例は無し）を一覧し、作者を検討してみる。

1. 夏虫の身をいたづらになす事もひとつ思ひに困りてなりけり（古今・恋一・544番・読人不知）
 2. 夜ゐの間もはかなく見ゆる夏虫にまどひまされる恋もする哉（古今・恋二・561番・紀友則）
 3. 夏虫を何かいひけむ心から我もおもひにもえぬべら也（古今・恋二・600番・凡河内躬恒）
 4. 八重葎しげきやどには夏虫の声より外に問う人もなし（後撰・夏・194番・読人不知）、
 5. つつめども隠れぬ物は夏虫の身よりあまれる思ひなりけり（後撰・夏・209番・読人不知）
- 詞書「桂のみこの「ほたるをとらへて」と言い侍ければ、わらはのかざみの袖につつみて」
6. 夏虫の身をたきすて玉しあらば我とまねばむ人目もる身ぞ（後撰・夏・213番・読人不知）、
 7. 夏虫の知るしるまどふ思ひをば懲りぬ悲しと誰か見

一見女性にも用例があるように見えるが、詳細に検討すると、実はそうではないことがわかる。5の詞書にある「桂のみこ」は宇多天皇皇女の内親王で、詠歌状況はこの内親王に、ある男あるいは詞書の男童が思いを寄せて詠みかけたと解釈するのが通説である。従って詠歌主体は男となる。また、7は唯一女性の詠んだ例だが、後撰集の配列で見ると、この歌は一首前の歌である967番の男性の送って来た歌に答えたものとなっており、「夏虫」はむくわれないことを承知で自分を恋する、そんな相手の男性を喩えたものであることがわかる。すなわち、古今・後撰の用例では、「夏虫」に比喻されるのは、7のように女性が男性を指した例も含めて、男性に限定されていることがわかる。

このことは、私家集に範囲を広げて用例検索を行なうと、さらに明確な結果を得ることができる。新編国歌大観CD-ROMでの検索結果では、女流歌人の家集で「夏虫」の用語が検出出来るのは、上記の伊勢の他では和泉式部に1例を検出出来るのみに過ぎない。和泉と同時代の代表的女流である紫式部、赤染衛門、伊勢大輔、相模などにも全く用例がなく、本院侍従、道綱母、馬内侍、出羽弁など、平安前期から末期に至るまで、女性歌人の家集での用例は検出されないのである。それに対して男性歌人達は、平安時代全体を通じて、平均的にこの語を詠んでいるのだが、この現象の背後にある平安時代の文学的な言語意識が、平安中期の歌学書『能因歌枕』に端的に述べられている。

夏虫（ナツムシ）とは 女によりて身をいたづらになす物にたとふ（『能因歌枕』）

ここでは歌語「夏虫」の比喻を、「女によりて身をいたづらになす」ことのたとえ、すなわち男が女によって身を破滅させる意味に限定して考えている。上記の古今以来の用語例とこの叙述はみごとに呼応するのであり、これが当時の文学に根ざした言語感覚と理解される。従来の研究や注釈は、この語を男女の性差に限定せず、単に人が恋に身を焼く意としてだけ捉えてきたが、そうした平安時代の言語感覚を正しく把握していなかったことになるだろう。出現頻度数2にすぎないことば、「夏虫」にも、以上のように、性差とことばに関わる有用な結果を得ることができた。このことはまた、今回得た男女の差分のデータ結果が、出現頻度の低い語彙に至るまで、有

意である可能性を示唆している。

1.10 結語

以上の検証を通じて、現代の男性語・女性語の別に相当するような区別が、平安和歌の用語にもあったことをあきらかにし、平安時代において、社会生活のあり方がことばの男女差に反映していくあり方や、いわゆる男ことば・女ことばの実態、更に男性が、女性的なるものをどのように文学のことばの型に定着させていったか、男性にとっての「恋」が、一つどのような型で認識されていたか、などその実態の一部を見た。特に最後の問題は、それが、広く日本文化において、女性らしさ・男性らしさなどという認識や型、またその認識をあらわすことばの形成と展開に、古今集をはじめとする和歌文学が大きな役割を果たしたことを予想させる。和歌という「言葉の制度」が備えていたジェンダー性について、具体的に解き明かすことが出来るであろう。

また研究の過程で特に実感したのが、平安時代のことばにあらわれた男女差を現代の語感で捉えるのがいかに困難であるか、結果として従来の研究が、いかにその多くを見落としてきたかという点である。今後、古今集以外の平安和歌集にも対象を拡大し、また前時代の万葉集も加えて比較していくことで、和歌のことばがもったジェンダー性の確立過程や変遷、またその意義を明確にしていくことができるであろうし、今回データ量の足りなかった女性側の問題についても、多くを明らかにしていくことが可能となると考える。

また、このような古典語の微妙な語感を抽出してゆくためのツールとしてnグラム統計の方法が、単純ながらきわめて効果の高いものであることも実証できたものと考えている。

2 <付説> n グラム統計処理によった類句・類歌の抽出について

本年度研究テーマとした和歌のことばと性差ということと直接連動するものではないが、今回、n グラム統計処理を和歌研究に応用する過程で、一つ有効な課題が浮かんできたので、簡単に紹介しておく。それは、長いn グラムをとった場合に、類句あるいは類歌が、非常に容易に抽出出来るということである。たとえば、男性歌で、n グラム飽和前の最も長い $n = 12$ （12文字列）のこ

とばとしては次の7種類が検出される。

1. をみなへしおほかるのべに（出現頻度2）
2. みよしののやまのしらゆき（同上）
3. ながらのはしのながらへて（同上）
4. さみだれのそらもとどろに（同上）
5. こしのしらやましらねども（同上）
6. かとのみぞあやまたれける（同上）
7. あかずしてわかるるなみだ（同上）

いずれも詠歌の断片というより、まとまりある類句と
いってよい。さてこの7種類の文字列はいずれも出現頻
度が2回であるのだが、該当する歌をあげると次のよう
になる。

1. をみなへし多かる野辺に やどりせばあやなくあだの
名をやたちなん（秋歌上・229・小野美材）
花にあかでなに帰るらむを みなへし多かる野辺に 寝
なましものを（秋歌上・238・平貞文）
2. みよしのの山の白雪 つもるらし古里さむく成りまさ
るなり（冬・325・坂上是則）
みよしのの山の白雪 ふみわけて入りにし人のをとづ
れもせぬ（冬・327・壬生忠岑）
3. 逢ふことを ながらの橋のながらへて 恋ひわたるまに
年ぞへにける（恋歌五・826・坂上是則）
……かくしつつ ながらの橋のながらへて 難波の浦
に立つなみのなみの皴にや……（雑体・1003「古
歌に加へて奉れる長歌」中の一句 壬生忠岑）
4. 五月雨の空もとどろに 郭公なにを憂しとか夜ただな
くらん（夏歌・160・貫之）
……春霞 思ひみだれて 五月雨の空もとどろに 小
夜ふけて山ほととぎす鳴くごとに（雑体・1002・
「古歌奉りし時の目録の、その長歌」貫之）
5. きみが行く 越の白山しらねども 雪のまにまに跡はた
づねむ（離別歌・391・藤原兼輔朝臣）
思やる 越の白山しらねども 一夜も夢にこえぬ夜ぞな
き（雑歌下・980・紀貫之）
6. み吉野の山辺にさけるさくら花
雪かとのみぞあやまたれける（春歌上・60・友則）
花見つつ人まつ時は白妙の

袖かとのみぞあやまたれける（秋歌下・274・
友則）

7. あかずして別るる涙 たきにそふ水やまさると下は見
ゆらむ（離別歌・396・兼芸法師）
…富士の嶺の燃ゆるおもひも あかずして別るる涙 藤
衣をれるこころも八ちぐさの（雑体・1002「古
歌奉りし時の目録の、その長歌」貫之）

1～7の類句はすべて万葉に無く、ここにあげた古今
時代の歌人が固有に展開したものである。このような長
いnの表現が、友則・貫之・是則・忠岑・兼輔・貞文など、
時代的にも近接した限られた歌人に集中してくるとは、
データをとる前には、予期せざる結果であった。4・6
のように同一歌人が同一の類句で詠じているのは、歌人
の歌風を考察する一助となるし、1・2・3・5のような
例は、歌人どうしの影響関係の実態を考える上で新しい
指標となる。これはひいては類似歌の問題にも拡大でき
るであろう。和歌の類似歌の自動抽出については、当特
定領域研究で、山崎真由美らが最長共通部分文字列（L
CS）を用いて高い成果をあげているが、本歌取りとい
う高次元の影響関係が発生する平安後期とは異なり、平
安前期和歌は、語彙やことば続きレベルの影響関係が一
般で、nグラム統計処理によってもかなりのことが解明
出来ると思われる。

nグラム統計処理は、本論で述べたように、和歌デー
タの分析には極めて簡便に適用できるので、これによっ
て、各時代の和歌の慣用表現を網羅的に析出することや、
近い歌人どうしの私家集を網羅分析して表現の影響関係
を語彙以外のレベルでも検証することなど、多くの成果
が期待出来、稿者としても今後その方向での研究も進め
ていきたい。

謝辞

本研究にあたっては、京都大学の長尾真先生、森信介
氏（現在 IBM 東京基礎研究所）のご好意により、両氏の
開発になるnグラム統計を高速に算出するソフトウェア
を利用させていただくことができた。また、特に長尾先
生からは、古典語研究にnグラム統計を用いることにつ
いてのご示唆をいただいた。ここに記して、深く感謝申
し上げます。