

# 天草版『エソポのハブラス』・天草版『平家物語』の 語彙の豊富さ, 類似度, 偏り

濱 千 代 い づ み

## The Tendencies of the Vocabularies of Aesop's Fables (Amakusa Version) and the Tale of the Heike (Amakusa V.) by Using Statistical Measures

Izumi HAMACHIYO

### Abstract

The object of this paper is to investigate the tendencies of the vocabularies of Aesop's Fables (Amakusa version) and the Tale of the Heike (Amakusa v.) by using statistical measures. The results are as follows.

1. According to corrected Type/Token Ratio (Carroll) and the characteristic K (Yule), the vocabulary of the Tale of the Heike (Amakusa v.) is richer than the vocabulary of Aesop's Fables.
2. According to the degree of similarity C (Miyazima), the vocabulary of the Tale of the Heike (Amakusa v.) resembles the vocabulary of the Tale of the Heike (Takano-bon). And the degree of the similarity between Aesop's Fables and the Tale of the Heike (Amakusa v.) is higher than the degree of the similarity between Aesop's Fables and the Tale of the Heike (Takano-bon).
3. According to the expected value deviation (Tanaka), the vocabulary of Aesop's Fables is average in 17 works, but the vocabularies of the Tale of the Heike (Amakusa v.) and the Tale of the Heike (Takano-bon) are removed from the common words of 17 works.

### Key words

Aesop's Fables (Amakusa v.), the Tale of the Heike (Amakusa v.), vocabulary richness, vocabulary similarity, vocabulary bias

### 1 はじめに

この研究の目的は, 統計上のいろいろな指標を利用して, 天草版『エソポのハブラス』と天草版『平家物語』の自立語の語彙の豊富さ・語彙の類似度・語彙の偏りをはかり, 作品の語彙の傾向を探ろうとするものである。平成12年に天草版『平家物語』の自立語の語彙, 平成21年に天草版『エソポのハブラス』の助動詞の語彙について, 各見出し語の使用率を計算し, 段階に分けて, 作品の骨格部分をなす語集団という観点から作品の語彙の特色を解明した<sup>(注1)</sup>。今回は天草版『エソポのハブラス』の語彙の概要を天草版『平家物語』との比較を通して述べ, 続いて統計上の指標を用いて数値で示すことで両作品の語彙の傾向を把握することに努める。

天草版『エソポのハブラス』と天草版『平家物語』は『金句集』と合わせ綴じた形で出版され、1593年の総序を持つ。ただし、天草版『エソポのハブラス』の扉の刊行年は1593年であるが、天草版『平家物語』の扉・序の刊行年は1592年である。現在、大英図書館に唯一本が伝存されている、天下の孤本である。天草学林（コレジヨ）で印刷され、ポルトガル語式の写音法によるローマ字で日本語が綴られている。このうち、『エソポのハブラス』はイソップの生涯と数々の寓話とで構成され、室町時代末期の話し言葉に翻訳してある。また、天草版『平家物語』は古典の平家物語を室町時代末期の話し言葉で書き直してあり、喜一検校と右馬の允の二人が対話する形式で進行する。ともにイエズス会の宣教師たちが日本語の学習をするためのテキストとして編集された。さらに、前者は布教時の引用の拠りに用い、後者は日本の歴史を学習する手立てにする目的があった。

天草版『エソポのハブラス』と天草版『平家物語』は、それぞれ「扉・序・物語の本文・目録」の四部によって構成されている。このうちの物語の本文に使われている自立語を対象とする。計量には次の文献を利用し、単語の認定の基準を一致させるようにした。

a 『エソポのハブラス本文と総索引』

b 『天草版平家物語語彙用例総索引』

略称として次のものを用いる。

〈エソポ〉〈エ〉…天草版『エソポのハブラス』

〈ヘイケ〉〈へ〉…天草版『平家物語』

## 2 『エソポのハブラス』の語彙の概要

### 2.1 『エソポのハブラス』の語彙の異なり語数・延べ語数

〈エソポ〉の自立語の語彙について各見出し語の使用度数を計量し、全体の異なり語数・延べ語数、品詞別の異なり語数・延べ語数を集計し、平均の使用度数を算出した。その結果を表に示すと次のようである。

表1 〈エソポ〉の自立語の語彙

品詞	異なり語数	比率(%)	延べ語数	比率(%)	平均
名詞	1515	521.69	5865	499.15	3.9
動詞	963	331.61	3759	319.91	3.9
形容詞	103	35.47	398	33.87	3.9
形容動詞	110	37.88	216	18.38	2.0
副詞	155	53.37	646	54.98	4.2
連体詞	12	4.13	651	55.40	54.3
接続詞	30	10.33	186	15.83	6.2
感動詞	12	4.13	23	1.96	1.9
その他	4	1.38	6	0.51	1.5
全体	2904	1000.00	11750	1000.00	4.0

名詞と副詞というように、2種以上の品詞にわたる機能を有する語は主たるものによって整理

した。また、動詞で自他の別によって活用の型が異なる場合は別語として扱った。

〈エソポ〉は異なり語数が2904、延べ語数が11750である。全体の平均度数は4.0になる。

〈エソポ〉の語彙を品詞別に見ると、名詞がもっとも多く、異なり語数・延べ語数ともに全体の半分を占めている。次に動詞が多く、異なり語数・延べ語数ともに全体の30%を超えている。平均の使用度数は連体詞が54.3で最も多く、次に接続詞が6.2で続いている。

データの散らばり具合を知るために分散（各データと平均値との差を二乗したものの総和をデータの数で除した平均）、および標準偏差（分散の正の平方根）を算出して調べよう。〈エソポ〉の見出し語は、使用度数のもっとも多いものから降順に「いふ（言）」378、「こと（事）」323、…「をなご」1、「をば（尾羽）」1の2904語である。

$$\{(378-4.0)^2 + (323-4.0)^2 + \dots + (1-4.0)^2 + (1-4.0)^2\} / 2904 = 209.50$$

$$\sqrt{209.50} = 14.47$$

〈エソポ〉の分散は209.50、標準偏差は14.47である。

ことばの集合の傾向を数値で示すとき代表値と散布度を用いる。本論では代表値として平均度数、散布度として標準偏差を用いることにする。〈エソポ〉の語彙の傾向について次のことが言える。

(a) 〈エソポ〉の語彙の傾向を数値で示すと、平均度数4.0、標準偏差14.47である。

## 2.2 天草版『平家物語』の語彙との比較

〈ハイケ〉の自立語の語彙について各見出し語の使用度数を計量し、全体の異なり語数・延べ語数、品詞別の異なり語数・延べ語数を集計し、平均の使用度数を算出した<sup>(註2)</sup>。その結果を表に示すと次のようである。

表2 〈ハイケ〉の自立語の語彙

品詞	異なり語数	比率(%)	延べ語数	比率(%)	平均
名詞	4697	632.93	23810	507.75	5.1
動詞	1931	260.21	16510	352.08	8.5
形容詞	233	31.40	1691	36.06	7.3
形容動詞	172	23.18	545	11.62	3.2
副詞	291	39.21	2452	52.29	8.4
連体詞	18	2.43	1119	23.86	62.2
接続詞	51	6.87	650	13.86	12.7
感動詞	22	2.96	107	2.28	4.9
その他	6	0.81	9	0.19	1.5
全体	7421	1000.00	46893	1000.00	6.3

〈ハイケ〉は異なり語数が7421、延べ語数が46893である。全体の平均度数は6.3になる。

〈ハイケ〉の語彙を品詞別に見ると、異なり語数・延べ語数ともに名詞がもっとも多く、次に

動詞が続いている。平均の使用度数は連体詞が62.2でもっとも多く、次に接続詞が12.7で続いている。

〈ヘイケ〉の見出し語は、使用度数のもっとも多いものから降順に「あり(有)」1000、「こと(事)」844、…「をりかく(折掛)」1、「をんなむしや」1の7421語である。

$$\{(1000-6.3)^2+(844-6.3)^2+\dots+(1-6.3)^2+(1-6.3)^2\}/7421=858.59$$

$$\sqrt{858.59}=29.30$$

〈ヘイケ〉の分散は858.59、標準偏差は29.30である。〈ヘイケ〉の語彙の傾向について次のことが言える。

- (b) 〈ヘイケ〉の語彙の傾向を数値で示すと、平均度数6.3、標準偏差29.30である。そして、〈エソボ〉の自立語の語彙を〈ヘイケ〉と比較すると次のことが言える。
- (c) 〈エソボ〉の語彙量を1.0とすると、〈ヘイケ〉の語彙量は異なり語数で約2.6、延べ語数で約4.0である。
- (d) 異なり語数の比率を品詞別で見ると、〈エソボ〉の名詞は〈ヘイケ〉より約10%少ない。また、〈エソボ〉の動詞・形容動詞・副詞は〈ヘイケ〉よりかなり多く、接続詞はやや多い<sup>(注3)</sup>。
- (e) 延べ語数の比率を品詞別で見ると、〈エソボ〉の形容動詞・連体詞は〈ヘイケ〉よりかなり多い。また、〈エソボ〉の動詞は〈ヘイケ〉よりかなり少なく、名詞はやや少ない<sup>(注4)</sup>。

### 3 語彙の豊富さ

#### 3.1 タイプ・トークン比を用いて

〈エソボ〉に用いられた語彙が〈ヘイケ〉と比較して豊富であるかどうかを知るのに、タイプ・トークン比 (Type / Token Ratio, TTR) という指標を用いることにする。一般的に、テキストの中に異なった単語が多く用いられていると、語彙が豊富であり、表現が豊かであるといえる。TTRは延べ語数 (Token) に対する異なり語数 (Type) の比率を示す指標である。数値が1に近くなるほど語彙が豊富である。〈エソボ〉と〈ヘイケ〉のTTRは次のようになる。

$$\langle \text{エソボ} \rangle \text{ の TTR} \quad 2904/11750=0.25$$

$$\langle \text{ヘイケ} \rangle \text{ の TTR} \quad 7421/46893=0.16$$

この結果によると〈エソボ〉の方が〈ヘイケ〉に比べてかなり語彙が豊富であるといえる。しかし、〈エソボ〉と〈ヘイケ〉では語彙量に大きな差があり、この結果ほどの違いがあるのか疑問が残る。TTRは延べ語数が多くなるほど数値が小さくなる傾向がある。そのため1000語単位で算出するとよいとも言われている。

そこで、キャロル (Carroll) が改良した TTR (TTR 2 とする) を利用して計算する。これも延べ語数の多さの影響から逃れられないが、その影響がもとの TTR よりもはるかに和らぎ、手順が簡便だという利点がある。TTR 2 は延べ語数に 2 をかけた平方根に対する異なり語数の比率を示す指標である。〈エソボ〉と〈ヘイケ〉の TTR 2 は次のようになる。

$$\langle \text{エソボ} \rangle \text{ の TTR 2} \quad 2904/\sqrt{11750 \times 2}=18.94$$

$$\langle \text{ヘイケ} \rangle \text{ の TTR } 2 \quad 7421 / \sqrt{46893 \times 2} = 24.23$$

TTR 2の結果によると、〈エソポ〉より〈ヘイケ〉の方が語彙が豊富であるといえる。

以上をまとめると次のようになる。

- (a) キャロルが改良したタイプ・トークン比を指標に用いると、〈エソポ〉より〈ヘイケ〉の方が語彙が豊富である。

### 3.2 K特性値を用いて

〈エソポ〉に用いられた語彙が〈ヘイケ〉と比較して豊富であるかどうかを知るのに、ユール (Yule) の K 特性値 (characteristic K) という指標を用いることにする。これは見出し語の使用度数を使う指標で、数値が小さいほど語彙が豊富であることを示す。延べ語数が N, 異なり語数が V の文章で、使用度数が m の見出し語の数を V (m, N) とした時、この値は次の式で定義されている。

$$K = \left\{ \sum_{m=1}^{m \text{ の最大}} m^2 V(m, N) - N \right\} / N^2 \times 10^4$$

〈エソポ〉の見出し語の数は、使用度数のもっとも多いものから降順に378が1語、323が1語、…2が461語、1が1608語である (表 I 参照)。さきに挙げた式に入れて計算すると次のようになる。

$$378^2 \times 1 + 323^2 \times 1 + \dots + 2^2 \times 461 + 1^2 \times 1608 = 655938$$

$$655938 - 11750 = 644188$$

$$644188 / 11750^2 \times 10^4 = 46.66 \quad \dots \langle \text{エソポ} \rangle \text{ の K 特性値}$$

〈ヘイケ〉の見出し語の数は、使用度数のもっとも多いものから降順に1000が1語、844が1語、…2が1228語、1が3580語である (表 II 参照)。

$$1000^2 \times 1 + 844^2 \times 1 + \dots + 2^2 \times 1228 + 1^2 \times 3580 = 6667887$$

$$6667887 - 46893 = 6620994$$

$$6620994 / 46893^2 \times 10^4 = 30.11 \quad \dots \langle \text{ヘイケ} \rangle \text{ の K 特性値}$$

算出した数値は〈エソポ〉より〈ヘイケ〉の方が小さい。K特性値の結果によると、〈エソポ〉より〈ヘイケ〉の方が語彙が豊富であるといえる。

以上をまとめると次のようになる。

- (b) K特性値を指標に用いると、〈エソポ〉より〈ヘイケ〉の方が語彙が豊富である。そして、この章をまとめると次のようになる。

- (c) 改良されたタイプ・トークン比を指標としても、K特性値を指標としても、〈エソポ〉より〈ヘイケ〉の方が語彙が豊富であると判断される。

## 4 語彙の類似度

〈エソポ〉の語彙が〈ヘイケ〉の語彙とどのくらい類似しているかをはかるのに、宮島 (1970)

で示された語彙の類似度  $C$  を用いることにする。これは作品  $A \cdot B$  に共通する 1 組の見出し語のうち、使用率の小さい方を取り出し、その使用率の総和をとったものである。ふたつの作品の使用率を  $P_i(A)$ 、 $P_i(B)$  とした時、この値は次の式で定義されている。

$$C_{AB} = \sum_i \min[P_i(A), P_i(B)]$$

$C$  の値が 1 に近くなるほど類似度は高くなる。〈エソポ〉と〈ヘイケ〉に共通する見出し語は 1559 語である。その使用率の小さい方の値の総和を計算すると 438.453% になる。類似度は 0.438 である。

この数値だけでは類似の程度がつかみにくいので、高野本『平家物語』（以下、〈高野本〉と略す。）の語彙との類似度も出してみよう<sup>(註5)</sup>。共通する見出し語の数、その使用率の小さい方の値の総和、類似度を表に示すと次のようになる。

表 3 3 作品の語彙の類似度

作品	共通する見出し語	使用率の小さい方の値の総和	類似度
〈エソポ〉〈ヘイケ〉	1559 語	438.453%	0.438
〈エソポ〉〈高野本〉	1644 語	354.886%	0.355
〈ヘイケ〉〈高野本〉	5561 語	641.222%	0.641

この結果から次のことが言える。

- (a) 〈エソポ〉と〈ヘイケ〉の方が〈エソポ〉と〈高野本〉よりも共通する見出し語の数は少ないが、語彙の類似度は高い。
- (b) 〈ヘイケ〉と〈高野本〉は共通する見出し語も大変多く、語彙の類似度もきわめて高い。
- (c) 〈エソポ〉と〈ヘイケ〉の語彙の類似度は、高いとも言えないが低いとも言えない。

## 5 語彙の偏り

### 5.1 3 作品の語彙の偏り

〈エソポ〉〈ヘイケ〉〈高野本〉の 3 作品すべてにわたって現れる見出し語は 1348 語である。この共通見出し語の各作品における現れ方の偏りをはかるのに、田中（1983）で示された期待値偏差を用いることにする。これは「語彙の偏りを、期待値に対する、期待値からの偏りの比率としてとらえようという考え方で」<sup>(註6)</sup>、期待値偏差は「作品相互間の偏りの比較にしか使えない、相対的な尺度である。」<sup>(註7)</sup>期待値偏差は次の手順で求める<sup>(註8)</sup>。

各作品の共通見出し語の使用度数を  $f$ 、全作品の共通見出し語の使用度数を  $\Sigma f$ 、各作品の延べ語数を  $n$ 、全作品の延べ語数を  $N$  とする。

$$\text{全体比率} : P = \Sigma f / N = 82013 / 158005 = 0.5191$$

$$\text{期待値} : f' = n \times P$$

$$\text{期待値偏差} : K.D. = (f - f') / f'$$

求めた数値を表にすると次のようである。

表4 3作品の共通見出し語の現れ方

作品名	延べ語数 n	共通語の使用度数 f	期待値 f'	f-f'	期待値偏差 K.D.
〈エソポ〉	11750	8371	6098.84	+2272.16	+0.3726
〈ヘイケ〉	46893	26456	24339.81	+2116.19	+0.0869
〈高野本〉	99362	47186	51573.85	-4387.85	-0.0851
	N 158005	Σf 82013			

全体比率（3作品全体での共通語群の使用率）は51.91%で、これが3作品全体を通じての平均像である。各作品の共通見出し語の使用度数と期待値が一致すると、期待値偏差の値がゼロになる。期待値偏差の絶対値が大きくなると、平均像からのズレが大きくなる。上の結果から、3作品における共通見出し語の現れ方の偏りに関して次のことが言える。

- (a) 〈エソポ〉の語彙は37.26%もプラスの方向に偏り、共通語群に大きく頼っている。
- (b) 〈ヘイケ〉の語彙は8.69%プラスの方向へ偏り、共通語群への傾斜がやや多い。
- (c) 〈高野本〉の語彙は8.51%マイナスの方向へ偏り、共通語群への傾斜がやや少ない。
- (d) 3作品の中では〈エソポ〉が平均像から大きく隔たっている。

## 5.2 17作品の語彙の偏り

期待値偏差は「作品相互間の偏りの比較にしか使えない、相対的な尺度」だという制約がある。5.1で得た結果は、日本語学習教材、または軍記という性格を持つ〈エソポ〉〈ヘイケ〉〈高野本〉の3作品の間でのみ通用する。そこで、調査対象の作品の範囲を広げ、3作品の語彙の偏りを確かめることにする。新しく「古典対照語い表」でとりあげている14作品を加え、合計17作品を用いる。「古典対照語い表」の14作品で共通の見出し語は137語である<sup>(註9)</sup>。この137語のうち、〈エソポ〉〈ヘイケ〉〈高野本〉の3作品と共通するのは124語である。それを「古典対照語い表」の付表の形式に従って示すと次のようになる（使用度数は表Ⅲ参照）。

表5 17作品に共通の124語

〔名詞〕 59語

(抽象) もの こと かひ よし; これ それ なに いづれ  
(とき) とき ころ いま のち もと むかし つね; 日 夜(よ) 年 春  
(ところ) ところ ここ あたり なか うち 方(かた)  
(数量) かず ほど かぎり みな ひとつ ひとり  
(人) 人 おや 子 国 世 世の中; われ たれ  
(動作・精神) わざ しるし 名 ころろ まこと  
(生産物) 道 かど  
(自然) 風 雲 水 かげ 音; 山 海  
(生物) 鳥 花; 身 手 足 目

〔動詞〕 41語

(関係) 似る 指す まさる  
(存在・消滅) あり るる  
(変化) 付く〔四〕 付く〔下二〕 とどむ なる 寄す

- (移動) 出づ 入る[四] 入る[下二] おく かへる のぼる のる ゆく わたる  
 (経過) 明く 暮る 過ぐ 経(ふ)  
 (精神) おもふ しる なげく わする; きく きこゆ みる みゆ  
 (行為) す なす; いふ 書く 問ふ 得(う) とる 着る 泣く 吹く  
 [形容詞] 9語  
 なし よし おなじ おほし; ちかし とほし たかし ふかし; くるし  
 [形容動詞] 3語  
 あはれ 如何(いか) 異(こと)  
 [副詞] 8語  
 あまた いと かく さらに しばし ただ なほ また  
 [連体詞] 4語  
 この その かの わが

この17作品の共通見出し語が、各作品にどのように現れたかを表にすると次のようである。なお、〈エソボ〉〈ヘイケ〉〈高野本〉の3作品の見出し語を、「古典対照語い表」の見出し語の注記を参考にして14作品に合わせ、使用度数を調整した<sup>(注10)</sup>。

表6 17作品の共通見出し語の現れ方

作品番号	作品	延べ語数 n	共通語の使用度数 f	期待値 $f'$	$f - f'$	期待値偏差 K.D.
1	エソボ	11750	3806	3710.65	+95.35	+0.0257
2	ヘイケ	46893	12458	14808.81	-2350.81	-0.1587
3	高野本	99362	21193	31378.52	-10185.52	-0.3246
4	徒然	17114	6468	5404.60	+1063.40	+0.1968
5	方丈	2527	889	798.03	+90.97	+0.1140
6	大鏡	29212	8665	9225.15	-560.15	-0.0607
7	更級	7243	2734	2287.34	+446.66	+0.1953
8	紫	8737	2466	2759.14	-293.14	-0.1062
9	源氏	207808	73227	65625.77	+7601.23	+0.1158
10	枕	32906	11330	10391.71	+938.29	+0.0903
11	蜻蛉	22398	9731	7073.29	+2657.71	+0.3757
12	後撰	11955	4607	3775.39	+831.61	+0.2203
13	土佐	3496	1426	1104.04	+321.96	+0.2916
14	古今	10015	3705	3162.74	+542.26	+0.1715
15	伊勢	6931	2729	2188.81	+540.19	+0.2468
16	竹取	5124	1946	1618.16	+327.84	+0.2026
17	万葉	50070	13741	15812.11	-2071.11	-0.1310
		N 573541	$\Sigma f$ 181121			

全体比率： $P = \Sigma f / N = 181121 / 573541 = 0.3158$

期待値： $f' = n \times P$

期待値偏差： $K.D. = (f - f') / f'$

全体比率（3作品全体での共通語群の使用率）は31.58%で、これが17作品全体を通じての平均像である。共通語群の占める割合は、蜻蛉日記が37.57%で最も高く、土佐日記・伊勢物語・後撰集が20%代で続いている。そして、〈高野本〉が-32.46%で最も低く、〈ヘイケ〉・万葉集・紫式部日記が-10%代で続いている。蜻蛉日記・土佐日記・伊勢物語・後撰集は共通語群に頼った作品であり、〈高野本〉・〈ヘイケ〉・万葉集・紫式部日記は共通語群への傾斜が低い作品だということになる。また、〈エソポ〉が2.57%で最もゼロに近く、+-10%以内のものに大鏡と枕草子がある。このあたりが平均像に近い作品になる。田中（1983）の「古典十四作品の語彙の偏り度」では「共通語群の占める割合は、蜻蛉日記において最も高く、万葉集で最も低い。」「源氏物語・方丈記・古今集・枕草子あたりが、いちばん平均像に近いと推定される」とある<sup>(註11)</sup>。〈エソポ〉〈ヘイケ〉〈高野本〉の3作品を加えた17作品で作品の語彙の偏りを見ると、〈高野本〉〈ヘイケ〉が万葉集よりも共通語群への傾斜が少なく、〈エソポ〉がどの作品よりも平均像に近いことが判明した。

上の結果から、17作品の共通見出し語の現れ方における3作品の語彙の偏りに関して、次のことが言える。

- (a) 〈エソポ〉の語彙は2.57%プラスの方向へ偏り、平均像に近い。
- (b) 〈ヘイケ〉の語彙は15.87%マイナスの方向へ偏り、共通語群への傾斜が少ない。
- (c) 〈高野本〉の語彙は32.46%マイナスの方向へ偏り、共通語群への傾斜がきわめて少ない。
- (d) 17作品の中では〈エソポ〉が最もゼロに近く、いちばん平均像に近い。そして、〈高野本〉〈ヘイケ〉が万葉集よりも共通語群への傾斜が少ない。

## 6 おわりに

統計上のいろいろな指標を利用して、天草版『エソポのハプラス』と天草版『平家物語』の自立語の語彙の傾向を探求し、次の点を記述した。

- (a) キャロルにより改良されたタイプ・トークン比、ユールのK特性値を指標に利用すると、〈エソポ〉より〈ヘイケ〉の方が語彙が豊富であると判断される。
  - (b) 宮島の語彙の類似度を指標に利用すると、〈エソポ〉と〈ヘイケ〉の方が〈エソポ〉と〈高野本〉よりも共通する見出し語の数は少ないが、語彙の類似度は高くなり、〈ヘイケ〉と〈高野本〉は共通する見出し語も大変多く、語彙の類似度もきわめて高くなる。
  - (c) 田中の期待値偏差を指標に利用して〈エソポ〉〈ヘイケ〉〈高野本〉の3作品の中で語彙の偏りを測ったときには、〈エソポ〉の語彙が最も大きく平均像から隔たり、共通語群に頼っているという結果を得た。しかし、古典14作品を合わせて17作品の中で語彙の偏りを測ったところ、〈エソポ〉の語彙が最も平均像に近く、〈高野本〉〈ヘイケ〉の語彙が古典14作品よりも共通語群への傾斜が少ないという結果になった。
- (c)の結果から次のようなことが推察できる。
- (d) 〈エソポ〉は古典でよく使われ、室町末期でも使った語彙を用いて翻訳された作品である。〈高野本〉はマイナス方向への偏りが大きく、他の古典作品と異なる語彙を多く使用している作品である。〈ヘイケ〉は〈高野本〉に見られるような平家物語の語彙を継承しつつ、古典でよく使われ、室町末期でも使った語彙を用いて書き直された作品である。

## 〈注記〉

- 注1 近藤政美・濱千代いづみ (2000) 「天草版『平家物語』の語彙の特色——『平家物語』〈高野本〉との比較による——」(愛知県立大学大学院『国際文化研究科論集』第1号)、濱千代いづみ (2009) 「天草版『エソポのハプラス』の助動詞の語彙——国字本『伊曾保物語』・天草版『平家物語』との比較を通して——」(『岐阜聖徳学園大学国語国文学』第28号)。
- 注2 近藤政美・濱千代いづみ (2000) の調査を利用し、〈エソポ〉と単語の認定の基準を一致させるように整理し修正した。
- 注3 有意差検定の結果のZスコアは名詞10.381, 動詞7.263, 形容動詞4.120, 副詞3.182, 接続詞1.792である。名詞・動詞・形容動詞・副詞は有意水準1%, 接続詞は有意水準10%で有意差がある。言いかえると、99%, 90%の確かさで〈エソポ〉と〈ヘイケ〉の比率の差は意味のある差であると言える。
- 注4 有意差検定の結果のZスコアは名詞1.667, 動詞6.557, 形容動詞5.790, 連体詞17.870, である。動詞・形容動詞・連体詞は有意水準1%, 名詞は有意水準10%で有意差がある。
- 注5 近藤政美・濱千代いづみ (2000) の調査を利用し、3作品の単語の認定の基準を一致させるように整理し修正した。
- 注6 田中章夫 (1983) の p.34, 14~15行
- 注7 同上の p.37, 9~10行
- 注8 同上の p.35で形容動詞の期待値を求める式が  $f' = f \times \sum f / N = f \times P$  と示してある。この  $f$  は各作品の形容動詞の使用度数であるので、このままでは期待値  $f'$  が求められない。この  $f$  は誤りで、各作品の延べ語数  $n$  が正しい。また、期待値偏差を求める式が  $K.D. = (f - f') / f'$  と示してあるが、p.36の6~7行の説明が「各作品の延べ語数と期待値の差をとって、それを期待値で割ると、ズレの率が出てくる」となっている。この説明の中の「延べ語数」は「形容動詞の使用度数」の誤りであろう。
- 注9 宮島達夫 (1971) に「14作品に共通の137語」として掲示してある。
- 注10 〈エソポ〉〈ヘイケ〉〈高野本〉で別語の扱いをしていた「あく(下二)」「あくる」, 「あはれ」「あはれなり」, 「いかなり」「いか」, 「こ」「こども」, 「こと(異)なり」「こと(殊)に」, 「まこと」「まことに」, 「もと(本)」「もと(旧)」, 「よ」「よに」を一語にした。
- 注11 田中章夫 (1983) の「表7 古典14作品共通の137語の現われ方」では古今集の共通見出し語の使用度数が3581語となっているが、「フロッピー版古典対照語い表」で調査したところ、3881語という結果を得た。この結果で計算すると、古今集は竹取物語に近くなり、いちばん平均像に近いとは言えない。

## 〈文献〉

- 伊藤雅光 (2002) 『計量言語学入門』 大修館書店発行
- 大塚光信・来田隆 (1999) 『エソポのハプラス本文と総索引』 清文堂出版発行
- 大塚光信 (1983) 『キリシタン版エソポのハプラス私注』 臨川書店発行
- 金明哲 (2009) 『テキストデータの統計科学入門』 岩波書店発行
- 近藤政美・池村奈代美・濱千代いづみ (1999) 『天草版平家物語語彙用例総索引』 勉誠出版発行
- 近藤政美・武山隆昭・近藤三佐子 (1996) 『平家物語〈高野本〉語彙用例総索引』(自立語篇) 勉誠社発行
- 田中章夫 (1983) 「作品の語彙の偏りを測る」『国語語彙史の研究』4 和泉書院発行
- 田中章夫 (2002) 『近代日本語の語彙と語法』 東京堂出版発行
- 宮島達夫 (1970) 「語いの類似度」『国語学』82
- 宮島達夫 (1971) 『古典対照語い表』 笠間書院発行
- 宮島達夫・中野洋・鈴木泰・石井久雄 (1989) 『フロッピー版古典対照語い表 および使用法』 笠間書院発行
- Carroll, J. B. (1964) *Language and Thought*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Yule, G. U. (1944) *The statistical study of literary vocabulary*. Cambridge University Press. Cambridge.

表I 〈エソボ〉の語彙の使用度数

使用度数	比率%	異なり語数	延べ語数
378	32.170	1	378
323	27.489	1	323
277	23.574	1	277
181	15.404	1	181
174	14.809	1	174
153	13.021	1	153
140	11.915	1	140
139	11.830	1	139
133	11.319	1	133
114	9.702	1	114
105	8.936	1	105
104	8.851	1	104
96	8.170	1	96
79	6.723	1	79
78	6.638	1	78
72	6.128	1	72
69	5.872	1	69
65	5.532	1	65
63	5.362	1	63
62	5.277	2	124
61	5.191	1	61
60	5.106	1	60
59	5.021	1	59
55	4.681	2	110
52	4.426	1	52
50	4.255	1	50
44	3.745	1	44
41	3.489	1	41
40	3.404	1	40
39	3.319	1	39
37	3.149	2	74
35	2.979	2	70
34	2.894	2	68
33	2.809	5	165
32	2.723	1	32

  

使用度数	比率%	異なり語数	延べ語数
31	2.638	3	93
29	2.468	3	87
28	2.383	2	56
27	2.298	3	81
26	2.213	5	130
25	2.128	3	75
24	2.043	7	168
23	1.957	3	69
22	1.872	4	88
21	1.787	1	21
20	1.702	9	180
19	1.617	3	57
18	1.532	6	108
17	1.447	9	153
16	1.362	8	128
15	1.277	7	105
14	1.191	7	98
13	1.106	10	130
12	1.021	20	240
11	0.936	13	143
10	0.851	21	210
9	0.766	22	198
8	0.681	39	312
7	0.596	45	315
6	0.511	74	444
5	0.426	85	425
4	0.340	137	548
3	0.255	242	726
2	0.170	461	922
1	0.085	1608	1608
全体		2904	11750

表Ⅱ 〈ハイケ〉の語彙の使用度数

使用度数	比率%	異なり語数	延べ語数	使用度数	比率%	異なり語数	延べ語数
1000	21.325	1	1000	113	2.410	1	113
844	17.998	1	844	112	2.388	2	224
822	17.529	1	822	110	2.346	1	110
754	16.079	1	754	109	2.324	1	109
716	15.269	1	716	108	2.303	1	108
545	11.622	1	545	106	2.260	3	318
507	10.812	1	507	102	2.175	2	204
427	9.106	1	427	101	2.154	1	101
408	8.701	1	408	98	2.090	2	196
388	8.274	1	388	95	2.026	1	95
384	8.189	1	384	93	1.983	1	93
350	7.464	1	350	90	1.919	1	90
341	7.272	1	341	88	1.877	1	88
328	6.995	1	328	87	1.855	2	174
287	6.120	1	287	86	1.834	2	172
272	5.800	1	272	84	1.791	3	252
258	5.502	1	258	83	1.770	2	166
230	4.905	1	230	82	1.749	2	164
204	4.350	1	204	81	1.727	1	81
200	4.265	1	200	80	1.706	1	80
183	3.903	1	183	79	1.685	3	237
169	3.604	1	169	78	1.663	1	78
162	3.455	2	324	77	1.642	1	77
161	3.433	1	161	76	1.621	1	76
158	3.369	1	158	75	1.599	1	75
153	3.263	1	153	74	1.578	1	74
147	3.135	2	294	73	1.557	3	219
145	3.092	1	145	72	1.535	2	144
136	2.900	1	136	71	1.514	1	71
133	2.836	1	133	70	1.493	2	140
127	2.708	1	127	69	1.471	2	138
123	2.623	1	123	68	1.450	2	136
118	2.516	1	118	67	1.429	2	134
117	2.495	1	117	66	1.407	3	198
115	2.452	1	115	63	1.343	4	252

使用度数	比率%	異なり語数	延べ語数
62	1.322	2	124
61	1.301	2	122
60	1.280	2	120
59	1.258	4	236
57	1.216	2	114
56	1.194	1	56
55	1.173	6	330
54	1.152	4	216
53	1.130	4	212
52	1.109	7	364
51	1.088	5	255
50	1.066	2	100
49	1.045	3	147
48	1.024	6	288
47	1.002	6	282
46	0.981	4	184
45	0.960	3	135
44	0.938	3	132
43	0.917	3	129
42	0.896	8	336
41	0.874	3	123
40	0.853	5	200
39	0.832	5	195
38	0.810	4	152
37	0.789	4	148
36	0.768	8	288
35	0.746	7	245
34	0.725	8	272
33	0.704	7	231
32	0.682	2	64
31	0.661	9	279
30	0.640	9	270
29	0.618	12	348
28	0.597	14	392
27	0.576	10	270

使用度数	比率%	異なり語数	延べ語数
26	0.554	14	364
25	0.533	13	325
24	0.512	15	360
23	0.490	10	230
22	0.469	17	374
21	0.448	30	630
20	0.427	21	420
19	0.405	14	266
18	0.384	27	486
17	0.363	29	493
16	0.341	26	416
15	0.320	30	450
14	0.299	31	434
13	0.277	45	585
12	0.256	60	720
11	0.235	71	781
10	0.213	71	710
9	0.192	92	828
8	0.171	135	1080
7	0.149	140	980
6	0.128	208	1248
5	0.107	270	1350
4	0.085	371	1484
3	0.064	592	1776
2	0.043	1228	2456
1	0.021	3580	3580
全体		7421	46893

表Ⅲ 17作品の共通見出し語の使用度数

番号	かな見出し	漢字表記	品詞	エソポ	ヘイケ	高野	徒然	方丈	大鏡	更級	紫	源氏	枕	蜻蛉	後撰	土佐	古今	伊勢	竹取	万葉	全体
1	あり	有	動ヲ変	181	1000	1469	490	51	397	169	110	4450	763	1016	219	103	176	219	118	753	11684
2	こと	事		323	844	989	538	45	720	98	139	4804	449	398	98	40	54	58	108	105	9810
3	ひと	人		104	350	559	572	46	326	134	189	3732	669	438	211	80	235	173	94	465	8377
4	す	為	動ヲ変	174	545	1007	360	47	455	128	146	3060	654	497	173	65	139	133	97	631	8311
5	なし	無	形	139	388	654	305	44	221	104	103	3346	284	332	212	31	153	75	61	514	6966
6	いと	甚	副	5	4	11	69	2	362	95	148	4237	715	436	9	19	4	70	31	13	6230
7	いふ	言	動四	378	822	704	354	17	223	120	69	1228	735	466	89	124	62	134	147	250	5922
8	おもふ	思	動四	78	341	494	129	4	155	74	64	2468	283	444	192	27	163	118	46	400	5480
9	ところ	心		39	118	239	168	29	129	49	87	3411	156	171	168	26	117	57	32	330	5326
10	もの	物・者		153	408	500	263	25	195	52	80	1942	535	225	174	41	124	57	76	302	5152
11	みる	見	動上一	62	272	422	111	11	163	97	59	1839	227	246	214	39	158	74	69	794	4857
12	この	此	連体	140	384	591	108	13	506	67	28	1624	108	111	19	61	13	65	74	179	4091
13	ほど	程		24	54	253	80	12	200	56	64	1762	253	330	32	4	6	16	18	8	3172
14	なる	成・為	動四	33	287	540	77	13	233	49	25	919	151	250	85	17	71	54	32	136	2972
15	その	其	連体	277	507	746	132	27	219	52	34	568	110	59	1	12	2	47	23	99	2915
16	よ	世		12	127	235	96	21	179	37	34	1570	55	62	61	1	50	22	15	71	2648
17	きこゆ	聞	動下二	6	127	250	21	3	96	16	16	1659	103	103	18	3	7	9	1	25	2463
18	わが	我	連体	61	51	129	35	7	40	18	19	422	37	72	145	9	140	28	12	970	2195
19	これ	此		105	427	746	73	21	98	24	13	368	128	88	7	21	6	20	35	12	2192
20	みゆ	見	動下二	12	112	237	60	4	73	45	56	890	181	129	71	15	53	26	13	200	2177
21	いま	今		31	158	340	29	11	155	28	12	836	78	100	57	16	49	25	13	181	2119
22	また	又	副	59	162	251	80	21	230	25	30	698	217	155	9	23	8	15	7	65	2055
23	しる	知・領	動四	26	102	181	127	21	58	25	16	591	133	78	112	16	86	31	21	282	1906
24	かた	方		8	84	155	30	4	58	27	34	1159	80	96	25	5	20	15	11	11	1822
25	いか	如何	形動	42	117	357	41	4	90	22	22	698	120	133	26	3	14	5	7	83	1784
26	かく	斯	副	3	13	72	37	9	111	9	16	1066	35	119	20	18	18	20	28	153	1747
27	ところ	所		62	147	324	83	14	49	52	20	598	148	135	3	44	4	30	17	1	1731
28	さく	聞	動四	35	162	194	39	7	82	35	13	535	121	153	67	17	31	25	30	129	1675
29	とき	時		114	153	352	98	18	127	12	14	225	33	59	46	15	51	33	20	231	1601
30	ただ	唯・直・徒	副	32	115	206	40	10	78	20	30	727	154	73	10	7	8	7	9	73	1599
31	み	身		55	98	186	59	19	29	12	18	685	26	77	140	2	80	25	11	54	1576
32	あはれ	哀	形動	2	63	132	37	1	66	44	15	944	87	99	16	5	20	11	4	9	1555
33	われ	我		72	84	157	19	2	49	18	16	356	38	75	81	7	81	29	10	440	1534
34	なほ	尚	副	7	52	106	29	5	48	11	10	826	148	85	24	20	13	18	14	29	1445
35	うち	内・内裏		27	133	251	45	7	73	22	33	623	69	46	5	8	9	7	9	41	1408

番号	かな見出し	漢字表記	品詞	エソポ	ハイケ	高野	徒然	方丈	大鏡	更級	紫	源氏	枕	蜻蛉	後撰	土佐	古今	伊勢	竹取	万葉	全体
36	よし	良	形	20	55	78	109	2	77	10	32	527	179	41	2	13	5	13	23	74	1260
37	なか	中・仲		33	95	229	28	8	70	17	15	448	61	41	12	9	17	14	19	25	1141
38	かの	彼	連体	63	35	145	20	3	33	2	11	730	21	24	1	4	2	24	20	2	1140
39	はな	花		1	31	61	22	2	19	20	4	273	64	23	132	6	146	23	3	308	1138
40	おほし	多	形	15	50	126	80	9	53	8	20	566	70	34	9	6	2	6	7	49	1110
41	ひ	日		8	52	119	30	4	83	19	16	240	84	141	18	26	27	10	6	221	1104
42	みな	皆		5	145	245	40	7	57	17	30	343	98	38	6	10	2	6	9	17	1075
43	いづ	出	動下二	4	36	207	14	5	60	17	16	352	78	73	20	14	33	26	10	100	1065
44	ゆく	行	動四	24	55	95	22	2	23	16	7	185	21	42	59	24	47	32	6	381	1041
45	のち	後		23	106	247	50	5	98	17	11	277	47	23	10	2	13	11	6	80	1026
46	やま	山		11	53	71	12	6	18	36	4	141	28	28	88	9	91	17	25	372	1010
47	よ	夜		1	67	123	18	2	27	53	23	262	49	64	54	10	45	28	4	173	1003
48	なく	泣・鳴	動四	5	52	63	4	2	15	16	2	216	32	47	77	4	99	26	9	330	999
49	なに	何		13	35	54	27	4	34	26	12	379	104	58	54	5	33	10	17	78	943
50	おなじ	同	形	6	90	285	19	6	87	8	15	293	45	38	16	14	4	2	3	3	934
51	つく	着・付	動下二	6	47	84	42	8	31	8	10	552	59	34	6	3	2	11	11	17	931
52	むかし	昔		1	48	144	18	4	33	16	5	375	16	17	20	9	15	149	5	13	888
53	とし	年		4	40	71	24	4	160	12	8	229	17	40	64	3	45	12	6	128	867
54	とる	取	動四	29	136	250	34	2	37	5	19	146	47	22	1	2	8	13	42	72	865
55	かぎり	限		3	31	41	11	1	32	17	17	530	38	36	23	2	15	8	13	9	827
56	それ	其		55	87	140	12	2	109	8	18	205	93	45	5	3	5	15	17	8	827
57	め	目		5	42	69	30	5	26	12	19	438	37	27	15	2	20	14	5	56	822
58	ふかし	深	形	8	52	84	21	7	9	11	11	475	16	25	30	2	13	1	4	18	787
59	わたる	渡	動四	5	29	110	7	1	42	16	8	375	32	41	23	9	15	6	1	61	781
60	こと	異	形動	2	13	52	21	6	13	5	27	543	46	19	5	1	7	4	1	3	768
61	かく	書	動四	4	71	118	29	2	29	13	18	247	96	73	1	5	2	16	9	1	734
62	かへる	帰	動四	24	66	124	15	2	23	25	2	168	51	88	39	8	13	16	16	41	721
63	ちかし	近	形	10	34	57	20	7	35	19	7	347	67	35	21	4	9	4	6	38	720
64	まこと	真・誠		19	91	117	37	2	59	4	6	233	78	5	2	2	1	2	8	18	684
65	いる	人・粟	動四	15	88	148	23	1	20	9	10	223	45	25	13	5	9	9	12	27	682
66	みち	道		16	41	79	61	7	28	14	2	198	21	19	14	2	20	17	4	135	678
67	つね	常		4	20	45	24	8	36	6	10	335	49	14	24	2	10	7	2	76	672
68	とふ	問	動四	26	53	74	32	1	43	17	1	142	56	48	35	7	17	12	9	71	644
69	こ	子・蚕		21	54	115	17	1	27	4	1	111	39	20	3	9	1	12	17	184	636
70	おく	置	動四	22	61	105	15	3	24	1	2	143	23	15	48	2	24	6	2	135	631
71	もと	元・本		15	54	113	16	5	32	12	11	149	94	20	4	4	5	45	13	22	614

番号	かな見出し	漢字表記	品詞	エソボ	ヘイケ	高野	徒然	方丈	大鏡	更級	紫	源氏	枕	蜻蛉	後撰	土佐	古今	伊勢	竹取	万葉	全体
72	すぐ	過	動上二	10	47	81	22	3	18	19	9	211	29	33	14	2	8	5	2	100	613
73	て	手		13	53	114	20	1	25	5	4	196	36	22	5	6	6	6	10	89	611
74	かぜ	風		3	34	66	5	8	8	15	4	179	30	36	50	15	50	11	9	86	609
75	ゐる	居	動上一	37	113	30	7	2	28	8	33	153	95	13	6	2	3	8	8	54	600
76	ここ	此処		20	70	64	11	2	9	6	4	251	22	71	2	7	5	5	6	41	596
77	よし	由		12	59	187	19	1	20	9	6	160	10	7	11	1	10	6	8	55	581
78	ふ	経	動下二	2	15	44	4	8	13	2	3	211	6	41	65	11	38	23	6	84	576
79	たれ	誰		12	30	64	15	4	22	8	5	224	49	20	28	2	24	6	2	47	562
80	はる	春		2	21	54	12	2	5	12	1	119	19	13	72	5	71	14	2	127	551
81	つく	着・付	動四	11	82	173	26	2	52	5	1	104	27	16	3	3	3	8	4	30	550
82	ころ	頃		4	41	73	27	6	26	4	7	245	36	21	10	1	8	2	2	13	526
83	な	名		4	27	56	25	1	15	3	4	124	37	5	43	7	40	11	6	104	512
84	くに	国		27	19	64	18	2	20	37	2	55	15	3	5	11	5	32	27	150	492
85	まさる	勝	動四	2	6	12	28	1	14	4	12	254	16	22	29	4	22	15	1	43	485
86	かひ	効		2	13	14	6	1	10	4	1	299	25	49	31	6	3	7	10	1	482
87	よのなか	世中		1	14	18	2	4	30	8	3	249	9	26	22	2	32	8	4	44	476
88	わざ	業		6	1	2	23	5	17	1	15	324	24	23	1	2	1	7	3	7	462
89	みづ	水		13	42	78	17	6	5	21	7	77	18	23	31	6	36	16	3	62	461
90	さらに	更	副	7	9	17	15	7	23	1	3	262	55	28	6	2	1	3	8	13	460
91	あまた	数多		14	29	48	15	1	29	2	7	196	44	32	4	1	6	7	7	10	452
92	あく	開・明	動下二	1	51	104	4	1	9	7	5	82	39	44	23	3	17	11	1	49	451
93	くるし	苦	形	1	21	24	7	2	4	6	4	238	13	28	9	5	11	3	3	51	430
94	きる	着	動上一	2	54	92	4	1	6	9	9	71	70	5	19	1	11	5	4	54	417
95	ひとつ	一		22	40	82	25	8	37	9	6	107	25	18	8	3	9	6	4	5	414
96	おと	音		2	19	39	7	3	3	19	12	115	37	36	19	1	18	4	4	54	392
97	なす	為	動四	33	38	114	26	7	28	6	1	78	12	10	11	1	4	3	5	14	391
98	いる	入	動下二	7	49	108	5	1	13	2	9	125	23	20	1	5	1	4	9	2	384
99	わする	忘	動下二	6	24	38	23	1	9	7	5	110	17	9	29	5	19	16	1	63	382
100	のほる	上	動四	10	73	113	8	1	32	17	4	52	28	14	1	6	1	4	8	9	381
101	ひとり	一人		3	6	13	24	3	38	10	4	104	22	24	15	3	15	7	5	74	370
102	ふく	吹	動四	5	15	27	5	5	7	12	1	54	23	21	30	14	34	8	11	92	364
103	のる	乗	動四	3	59	124	7	3	21	4	7	29	30	18	4	5	1	4	11	25	355
104	とどむ	止	動下二	1	11	42	3	2	14	5	4	197	16	12	7	1	14	4	5	12	350
105	しばし	暫	副	1	37	57	10	3	8	2	2	160	24	30	3	5	2	1	2	1	348
106	やす	寄	動下二	2	42	78	4	2	4	4	4	91	7	20	6	6	3	1	3	56	333
107	たかし	高	形	4	20	45	10	2	14	6	5	93	39	8	11	1	12	5	2	49	326



