

源氏物語におけるカバー率とジップ則

Cover Ratio and Zipf's Law on the Words in Genji Monogatari

亀山 寛, 石川 貴恵^a

Hiroshi KAMEYAMA and Kie ISHIKAWA

（平成21年10月6日受理）

In order to search the origin of the lower cover ratio (86.0% at 5,000 words) in Japanese words of the present ages than the cover ratio of English words, we study relative frequency of words with such the attached words as an auxiliary verb and a particle by analyzing "The Key Word Index in Genji Monogatari" (edited by H.Ueda et al.) which are constructed with few ratio of Chinese words. The next conclusions were obtained.

1) In "Genji Monogatari", the high cover ratio (90.8% at 5,000 words) is obtained and is found to be almost the same with the cover ratio of English words. Although the classifying method of words are not lemmatized, the higher cover ratio is obtained.
2) In "Genji Monogatari", Zipf's law is also applicable to the words including the attached words.

1. はじめに

我々は前に英単語における使用度数の分布則であるジップの法則と比較できる資料を得るために、『現代雑誌90種の用語・用字』¹⁾と『現代雑誌の語彙調査：1994年発行70誌』²⁾（国立国語研究所）を対象として、非自立語（付属語）を含めた場合の正確な日本語使用度数を算出し、次の結果を得た^{3, 4)}。

1) 非自立語を含めた日本語の上位5,000語のカバー率は『現代雑誌90種』で87.8%、『現代雑誌70誌』で86.0%であり、従来の自立語のみの結果より、6.1%もしくは4.3%高い値が得られたが、英語でのカバー率と比較すると依然として低い値に留まっている。

2) 延べ語数のサンプル数1,000個から5,000個において、単語の使用率と順位との関係で反比例分布に近く、べき指数は『現代雑誌90種』で、-1.031から-1.065までの、『現代雑誌70誌』で-1.009から-1.056までの範囲の値をもち、共に-1に近い値であり、ジップの法則が日本語単語に対しても成立する。

1)のカバー率の低さの要因として、漢語の影響を指摘した。本研究では、付属語（非自立語）も含めた「和語」においてはどうかという観点から和語の構成比率が現代日本語に比較してはるかに大きい源氏物語を用いて同様な分析を行う。ちなみに、源氏物語では自立語のみ

^a 富士ソフトABC

の延べ語数において、漢語の割合は3.4%である⁵⁾。そして、カバー率とジップの法則を求め、日本語の比較、分析を行うことを目的としている。

源氏物語を選択した理由は、1)平安時代に書かれ、漢語の影響を多く受けていない。2)平安時代に書かれた中で一番の長編作品であり、単語数が多い。3)精密な語彙調査が行われ、その結果が出版され、利用できる。4)精密な語彙調査とそれに関する論文から付属語の使用度数を算出することができる。などである。特に平安時代の物語などの語彙索引は自立語に主眼がおかれ、付属語の使用度数を得るのが困難な場合が多いので、4)の視点は重要である。

2. 単語の使用度数表の算出

分析の対象としたのは『源氏物語語彙用例総索引』であり、『自立語篇5巻』⁶⁾と『付属語篇5巻と付属語篇別冊』⁷⁾の11巻から構成されている。同書は文書の計量分析の観点から、池田亀鑑編著『源氏物語大成』⁸⁾の索引篇を基にして作成された用例付き索引である。同書を選定した視点は付属語の計量化がなされ、記載されているからである。一方『語彙用例総索引』の性格上、自立語のレマ化(基本形にまとめる)が図られてない。この点に関しては分析の中で吟味を行う。

自立語の使用度数は『自立語篇』の逆引き表から得られる。一方付属語の数量は別冊で断り書きがあるように、自立語中の付属語も重複してカウントしてある。別冊の数表2から4までリストアップされている付属語の総数は、212,671個になり、31%程重複してカウントされていることになる。従って『付属語篇別冊』の数表2から4に掲げられている付属語データは使用出来なかった。代わりに『源氏物語語彙用例総索引』のデータベース作成に関連して研究された助動詞と助詞の計量分析のデータを用いた。このようにして求めた源氏物語の自立語と付属語データを以後『源氏物語』と略称する。

2-1 自立語の使用度数表

自立語に関しては、『自立語篇』に掲載されている逆引き表を使用する。この表では単語とその使用度数が記されているため、データが直接利用できるからである。逆引き表が211ページに渡って掲載されている。次の手順で単語の使用度数(頻度数)の表を作成した。1)スキャナー

表1 『源氏物語』における上位20位までの自立語の順位と使用度数

順位	単語	使用度数	順位	単語	使用度数
1	(副)いと	4,220	11	あら	1,339
2	人	3,363	12	給は	1,301
3	事こと	2,559	13	ほと	1,240
4	[命](動)給へ	2,449	14	[用](動)み	1,203
5	給	2,443	15	さま	1,188
6	[終]賜・給給	1,990	16	(副)かく	1,053
7	[用](補動)きこえ	1,841	17	(動)し	960
8	[体]賜・給給	1,470	18	心	901
9	事事	1,387	19	(動ラ変)ある	813
10	この	1,369	20	[用]あり	803

でページを読み取った自立語の総数は213,169個であった。『付属語篇別冊』の表1では自立語の延べ総数は213,096個であり、少数の違いを生じた。この少数の違いは今回スキャナーを使用して読み通った値の正確さを示している。本研究では自立語の総数は計算の都合上、今回算出した値を用いる。自立語の異なり語数は26,492語であった。自立語の使用度数上位20位までを表1に示す。他の調査の自立語の見出し語は終止形で求められる場合が多いが、用例索引を重視している結果、終止形でない語彙が見られるのが特徴である。

2-2 助動詞の使用度数の算出

本研究には付属語の使用度数値が必要である。最初に付属語の中の助動詞の使用度数を算出する。「源氏物語の助動詞の計量分析」⁹⁾によれば、源氏物語の助動詞は26個で総数43,617個である。同論文に掲載されている54巻毎に各助動詞の出現率 (relative frequency) 表から、巻毎の助動詞の使用度数を算出した後、全巻の助動詞を足し合わせることによって、助動詞の使用度数を算出した。なお出現率は従来使用率とも呼ばれてきたが、本論文では上記論文の引用以外は使用率 (relative frequency) という用語を用いる。このようにして算出した助動詞の全合計は43,617個になり、一致を示し、各助動詞の使用度数の算出に誤りがないと云える。以上のようにして求めた助動詞の使用度数に対して、上位10位までを表2に示す。使用度数では表1の自立語の上位と同程度の大きさである。上位10位までの助動詞は1文字と2文字で占められている。

表2 助動詞 (10位)

助動詞	使用度数
ず	5,576
む	4,761
たり	4,312
けり	3,637
なり	3,536
り	3,378
ぬ	3,219
き	2,997
べし	2,895
つ	1,485

2-3 助詞の使用度数の算出

「源氏物語の計量分析(4)」¹⁰⁾に掲載されている助詞56個の使用率を使用して、助詞の使用度数を求めた。算出した助詞の総数は119,639個であった。助動詞+助詞 = 43,617 + 119,639 = 163,256となり、別冊の数表1掲載の付属語数162,958とほぼ同じである。本論で用いた付属語数は162,256である。今回算出した助動詞の使用度数に対して、上位10位までを使用率と共に表3に示す。上位の助詞の使用度数は自立語の使用度数に比較して、はるかに大きいのが特徴であり、単語全体の使用度数表に大きな影響を与える。上位10位までの助詞は9個が1文字、2文字が1個で占められている。

表3 助詞 (上位10位)

助詞	使用度数	使用率
に	18,336	0.04876
て	16,035	0.04264
の	14,305	0.03804
も	12,098	0.03217
と	10,672	0.02838
を	9,608	0.02555
は	8,954	0.02381
ば	5,505	0.01464
など	3,847	0.01023
ど	2,433	0.00647

2-4 全語彙の使用度数表

以上のようにして求めた付属語の延べ総数N2は163,256個であった。一方自立語の延べ総数N1は213,169個であるので延べ総語彙数Nは376,425個となった。自立語と付属語(助詞と助動詞)の使用度数表を合体し、使用度数順にソートして、全単語の使用度数表を得た。この表から、カバー率とジップ則の検討を行う。付属語も含めた全単語の上位40位の使用度数と使用率を表4に示す。34位は漢字で表記された事を意味する。ここでの使用率は全語彙数Nで割った値

を示してある。上位を付属語すなわち、助詞や助動詞が占めている点は特徴的である。この点と付属語の割合が40%であることの2点が、以下の計量分析に付属語が大きな役割を果たす要因となる。

表4 『源氏物語』における付属語を含めた語彙における上位40位までの使用度数と使用率

順位 n	単語	使用 度数	使用率 p	順位 n	単語	使用 度数	使用率 p
1	に	18,336	0.0488	21	事こと	2,559	0.0068
2	て	16,035	0.0426	22	[命](動)給へ	2,449	0.0065
3	の	14,305	0.0380	23	給	2,443	0.0065
4	も	12,098	0.0322	24	ど	2,433	0.0065
5	と	10,872	0.0284	25	や	1,993	0.0053
6	を	9,808	0.0255	26	[終]賜・給給	1,900	0.0053
7	は	8,954	0.0238	27	こそ	1,880	0.0050
8	ず	5,576	0.0148	28	[用](補動)きこえ	1,841	0.0049
9	ば	5,505	0.0146	29	なむ	1,801	0.0048
10	む	4,761	0.0127	30	ぞ	1,700	0.0045
11	たり	4,312	0.0115	31	つ	1,485	0.0039
12	(副)いと	4,220	0.0112	32	[体]賜・給給	1,470	0.0039
13	など	3,847	0.0102	33	る	1,453	0.0039
14	けり	3,837	0.0097	34	事事	1,387	0.0037
15	なり	3,536	0.0094	35	この	1,360	0.0038
16	り	3,378	0.0090	36	より	1,346	0.0038
17	人	3,363	0.0089	37	あら	1,330	0.0036
18	ぬ	3,219	0.0086	38	給は	1,301	0.0035
19	き	2,997	0.0080	39	す	1,267	0.0034
20	べし	2,895	0.0077	40	ほと	1,240	0.0033

3. 『源氏物語』の単語におけるカバー率

3-1 自立語のみのカバー率

語彙の計量化は自立語のみに対象として行われることが多く、自立語のみの計量データが得やすいので、最初に『源氏物語』に関連した自立語のみのカバー率の算出を行い、問題点の整理を行う。

本研究で使用した『源氏物語』の自立語を対象とし、自立語延べ総語彙数 $N_1=213,169$ 個を利用して求めた自立語のカバー率を表示した。また、同グラフには池田亀鑑編『源氏物語大成索引』の結果を宮島等によるまとめ¹⁴⁾に従って、自立語延べ総語彙数207,808個を採取し(『源氏物語大成』と呼ぶ)、求めた自立語のカバー率も示してある。『源氏物語大成』の自立語のみのカバー率は上位5,000語で、93.7%の大きさになっている。一方本研究で分析対象とした『源氏物語』の自立語のカバー率は、『源氏物語大成』に比較してかなり小さく、上位5,000語で、83.9%である。『源氏物語』と『源氏物語大成』の違いは、レマ化が行われてあるかどうかである。レマ化が行われている『源氏物語大成』の自立語延べ総語彙数は少なくなる。そして、カバー率においては図1からわかるようにレマ化を行うと大きいカバー率が得られる。

『源氏物語』という一作品でなく、複数の作品、著者による影響の有無を調べるために次のデータの検証を行った。平安時代において、『源氏物語』も含めた多数の著者の作品の語彙の使用度数(頻度数)の調査として、大野による研究¹⁵⁾がある。自立語に限定しているが平安時代の10

作品の語彙（以後『平安時代』と略称）を収集したものであり、自立語延べ総語彙数は410,085個で、『源氏物語大成』の2倍の規模である。レマ化は行ってある。その中で上位1,321語の順位と使用度数を公にしており、1,321語の延べ語数は324,286個としている。今回カバー率を算出するために1,321語の使用度数をすべて読み取ったが、1,321語の延べ語数は323,810個であり、0.15%の違いがあった。使用度数の累積語彙数を自立語延べ総語彙数で割ることによって、自立語のカバー率を求めることができ、図1に併せて示した。また、同グラフには池田亀鑑編『源氏物語大成索引』の結果を宮島等によるまとめ¹²⁾から、自立語延べ総語彙207,808個を採取し（『源氏物語大成』と呼ぶ）、計算して求めた自立語のカバー率も示してある。

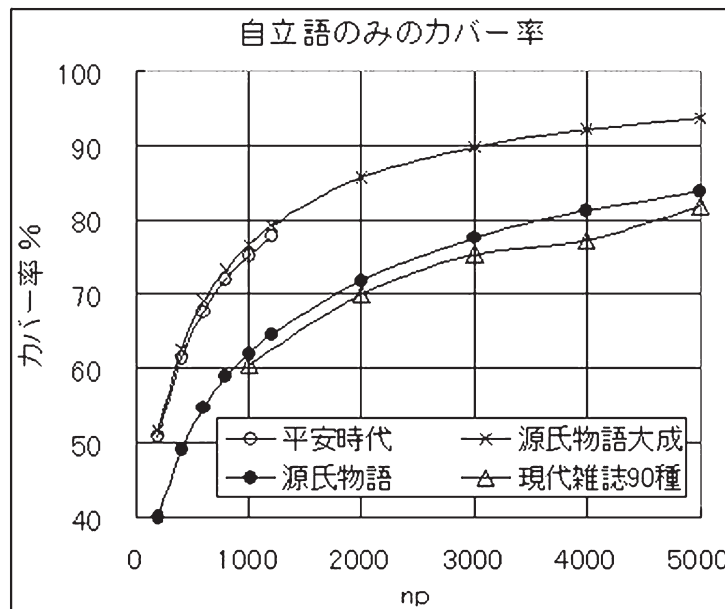


図1 『平安時代』、『源氏物語大成』、『源氏物語』、『現代雑誌90種』における自立語のみのカバー率

『平安時代』は10作品の合計であり、語彙数規模は『源氏物語大成』の2倍程であるが、カバー率の振る舞いは『源氏物語大成』とほぼ同じである。このことから、『源氏物語大成』や『源氏物語』の一著者一作品によることが、カバー率に与える影響はそれほど大きくないと考えられる。

図1には現代日本語『現代雑誌90種』の自立語のみのカバー率の文献値（後掲の表6中における日本語Aデータ）も示した。『現代雑誌90種』はレマ化が行われているので、比較はレマ化済みの『源氏物語大成』や『平安時代』と行うべきである。カバー率が、一様に10%程低く出ていることがわかる。なお、英語などのカバー率のデータは付属語に相当するof, the, andの語彙を含むので、自立語のみのカバー率は英語などのカバー率とは比較の対象とはならない。

3-2 付属語を含めたカバー率

『源氏物語』の付属語を含めた場合において、1位から上位5,000語までの累積単語数を算出することによってカバー率を求め、表5に示した。付属語を含めた場合において、外国語のカバー率の比較が可能となる。『源氏物語』におけるカバー率は上位5,000語で、90.8%であり、現代日本語のカバー率より大きく、外国語に比較して、やや低いがほぼ同程度のカバー率を示して

表5 『源氏物語』におけるカバー率

語数	1000	2000	3000	4000	5000
累積単語数	293,278	315,343	327,724	335,834	341,702
カバー率(%)	78.0	83.8	87.1	89.3	90.8

いる。このことを明確に示すために、現在までに得られている現代日本語⁴⁾や外国語¹²⁾のカバー率を表6にまとめた。表6中の『日本語A』は自立語のみの場合で文献値¹⁾であり、『現代雑誌90種』と『現代雑誌70誌』は我々の研究⁴⁾で得られた結果であり、付属語をも含む場合である。付属語を含めて算出を行った『現代雑誌90種』と『現代雑誌70誌』の場合においても、外国語に比較すると依然として低い値に留まっているが、『源氏物語』において、ほとんど、外国語の値と同レベルの値となっていることがわかる。

表6 全単語使用度数に占める上位5,000語までのカバー率の比較(日本語Aは自立語のみのカバー率)、雑誌90種と雑誌70誌は『現代雑誌90種』と『現代雑誌70誌』である

語数	英語	フランス語	スペイン語	中国語	韓国語	日本語A	雑誌90種	雑誌70誌	源氏物語
1-1000	80.5	83.5	81.0	73.0	73.9	80.5	74.8	71.5	77.9
1-2000	86.6	89.4	86.6	82.2	81.2	70.0	80.9	78.0	83.8
1-3000	90.0	92.8	89.5	86.8	85.0	75.3	84.2	81.7	87.1
1-4000	92.2	94.7	91.3	89.7	87.5	77.3	86.5	84.2	89.3
1-5000	93.5	96.0	92.5	91.7	89.3	81.7	88.2	86.0	90.8

以上の傾向をさらにはっきり示すために、図1に『源氏物語』と英語のカバー率の比較を示した。英語のカバー率のデータ¹³⁾は源データが『ソーンダイクとロージ』による¹⁴⁾のものであると、前研究⁴⁾で同定した。この図をみると上位1,000語までは英単語の場合とほぼ同様な値を持っていることがわかる。1,000語以上でやや小さい値を示している。『源氏物語』の語彙はレマ化を行ってなく、前節の自立語の場合の分析から、レマ化を行わないとカバー率の値が低くなるから、このことがかなり影響を与えていると推定される。

レマ化が図られてないことを考慮しても、『源氏物語』のカバー率は『現代雑誌70誌』より大

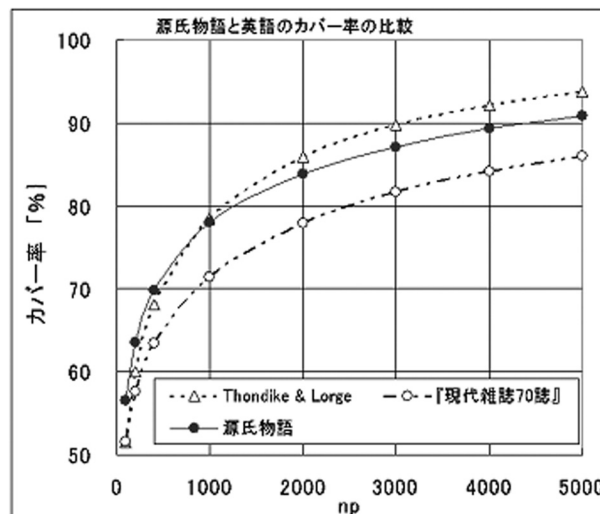


図2 『源氏物語』と英語のカバー率の比較

きいといえる。『源氏物語』は漢語の影響が少なく、これはほとんど和語からなるため、外国語と同様なカバー率を持っていたと推定できる。

4. 『源氏物語』の単語におけるジップ則

ジップ則とは単語の使用度数（頻度数）を r 、単語の使用度数順位を n とすれば、

$$r = C/n \quad C: \text{定数} \quad (1)$$

なる式が成立することである。単語の使用率(relative frequency) p ($=r/N$, N : 総延べ単語数)で考えると、

$$p = C_0/n \quad C_0: \text{定数} \quad (2)$$

となる。使用率を採用すると単語のサンプル数によらない結果が得られる。(1)式を一般化し、

$$p = C_1/n^b \quad (3)$$

としたべき則でジップ則が論じられることが多い。(3)式の両辺の対数をとると、

$$\log(p) = a \times \log(n) + \log C_1 \quad (a = -b) \quad (4)$$

となり、 $\log(p)$ は $\log(n)$ に比例し、比例係数は a ($= -b$)となる。以上の式との関連で、 $\log(p)$ - $\log(n)$ プロットで取った場合のグラフと分析を以下に示す。対数の底は10である。自立語のみの場合と付属語を含む場合に対して、延べ語数の1位から数えて1,000個、3,000個での $\log(p)$ - $\log(n)$ プロットを図3、図4に示した。これらの図からわかるように $\log(p)$ - $\log(n)$ プロットはほぼ次の直線に載っている。今

$$y = ax + b \quad (5)$$

とし、最小自乗法によって傾き a と y 切片 b を算出する。この傾き a は(3)式中の指数べき a の値と一致している。この際、エクセルに付属する分析ツールの回帰ツールを用いた。このようにして求めた回帰直線と直線式の各値を各図中に示した。図中に示した決定係数 R^2 は1に近いほど得られた回帰直線は正確であることを示すものであり、共に0.98から0.99のほぼ1に近い値を示している。いずれのグラフにおいても、上位10位程のデータを除き、直線上に並んでいる。

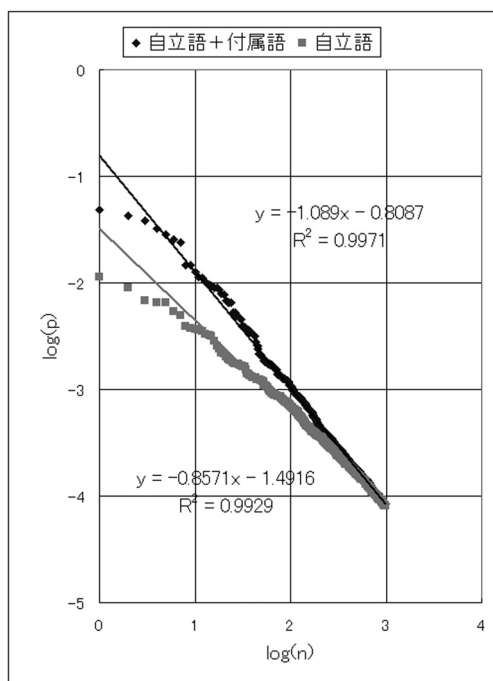


図3 延べ語数上位1,000個における $\log(p)$ - $\log(n)$

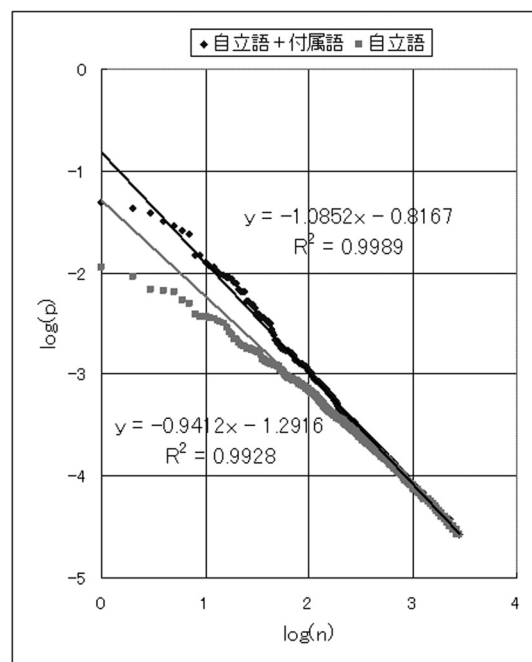


図4 延べ語数上位3,000個における $\log(p)$ - $\log(n)$

表8 延べ語数の各 $\log(p) - \log(n)$ 直線における指数aの値

個数	自立語+付属語	自立語
1000	-1.089(0.997)	-0.857(0.993)
2000	-1.082(0.998)	-0.907(0.994)
3000	-1.085(0.999)	-0.941(0.993)
4000	-1.101(0.999)	-0.977(0.991)
5000	-1.119(0.998)	-1.134(0.980)

表9 延べ語数の各 $\log(p) - \log(n)$ 直線におけるy切片bの値

個数	自立語+付属語	自立語
1000	-0.809	-1.492
2000	-0.828	-1.378
3000	-0.817	-1.292
4000	-0.775	-1.198
5000	-0.724	-0.734

そして、「自立語+付属語」の場合の方が、自立語のみの場合よりも指数aが-1に近く、ジップ則によく従っていることがわかる。付属語を含めた分析の重要性を示している。

このような手続きで求めた指数aとy切片bとの値を各個数に対してまとめて、表にしたのが表8と表9である。表8と表9における値の大きさと傾向は、『現代雑誌70誌』と同様である。一方、表8からわかるように、「自立語+付属語」の場合の指数aの値は3,000個までの個数に対して、-1に近く、ジップ則に従っているが、4,000個以上では-1から少しずれている。『現代雑誌70誌』の場合でも、個数が大規模になると-1からのズレが見られる。『源氏物語』において、4,000個のレベルで大規模化による現象が生じている可能性もある。すでに前研究で、付属語を含めれば、現代日本語の『現代雑誌90種』と『現代雑誌70誌』で共にジップの法則が成立することを明らかにした。今回、レマ化が図られてない『源氏物語』においても、ジップの法則が成立していることがわかった。レマ化を図れば、べき指数が多少大きくなることが予想される。

5 討論

本研究において、付属語の数量的データが得られるという観点から、『源氏物語語彙用例総索引』を用いて、語彙の数量的解析を行った。『語彙用例総索引』の性格上、自立語のレマ化が図られてない。付属語の『源氏物語』(逆引き表)に基づいて、付属語(非自立語)も含めた語彙の使用度数を算出し、次の結果を得た。

1) 単語のカバー率は現代日本語のカバー率の低さ(86.0%)と比較して、より英語などの外国語のカバー率に近い振る舞いを示した。上位5,000語において90.8%であり、英語のカバー率に近い値を示した。

2) 『源氏物語』において、付属語を含めればジップの法則は成立する。

1)に関連して、自立語のみの分析結果(図1参照)において、レマ化を図るとカバー率は更に

表7 『源氏物語』における語彙に対する語種割合(自立語データは『古典対照語い表』による)

	自立語		自立語+付属語	
	異なり		異なり	
	度数	%	度数	%
和語	9,953	87.1	10,035	87.2
漢語	1,008	8.8	1,008	8.8
混種	462	4.0	462	4.0
	延べ		延べ	
	度数	%	度数	%
	和語	198,684	95.4	361,642
漢語	7,116	3.4	7,116	1.9
混種	2,008	1.0	2,008	0.5

数%は高い値になることを示した。漢語の影響が少ない状況ではカバー率は高いといえる。当時の漢語の割合がどの程度であったかは興味あることなので、付属語を含めた延べ語彙数における漢語の割合を算出してみよう。自立語のみの場合は『古典対照語い表』⁵⁾で求められている。助詞と助動詞からなる付属語はすべて和語であるので、付属語の延べ総語数162,958等を用いると付属語を含む場合の語種割合も算出でき、表7にまとめて示した。付属語を含めると、延べ語彙で和語割合が97.5%とより大きな割合になることがわかる。このことから、『源氏物語』における単語の数量的分析は、100%和語でなくて、97.5%和語からなる数量的分析を示しているといえる。そして漢語の影響は混種語も含めて、延べ総語彙数で2.5%ある。付属語を含めると、延べ語彙で和語割合が97.5%とより大きな割合になることがわかる。

2)に関連して、すでに前研究で、付属語を含めれば、現代日本語の『現代雑誌90種』と『現代雑誌70誌』で共にジップの法則が成立することを明らかにした。今回、延べ語彙で97.5%と和語割合が大きい『源氏物語』においても、レマ化が図られなくても、付属語を含めれば、ジップの法則が成立していることがわかった。ジップの法則は、語種割合に関わらず、成立する強固な(robust)法則であるといえる。

6 まとめ

英語などの外国語に比較して現代日本語のカバー率の低さ(5,000語で86.0%)の要因を探るために、現代日本語に比べて漢語割合が低く、和語割合が高い『源氏物語』を対象として、付属語(非自立語)も含めた語彙の使用度数を算出した。『源氏物語』はレマ化が図られてないにもかかわらず、次の結果を得た。

- 1) 『源氏物語』において、単語のカバー率は上位5,000語において90.8%であり、英語のカバー率に近い値を示した。
- 2) 『源氏物語』において、付属語を含めればジップの法則は成立する。

文献

- 1) 国立国語研究所：国立国語研究所報告 121 『現代雑誌の語彙調査：1994年発行70誌』
- 2) 国立国語研究所：『現代雑誌九十種の用語・用字1巻』, p. 26, 314, 秀英出版, 1962年
- 3) 亀山寛：日本語単語頻度数におけるジップ則, 計量国語学26巻4号, pp. 123-138, 2008年
- 4) 亀山寛, 長谷川哲子：日本語と英語の語彙における使用率とカバー率, 計量国語学26巻6号, pp. 196-210, 2008年
- 5) 宮島達夫編：『古典対照語い表』, p. 336, 笠間書院, 1971年
- 6) 上田英代, 村上征勝, 今西祐一郎, 樺島忠夫, 上田裕一：『源氏物語語彙用例総索引—自立語編—全5巻』, 勉誠社, 1994年
- 7) 上田英代, 村上征勝, 今西祐一郎, 樺島忠夫, 藤田真理, 上田裕一：『源氏物語語彙用例総索引—付属語編—全5巻, 付属語篇 別冊』, 勉誠社, 1997年
- 8) 池田亀鑑編著：『源氏物語大成—全8巻』中央公論社, 1953年
- 9) 村上征勝, 今西祐一郎：源氏物語の助動詞の計量分析, 情報処理学会論文誌40(3), pp. 774-782, 1999年

- 10) 上田英代, 村上征勝, 今西祐一郎, 樺島忠夫, 藤田真理:源氏物語の計量分析(4), 日本行動計量学会, 1996年
- 11) 大野晋:平安時代和文脈系文学の基本語彙に関する二三の問題, 『文法と語彙』, 岩波書店, pp. 255-280, 1987年
- 12) 宮島達夫, 中野洋, 鈴木泰, 石井久雄編:『フロッピー版古典対照語い表および使用法』, 笠間書院, 1989年
- 13) 玉村文郎:語彙教育の基本, 『日本語教育ハンドブック』, 大修館書店, pp. 342-355, 1990年
- 14) E. L. Thorndike and I. Lorge; *The Teacher's Word Book of 30,000 Words*, Bureau of publications, 1944