

文体から見た『今昔物語集』の語彙

—『日本語歴史コーパス 平安時代編』と比較して—

田中 牧郎 (国立国語研究所言語資源研究系)[†]

The Vocabulary of *Tales of Times Now Past (Konjaku Monogatarishū)*:

Comparison with *the Heian Period Series of the Corpus of Historical Japanese*

TANAKA Makiro (National Institute for Japanese Language and Linguistics)

1. はじめに

『日本語歴史コーパス』¹の設計と構築においては、江戸時代までの口語性の強い資料群を優先してコーパス化のための基礎的研究を行っている。一方、文語性の強い資料も、日本語史研究においては重要であり、そのコーパス化も望まれる。現在、国立国語研究所コーパス開発センターで取り組んでいる日本語史資料のコーパス化は、平安時代の和文資料(2013年度完成公開予定) 室町時代の狂言資料、平安鎌倉時代の和漢混淆文資料、江戸時代の洒落本資料の順で作業に着手している。

そのうち、¹、²、³が口語性の強い資料群、⁴が文語性の強い資料群である。については、『今昔物語集』『宇治拾遺物語』『方丈記』『徒然草』などのコーパス化を試行しているが、その中で量的にも大部で、コーパス化にあたって技術的な問題も多い、『今昔物語集』の研究に力を入れている。本稿は、試作中の『今昔物語集』のコーパスのデータの一部を用いて、その語彙の特徴を、平安時代の和文資料を対象としている『日本語歴史コーパス 平安時代編』(先行公開版)の語彙と比較しながら、文体的な観点で述べていきたい。

2. 『今昔物語集』のコーパス化

コーパス化を行う『今昔物語集』(以下、『今昔』と略称)の本文は、小学館の「新編日本古典文学全集」の『今昔物語集 1~4』(馬淵和夫・国東文麿・稲垣泰一校注)により、コーパス構築のために小学館から国立国語研究所に提供された電子テキストを利用している。新編全集の『今昔』は、巻1~10の天竺震旦部は収録しておらず、巻11~31の本朝部のみを収録しており、コーパス化の対象もこの範囲になる。新編全集の底本は、巻12・17・27・29の4巻は『今昔』の最古の写本である鈴鹿本(現在は、京都大学図書館蔵、鎌倉時代書写) 巻11・13・14・15・16・19・20・22・24は実践女子大学本、巻23・25・26・28・30・31は東京大学国語研究室本である。コーパスの試作は、鈴鹿本現存巻である巻12から着手しており、本稿でも巻12のデータを用いる。

『今昔』に対する形態素解析は、まず「中古和文 UniDic」を用いて自動形態素解析を行

[†] mtanaka@ninjal.ac.jp

¹ http://www.ninjal.ac.jp/corpus_center/chj/

い、未知語となった語を辞書登録して、『今昔』を解析するための UniDic を整備しているところである。第一段階の作業として『今昔』巻 12 に対して形態素解析を実施し、その結果を目視で確認して、誤解析の修正と揺れの統一を手作業で行った。その作業での単語の認定基準は、小木曾・小椋・須永(2012)が、中古和文を対象として定めた、短単位規程に従うが、『今昔』に適用するにあたって一部変更したところがある。

3. 『今昔』巻 12 の性格

『今昔』は、和漢混淆文の作品であるが、全 31 巻(うち 3 巻は欠巻のため、実際は 28 巻)のうち、はじめの方の巻は漢文訓読体としての性格が強く、後の巻に進むにつれて漢文訓読体としての性格は弱まり和文体としての性格が強まっていくという性質を持っている。巻 12 はその半ばよりも少し前に位置付いており、漢文訓読体がまだかなり強いが、説話によっては和文体の要素を含んでいるものも交じっている。巻 12 には、塔の建立、法会の起源、諸仏の靈驗、法華經の靈驗などの説話 40 話が集められている。『日本靈異記』を依拠資料とするものが 16 話あるほか、『法華驗記』、『三宝絵』を依拠資料とするものがそれぞれ 10 話、6 話あり、これらは漢文系の資料で、『今昔』の説話も漢文訓読体である。一方、和文体の『古本説話集』または『宇治拾遺物語』と同文的であり、それらの共通母胎である散逸した『宇治大納言物語』を依拠資料とする説話が 3 話あり、『今昔』の説話もやや和文に近づいた文体になるが、『今昔』の後半の巻に多い和文的な説話に比べるとかなり硬い。このほか、依拠資料が未詳のものが 5 話ある。以上を総合すると、部分的に和文体に少し近い軟らかい文体も交じっているものの、巻 12 全体では漢文訓読体としての性格を持つ硬い文体であると言える。

『今昔』巻 12 の語彙は、記号類と付属語を除くと、短単位で集計して、延べ語数 17,685 語、異なり語数 2,578 語である。比較対象に用いる、『日本語歴史コーパス 平安時代編』は、延べ語数 378,106 語、異なり語数 11,139 語である。なお、同語か異語かの判別は、『中納言』で取得できる UniDic による付加情報のうち「語彙素読み」「語彙素」「語彙素細分類」「品詞」「活用型」の五つのうちいずれかが違っていれば、異なる語と認定する基準を立てて行った。

4. 『今昔』巻 12 の語彙と平安和文の語彙との比較—品詞と語種—

4.1 品詞

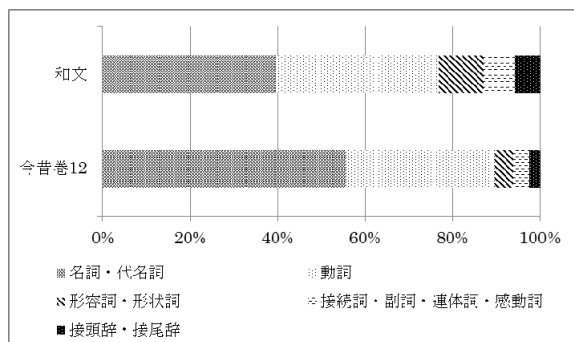


図1 今昔巻12 と和文の品詞構成(延べ語数)

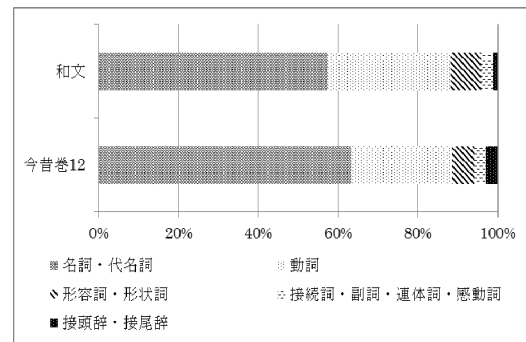


図2 今昔巻12 と和文の品詞構成(異なり語数)

まず品詞構成を比較しよう。図1は延べ語数、図2は異なり語数の比較である。品詞の分類枠は、UniDicの大分類をもとに、「名詞・代名詞」「動詞」「形容詞・形状詞」「接続詞・副詞・連体詞・感動詞」「接頭辞・接尾辞」の5つにまとめた。

延べ語数(図1)では、『今昔』巻12は、和文に比較して、名詞・代名詞の比率がかなり高くなっており、形容詞・形状詞の比率は大幅に低くなっている。これは、一般に、漢文訓読文は主語や目的語など名詞に相当する語句を表示しやすく、和文は人物の状態や感情など形容詞・形状詞に相当する語句を表示しやすいという、それぞれの文体で書かれる文章の性質によるものだと考えられる。また、全体に占める分量は多くないが、図1では、接続詞・副詞・連体詞・感動詞、および接頭辞・接尾辞の比率も、和文で高く『今昔』巻12で低くなっている。これらの分類には、多様なものを含めてしまっているため、さらに細分類して、このような差が出る理由について、今後、よく分析していく必要がある。

そして、異なり語数(図2)では、『今昔』巻12と和文との違いは、延べ語数の場合ほどには大きくない。それでも、『今昔』巻12は和文に比べて、名詞・代名詞の比率が高く、動詞、形容詞・形状詞の比率が低い。延べ語数では差がなかった動詞にも、差が見られるようになってきている。そして、接頭辞・接尾辞の比率は、延べ語数の場合と反対に、『今昔』巻12の方で高くなっている。このような異なり語数のデータに見られる、品詞構成上の特徴が何を意味するのかについても、各品詞の中に含まれる語彙の内訳を見て、よく分析していく必要があるだろう。

4.2 語種

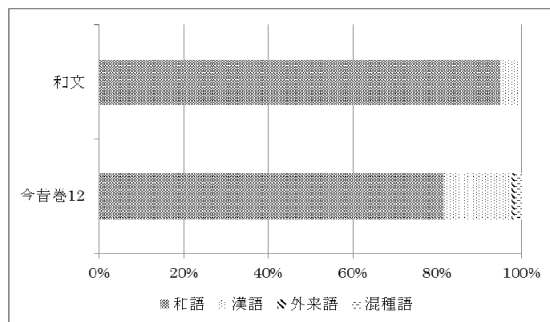


図3 今昔巻12と和文の語種構成(延べ語数)

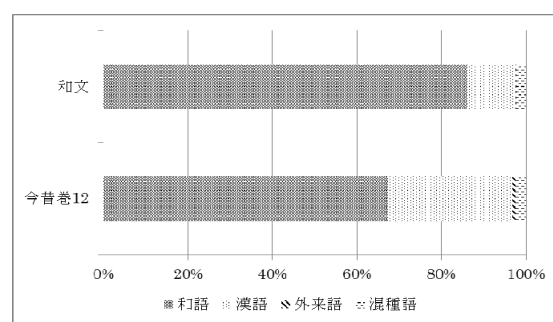


図4 今昔巻12と和文の語種構成(異なり語数)

次に語種分布を比較しよう。延べ語数(図3)でも異なり語数(図4)でも、『今昔』巻12は和文に比べて、漢語の比率が非常に高くなっており、その分和語の比率が大幅に低下していることが明らかである。『今昔』巻12の漢語の比率が、延べ語数よりも異なり語数の方でより高いのは、漢語は、繰り返し用いられる基本的な語には少なく、何度も用いられない周辺的な語には多いことを意味していよう。漢文訓読文としての性格が強い『今昔』巻12に漢語が多いことは当然のこととも言えるが、見方を変えれば、『今昔』巻12におけるその比率は、延べ語数で16%程度、異なり語数でも30%に達しておらず、語彙の大勢は和語が占めていることも確認できる。

前田(1984)は、漢文訓読資料や和漢混淆文資料についての諸調査をもとに、語種比率を表にまとめているが、そこで示された漢語比率と今回の『今昔』巻12の調査結果とを一つの表にして示すと表1のようになる。1が漢文訓読資料、2～7は和漢混交文資料である。

表1 平安時代から室町時代の漢文訓読文・和漢混淆文の語種比率

	文献資料	成立期	調査者	漢語比率(延べ語数)	漢語比率(異なり語数)
1	興福寺本大慈恩寺三蔵法師伝	11世紀末	築島裕	43.4%	85.8%(混種語含む)
2	今昔物語集全巻	12世紀初	有賀嘉寿子	—	41.9%(混種語含む)
3	今昔物語集巻12	12世紀初	田中牧郎	16.2%	29.4%
4	保元物語	13世紀初	西田直敏	—	27.8%
5	平治物語	13世紀後	西田直敏	—	34.5%
6	平家物語	13世紀後	白井清子	23.1%	49.4%
7	史記抄	15世紀初	柳田征司	—	58.5%

表1にあがる調査は、混種語を漢語に含めるか否かや、一語と認定する単位などが、それぞれの基準に基づいているので、そのままでは比較できない。表1の2番にあがる『今昔』の全巻を調査した有賀(1982)のデータと、本稿のデータの差の要因も、全巻と巻12の違いによる面と、調査単位の違いによる面の両方があるだろう。このように表1の諸データは単純に比較できるものではないが、どの文献資料においても、平安和文に比較して漢語の比率は非常に高い。今後は、表1にあるような代表的な漢文訓読資料・和漢混淆文資料に対して、UniDicの斉一な単位認定基準にしたがった語種調査を行っていき、この系統の文体での漢語比率の変化とその背景などを跡づけていくことが望まれよう。

5. 『今昔』巻12と和文との共通語彙

5.1 頻度による語彙のレベル分け

4節では、品詞や語種の観点で語彙の全体的な比較をしたが、本節と次節では、個別の語について考察していこう。まず、『今昔』巻12と和文とで共通する語彙を見ていこう。ここでは、どちらにもよく用いられる共通の高頻度語彙を抽出する。

まず、『今昔』巻12の語彙を度数順に並べ、度数10以上の高頻度語彙(338語)、度数5～9の中頻度語彙(330語)、度数4以下の低頻度語彙(1909語)の三つのレベルに分けた。『今昔』巻12において、延べ語数17,685の中で、高頻度語彙の累積度数(12,342)が占める比率(カバー率)はちょうど70%となり、同じく高頻度語彙と中頻度語彙を合わせたもの(14,446)のカバー率は82%となる。この70%、82%という数字を基準にして、和文の語彙についても、同じように、高頻度語彙、中頻度語彙、低頻度語彙の三つのレベルに分けた。その概要をまとめたのが、表2である。

表2 頻度による語彙のレベル分け

	今昔巻12		和文		カバー率
	度数区間	語数	度数区間	語数	
高頻度語彙	10～446	338	131～17,272	461	～70%
中頻度語彙	5～9	330	52～130	562	70～82%
低頻度語彙	1～4	1,909	1～51	10,116	82～100%
全体	1～446	2,578	1～17,272	11,139	—

5.2 共通の高頻度語彙

表2の「高頻度語彙」に配属された、『今昔』巻12の338語と、平安和文の461語を突き合わせると、共通するものは168語となる。この168語は、和文にも漢文訓読文にもよく使われる、平安時代の基本語彙であると見てよいものと思われる。先に分類した五つの品詞に分けて、次に掲げる。

● 名詞・代名詞

尼、哀れ、家、如何、命、今、院〔名〕、院〔接尾〕、上、内、女、彼、方、形、守、川、木、君〔名〕、君〔代名〕、国、車、是、声、心、事、此れ、先、里、十、其、空、其れ、為、使い、常、罪、手、時、所、年、年頃、名、中、何、涙、日、後、母、日、一、人、一人、他、程、仏、前、真(マコト)、身、水、自ら、道、皆、昔、元、者、物、山、故、夢、世、夜、様(ヨウ)、由、童、我

● 動詞

会う、合う、明ける、有る、言う、行く、出でる、居る、入る、入れる、失せる、得る、置く、覚える、思う、おわします、返る、掛ける、語る、聞く、聞こえる、来る、籠もる、然り、従う、知る、過ぎる、捨てる、住む、為る、絶える、立つ、奉る、立てる、給う〔四段〕、給う〔下二段〕、遣わす、付く、作る、問う、取る、泣く、嘆く、成す、成る、宣う、上る、乗る、始める、引く、臥す、経る、参る、見える、見る、召す、申す、詣でる、持つ、止む、遣る、許す、読む、寄る、渡る

● 形容詞・形状詞

多い、恐ろしい、同じい、限り無い、高い、無い、長い、久しい、深い、やんごとない

● 接続詞・副詞・連体詞・感動詞

斯く、必ず、更に、少し、唯〔副詞〕、唯〔形状詞〕、猶、先ず、又〔接続詞〕、又〔副詞〕、良く

● 接頭辞・接尾辞

御、日(カ)、大(ダイ)、つ、殿(ドノ)、共(ドモ)、御(ミ)

『今昔』巻12の高頻度語の約半数は和文でも高頻度語彙で、基本的な語彙の半分程度は重なることが分かる。一方、『今昔』巻12の中頻度語彙330語のうち、和文でも中頻度語彙であるものは65語に止まる。中頻度語彙は、『今昔』巻12と和文とであまり重ならない。今後は、中頻度語彙、低頻度語彙も含めた、語彙の詳細な比較も必要になっていこう。

6. 『今昔』巻12に特有の語彙

6.1 『今昔』巻12の高頻度語彙で和文では使われない語彙

次に、和文と比較した際に『今昔』巻12に特有の語彙となるものを抽出し、それがどのような性質のものか考えてみよう。

まず、『今昔』巻12では高頻度語彙でありながら、和文には全く使われていない語を抽出すると、次の26語となった。

● 名詞・代名詞

安置、願主、奇異、ゲンシン(源信:人名)、金堂、聖人、織冠、書生、像、読誦、父母、法会、放生、維摩、靈驗、礼拝

● 動詞

与える、哀れむ、降りる、住する、在(マシマ)す

● 形容詞・形状詞

専(もはら)

●接統詞・副詞・連体詞・感動詞

極めて、暫く、既に

●接頭辞・接尾辞

歳(サイ)

下線を付したのは、漢語または梵語あるいは混種語で、名詞・代名詞と接頭辞・接尾辞には、和語はない。動詞、形容詞・形状詞、接統詞・副詞・連体詞・感動詞には和語が多いが、それらはすべて、山田(1935)、築島(1965)、築島(1969)などによって、漢文訓読語であると指摘されているものである。『今昔』巻12における基本的な語彙でありながら和文では全く使われていない語彙は、人名以外では、漢語、梵語、混種語、漢文訓読語のいずれかであるということになる。

6.2 『今昔』巻12の高頻度語彙で和文では特に頻度が低い語彙

『今昔』巻12の高頻度語彙のうち、和文に使われていても、その使用頻度が極めて低いものは、6.1で扱った語彙に準じる扱いをしてよい語であると思われる。表2において「低頻度語彙」とした和文の語彙は10,116語もある。量の少ない『今昔』巻12の語彙の分類に合わせて、カバー率82%以上のところに位置付く語彙がすべてここに分類されているが、量の多い和文の語彙については、これらをさらに段階に分けて、さらに低頻度の語彙だけを取り出すこともできる。そこで、カバー率97%以上という基準を立てると、度数区間が1~4の特に頻度の低い語彙6,570語を抽出することができた。

『今昔』巻12で高頻度語彙でありながら、和文では特に頻度の低い語彙となるものを取り出すと、28語になった。その28語の用例を、『中納言』を用いて、『日本語歴史コーパス平安時代編』で検索し、検索結果画面に表示される、地の文か発話部分か、和歌・詞書・手紙・序文などのその他の部分かといった、使用箇所を確認して、表3に整理した。表3の「語」の列に*を付けた語は、漢語・梵語・混種語のいずれかのもので、*の付いていないものは和語である。作品名は頭文字で示し、数字は使用件数で数字のないものは1件であることを意味している。

この表から、これらの語は、女性発話の部分に用いられないことが分かる。「性不明発話」としたところは、登場人物名が『中納言』の検索結果には表示されないものだが、該当箇所の作品本文を読んで確認をしていくと、女性であるものが3件見つかる。次の通りである。

大悲^査には、他事(ことごと)も申さじ。あが姫君、大貳の北の方ならずは、当国の受領の北の方になしたてまつらむ。三条らも、随分にさかえて返し申しは仕うまつらむ。(源氏物語・玉鬘・三条という女房の発話)

「一乗の^法ななり」など人々も笑ふ事の筋なめり。(枕草子・「御方々、君達、上人など、御前に」・女房達の発話)

昨夜『物言はむ』とて来^りしを、(落窪物語・巻一・あこぎの発話)

表3 『今昔』巻12の高頻度語で和文で特に頻度の低い語 和文における使用箇所

語	品詞	地の文	男性発話	女性発話	性不明発話	その他
希有*	名詞		1(源:僧都)			
持経*	名詞	2(源2)				
釈迦*	名詞	3(源、枕2)			1(源)	
舍利*	名詞					1(古:詞書)
修行*	名詞	2(伊、枕)	1(源:光源氏)			
大門*	名詞	1(枕)				
天皇*	名詞	4(伊、源3)				
塔*	名詞	2(源、紫)				
童子*	名詞	2(落、枕)				
女人*	名詞		2(源:律師・僧都)			
法(ホウ)*	名詞	2(源2)			1(枕)	
房(ボウ)*	名詞	1(大)				
諸々	名詞					1(古:序)
薬師*	名詞	4(源2、枕2)				
致す	動詞	1(伊)	1(源:律師)			
終わる	動詞	1(伊)	2(源2:入道・律師)			
来たる	動詞				1(落)	1(源:和歌)
講ずる*	動詞	4(源4)				
叫ぶ	動詞					1(古:詞書)
然り	動詞	1(土)			1(竹)	2(古:和歌、竹:手紙)
尊ぶ	動詞	1(源)				
説く	動詞	3(枕2、紫1)	1(源:光源氏)			
在(マシマ)す	動詞	1(大)	1(源:供人)			1(古:詞書)
速やか	形状詞				1(土)	
但し	副詞				2(竹)	
漸く	副詞	1(土)				2(古:詞書2)
会(エ)*	接尾辞		1(源:使)			1(古:詞書)
者(シャ)*	接尾辞	2(伊、枕)	1(源:光源氏)		1(源)	

第一例の『源氏物語』玉鬘の三条という女房は田舎者として造型されており、引用箇所の発話部分には「者」のほか、「大悲」「当国」「受領」「随分」など漢語が多く用いられている。「受領」以外は、和文で使われることは極めて珍しいもので、この発話が異常な言葉遣いとして描かれているところである。第二例の『枕草子』での女房達の発話は法華經の引用文である。そして、第三例の『落窪物語』のあこぎの発話は、通常の発話部分であるが、これは「来たる」という動詞ではなく、「来(く)」(動詞) + 「たり」(助動詞)と認定すべきものが誤って動詞「来たる」と認定されたものであろう。こう見てくると、通常の女性の発話で用いられた例は皆無ということになる。男性の発話者も、僧都、阿闍梨、律師といった仏教関係者が目立つ。これらのことから、この28語はかなり硬い文体的価値を持った語であると見ることができよう。実際、漢語が多く、和語のうち、「諸々」「致す」「終わる」「来たる」「然り」「尊ぶ」「説く」「在す」「速やか」「但し」「漸く」などは、6.1で見た諸語と同じく、山田(1935)、築島(1965)、築島(1969)などで漢文訓読語とされているものである。

「叫ぶ」だけは、先行研究で漢文訓読語とされていないものであるが、その和文での使

用例は次のものである。

法皇、西川におはしましたりける日、「猿、山の峽に叫ぶ」といふことを題にて歌よ
ませ給うける
わびしらに猿（ましら）な鳴きそあしひきの山のかひある今日にやはあらぬ
(古今和歌集・雑体・一〇六七・詞書)

詞書に用いられる「猿、山の峽に叫ぶ」は、新編日本古典文学全集の注によると、漢詩の題詞に「猿叫峽」とあったものであり、「叫ぶ」はその訓読と考えられる。これに相当する内容は、和歌では「鳴く」が用いられている。したがって、「叫ぶ」も、漢文訓読語であると見ることができよう。

以上のように、『今昔』巻 12 の高頻度語彙のうちで、和文の語彙と比較した際の特有語は、漢語・梵語・混種語もしくは漢文訓読語のいずれかであると見ることができる。「叫ぶ」のように従来は見逃されていた漢文訓読語を特定していくこともできる。本稿では、比較対象とした和文で特に頻度の低い語彙を度数 4 以下の語としたが、度数 5 以上の語にも、和文での用例の出方が偏っているものもある。比較対象とする和文の語彙の範囲を拡大すれば、「叫ぶ」のような漢文訓読語を、さらに数多く発見していくこともできると考えられる。

7. おわりに

本稿で報告したのは『今昔』のうち巻 12 だけのデータに基づくものであったが、『今昔』の語彙と和文の語彙を比較することで、文体的観点から、語彙を分類していくことが様々な可能になることを示した。『今昔』の他の巻のデータを整備し、『日本語歴史コーパス 平安時代編』と比較していくことで、文体から見た語彙の分析を総合的に進めていくことができるようになるだろう。

付記

本研究は、国立国語研究所共同研究プロジェクト「通時コーパスの設計」(プロジェクトリーダー：近藤泰弘) 及び、日本学術振興会科学研究費基盤研究(B)「和漢の両系統を統合する平安・鎌倉時代語コーパス構築のための語彙論的研究」(24320086、研究代表者：田中牧郎) による成果の一部です。

文献

- 有賀嘉寿子 (1982) 「今昔物語集の語彙」(『講座日本語の語彙 3 古代の語彙』明治書院)
 小木曾智信・小椋秀樹・須永哲矢 (2012) 「中古和文 UniDic 短単位規程集」(科研費報告書)
 築島裕 (1965) 『平安時代の漢文訓読語につきての研究』(東京大学出版会)
 築島裕 (1969) 『平安時代語新論』(東京大学出版会)
 前田富祺 (1984) 「語種構造の漸移相」(『日本語学』3-9、明治書院)
 山田孝雄 (1935) 『漢文の訓読によりて伝えられたる語法』(宝文館)