

中古散文22作品の分類についての一試み

前川 武*

An Attempt to Classify 22 Old and Medieval Prose Works.

Takeshi Maekawa *

キーワード

分類、語彙、形容詞、中古散文作品

I はじめに

筆者は、以前に、中古散文資料 22 作品¹⁾ 相互の類似度を測定する試みを行った。²⁾ その際用いた手法は宮島達夫氏の考案によるもので、見出し語の使用率から 2 つの作品間の類似度を測定するという方式である³⁾。

宮島氏は、その測定法を用いて、万葉集から徒然草に至る 12 の古典作品(歌集 3、散文 9) 相互の類似度を求め、その類似度を距離に見立ててデンドログラムのような図を作成し、大きく歌集のグループと散文のグループに分かれることを示した⁴⁾。

また、進藤義治氏は、王朝和文文芸 17 作品における五千自立語毎の形容詞類語彙の使用回数を比較し、形容詞類語彙の多様傾向が年代の進行につれて増すことを論じ、その過程で 17 作品をいくつかのグループに分類している⁵⁾。

さらに、田中章夫氏は、「作品の語彙の偏り」という概念を用いて古典 14 作品を比較し、各作品の位置づけを明らかにしようとした⁶⁾。

これらの手法はいずれも、語彙の使用頻度や出現数が元になっており、一語一語は全く同等に扱われている。

筆者は、中古散文資料 22 作品相互の類似度の測定を行った際、考察の最後に、形容詞の持つ属性と用いられている作品との関係に着目する作品分類の可能性について言及した。

古代語の形容詞の持つ属性については、村田菜穂子氏の『形容詞・形容動詞の語彙論的研究』⁷⁾ に詳しい。村田氏は、従来の索引データ(見出し語、使用度数など)に、語構造や造語論に関わる属性を新たに付加して、単なる使用頻度や出現数ではなく、これらの

*まえかわ たけし：大阪国際大学短期大学部教授 (2009.12.17受理)

新しい要素を加味した様々な考察を行っている。しかし、村田氏の分析も、多変量解析の手法を用いていないという観点では、単一側面からの分析の域を出ないと言える。

そこで、今回、同書に別表一として掲載されている「古代語形容詞の語構成」のデータを基に、いくつかの属性を変数とした多変量解析を行い、中古散文資料 22 作品を分類しようと試みた。

Ⅱ 解析対象としたデータ

上記別表によると、形容詞の語構成に関する要素は、①活用、②単位数、③結合型、④結合タイプ、⑤階層構造、⑥造語型、⑦造語形式である。

それぞれの要素の意味は以下のとおりである。

- ①活用・・・ク活用またはシク活用のいずれかを示したもの
- ②単位数・・・語を構成する最小単位である語基または接辞を単位としていくつ結びついているかその数を表したもの
- ③結合型・・・「④結合タイプ」をさらに 3 種類に大別したもの
- ④結合タイプ・・・語基と接辞がどのような形で結びついているかを基準に 38 種類に分類したもの
- ⑤階層構造・・・形容詞を発達段階によって 3 種類に分類したもの
- ⑥造語型・・・「⑦造語形式」をさらに 9 種類に大別したもの
- ⑦造語形式・・・語がどのような造語成分から組み立てられているかを基準に 68 種類に分類したもの

解析対象の要素としては、上記の①～⑦が考えられるが、②単位数は④結合タイプに内包されること、③結合型と④結合タイプは同義であること、⑥造語型と⑦造語形式は同義であることから、①活用、④結合タイプ、⑤階層構造、⑦造語形式の 4 つに限定した。

次に、作品毎のそれぞれの要素を数値化することとし、その尺度をどうするかを考えた。

『形容詞・形容動詞の語彙論的研究』別表三「中古散文作品の形容詞対照語彙表」には、形容詞の各見出し語の作品毎の出現数が掲載されている。

作品間の規模による差を吸収するためには、出現数などの量そのものではなく、出現率などの全体に対する比率を用いる必要がある。

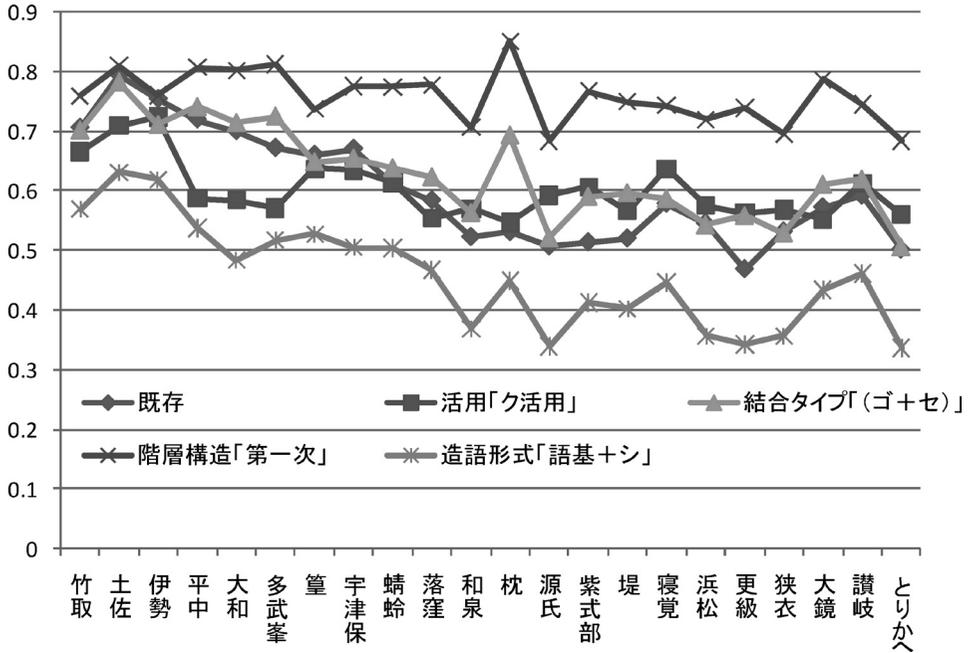
そこで、例えば、④結合タイプの場合、ある作品で使用されている全形容詞の出現数に対する特定の結合タイプに属する形容詞の出現数の割合という形ですべての結合タイプについて比率を算出した。比率を算出する際には、分母にあたるものの数値がある程度大きくないと意味がないが、今回はどの要素についても特に問題はなかった。

このようにして、作品毎に、①活用、④結合タイプ、⑤階層構造、⑦造語形式のそれぞれについて、出現するすべての型の比率を求めると、どの作品にも共通して、それぞれ出現する型の中で使用比率が最も高い値を示す基本型とも言うべきものが存在することが判明した。

具体的には、①活用においては「ク活用」、④結合タイプにおいては「(ゴ+セ)」、⑤階

層構造においては「第一次」、⑦造語形式においては「語基+シ」である。

これらの値を「既存（上代に使用されていたもの）」の割合とともにグラフに表すとグラフ1「語構成要素の基本型データの作品別推移」のように特徴的な傾向を示す。



グラフ1 語構成要素の基本型データの作品別推移

そこで、作品毎の①活用、④結合タイプ、⑤階層構造、⑦造語形式の4つの要素の値を、上記に述べた基本型の比率で代表させることとし、これに、既存の比率を加えた5つの数値を最終的に解析対象データとすることにした。

Ⅲ 多変量解析の試み

多変量解析の手法にはいくつかあるが、ここでは、上記5つの数値を変数として統計分析ソフト SPSS12.0J を用いてクラスター分析を行った。

距離の測定方法は、平方ユークリッド法を用い、クラスタ化の手法は、最短距離法、最長距離法、メディアン法、重心法、群平均法、可変法、Ward法のすべてを試してみた。

クラスタ化の手法によってかなりの差が出るのではと懸念していたが、実行してみると、7つの手法の内の5つの手法でほとんど同じようなデンドログラムを得ることができた。以下にこの5つのデンドログラム（図1～図5）を示す。

Dendrogram using Complete Linkage

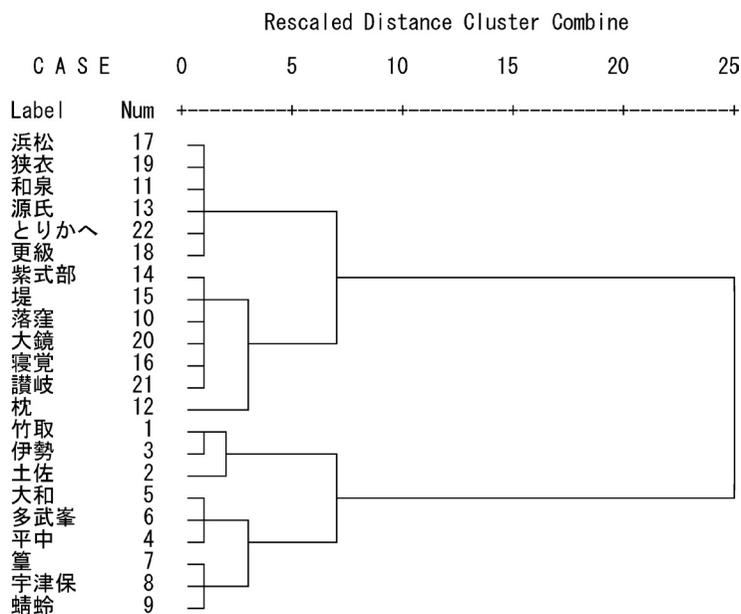


図1 デンドログラム (最長距離法による)

Dendrogram using Average Linkage (Between Groups)

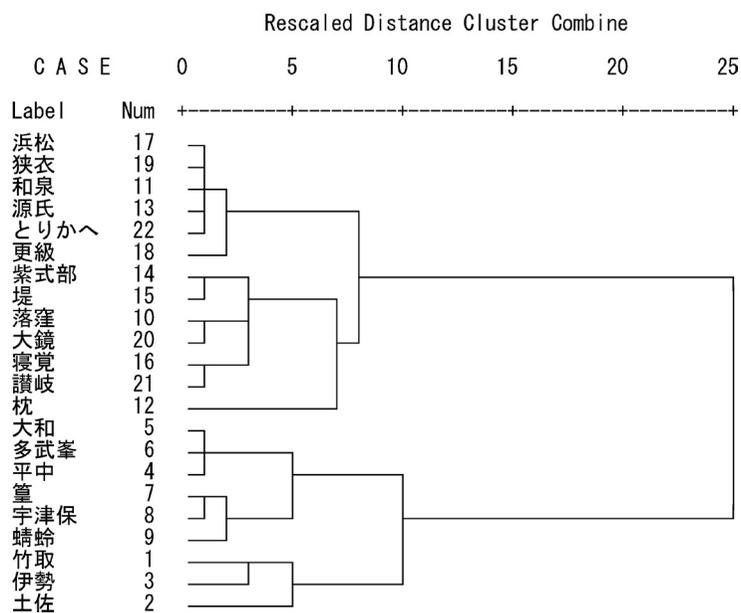


図2 デンドログラム (群平均法による)

中古散文22作品の分類についての一試み

Dendrogram using Centroid Method

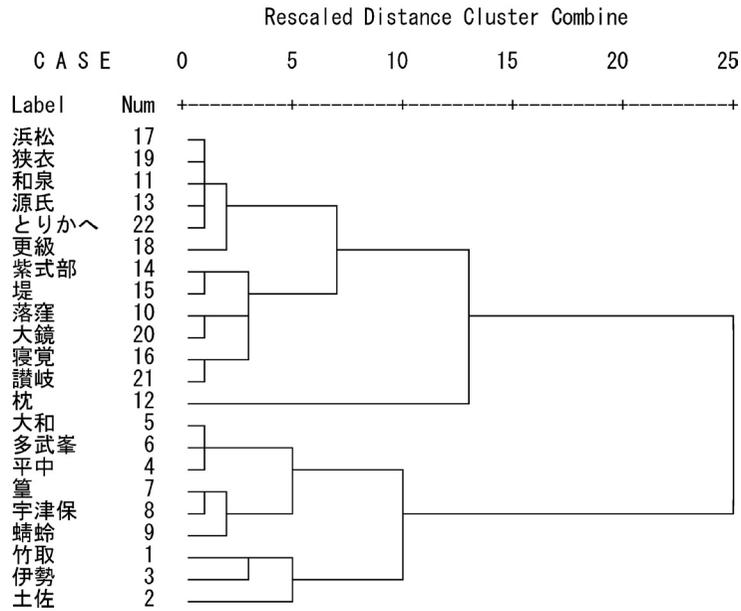


図3 デンドログラム (重心法による)

Dendrogram using Median Method

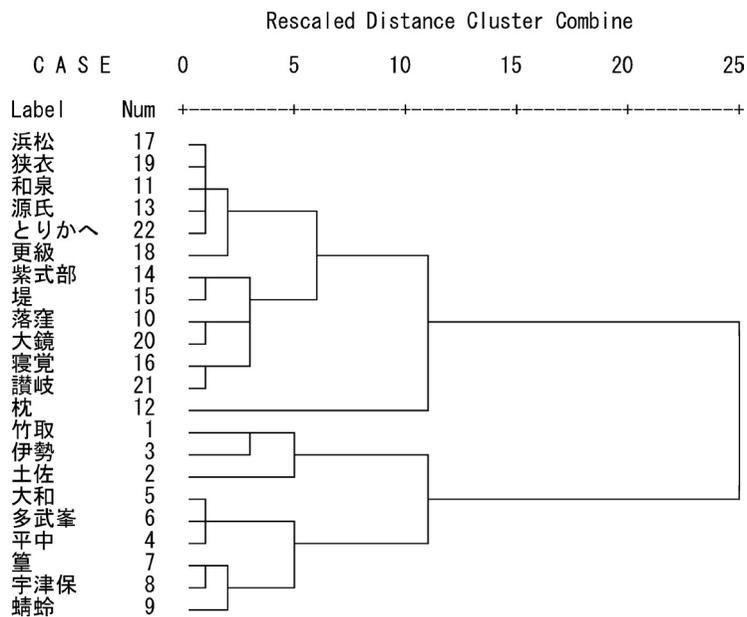


図4 デンドログラム (メディアン法による)

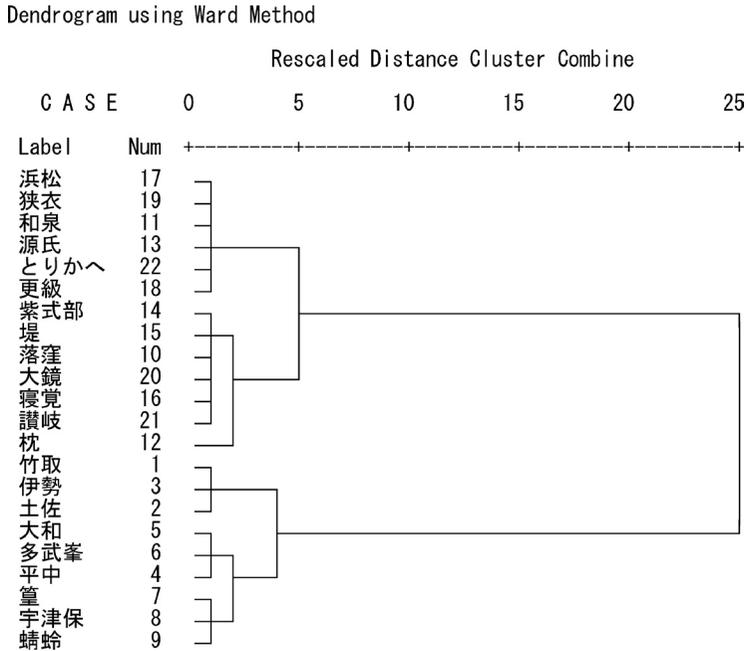


図5 デンドログラム (Ward 法による)

Ⅲ 考察

各デンドログラムで示されている2列目の数値は成立年代の古い順(通説に従う)につけた連番を表わしている。

クラスター分析の結果、大別すると、①平安初期の竹取物語から蜻蛉日記のグループ(以下グループ①と呼ぶ)と②平安中期以降の落窪物語からとりかへばや物語のグループ(以下グループ②と呼ぶ)に分かれることがわかった。

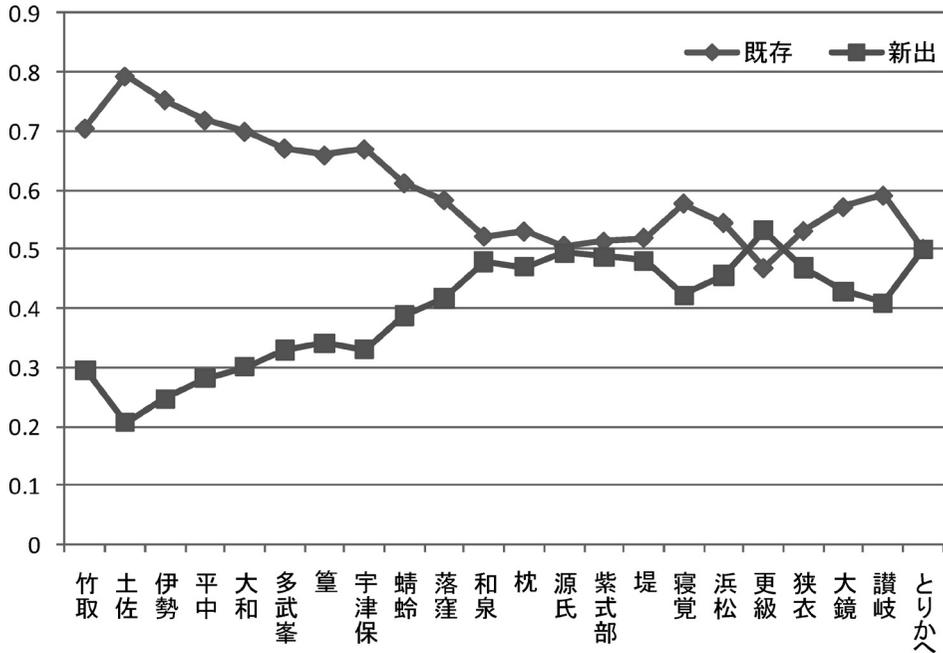
それぞれのグループをみると、グループ①では成立年代に疑問が存する篁物語、平中物語においても通説に従った結果になっているが、グループ②では、通説上の成立年代と前後する部分があることが伺える。

前出の著作で、村田氏は以下のように述べている。

平安初・中期の古い作品ほど既存形容詞の使用率の方が高く、既存形容詞と新出形容詞の使用率の差は大きくなっているが、平安後期に向かって徐々に既存の語の使用率は減少し、新出の語の使用率が増加してくるようになる。ところが、『和泉』あたりを境に、既存の語と新出の語それぞれの使用率は、同程度で一定するようになり、作品の規模等にかかわらず安定した状況が保たれるようになる。

この状況を表わしたものがグラフ2「中古散文 既存・新出(延べ語数)比率」である。

中古散文22作品の分類についての一試み



グラフ2 中古散文 既存・新出 (延べ語数) 比率

(出典：村田菜穂子、『形容詞・形容動詞の語彙論的研究』、初版、和泉書院、2005年)

このグラフ2を図1～5と比べるとデンドログラムの中央グループの『讃岐』、『寝覚』、『大鏡』、『落窪』、『堤』、『紫式部』の内、前者4作品はグラフ2上では周りから突出して同じ傾向を示している。

また、デンドログラムの最上グループの『浜松』、『狭衣』、『和泉』、『源氏』、『とりかへ』の内、後者3作品はグラフ2上ではほぼ中央付近に位置し同じ傾向を示している。

このことは、「既存」かどうか、「活用」、「結合タイプ」、「階層構造」、「造語形式」それぞれの基本型とも言うべきものと密接な関係にあることを示している。

IV おわりに

今回、中古散文22作品を分析するための変数として数値化できるものを用いてクラスター分析を行ったところ、複数の手法ではほぼ妥当と思われる分類結果が得られ、当初の目的は一応達成されたと言える。

しかし、この分類結果は考察のスタートに過ぎず、この分類結果をどのように解釈するかが重要であり、詳細については次回の課題としたい。

また、前述したように、宮島氏の提唱した手法に基づいて形容詞および形容動詞の見出し語の使用率から中古散文22作品の距離感を測定して図示する試みを行ったが、その際、距離感を一つの平面上に図示することはできたものの、距離感に矛盾が生じることも明ら

かになった。

この矛盾を解決する方法として三次元空間の中で点をプロットすることを提唱したが、その後、数量化理論Ⅳ類を使えば解決するのではないかと考えた。本稿には間に合わなかったが、次回は、数量化理論Ⅳ類による分類結果と今回の結果とを比較するなど、さらなる考察を行いたい。

【付記】

本稿は、日本学術振興会平成19-22年度科学研究費補助金（基盤研究（C）課題番号19520407）による研究成果の一部である。

注

- 1) 『竹取物語』、『土佐日記』、『伊勢物語』、『平中物語』、『大和物語』、『多武峯少将物語』、『篁物語』、『宇津保物語』、『蜻蛉日記』、『落窪物語』、『和泉式部日記』、『枕草子』、『源氏物語』、『紫式部日記』、『堤中納言物語』、『夜の寝覚』、『浜松中納言物語』、『更級日記』、『狭衣物語』、『大鏡』、『讃岐典侍日記』、『とりかへばや物語』
- 2) 『大阪国際大学紀要国際研究論叢』、21-1、2007年
- 3) 宮島達夫、「語いの類似度」、『國語學』、82、1970年
- 4) 宮島達夫、「語いの類似度」、『國語學』、82、1970年
- 5) 進藤義治、『源氏物語 形容詞類語彙の研究』、笠間書院、1978年
- 6) 田中章夫、「作品の語彙の偏りを測る」、『国語語彙史の研究』、4、1983年
- 7) 村田菜穂子、『形容詞・形容動詞の語彙論的研究』、初版、和泉書院、2005年

参考文献

- 1) 村上正康・田栗正章 訳、『多変量解析の基礎』、1992年
- 2) 村上征勝・金明哲、『講座 人文科学研究のための情報処理[第5巻 数量的分析編]』、初版、尚学社、1998年
- 3) 水谷静夫、「調査報告 計量語彙論から見た明星派と根岸派」、『計量国語学』、第11巻1号、1977年