

# 多変量解析を用いた万葉歌の 筆録者同定の可能性試論

村田 右富実  
川野 秀一

## 1. はじめに

『万葉集』に見える1音節を表す文字<sup>(1)</sup>は、時に歌の書記者（歌作者ではなく）の個性を表しているように見える。たとえば、稲岡耕二氏『万葉表記論』（塙書房 1976年）に代表される用字研究は、巻五の憶良歌や旅人歌の用字について大きな成果をあげている。しかし、その一方、ほんの僅かな用字の違いを書記者決定の根拠としようとする傾向も見られ、用字と書記者との係属についての客観的な基準の作成は事実上棚上げされて来たといっても過言ではない<sup>(2)</sup>。そこで本稿では、そうした係属の有無を個別論の積み重ねではなく、多変量解析を用いて『万葉集』全体から把握することを目的とする。

## 2. データについて

万葉歌の底本には、鶴久氏・森山隆氏『万葉集』（おうふう 1972年）を用いた。また、「1字1音の文字」のデータ作成にあたり、音仮名のみならず「1字1音の文字」全てを対象とした<sup>(3)</sup>。なお、稿者はその立場を取るものではないが、研究史上の慣習として音仮名<sup>(4)</sup>のみを対象とすべきとする見方も存在すると考え、1字1音の音仮名のみをデータとした解析も行った。

また、『万葉集』中に初句から結句まで完全な形で残っている歌は4,536首存在しており、これらを一首ずつ1データとしている。異伝については、1デ

---

(1) 以下、「1字1音の文字」と記す。

(2) 巻五における用字と筆録者との係属についての先行研究については、原田貞義氏「松浦佐用姫の歌群」（『セミナー万葉の歌人と作品 第四巻』（2000年5月）に適切なまとめがある。

(3) たとえば「し」は、音仮名（之、思等）、正訓字（為等）、訓仮名（石等）、音訓の区別の曖昧なもの（師、四等）がある。これら全てに明確な線引きを施すことは不可能と思われる。

(4) 以下、音仮名は、「濫、藍、覽（らむ）」、「南（なむ）」等の二音節を示す音仮名を除き、1字1音のものを指し示すこととする。また、汎用性を高めるために、「に=尔」、「を=乎、叫」なども音仮名として扱っている。

ータとして認めることはできないものの、全体の解析に参画させる必要があるため、全てを統合して1データとして取り扱い、解析の結果から除外している。また、「1字1音の文字」を含まない歌が9首あるため、これを除き、都合4,528データを解析の対象とした<sup>(5)</sup>。

### 3. データの重み付けについて

具体的な解析に入る前に、データの重み付けについて触れておく必要がある。たとえば、集中に「1字1音の文字」によって記される「あ」は1,217個存在する。そのうち、音仮名の「<sup>あ</sup>安」は940個を数えるのに対し、音仮名の「<sup>あ</sup>阿」は227個と、音仮名の「<sup>あ</sup>安」のほぼ1/4しか存在しない。また、正訓字の「<sup>あ</sup>妾」は1個しかないため、音仮名の「<sup>あ</sup>安」の1/940となる。同じ「あ」をあらわしている場合でも、音仮名の「<sup>あ</sup>安」は一般的であり、正訓字の「<sup>あ</sup>妾」は珍しく、解析にあたり、それぞれに応じた重み付けをする必要が生じる。

そこで、一つの音を表す複数の文字に対して、それぞれの文字に応じた重み付けを加え、1文字あたりの数値を算出することとし、その式には、

その音の合計数 / その用例の合計数 = 1文字あたりの数値

を用いた。上の「あ」を例に取れば、下のゴチック部分の数値が該当する。

音仮名の「<sup>あ</sup>安」 →  $1,217/940 \doteq 1.29 \rightarrow 1.29 * 940 = 1,217$

音仮名の「<sup>あ</sup>阿」 →  $1,217/227 \doteq 5.36 \rightarrow 5.36 * 227 = 1,217$

正訓字の「<sup>あ</sup>妾」 →  $1,217/1 = 1,217 \rightarrow 1 * 1,217 = 1,217$

つまり、一つの音の中でその文字の使われる用例数が少なければ少ないほど数値が高くなるようにし、かつ、その文字の数値の合計がその音の合計数になるようにした<sup>(6)</sup>。

### 4. 1Class Support Vector Machine について

次に、本稿が用いた多変量解析の手法について述べる。Support Vector Machine は、ウラジミール・ヴァプニクによって提唱された多変量解析の手

(5) 解析の対象とするデータの概略は以下の通り。

	1字1音	音仮名
総文字数	79,625	58,301
異なり字度数	1,286	552
総データ数	4,528	4,528

(6) 処置を施したデータの一部は以下の通り。

	あ	阿	安	我	吾	妾	足	網	有	足	余	合計
合計(a)	227	940	3	26	1	9	6	1	3	1	1217	
重み(b)	5.36	1.29	405.67	46.81	1217	135.22	202.83	1217	405.67	1217		
=a*b	1217	1217	1217	1217	1217	1217	1217	1217	1217	1217		

法<sup>(7)</sup>の一つであり、これを改良したものが、1Class Support Vector Machine (以下、1C-SVM) である。これは全体の中から外れ値を探し出す手法であり、例えば、数多くのメールの中から、外れ値として迷惑メールを選び出すために用いることができる。以下、本手法の最も基本的な点のみ記しておく<sup>(8)</sup>。

次表は、A～Eの音がイ～ニの歌にどのように分布しているかを例示している。また、その数値は重み付けされているものとする。

	A	B	C	D	E
イ歌	1	3	2	1	2
ロ歌	2	1	4	1	2
ハ歌	1	2	1	4	1
ニ歌	3	3	1	3	3

まず、この表のデータ同士（イ～ニ）が似ているか否かについて、データ間の距離（類似度）を計算することから始める。データが似ている・似ていないの定義は種々考えられるが、ここでは一番簡単なデータ間の距離をデータ間の類似度として考えることとする。計算した値が小さければデータ間の距離が近いと判断できるため、データ間の類似度は高くなる。一方、計算した値が大きいとデータ間の距離が遠いと判断できるためデータ間の類似度は低いことになる。

具体的に記せば、イ歌のデータとロ歌のデータ間の距離であれば、

$$(\text{イ歌のA音} - \text{ロ歌のA音})^2 \rightarrow (1-2)^2 = 1$$

$$(\text{イ歌のB音} - \text{ロ歌のB音})^2 \rightarrow (3-1)^2 = 4$$

$$(\text{イ歌のC音} - \text{ロ歌のC音})^2 \rightarrow (2-4)^2 = 4$$

$$(\text{イ歌のD音} - \text{ロ歌のD音})^2 \rightarrow (1-1)^2 = 0$$

$$(\text{イ歌のE音} - \text{ロ歌のE音})^2 \rightarrow (2-2)^2 = 0$$

と、それぞれの列の差を求めて二乗することによって計算でき、この合計がイ歌とロ歌の距離（ここでは「 $1+4+4+0+0=9$ 」）として計算される（次式参照）。

$$(1-2)^2 + (3-1)^2 + (2-4)^2 + (1-1)^2 + (2-2)^2 = 1+4+4+0+0=9$$

次に、この操作をすべての場合において行う。つまり、イ歌とロ歌、イ歌と

(7) Vapnik, V. N. (1998) *Statistical Learning Theory*. Wiley, New York.

(8) 1C-SVM の詳細については

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J. and Williamson, R. C. (2001) "Estimating the support of a high-dimensional distribution." *Neural Computation* 13, 1443-1471

赤穂昭太郎氏『カーネル多変量解析—非線形データ解析の新しい展開—』（岩波書店 2008年）を参照されたい。

ハ歌、イ歌とニ歌、ロ歌とハ歌、ロ歌とニ歌、ハ歌とニ歌の全6パターンについて距離（類似度）を計算する。結果は以下の通り。

イ歌とロ歌： $(1-2)^2 + (3-1)^2 + (2-4)^2 + (1-1)^2 + (2-2)^2 = 1+4+4+0+0=9$

イ歌とハ歌： $(1-1)^2 + (3-2)^2 + (2-1)^2 + (1-4)^2 + (2-1)^2 = 0+1+1+9+1=10$

イ歌とニ歌： $(1-3)^2 + (3-3)^2 + (2-3)^2 + (1-3)^2 + (2-3)^2 = 4+0+1+4+1=10$

ロ歌とハ歌： $(2-1)^2 + (1-2)^2 + (4-1)^2 + (1-4)^2 + (2-1)^2 = 1+1+9+9+1=21$

ロ歌とニ歌： $(2-3)^2 + (1-3)^2 + (4-1)^2 + (1-3)^2 + (2-3)^2 = 1+4+9+4+1=19$

ハ歌とニ歌： $(1-3)^2 + (2-3)^2 + (1-1)^2 + (4-3)^2 + (1-3)^2 = 4+1+0+1+4=10$

最も距離（類似度）に近い（高い）のはイ歌とロ歌であり、最も距離（類似度）が遠い（低い）のがロ歌とハ歌であることがわかる。このようにして求められた類似度を用いて、『万葉集』全体の中から距離の遠い外れ値を検出して行く<sup>(9)</sup>。

また、計算結果についての外れ値の認定については、閾値（以下、NUと記す）をどのように設定するかによって変化する。たとえば、NUを低く設定すればするほど、外れ値については厳しく判定され、外れ値の歌数は少なくなる。確実に外れ値の歌を検出できる一方、外れ値の歌であるにも関わらず、外れ値ではないと検出する率は高まってしまう。一方、NUの値を上げてゆくと、徐々に外れ値の歌数は増えてくる。数多くの外れ値の歌が検出される一方、外れ値ではない歌も外れ値として検出する率も高まってしまうことになる。

この点を踏まえて、以下、外れ値の認定について記す。

## 5. 解析結果の絞り込みについて

実際に万葉歌を解析すると、長歌や正訓字主体の歌々があるため、1データ当たりの文字数のばらつきが極めて大きく、たとえば、長歌は文字数が極端に多いことが理由で外れ値として検出されてしまう。そこで、解析結果から、いわゆる仮名書き歌巻（五、十四、十五、十七、十八、二十）の短歌を抜き出し（984首）、この984首に対して、統計学において一般的な5%を有意水準として設定した（ $984 \text{ 首} * 0.05 = 49.2 \rightarrow 49 \text{ 首}$ ）。すなわち、NUの値の低い方（万葉歌全体の中でより一般的でないもの）から順に49首を万葉歌全体から見たときの外れ値と認定した。

---

(9) なお、実際の解析においては、この距離測定法は使っておらず、以下のガウスカーネルと呼ばれる類似度を指標として使っている。このガウスカーネルでは、値が大きいと類似度は高くなり、値が小さいと類似度は低くなる。

$$\exp\left(-\frac{\|x_i - x_j\|^2}{\sigma}\right)$$

ここで、 $x_i$  は  $i$  番目のデータ、 $x_j$  は  $j$  番目のデータ、 $\sigma$  は全データ間で計算したユークリッド距離の中央値である。

## 6. 解析結果（「1字1音の文字」をデータとした場合）

次の表は外れ値と認定された49首をまとめたものである<sup>(10)</sup>。

### ○巻毎

巻	全歌数	短歌数	外れ値	検出率	備考
5	114	104	21	20.19%	梅花宴歌6首／吉田宜4首(864～867)他
14	238	238	8	3.36%	
15	208	200	2	1.00%	作者不明1首(3687 本文に問題あり)／ 阿倍継麻呂1首(3708 本文に問題あり)
17	142	127	2	1.57%	平群女郎1首(3939)／大伴家持1首(3992)
18	107	97	4	4.12%	補修部3首(4047,4049,4082)／ 大伴池主1首(4130)
20	224	218	12	5.50%	防人歌11首／ 大伴家持1首(4400 本文に問題あり)
総計	1,187	984	49	4.98%	

### ○歌のグループ毎

	全歌数	短歌数	外れ値	備考
吉田宜	4	4	4	
巻五・梅花宴歌	32	32	6	
巻十八・補修部	18	15	3	
巻二十・防人歌	93	92	11	
宅守・弟上娘子	63	63	0	
巻十八（補修部を除く）	89	82	1	4130（池主の書簡）
巻二十（防人歌を除く）	131	126	1	4400（本文に問題あり）
大伴家持	472	284	3	3992／4082（補修部）／ 4400（本文に問題あり）

巻別に見た時、巻五の突出した検出率が目を引く。とくにそもそも四首しかない吉田宜歌は、その全てが外れ値として検出される上、宜の歌は前後の記述から書簡の一部であることが判明しており、積極的に書記者を宜自身と認定してよい。また、梅花宴歌は全32首中6首が外れ値として検出されている。梅花宴歌に複数の書記者がいたことも認定してよいであろう（ただし、32人の書記者がいたことを証することにはならない）。これら二点は、従来たびたび指摘されており、今回の解析結果はこれまでの研究成果と背馳しない。巻五についていえば、歌の筆録者の文字が残存している可能性が極めて高いと思われる。この点は、巻五の前半部が書簡集であるとする拙稿『『万葉集』巻五の前半部の性質について』（『万葉集研究』第34集2013年10月）の結論を別の側面から支える<sup>(11)</sup>。

(10) 外れ値の歌々の概要については末尾に別表を付した。

一方、巻十四、巻十五、巻十七の検出率の低さは、これらの巻々が均質な文字列によって記されていることを証する。しかも巻十五に至っては後述するように本文に問題を含む二首以外は検出されておらず、極めて安定した様相を見せる。さらに、巻十八のいわゆる補修部<sup>(12)</sup>、巻二十の防人歌における検出率が高く、この点も巻十八の補修部の存在や、防人歌の文字列が書記者のそれを保存しているとするこれまでの研究を新しい視点から補強するものである。

今回の解析結果は、①巻十四、②巻十五、③家持以外の書簡を除く巻十七、④家持以外の書簡といわゆる補修部を除く巻十八、⑤防人歌を除く巻二十（以下、①～⑤をまとめてA群と称する）の書記者が単独である、あるいは共通の書記圏に属する少数の人々であることを否定しない。具体的にいえば、家持とその周辺の人々がこのA群の書記者として想定される。従来、漠然と巻十四、巻十五の書記者としてそうした人を想定していたが、今回の解析結果はその蓋然性を高める結果となっている。

しかし、この解析結果が物語るのは今述べてきたような研究史の追認や蓋然性の向上に留まらない。上記の表中、15・三六八七、15・三七〇八、20・四四〇〇に「本文に問題あり」と記した。以下、この点について、歌番号順に述べて行く。

## 7. 本文に問題ありの三首について

第一例、15・三六八七の原文は以下の通り。

安思必寄能 山等妣古由留 可里我祢波 美也故尔由加波 伊毛尔安比豆  
許祢  
あしひきの やまとびこゆる かりがねは みやこにゆかば いもにあひ  
てこね

問題となるのは、第一句の四文字目「寄」である。諸本の状況と近年の注釈書等の判断は以下の通り<sup>(13)</sup>。

廣・類／温・陽／／新編・埴CD → 「奇」。

- 
- (11) なお、最後にも触れるが、憶良歌についていえば、文字数が多いことと、巻五とそれ以外の巻の憶良歌とはあまりにも文字列の様相が違い、『万葉集』全体から見た時に外れ値とはならないと思われる。別の解析方法を考える必要がある。
- (12) 巻十八の補修部については、乾善彦氏「万葉集巻十八補修説の行方」（『高岡市万葉歴史館紀要』14号・2004年3月）に先行研究のまとめがある。また、同論によれば、補修部そのものが層を成している可能性も否定できない。ただし、今は可能な限り客観性を担保するため、旧大系の認定をもって巻十八の非一般的な部分と理解しておく。従って補修部という名称も適切でない面を持つが、通用語として使用しておく。
- (13) 写本、および注釈書等の略称は通称に従った。

天／細・宮／西・紀・矢・近・京／無・附・寛／／和泉・釈注・新大系・  
全歌講義・全解・和歌大系→「寄」。

廣瀬本の出現以前は、天治本の「寄」を採用するのが当然であったろうが、現在では、廣瀬本、類聚古集の二本が「奇」である点から、ここは「奇」の可能性が高いと考えられる。勿論こうした個別論的な校異作業の結果は重要であるが、それだけではなく、この歌の他の音仮名が集中にあってごく一般的なものであるのに対し、「き」の乙類を「寄」で記すのは、集中この一例のみである点も考慮すべきである。さらに巻十四、十五、十七、十八、二十において、「き」の乙類を示す音仮名は69例（当該例を除く）を数えるが、うち64例までが「奇」で記される<sup>(14)</sup>。そもそも「寄」字は歌中に用いられることが極めて少ない文字であり<sup>(15)</sup>、これらの点からも当該文字は「奇」である蓋然性が高い。

IC-SVMを利用することによって、『万葉集』全体を文字群として把握し、その全体から個別の文字を見た結果、こうした例を剔出できる点に注目したい。

第二例は、15・三七〇八。

毛能毛布等 比等尔波美要縹 之多婢毛能 思多由故布流尔 都奇曾倍尔  
家流

ものもふと ひとにはみえじ したびもの したゆこふるに つきそへに  
ける

第二句の最後の文字「縹」は集中に孤例であることは勿論、記紀風土記にまで範囲を広げても、『日本書紀』に一例のみ存在する稀少字である<sup>(16)</sup>。第一例と同じように写本などの状況を記すと以下の通り。

廣→「縹」。

天→「倭」右ニ「可□坎」ノ如キ書入アリ。

類→「綏」。

西・矢・京→「織」。

近→「溜」。

細・宮／紀・温・陽／無・附・寛／／第一例に記した全ての注釈書類→  
「縹」。

非仙覚本の三本が一致せず、仙覚本系統の諸本においても、三種類の文字が存在している。この点について、たとえば『旧大系』は、

(14) 残る5例は、「紀」が3例→17・三九五七（家持）、17・三九六九（家持）、20・四三四九（上総国防人歌）と、「綺」が2例→17・四〇〇三（池主）、18・四〇四〇（福麻呂）。

(15) 「何所可將寄」(3・四八〇)、「寄濱部丹」(13・三三三九)の二例のみ。

(16) 天智五年是冬条。

緇は、ここ以外に用例がない。～中略～（この字は一引用者注）清音シにあたる。それをジにあてて使っている。使用者が字音を精しく知らずに、字画の複雑さから、濁音に使う文字と違って使用したものであろう。（『日本古典文学大系 万葉集 四』）

と、苦しい説明を加える。また、『旧全集』は「この字をジの仮名に用いた理由は不明。あるいは誤字か。」と述べる。先の歌同様、この「緇」以外は一般的な音仮名の羅列なだけに、『旧全集』の指摘を受け、原字不明とすべきであろう。

第三例は 20・四四〇〇。

伊弊於毛負等 伊乎祢受乎礼婆 多頭我奈久 安之弊毛美要受 波流乃可須美尔  
いへおもふと いをねずをれば たづがなく あしべもみえず はるのかすみに

初句の五文字め「負」を「ふ」の音仮名として使用する点が珍しく、他には、「ものふ」を表す「物乃負」（3・四七八）、「物負」（6・一〇四七）の二例を見るのみである。諸本、注釈書の状況は以下の通り。

元・廣／／新編・和泉・全解・塙CD・和歌大系→「布」。

類／細・宮／西・紀・温・陽・矢・近・京／無・附・寛／／新大系・釈注・全歌講義→「負」。

仙覚本系諸本が一貫して「負」ではあるものの、元暦校本と廣瀬本が「布」である点は考慮されるべきである。第一例同様、ここは積極的に「布」に校訂すべきであろう。

繰り返しになるが、これら3例は、特定の1字が珍しいという点ばかりでなく、極めて一般的な文字列環境の中にある外れ値として検出されている点に注目したい。IC-SVMはこの双方の観点から当該歌を外れ値として検出しているのである。そのため、試みにこれら三首を、上述のように本文校訂を施した上で、あらためてIC-SVMを掛けると、三首共に外れ値として検出されなくなる<sup>(17)</sup>。

我々には『万葉集』全体の書き様を感覚的に理解することはできても、一首の歌の文字列が『万葉集』全体の中にあつてどの程度一般的であるか否かを論理的に外化することはできない。本稿はこの点について外化を試みたものである。

続いて、音仮名を解析対象とした場合の結果について述べて行く。

---

(17) 本文校訂を施し再計算させた順位は、15・三六八七 = 348位、15・三七〇八 = 563位、20・四四〇〇 = 338位となり、いずれも外れ値にならない。



## 8. 解析結果（音仮名をデータとした場合）

○巻毎（外れ値の括弧内は「1字1音の文字」の解析結果の検出歌数）

巻	全歌数	短歌数	外れ値	検出率	備考
5	114	104	22(21)	21.15%	梅花宴歌7首／吉田宜4首(864～867) 他
14	238	238	7(8)	2.94%	
15	208	200	2(2)	1.00%	作者不明1首(3687 本文に問題あり)／ 阿倍継麻呂1首(3708 本文に問題あり)
17	142	127	2(2)	1.57%	平群女郎1首(3939)／大伴家持1首(3992)
18	107	97	4(4)	4.12%	補修部3首(4047,4049,4082)／ 大伴池主1首(4130)
20	224	218	11(12)	5.05%	防人歌10首／ 大伴家持1首(4400 本文に問題あり)
総計	1,187	984	48(49)	4.88%	

○歌のグループ毎

	全歌数	短歌数	外れ値	備考
吉田宜	4	4	4	
巻五・梅花宴歌	32	32	7	
巻十八・補修部	18	15	3	
巻二十・防人歌	93	92	10	
宅守・弟上娘子	63	63	0	
巻十八（補修部を除く）	89	82	1	4130（池主の書簡）
巻二十（防人歌を除く）	131	126	1	4400（本文に問題あり）
大伴家持	472	284	3	3992／4082（補修部）／ 4400（本文に問題あり）

「1字1音の文字」と音仮名の解析結果は、ほとんど変わらない。総数が1首少ないのは、49位が3首あり、これを加えなかったためであるが、その3首も「1字1音の文字」の51～53位<sup>(18)</sup>に位置している。

また、「1字1音の文字」で検出されながら音仮名では検出されなかった二首(14・三五〇五、20・四三二七)は「眠(ぬ)」(14・三五〇五)、「晝(ぬ)」(20・四三二七)と、集中唯一の正訓字を持つため、検出されなかったことがわかる。その裏返しになるのが、音仮名で検出されているながら、「1字1音の文字」では検出されなかった一首(5・八三八)であるが、こちらは「1字1音の文字」の次点(50位)である。「1字1音の文字」と音仮名との解析結果にはほとんど違いはないといってよい。さらに、先に詳細に述べた三首は音仮名を対象とした解析においても検出されている<sup>(19)</sup>。

(18) 音仮名の49位の3首の「1字1音の文字」における順位は以下の通り。14・三四三〇＝51位、14・三四八九＝52位、14・三五〇三＝53位。

やはり、本稿に示した約 50 首の歌々は万葉歌の中にあつて、「1 字 1 音の文字」や「音仮名」から見た時に一般的ではない。

## 9. 小結

以上、1C-SVM を用いて、『万葉集』における「1 字 1 音の文字」と音仮名とを対象に、外れ値の検出を行ってきた。以下、結果をまとめる。

### ○卷五

#### 1. 大伴旅人歌(806-807)、藤原房前歌(812)、吉田宜歌(864-867)

これらの歌は、書簡に記された歌でもあり、その書記者は歌作者である可能性が極めて高い。特に宜歌は 4 首全てが検出されており、宜歌については宜の文字遣いであると認定してよいであろう。

#### 2. 梅花宴歌(815-846)

複数の書記者の存在が想定される。中には小字注に記された歌作者が書記者である場合もあろうが、複数であるという点に留めおくべきであろう。

#### 3. 憶良歌(796、818、890、906)

それぞれ以下の理由で、その書記者を憶良とすることを保留する。

(ア)796—以前から憶良による文字列らしくないと指摘されている。たしかに音仮名「伴(は)」は集中に 9 例しか見えず、音仮名「別(べ)」(2 例)はこの歌にしか存在しない。しかし、その一方、検出順位はそれほど高くなく(1 字 1 音 = 40 位、音仮名 = 36 位)、確定的なことはいえない。

(イ)818—梅花宴歌中の一首として見るべきであり、書記者の具体には触れるべきではあるまい。

(ウ)890—甲類の「ぞ」自体が極めて珍しいために検出された可能性が否定できない。

(エ)906—「布施」は字音語であるため、他に書き記す方法がない。たしかに「せ」を表す「施」は、集中に 1 例のみであるが、字音語に用いられている「施」を音仮名としてよいかの問題も残る。この点を考慮すれば、書記者を憶良と即断することは躊躇される。

(オ)以上見てきたように、外れ値として検出された 4 首の憶良歌中 2 首に、推定できる検出理由があり、『万葉集』全体から見たとき、巻五の憶良歌の文字列が特殊であるとはいえない。

ただし、巻五の憶良歌はそれ自体、文字数が多いため、外れ値として検出されないという可能性は否定できない。また、本論の解析方法では、

---

(19) 本文校訂を施し再計算させた順位は、15・三六八七 = 311 位、15・三七〇八 = 556 位、20・四四〇〇 = 304 位。いずれも外れ値にならない。

他巻の憶良歌と巻五の憶良歌とを区別することができないため、この点については、今後の課題としたい。

○巻十四

外れ値として検出された巻十四の8首中5首（音仮名の場合は7首中5首）は、いわゆる甲乙の違例による検出と推測される。これらは東国の音韻を表現しようとしたものと理解することも可能であり、巻十四は『万葉集』にあって、一般的な文字列によって構成されていると違って差し支えない。

○巻十五

本文に問題のある15・三六八七と15・三七〇八は、外れ値から除いてよいと考えられる。すると、巻十五には外れ値の歌は存在しなくなり、巻十四同様、一般的な文字列であると判断してよい。

○巻十七

17・三九三九は平群女郎の歌であるが、歌群の左注に「右の件の十二首の歌、時々便使に寄せて来贈せり。一時に送る所にあらず。」とあり、書簡に記されていた文字列を残している可能性が高い。また、17・三九九二は家持作歌の唯一の検出歌であり、家持らしからぬ文字列ということになるが、「偲（し）」が選ばれた理由は不明であり、例外とするよりない。それでも、巻十七も一般的な文字の集合体であることは間違いない。

○巻十八

補修部の高検出率(3/15)は、補修部が『万葉集』の中にある異質であることを示す。やはり補修部として認定してよいだろう。また、大伴池主歌(4130)は書簡であり、池主をその書記者として比定できよう。巻十八は、補修部を除けば、一般的な文字列の集合である。

○巻二十

防人歌については、以前から書記者の違いが指摘されており、ここでもそれを追認できる。また、先述の大伴家持歌(20・四四〇〇)は、本文「負」を「布」とすべきである。すると、巻二十は、防人歌を除けば一般的な文字列によって記されていることになる。

○全体を通して

巻五の文字多様性は、『万葉集』における非一般性を示すが、それは当時の書記者の表記する文字の多様性の反映であろう。また、梅花宴歌の多様性は、何人の書記者が関わっていたかを具体的に述べることはできないものの、自ら記した歌を持ち寄った者がおり、その文字の残っている可能性が高い。そして、家持の歌は、補修部、本文の異同を考慮すると、284首中外れ値は1首(17・三九九二)のみとなる。これらは家持歌の書記者を家持自身として考えるこれまでの研究史をそのままに受け継ぐ。とすれば、やはりA群の書記者は、家持の文字運用の具体と、極めて近い文字運用をしていた者とい

うことになる。統計学の立場からいってしまえば、A群の書記者を家持に比定しても問題は発生しない。

## 10. むすび

以上、1C-SVMを利用して、万葉短歌の外れ値を検出してきた。その結果は、これまでの先行研究の追試験としても有効であり、かつ、新たな議論の可能性を導き出すものであった。全体から見た個別論の結果については上に述べたとおりだが、今回の解析結果は、A群の均一性をとも証している。文学研究の立場からいえば、A群の書記者を家持に比定してしまうことには躊躇を覚える。しかし、書記者が複数いたとしても彼らが家持の文字運用方法に極めて近い人物であることは間違いないだろう。それは、A群以外の歌々の中でも、平群娘子や池主などの書簡の歌々は、それぞれ若干の外れ値の歌が検出される程度であり、彼らが家持と近い文字運用者であったことから推測できる。勿論、この論立ては、上代文献にこれほど大量の「1字1音の文字」や音仮名は他にないため、必然的に巻五との相対的な比較の中でしかいえないものであることも事実である。ために、その実体推定については、なお慎重であるべきと考える。ただし、木簡などの1字1音の文字を視野に入れれば、『万葉集』の「1字1音の文字」は当時の一般的なありようを必ずしも反映していない可能性が高いように思われる。

今回は、一首単位で1C-SVMを使用したため、短歌を中心に論じることを余儀なくされたが、今後、長歌も本解析に積極的に参画させるために、長歌を一定の文字数に分割することを視野に入れている。しかし、本稿において試みた学際的な研究は、研究の端緒についたばかりである。いたずらに歌を分解することに否定的な面が存在する可能性を排除しきれず、先を急がずに論を組み立てた。試論と題した所以である。

1C-SVM による外れ値の検出状況

巻	歌	1_字	O_字	稀少字など	作者など	1_順	O_順	1_検	O_検
5	796	31	31	別(べ) * 2	山上憶良	40	39	○	○
5	806	31	30	愛(え)、帝(て)、丹(た)	大伴旅人	1	1	○	○
5	807	31	31	昧(め)	大伴旅人	12	12	○	○
5	812	32	32	地(ち)、移(や)	藤原房前	2	1	○	○
5	817	30	30	僧(そ)	粟田人? (梅花宴)	43	39	○	○
5	818	30	30	登(と違)	山上憶良 (梅花宴)	34	32	○	○
5	837	31	31	隅(ぐ)	志紀大道 (梅花宴)	4	4	○	○
5	838	31	31	肥(ひ)	櫻? 鉢麻呂 (梅花宴)	50	44	×	○
5	841	31	31	遇(ぐ)	高? 老 (梅花宴)	13	14	○	○
5	844	30	30	陞(へ)	小野国堅 (梅花宴)	22	19	○	○
5	846	31	31	卑(ひ)	小野淡理 (梅花宴)	13	12	○	○
5	848	31	31	微(び)	(大伴旅人)	22	25	○	○
5	849	31	31	棄(き)	(大伴旅人)	6	4	○	○
5	864	31	31	殊(ず)	吉田宜	13	14	○	○
5	865	29	29		吉田宜	34	33	○	○
5	866	29	29		吉田宜	17	18	○	○
5	867	31	31		吉田宜	21	19	○	○
5	872	31	31		旅人? 憶良?	25	25	○	○
5	873	31	31	嬪(ひ)	旅人? 憶良?	9	6	○	○
5	883	32	30	容(よ)	三嶋王	10	9	○	○
5	890	25	24	俗(ぞ)	山上憶良	29	29	○	○
5	906	27	26	施(せ)	山上憶良	28	44	○	○
14	3419	24	24	欲(よ違)	東歌 / 上野	45	42	○	○
14	3431	31	31	故(こ違)	東歌 / 相模	36	33	○	○
14	3437	32	32	馬(め違)	東歌 / 陸奥	46	44	○	○
14	3450	31	31	敵(へ違)	東歌	25	23	○	○
14	3458	31	31	恥(ち)	東歌	39	35	○	○
14	3499	31	31	其(こ違)	東歌	47	43	○	○
14	3505	31	29	眠(ぬ訓)	東歌	33	256	○	×
14	3564	31	30	吾(こ)	東歌	29	29	○	○
15	3687	30	30	寄(き本文)	作者不明	38	35	○	○
15	3708	31	31	緇(じ本文)	阿倍継麻呂	49	44	○	○
17	3939	31	31	連(れ)	平群女郎	24	23	○	○
17	3992	32	31	僂(し)	大伴家持	47	48	○	○
18	4047	31	29	野(の違)	遊行女婦土師(補修部)	19	19	○	○
18	4049	33	31	野(の違)、度(ど違)	田辺福麻呂(補修部)	11	14	○	○
18	4082	32	32	度(ど違)	大伴家持(補修部)	20	19	○	○
18	4130	32	32	應(お)、授(ず)	大伴池主	4	6	○	○
20	4327	30	29	晝(ゑ訓)	防人歌(遠江)	40	371	○	×
20	4337	27	27		防人歌(駿河)	42	39	○	○
20	4348	30	30	非(び違)	防人歌(上総)	43	38	○	○
20	4349	31	31	紀(き違)	防人歌(上総)	32	29	○	○
20	4353	29	29	遲(ち)	防人歌(上総)	17	14	○	○
20	4370	30	30	米(め違)	防人歌(常陸)	29	27	○	○
20	4386	27	27	枳(き違)	防人歌(下総)	13	11	○	○
20	4389	31	31		防人歌(下総)	7	9	○	○
20	4390	30	30	里(ろ)	防人歌(下総)	3	3	○	○
20	4400	32	32	負(ふ本文)	大伴家持	36	35	○	○
20	4402	30	30	布(ふ)	防人歌(信濃)	27	27	○	○
20	4431	30	30	久(く)	防人歌(昔年)	7	6	○	○

○1\_字、O\_字は、1字1音、音仮名、それぞれについての文字数を示す。

○1\_順、O\_順は、1字1音、音仮名、それぞれについて外れ値の順位を示す。

○順位は数値が少ないほど外れ値の度合いが高い。

○1\_検、O\_検は、1字1音、音仮名、それぞれについて外れ値として検出されたか否かを示す。

○O\_検において、検出されなかった二首は、ともに、その音韻を示す正訓字として集中に唯一の例を持つ。

○「違」は特殊仮名遣いの違例を、「訓」は正訓字を、「本文」は本文に問題のあることを示す。