

「西行物語」の語彙

—『西行物語コーパス』を活用して—

富士池優美

要 約

「西行物語」5系統8種類の諸本を対象とした『西行物語コーパス』に基づき、語彙量とTTR、語種比率、品詞比率（名詞率とMVR）、高頻度語、共通度の5観点から、「西行物語」諸本の比較、『日本語歴史コーパス 鎌倉時代編』所収の同時代資料との比較を行った。その結果、采女本系3本の類似が数値として明らかになった。また、同時代資料を比較すると、「西行物語」諸本は類似したテキスト群であることが明らかになった。

キーワード：西行物語、諸本分類、コーパス、類似性、語彙の量的構造

はじめに

筆者らは、「西行物語」の諸本をもとにしたコーパス（以下、『西行物語コーパス』とする）を構築した。『西行物語コーパス』は、「西行物語」の諸本を比較して類似性や文体差を客観的に求めることを目的としており、諸本それぞれの本文に対して、語の見出しや品詞、語種といった形態論情報を付与した。本稿では、『西行物語』の諸本5系統8種類を対象に構築された『西行物語コーパス』から見る、「西行物語」の語彙の量的構造について示す。

1. 『西行物語コーパス』データの概要

「西行物語」は「西行の生涯を多数の歌をまじえて記した、鎌倉時代の物語。作者未詳。《西行一生涯草紙》《西行四季物語》《西行一代記》《西行記》とも称され、絵巻の形で伝わる」¹⁾ 作品群である。蔡(2021)に「『西行物語』は西行没後五十年ほどに成立したと推定されている。そして、時代の流れとともに、今に見る広本系・略本系・采女本系・中間本系（永正本寛永本系）といった、内容が少しずつ異なるテキストに改編された。」と説明され、蔡(2012)には陸奥への旅の経路が本の系統によって異なることが指摘されている。

『西行物語コーパス』構築にあたっては、『西行物語』の5系統の諸本から代表的なテキストを選定した。選定されたテキストは以下の8種類である。

- | | | |
|--------|-------------------------|---------------|
| ① 広本系 | 書陵部所蔵『西行物語』文明十二年本 | 文明本 |
| ② 略本系 | 島原図書館松平文庫所蔵『西行聖人物語』 | 島原本 |
| ③ 略本系 | 静嘉堂文庫蔵『西行物語』 | 静嘉堂文庫本(伝阿仏尼本) |
| ④ 中間本系 | 岩国徴古館所蔵『西行絵詞』 | 岩国本 |
| ⑤ 松平本系 | 島原図書館松平文庫所蔵『西行発心物語』 | 松平本 |
| ⑥ 采女本系 | 今治市河野美術館所蔵『西行四季物語』宝永五年本 | 宝永本 |
| ⑦ 采女本系 | 篠山市青山歴史村蔵『西行之絵草紙』 | 篠山本 |
| ⑧ 采女本系 | センチュリーミュージアム蔵『西行法師行状絵巻』 | センチュリー文化本 |

以下、この8種類のテキストをそれぞれ「文明本」「島原本」「静嘉堂文庫本」「岩国本」「松平本」「宝永本」「篠山本」「センチュリー文化本」の略称で示す。

これらの選定したテキストに対して、翻刻、本文整形（濁点、句読点の付与等）、形態素解析、形態論情報（単語区切り、品詞等）の人手修正という手順でコーパス化を進めた²⁾。本文整形の方針としては、テキストの変更を最低限に抑えたものとした。コーパス作成手順や人手修正の詳細については富士池・鴻野（2018）を参照されたい。

『西行物語コーパス』における単位（単語の区切り方）は『日本語歴史コーパス 鎌倉時代編』の「短単位」を採用し、同じ基準で語彙素・語彙素読み・品詞・活用型・活用形・語種等の形態論情報を付与している。形態論情報の付与にあたっては、形態素解析支援アプリケーション「Web茶まめ」を使用し、形態素解析辞書UniDicのうち「中世文語」の辞書を選択した³⁾。形態素解析によって付与された形態論情報を作業者が確認し、誤解析部分を人手で修正した。表1はコーパス内の情報の具体例である⁴⁾。書字形出現形が本文、語彙素は見出し語（現代語形）の代表表記、語彙素読みは見出し語の読みを示す。

『西行物語コーパス』のデータに基づき、富士池・鴻野（2018）では、島原本・宝永本・岩国本の3本を対象に共通の場面を取り上げ、場面構成、内容の異同、描写（情報量）の違いが

表1 コーパス内の情報の具体例

書字形出現形	語彙素	語彙素読み	品詞	活用型	活用形	語種
ほんがく	本覚	ホンガク	名詞-普通名詞-一般			漢
しんによ	真如	シンニョ	名詞-普通名詞-一般			漢
の	の	ノ	助詞-格助詞			和
月	月	ツキ	名詞-普通名詞-助数詞可能			和
いで	出でる	イデル	動詞-一般	文語下二段-ダ行	連用形-一般	和
がたく	難い	ガタイ	接尾辞-形容詞的	文語形容詞-ク	連用形-一般	和

あることを報告した。西行の旅のうち、天竜川の渡りから岡部までの場面は、島原本と岩国本は場面の構成が同一であるのに対して、宝永本は「さやの中山」の後、岡部の場面がない。天竜川の渡りの冒頭の一文のみを見ても、島原本は39語から成るのに対し、宝永本は23語、岩国本は99語と、文と文章の構成に大きな違いがあることを指摘した。「西行物語」諸本とは言っても表現の異同という表現では収まらない内容の異同がある。場面構成や描写（情報量）の程度が異なる作品群である「西行物語」を対象としていることが『西行物語コーパス』の大きな特徴となっている。

2. 『西行物語コーパス』から見る語彙の量的構造

形態論情報を付与したコーパスのデータを集計することによって語彙表が得られる。ここでは、語彙表⁵⁾に基づく分析の結果を示す。

(1) 語彙量とTTR

表2に「西行物語」諸本の延べ語数・異なり語数・TTRを示す。集計にあたり、補助記号（主に句読点）、題・巻・章段、解釈不明箇所、経文や漢籍の引用、見せ消し、刊記・奥書を除いた。

表2 「西行物語」諸本の異なり語数・延べ語数・TTR

	異なり語数	延べ語数	TTR
文明本	2065	14123	0.146
島原本	1871	10264	0.182
静嘉堂文庫本	1159	4402	0.263
岩国本	1873	12324	0.152
松平本	2938	20446	0.144
宝永本	1391	6713	0.207
篠山本	1085	4904	0.221
センチュリー文化本	1652	8043	0.205

延べ語数・異なり語数は資料規模を表す指標である。TTR (Type/Token Ratio) は異なり語数を延べ語数で割ったもので、語彙の豊かさを表す指標としてよく用いられる。TTRの数値が高いほど、見出し語の種類が多く、語彙が豊かであるとされる。TTRの高さの要因としては、構成場面が多く、場面ごとに異なる語が用いられることや、情報量が多い、言い換えると描写が密であることが想定される。

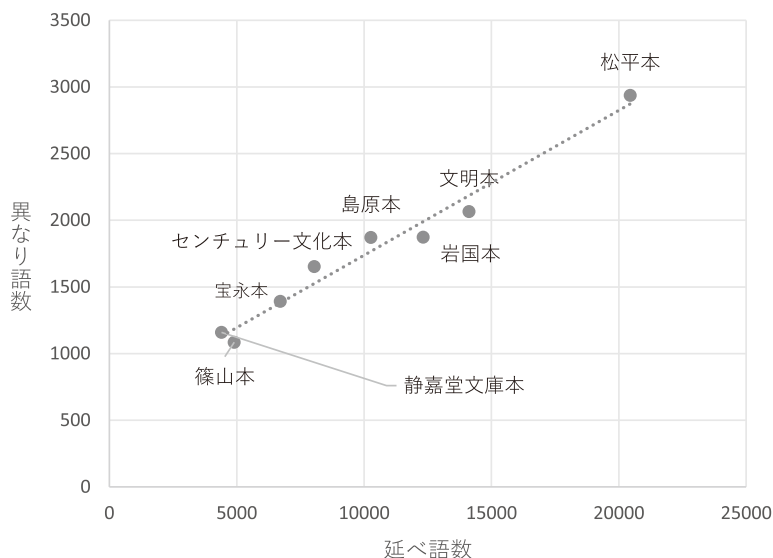


図1 「西行物語」諸本の延べ語数と異なり語数の相関

横軸に延べ語数、縦軸に異なり語数をとって今回の結果を見る（図1）と、「西行物語」の系統とは無関係に、延べ語数が多くなるにつれ異なり語数も多くなるという一般的な傾向が見られる。富士池（2021）では5本の比較の中で、松平本は資料規模と比較してTTRが低く、描写の疎密について注意してみる必要があるテキストだと指摘した。しかし本の種類を増やしたところ、資料規模が大きい松平本と文明本はTTRが同程度であった。「西行物語」の諸本は大まかな内容は共通している。資料規模が大きい本では「西行物語」の語彙が見出し語レベルで出揃っており、基礎的な語が繰り返し用いられているためにTTRが一定に保たれている可能性がある。

(2) 語種比率

「西行物語」諸本について、語種比率を比較する。UniDicにおける語種は大きく和（和語）・漢（漢語）・外（外来語）・混（混種語）・固（固有名）の5種類に分類される⁶⁾。外来語には「摩訶」のようなサンスクリット語由来の語が含まれる。混種語は主に「報ず」「具す」のような一字漢語サ変動詞である。

表3に延べ語数による「西行物語」諸本の語種比率を示す。どのテキストも和語が多く、90%前後と同程度である。宝永本では漢語の割合が比較的低いものに対して、松平本ではやや高くなっているが、大きな差ではない。

比較材料として、表4に『日本語歴史コーパス 鎌倉時代編 I 説話・随筆』『日本語歴史コーパス 鎌倉時代編 II 日記・紀行』所収10作品の語種比率⁷⁾を示した。『西行物語コーパス』と同様に和語・漢語・外来語・混種語・固有名の5種類を対象とした。和語の比率を見ると『十

「西行物語」の語彙

表3 「西行物語」諸本の語種比率

	和	漢	外	混	固
文明本	90.6%	7.1%	0.2%	0.3%	1.8%
島原本	90.1%	7.2%	0.1%	0.4%	2.1%
静嘉堂文庫本	89.6%	7.3%	0.1%	0.6%	2.4%
岩国本	90.2%	7.3%	0.2%	0.6%	1.6%
松平本	89.0%	8.0%	0.2%	0.2%	2.6%
宝永本	91.4%	6.4%	0.1%	0.3%	1.9%
篠山本	89.8%	7.5%	0.2%	0.4%	2.1%
センチュリー文化本	90.0%	7.4%	0.2%	0.4%	2.0%
総計	90.0%	7.3%	0.2%	0.4%	2.1%

訓抄』『海道記』では約60%と低く、『建礼門院右京大夫集』『十六夜日記』では85%を超えている。これに対応するように漢語比率が『海道記』『十訓抄』は高く、『建礼門院右京大夫集』『十六夜日記』は低くなっている。『日本語歴史コーパス 鎌倉時代編Ⅰ 説話・随筆』『日本語歴史コーパス 鎌倉時代編Ⅱ 日記・紀行』所収の鎌倉時代の和文・和漢混淆文の作品を平均すると、和語が約70%、漢語が約20%、外来語や混種語があわせて2%程度、固有名が約8%といった語種比率になっているが、作品による差が大きいことがわかる。

これらと比較すると、「西行物語」諸本の語種比率の差は非常に小さいことがわかる。ここから、「西行物語」諸本には内容の異同があり、場面構成や描写（情報量）の程度が異なることがすでに明らかになっているが、語種比率は一定していると見ることができる。

表4 『日本語歴史コーパス 鎌倉時代編Ⅰ』『日本語歴史コーパス 鎌倉時代編Ⅱ 日記・紀行』所収10作品の語種比率

	和	漢	外	混	固
今昔物語集	74.5%	17.8%	0.4%	1.7%	5.6%
方丈記	79.1%	16.2%	0.2%	1.1%	3.4%
宇治拾遺物語	72.5%	17.9%	0.4%	1.9%	7.4%
海道記	60.0%	33.1%	0.4%	1.4%	5.1%
建礼門院右京大夫集	89.6%	5.8%	0.1%	0.8%	3.6%
東関紀行	77.2%	12.4%	0.3%	0.5%	9.6%
十訓抄	56.9%	28.6%	0.3%	1.8%	12.4%
十六夜日記	87.2%	5.1%	0.1%	0.2%	7.4%
とはずがたり	72.1%	19.9%	0.3%	1.4%	6.3%
徒然草	70.1%	22.1%	0.3%	1.6%	5.9%

(3) 品詞比率

自立語の品詞比率は文体を示す指標として用いられる。ここでは、延べ語数の品詞比率に基づきテキストの特徴を示す指標として、名詞率とMVRを用いる。名詞の比率は文章の特質を表し、名詞の比率に応じて他の品詞もある傾向を持って変化する、つまり文章のジャンルによって品詞の割合が決定されると考えられる。



図2 名詞率とMVRの組み合わせから見出せる文体的特徴

樺島忠夫（1955）は名詞の比率は文章の特質を表し、名詞の比率に応じて他の品詞もある傾向を持って変化する、つまり文章のジャンルによって品詞の割合が決定されることを指摘している。名詞率が高い文章は「要約的な文章」とされる。逆に、名詞率が低い文章は「冗長な文章」となる。また、樺島・寿岳（1965）は、自立語について品詞をその機能によって体（名詞）・用（動詞）・相（形容詞・形容動詞・副詞・連体詞）・他（接続詞・感動詞）の四つに分類したとき、体の類と、用・相それぞれの類の関係を見るにあたり、MVRという「 $100 \times$ 相の類の比率 / 用の類の比率」の式で表される指標を提案した。

UniDicの短単位品詞体系では、体の類に「名詞-普通名詞-{一般・サ変可能・サ変形状詞可能・副詞可能・助数詞可能・形状詞可能}」「名詞-固有名詞-{一般・人名-一般・人名-名・人名-姓・地名-一般・地名-国}」「名詞-数詞」「代名詞」が、用の類に「動詞-{一般・非自立可能}」が、相の類に「形容詞-{一般・非自立可能}」「形状詞-{一般・タリ}」「副詞」「連体詞」が、他に「感動詞-一般」「接続詞」が含まれる。

名詞率とMVRの組み合わせから見出せる文体的特徴として、名詞率が高くMVRが小さいものを「要約的」な文章、名詞率が低くMVRが大きいものを「ありさま描写的」な文章、名詞率が低くMVRも小さいものを「動き描写的」な文章と位置づけられている。図2に名詞率とMVRの組み合わせから見出せる文体的特徴を整理したものを示す⁸⁾。

「西行物語」諸本の品詞比率を同時代の資料の品詞比率と比較するとどのような位置づけになるのか、名詞率とMVRの組み合わせに基づき検討する。「西行物語」諸本5系統8種類と『日本語歴史コーパス 鎌倉時代編Ⅰ 説話・随筆』『日本語歴史コーパス 鎌倉時代編Ⅱ 日記・紀行』

「西行物語」の語彙

表5 「西行物語」諸本5系統8種類と『日本語歴史コーパス 鎌倉時代編』所収10作品の名詞率・MVR

略称	資料	名詞率	MVR	略称	資料	名詞率	MVR
今昔	今昔物語集	57.4	41.1	文明	文明本	57.5	26.2
方丈	方丈記	58.6	38.5	島原	島原本	56.9	30.7
宇治	宇治拾遺物語	61.6	43.8	静嘉	静嘉堂文庫本	59.3	27.0
海道	海道記	66.3	29.8	岩国	岩国本	56.3	26.7
建礼	建礼門院右京大夫集	55.0	51.9	松平	松平本	57.4	27.7
東関	東関紀行	64.0	39.7	宝永	宝永本	55.6	24.7
十訓	十訓抄	68.4	38.1	篠山	篠山本	55.8	23.7
十六	十六夜日記	58.6	44.1	セ文	センチュリー文化本	57.1	23.8
とは	とはずがたり	65.1	47.4				
徒然	徒然草	59.7	46.9				

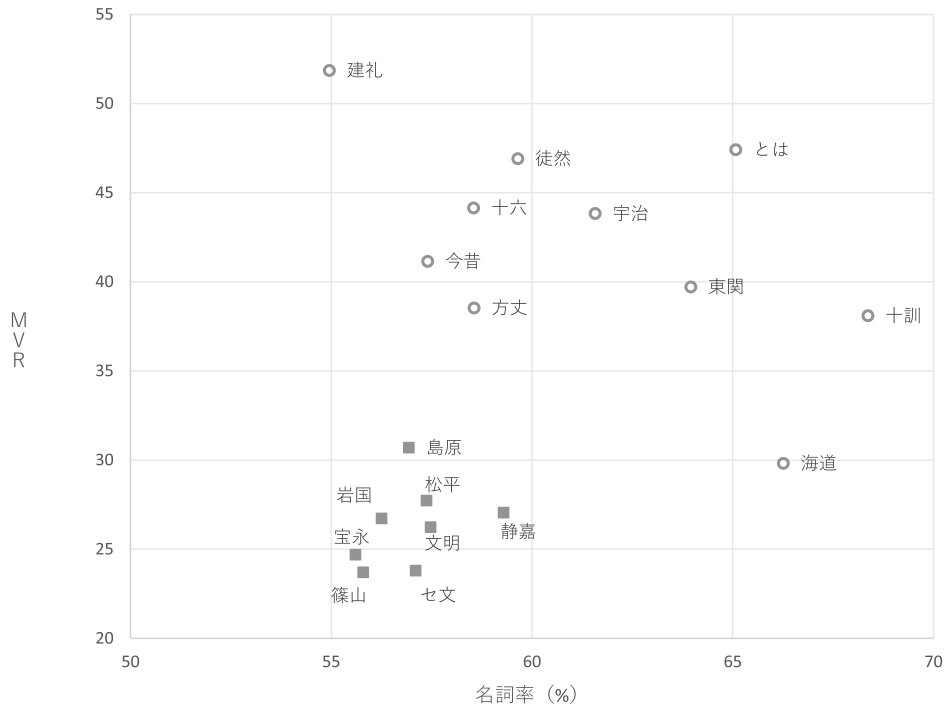


図3 「西行物語」諸本5系統8種類と『日本語歴史コーパス 鎌倉時代編』所収10作品の名詞率・MVR

の名詞率・MVRを表5・図3に示す。横軸が名詞率(%)、縦軸がMVR、■は「西行物語」諸本、○は『日本語歴史コーパス 鎌倉時代編』所収10作品である。図3では資料名を表5にある略称で示している。

「西行物語」諸本は図3の右下に位置しており、名詞率もMVRも低い「動き描写的」な文章としての特徴を持つことが明らかになった。

「西行物語」諸本がまとめて位置づけられていることは、タイトルが違い、内容に異同があるテキスト群である「西行物語」諸本が同一ジャンルであることを意味している。

「西行物語」諸本を比較してみると、名詞率よりもMVRの差が大きいように見える。諸本を比較すると、篠山本・センチュリー文化本といった絵巻・絵草紙はMVRが低い傾向がある。本の系統を合わせて考えると、島原本・静嘉堂文庫本といった略本系は本による差が大きく、広本系・中間本系・松平本系はMVRが中間にあり、宝永本・篠山本・センチュリー文化本の采女本系3本は共通してMVRが低い様子が見てとれる。「西行物語」諸本はテキストによって描写の方向性が異なり、采女本系のように本の系統との関連が認められる、類似しているテキストがあることが明らかになった。

『日本語歴史コーパス 鎌倉時代編』所収10作品を用いて同時代の資料と比較すると、『海道記』が「西行物語」諸本と比較的似ている⁹⁾と言えるだろう。前節で指摘したとおり、『海道記』は漢語が比較的多く用いられているが、「西行物語」諸本は和語が90%を占める和文体である。語種比率に差があるにも関わらず、共に歌枕を探訪する紀行文である『海道記』と「西行物語」諸本の文体が比較的似ているように見えることは特筆に値する。まずは歌と紀行文との量的バランスを確認する必要がある¹⁰⁾。そして紀行文の文体の中でもMVR、つまり動詞と形容詞・形容動詞とのバランスについて、今後さらに検討していく必要があるだろう。『海道記』『東関紀行』『十六夜日記』は中世三大紀行文とされるが、漢文要素が強いとされる『海道記』が「要約的」な文章の特徴を見せている一方で、和文要素が強いとされる『十六夜日記』は「ありさま描写的」な特徴を見せており、『東関紀行』は両者の中間に位置する。名詞率のみを見ると『十六夜日記』と「西行物語」諸本は同程度である。「西行物語」諸本の特徴はMVRの低さにあり、動詞を多用する文体であると言える。

ここから、「西行物語」諸本はテキストによって描写の方向性が異なるが、同一ジャンルの文章ではあることが明らかになった。

(4) 高頻度語

「西行物語」諸本における使用頻度の高い自立語10語を見る。表6に高頻度語を示す。諸本について、上位から見出し語・用例数を示している。見出し語は古語形である語形代表形を示した。いずれのテキストでも名詞「人」「事」「心」、動詞「す」「有り」「見る」「思ふ」といった基本的な語が上位にある。その一方で、文明本・宝永本では「申す」が上位に入っているの

表6 「西行物語」諸本の高頻度語

文明本		島原本		静嘉堂文庫本		岩国本	
思ふ	112	思ふ	91	有り	45	有り	137
す	110	有り	89	思ふ	40	思ふ	111
心	109	事	85	事	33	す	94
見る	104	す	84	す	31	見る	93
人	99	人	65	是	29	事	83
有り	86	心	55	言ふ	27	人	81
事	82	見る	52	心	21	心	72
月	78	言ふ	51	無し	21	此れ	71
哀れ	65	無し	51	人	21	絵	59
申す	65	月	46	見る	19	然々	56

松平本		宝永本		篠山本		センチュリー文化本	
有り	216	思ふ	67	す	49	す	76
事	212	す	59	思ふ	45	思ふ	69
す	172	事	56	事	43	事	67
言ふ	153	人	47	侍り	39	見る	53
思ふ	151	見る	46	見る	38	人	48
人	120	侍り	42	人	34	心	44
心	109	申す	36	申す	26	有り	43
給ふ	102	有り	34	有り	25	侍り	43
見る	99	心	33	心	24	無し	38
無し	98	無し	33	無し	24	言ふ	37
						申す	37

に対して島原本・松平本では「言ふ」が入っており、センチュリー文化本では「言ふ」と「申す」の双方が上位に入っている点、松平本で「給ふ」、宝永本で「侍り」が入っている点については、待遇表現のあり方が異なる可能性を示唆している。岩国本では「此れ」「絵」「然々」といった他のテキストにない語が上位に入っているが、これは「しかじかのゑこれあり」という絵の説明が繰り返し現れることによる。絵詞という資料性が高頻度語に現れたものである。

文明本と島原本で高頻度語に「月」が挙がっている。実は他のテキストでも「月」は上位20位程度に入る語である。最も有名な西行歌である「闇晴れて心の空に澄む月は西の山辺や近くなるらん」(新古今和歌集 1978番歌¹¹⁾)にも「心」と「月」が用いられている。「西行物語」において「心」と「月」はキーワードと言えるだろう。しかし、「西行物語」の諸本全てで「月」がキーワードとなるわけではない。篠山本、センチュリー文化本といった絵巻・絵草紙では「月」は上位に現れず、月よりも上位に「花」が現れる。テキストと絵との関連について、今後検討する必要があるだろう。

次に、同時代の資料として『日本語歴史コーパス 鎌倉時代編 I 説話・随筆』『日本語歴史コー

パス 鎌倉時代編Ⅱ日記・紀行』所収10作品と比較してみよう。すると、様相を異にしていることが明らかになる。『日本語歴史コーパス 鎌倉時代編Ⅰ説話・随筆』の使用頻度の高い自立語10語は「す、有り、此れ、斯く、下、我、上る、他、其れ、持つ」である。ここには資料規模が非常に大きい『今昔物語集』を含まれるため、高頻度語に資料規模の影響が強く現れている。『日本語歴史コーパス 鎌倉時代編Ⅱ日記・紀行』の使用頻度の高い自立語10語は「返る、行く、事、言ふ、会ふ、有り、置く、見る、内、立つ」である。上位の動詞には紀行文らしさが現れている。これらと比べると、「西行物語」諸本で基礎的な名詞・動詞と捉えていた「人」「心」「見る」「思ふ」は特徴的な語と考えることができるだろう。

(5) 共通度

『西行物語』諸本の語彙はどれだけ似ているのかを見る指標として、2作品間の共通度を用いる。この共通度は、作品A・Bにおけるある単語の使用率の低い方を合計した値である。『古典対照語い表』では、万葉集・竹取物語・伊勢物語・古今和歌集・土佐日記・後撰和歌集・蜻蛉日記・枕草子・源氏物語・紫式部日記・更級日記・大鏡・方丈記・徒然草の14作品の自立語を対象として共通度を算出している。例えば、古今和歌集と後撰和歌集という時代の近い勅撰和歌集どうしでは共通度が0.799と高く、同じ作者の源氏物語と紫式部日記では0.567と中程度、奈良時代の和歌集である万葉集と平安時代後期の歴史物語である大鏡では0.251と低くなっている。

富士池（2021）では「西行物語」諸本のうち5本を対象に調査を行い、共通度は概ね0.65前後と比較的高く、「西行物語」諸本の語彙は共通度が高いことを指摘した。

表7に「西行物語」諸本の共通度を示す。共通度が最も高いのは宝永本とセンチュリー文化本の0.849、最も低いのは静嘉堂文庫本と篠山本の0.507となった。宝永本とセンチュリー文化本のほか、宝永本と篠山本、篠山本とセンチュリー文化本も共通度が高い。これらの3本は共に采女本系である。本の系統と語彙との関連があることが示唆される。共通度の平均は0.644

表7 「西行物語」諸本の共通度

	文明本	島原本	静嘉堂 文庫本	岩国本	松平本	宝永本	篠山本	センチュリー 文化本
文明本		0.680	0.568	0.728	0.651	0.682	0.629	0.697
島原本	0.680		0.682	0.670	0.650	0.634	0.591	0.647
静嘉堂文庫本	0.568	0.682		0.571	0.562	0.556	0.507	0.575
岩国本	0.728	0.670	0.571		0.637	0.684	0.626	0.684
松平本	0.651	0.650	0.562	0.637		0.579	0.528	0.598
宝永本	0.682	0.634	0.556	0.684	0.579		0.802	0.849
篠山本	0.629	0.591	0.507	0.626	0.528	0.802		0.768
センチュリー文化本	0.697	0.647	0.575	0.684	0.598	0.849	0.768	

となる。本の種類を増やしても「西行物語」諸本の語彙は共通度が高いことがわかる。

おわりに

本稿では、いったんの完成を見た5系統8種類の諸本を対象とした『西行物語コーパス』について紹介した。この『西行物語コーパス』から得られた語彙表のデータを用いて、語彙量とTTR、語種比率、品詞比率（名詞率とMVR）、高頻度語、共通度の5観点から、「西行物語」諸本の比較、『日本語歴史コーパス 鎌倉時代編』所収の同時代資料10作品との比較を行った。その結果、「西行物語」諸本については、指標により類似・相違のパターンは異なるものの、宝永本・篠山本・センチュリー文化本の采女本系3本の類似が数値として明らかになった。また、同時代資料を比較すると、「西行物語」諸本は類似したテキスト群であることが改めて明らかになった。

『西行物語コーパス』によって、同一指標によるテキストの比較が可能になり、諸本の描写の疎密や方向性の違いを示唆する結果が出た。これらは諸本の類似性を検討する上での客観的な材料とすることができる。「月」と「花」から見える絵とテキストの関連、待遇表現の現れ方等、本の種類を増やすことによって新たに見えてきた課題もある。各指標をもとに、諸本の表現を改めて比較検討することが今後の課題となる。

附記

本研究はJSPS科研費JP19K00337『西行物語』諸本の本文の横断的研究（橋本美香・山口眞琴・富士池優美・鴻野知暁）の助成を受けたものです。

注

- 1) Japan Knowledge Lib『世界大百科事典』「西行物語」より引用。
- 2) 『西行物語コーパス』の当初の構想では章段の見出しや地の文・和歌といった本文種別の情報を付与する予定であった。しかし現段階では章段の見出しは諸本の一部について別表で管理されておりコーパスと連動しておらず、本文種別は付与されていない。
- 3) UniDicのバージョンは不明。
- 4) 富士池（2021）より再掲。
- 5) 2023年1月2日時点のデータを使用した。
- 6) 日本語歴史コーパス統合語彙表によると、和語・漢語・外来語・混種語・固有名のほか、不明と空白（データなし）がある。『西行物語コーパス』でも不明としているものがあるが、ここでは除外した。
- 7) 日本語歴史コーパス語彙統計バージョン2022.03のうち、統合語彙表（奈良時代編から江戸時代編までの全ての語彙をまとめた短単位語彙表）を利用し集計した。
- 8) 富士池（2021）より再掲。

- 9) 名詞率とMVRを使ったこの手法は、対象とする資料の中で相対的な位置づけを見るものである。富士池（2021）では同じ手法で集計・作図したところ、異なる結果が出ている。これは『日本語歴史コーパス 鎌倉時代編』のデータバージョンが異なることに起因すると考えられる。
- 10) 『西行物語コーパス』では『日本語歴史コーパス』と同様に、歌かどうかといった情報を付与する予定であったが、現時点では実装していないため、比較ができない。
- 11) Japan Knowledge Lib『新古今和歌集』新編日本古典文学全集43, 573ページ

参考文献

- 青木博史・岡崎友子・小木曾智信（2022）『コーパスによる日本語史研究 中古・中世編』ひつじ書房
- 樺島忠夫（1955）「類別した品詞の比率に見られる規則性」『国語国文』24（6），京都大学文学部国語学国文学研究室，中央図書出版社，385-387
- 樺島忠夫・寿岳章子（1965）『文体の科学』綜芸舎
- 樺島忠夫（1979）『日本語のスタイルブック』大修館書店
- 蔡佩青（2012）「『西行物語』の方法 一東海道を歩む西行一」『国際日本文学研究会会議録』35, 49-66
- 蔡佩青（2021）「近世における「西行物語」の継承と展開 一二種の『西行法師一代記』一」『西行学』第12号，西行学編集委員会，西行学会，128-143
- 橋本美香（2021）「『西行物語』の選歌意識—吉野の桜をめぐって—」『西行学』第12号，西行学編集委員会，西行学会，62-73
- 富士池優美・鴻野知暁（2018）「『西行物語』の語彙—コーパスを用いた予備的分析」『西行学』第9号，西行学編集委員会，西行学会，83-92
- 富士池優美（2021）「『西行物語コーパス』から見る語彙の量的構造」『西行学』第12号，西行学編集委員会，西行学会，119-127
- 宮島達夫（編）（1971）『古典対照語い表』笠間索引叢刊4（笠間書院）
- Web茶まめ <https://chamame.ninjal.ac.jp/>（2023年1月6日最終閲覧）
- Japan Knowledge Lib『新古今和歌集』新編日本古典文学全集43 <https://japanknowledge.com/library/>（2023年1月11日最終閲覧）
- Japan Knowledge Lib『世界大百科事典』「西行物語」 <https://japanknowledge.com/library/>（2023年1月6日最終閲覧）
- 日本語歴史コーパス語彙統計 バージョン2022.03 <https://clrd.ninjal.ac.jp/chj/chj-wc.html>（2023年1月6日最終閲覧）

（ふじいけ ゆみ）

Quantitative Structure of the Lexicon of “Saigyō Monogatari”: Based on the “Saigyō Monogatari Corpus”

Yumi FUJIIKE

Abstract

Based on the “Saigyō Monogatari Corpus”, which includes 8 different “Saigyō Monogatari” books in 5 lines, we compared the “Saigyō Monogatari” books in terms of lexical volume and TTR, word variety ratio, part-of-speech constitution ratio (noun rate and MVR), high frequency words, and commonality, and compared them with the contemporary materials in the “Japanese Historical Corpus: Kamakura-period Edition. As a result, I clarified the degree of similarity between the three “Unemebon” line books. Through the comparison of contemporaneous materials, I also revealed that the “Saigyō Monogatari” texts are similar.

Keywords: Saigyō Monogatari, categorization of various books, corpus, similarity, quantitative structure of vocabulary