

源氏物語と宇津保物語における語の使用傾向について

土山 玄
同志社大学大学院 博士前期課程

村上 征勝
同志社大学 文化情報学部

文体は作者や作品によって相違すると考えられるが、相違点を明示することは容易ではない。本研究はこの問題を考えるため、源氏物語と宇津保物語を対象に語彙の使用頻度を品詞別に求め、計量分析をおこなう。主成分分析とクラスター分析より、名詞、動詞、形容詞、形容動詞、副詞、助詞、助動詞に使用傾向の相違が認められた。次に、ランダムフォレストより、使用傾向に顕著な相違が見られる語彙をもとめた。結果、源氏物語に比べ宇津保物語は漢文訓読文体の特徴を有していると言える。

The Usage Trend of Vocabulary in “*The Tale of GENJI*” and “*The Tale of UTSUHO*”

Gen Tsuchiyama
Graduate School of Culture and Information Science
Doshisha University

Masakatsu Murakami
Faculty of Culture and Information Science
Doshisha University

We will investigate some differences in the style between “*the Tale of GENJI*” and “*the Tale of UTSUHO*” by the quantitative analysis. So, we will analyze two literary works quantitatively to reveal the difficulty. In analyzing, we have two viewpoints: the first is showing that the tales have different usage trend of the vocabulary; then, the second is searching for words, which have remarkably different frequency of use. As the result, we will conclude that “*the Tale of UTSUHO*” is more strongly influenced by Chinese classics than “*the Tale of GENJI*.”

1. はじめに

作者や作品によって、文体が相違するということは想像に難くない。しかし、文体とは多義的な概念であり、いかなる要素が文体という概念を規定するのか、明確に定義されることは少ない。本研究では語の出現頻度に着眼し、二つの古典作品について計量的な分析の比較を試みることによって、この問題にアプローチしたい。

現代文を対象とした文体研究では、語の出現頻度を計量的に分析した結果、助詞の用い方に作者の特徴が現れると報告されている（金 1997, 2002）。他方、古典文を対象とした計量的な文体研究は現代文に比べ非常に困難であることが多い。これは主に電子化された古典文のテキストが少ないこと、および古典文の形態素解析が容易ではないことに起因する。

本研究では、電子化されたデータベースが構築されている古典作品、すなわち源氏物語と宇津保物語を対象とし、これらの文体について計量的な検討を加える。本研究の目的は、古典文において作者が異なる場合、語の使用傾向の相違が認められる品詞を明らかにすること、次いで使用傾向の相違が認められた品詞については、どのような語彙の出現頻度が顕著に相違するか明らかにすることである。なお分析は品詞別に

行う。分析の対象となる品詞は名詞、動詞、形容詞、形容動詞、副詞、補助動詞、助詞、助動詞の8品詞である。

比較が意味を持つためには、相互に共通の基盤をもたなければならない。源氏物語と宇津保物語は成立が概ね同時期であり、ともに物語文学というジャンルに属し、かつ和文体に類別される。

2. データベース

分析に用いたデータベースについて、源氏物語の場合は『源氏物語語彙総索引』（上田他 1994）を電子化したものである。これは主に大島本を底本とする『源氏物語大成』（池田 1984）をもとに、54帖すべての本文を単語ごとに分割し、これについて仮名読み、終止形、品詞コード、活用形コードなどが付与されている。また宇津保物語については、尊経閣蔵前田家十三行本を底本とする『うつほ物語 全改訂版』（室城 1995）を『源氏物語語彙総索引』と同一規格で形態素解析を行い、電子化したものである。

なお、これらのデータベースでは、句読点などの情報は削除されている。また、たとえば「飽く」と「開く」というように品詞が同一であり、本文において「あく」と同一の仮名で表記されている場合、意味番号が付与されており、意味の相違による識別が可能である。

3. 分析

3.1. 分析対象

本研究では源氏物語および宇津保物語の各帖を分析対象とする。源氏物語は全 54 帖、宇津保物語は全 20 帖によって構成されるため、分析対象の総数は 74 となる。しかし、総語数が 2 万語を超える帖がある一方で、1000 語を下回る帖もあり、総語数のばらつきは大きい。一般に、テキストデータに対し多変量解析を行う場合、総語数の少ない対象が他の対象に比べ、安定した分析結果を示さないことが多い。このため、本研究においては、源氏物語の第 11 巻「花散里」の総語数が 724 語、第 16 巻「関屋」が 934 語、第 27 巻「篝火」が 653 語であるから、これら分析対象から除外した。よって、分析対象の総数は 71 となる。

3.2. 分析項目

名詞、動詞、形容詞、形容動詞、副詞、補助動詞、助詞、助動詞に属する語を変数とし、帖ごとにすべての語の相対頻度を求め、これを分析に用いる。相対頻度は、帖ごとに集計された各語の出現頻度の総語数に対する割合である。

また、両作品に現れるすべての語を変数として用いず、品詞ごとに変数選択を行う。変数を選択する基準は、総出現頻度が低く、かつ 71 の分析対象の半数以下にしか出現しない語を分析から除外する、というものである。なお、総出現頻度は帖ごとに集計された出現頻度の総和である。

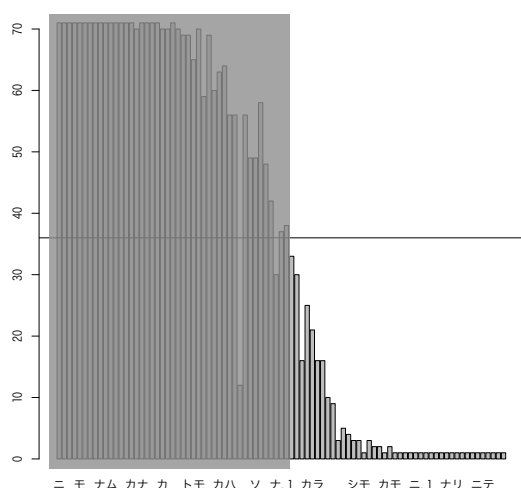


図1 変数選択 (助詞)

助詞を例とし、棒グラフにより図示すると図1のようになる。分析に使用した語は図1の灰色の領域にある語である。図1の横軸は総出現頻度を降順にソートした語を、縦軸は語の出現

する対象数を表している。したがって、図1において棒が高い語ほど多くの対象で用いられており、横軸について左にある語ほど総出現頻度が高いことを意味する。このように決定された各品詞の分析に用いる語数と各品詞の異なり語の総数は表1のようになる。

表1 語数

	異なり語数	使用語数
名詞	9594	273
動詞	6737	242
形容詞	969	109
形容動詞	769	45
副詞	360	51
補助動詞	34	6
助詞	87	45
助動詞	38	22

3.3. 分析手法

分析においては主成分分析、クラスター分析、そしてランダムフォレストを用いる。まず、源氏物語と宇津保物語の間に語の使用傾向の相違が認められるか否かについて検討するため、主として主成分分析を、補助的にクラスター分析を用いる。次いで、両作品の間に語の使用傾向が認められた場合、ランダムフォレストを用い、これにより顕著に使用傾向の相違があらわれる語を指摘する。

主成分分析は多次元データを低次元に縮約することでデータの特徴を理解しやすくする手法である。クラスター分析は類似した特徴を有する分析対象から段階的にクラスターを形成することで、分析対象の階層構造を示す手法である。本研究ではユークリッド距離とワード法による階層的なクラスター分析を用いる。

ランダムフォレストはアンサンブル学習と称される手法のひとつで、学習データを用い分類器を構築し、未知のデータの母集団を判別する手法である。母集団を判別する手法としてはバギング、ブースティング、サポートベクターマシン (SVM) などがあるが、文章の母集団の判別、すなわち文章の書き手の識別についてはこれらの手法に比べ、ランダムフォレストの精度が最も高いとされる (金・村上 2007)。

くわえて、ランダムフォレストは分類器を構築する際に、分類における変数の重要度の推定を行う。本研究において、重要度の高い変数と

は、源氏物語と宇津保物語のどちらかの作品に頻出し、他方の作品にはおよそ現れない語を指す。

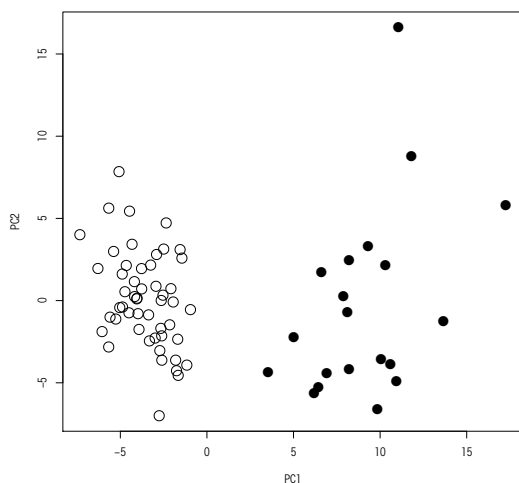
また、ランダムフォレストは分類器の構築の際に、乱数を発生させることで精度の向上を図っている。しかし、そのために複数回のランダムフォレストを行うと、同一の結果を得られることはおよそない。したがって、本研究ではランダムフォレストを 1000 回繰り返す。変数の重要度の推定値には 1000 回推定された値の平均を用いる。

3.4. 分析結果

3.4.1. 名詞

名詞における主成分分析の結果は図 2 に示すようになり、源氏物語の各帖と宇津保物語の各帖は横軸の、すなわち第 1 主成分の正負に分離して配置される。クラスター分析においても、両作品の各帖はおよそ作品別に 2 つのクラスターに分けられる。したがって、両作品の間において名詞の語彙の使用傾向に相違が認められる。

次に、図 3 に示すように、ランダムフォレストより両作品の間に顕著な使用傾向の相違が認められた語は「けはひ」、「ありさま」、「さま」である。これら 3 語は宇津保物語に比べ、源氏物語に頻出する。『日本語語彙体系 1 意味体系』(池原他 1997)によれば、これら 3 語は「様相」という範疇に属する語であるから、源氏物語の作者は宇津保物語に比べ人物や物事の在りように関心を払っていたと考えられる。



○：源氏物語 ●：宇津保物語

図 2 名詞の主成分分析

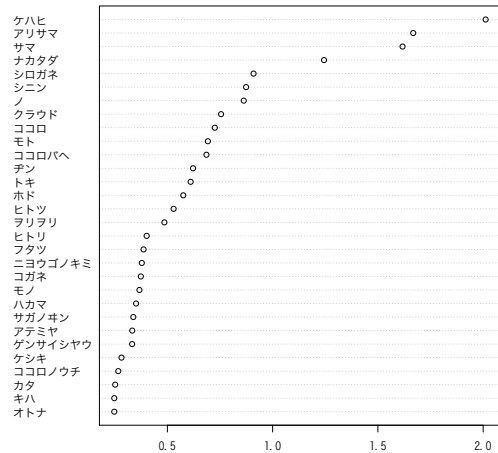
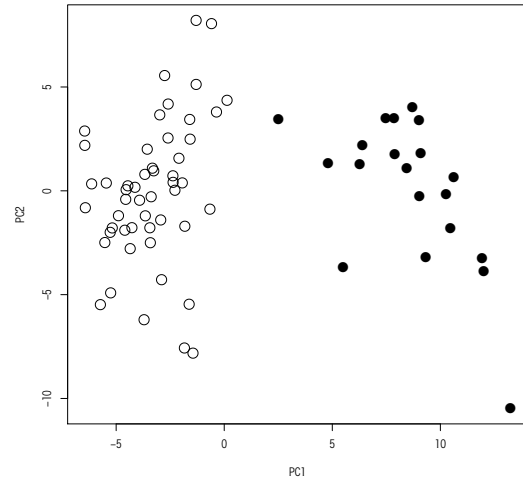


図 3 名詞のランダムフォレスト

3.4.2. 動詞

動詞における主成分分析の結果は図 4 に示すように、両作品は横軸の正負に分離し、クラスター分析においても作品別に 2 つのクラスターに分類される。すなわち、動詞においても源氏物語と宇津保物語との間に語の使用傾向に相違が認められる。



○：源氏物語 ●：宇津保物語

図 4 動詞の主成分分析

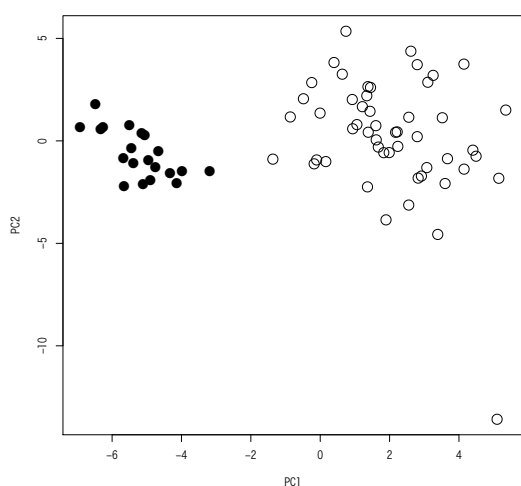
次に、ランダムフォレストより、両作品において顕著に使用傾向に相違が認められる語は「もてなす」、「たてまつる」、「たまふ」などである。

「もてなす」は相対的な源氏物語に頻出し、「たてまつる」、「たまふ」は宇津保物語に頻出する語である。これら 3 語のうち、「もてなす」は精神的行動の範疇に含まれ、「たてまつ

る」および「たまふ」は物理的行動の範疇に含まれる語であるから、ここに源氏物語と宇津保物語の文体上の相違を指摘できる。

3.4.3. 形容詞

形容詞における主成分分析の結果は図5に示すように横軸の正負に作品別に分離することから、両作品の間に語の使用傾向の相違があると考えられる。しかし、クラスター分析では図6に示すように源氏物語の9帖が一群を作り、宇津保物語の20帖と同一のクラスターを形成する。



○：源氏物語 ●：宇津保物語
図5 形容詞の主成分分析

次に、形容詞のランダムフォレストでは両作品において大きく出現頻度に相違が見られる語は「いたし」、「こころぐるし」、「なつかし」、「をかし」などで、これらは源氏物語に頻出する。このうち「いたし」、「こころぐるし」、「なつかし」は感情の状態に関わるものであるから、源氏物語は宇津保物語に比べ、より感情を指示する語を多用することが注目される。

3.4.4. 形容動詞

形容動詞は主成分分析において、図7に示すようにおよそ作品別に分離し、2群を形成していると見られる。ただし、宇津保物語の第19巻「楼の上上」と第20巻「楼の上下」の2巻は源氏物語の語彙の使用傾向に親近性が認められると言える。これについて、クラスター分析の結果も図9に示すように同様である。

次に、ランダムフォレストより両作品において大きく出現頻度に相違が見られる語は「さすがなり」、「あはれなり」、「おほいなり」などで、「さすがなり」、「あはれなり」は源氏物語に頻出し、「おほいなり」は宇津保物語に頻出する。「あはれなり」は形容詞の「をかし」と並び、美を表す語である。したがって、源氏物語は宇津保物語に比べより美を形容する語が多用されるといえる。なお、「おほいなり」は「おほきなり」の音便であると考えられ、これ

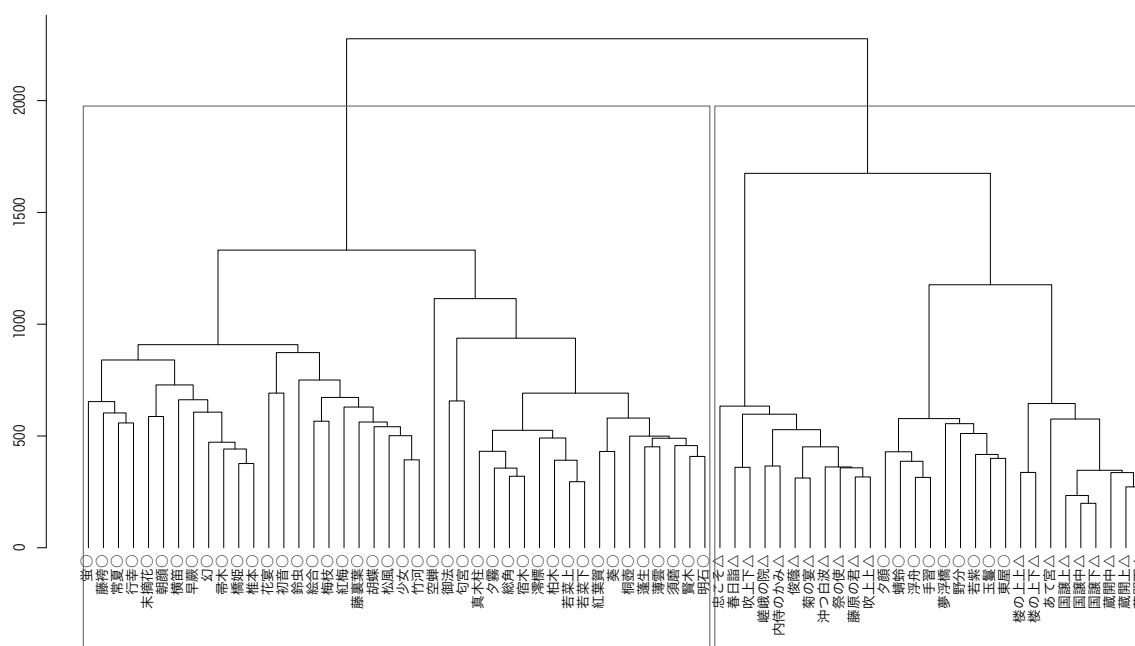
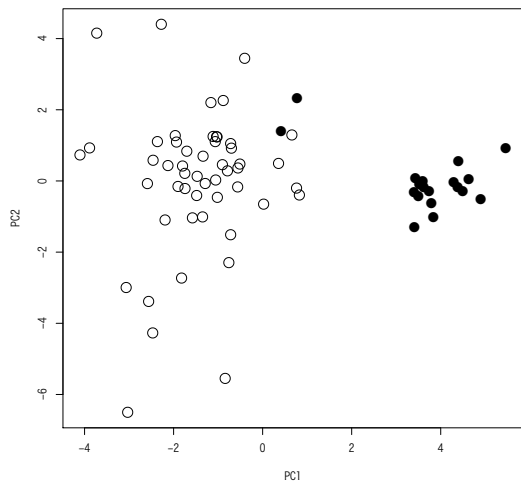


図6 形容詞のクラスター分析

は宇津保物語に40例認められるが、源氏物語においては第2巻「帚木」に1例確認されるのみである。くわえて、「おほいなり」の用例は源氏物語および宇津保物語においてすべて連体形の「おほいなる」の形であらわれる。

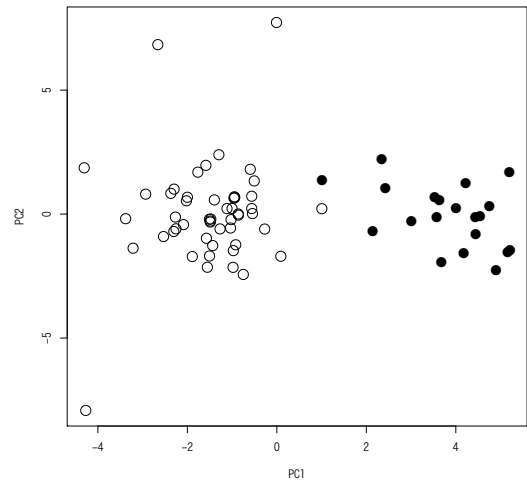
したがって、宇津保物語は源氏物語に比べ、あるいは本研究において用いた宇津保物語のデータベースである尊経閣蔵前田家十三行本の特徴として、「おほいなる」という音便形の使用が好まれると言える。



○：源氏物語 ●：宇津保物語
図7 形容動詞の主成分分析

3.4.5. 副詞

副詞は主成分分析において、図8に示すように作品別に分離し2群を形成していると思われる。クラスター分析においても、宇津保物語の数帖が源氏物語と混在するが、およそ作品別にクラスターが形成される。



○：源氏物語 ●：宇津保物語
図8 副詞の主成分分析

次に、ランダムフォレストより両作品において大きく出現頻度に相違が見られる語は「すなはち」、「なかなか」、「おのづから」そして

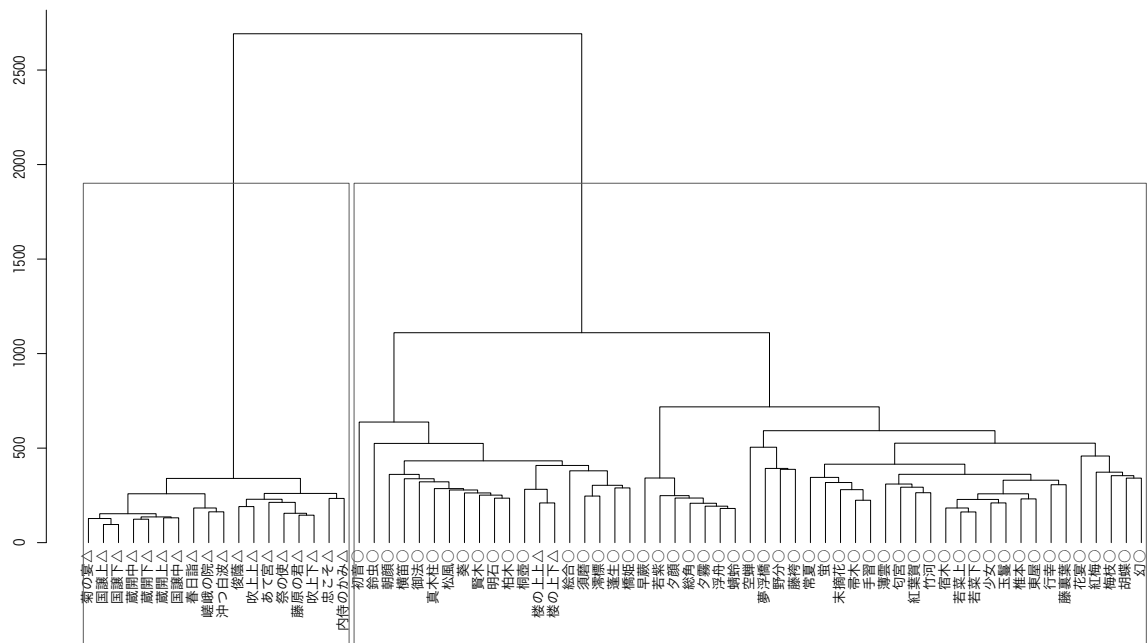


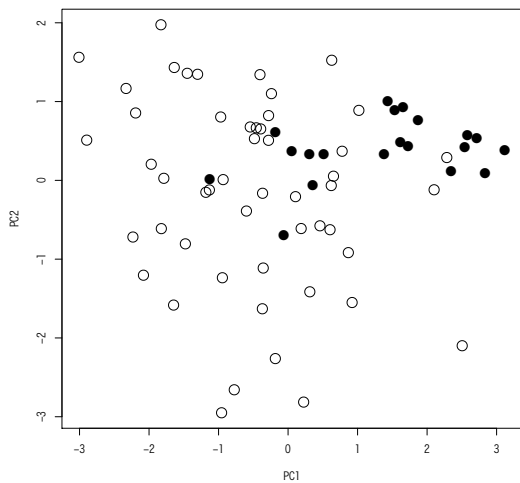
図9 形容動詞のクラスター分析

「すこし」である。このうち「すなわち」は宇津保物語に頻出し、「なかなか」、「おのづから」、「すこし」は源氏物語に頻出する語である。「なかなか」は期待されることが生起しない状況を述べる際に用いられる語であり、同じく源氏物語に頻出する形容動詞の「さすがなり」も期待に反する状況で用いられる語であることを考え合わせると、源氏物語にはこのような期待とは異なるときに用いられる副詞が宇津保物語に比べて多いと言える。なお、宇津保物語に頻出する「すなわち」は訓読語であるから（築島 1963）、宇津保物語は源氏物語に比べ漢文訓読文体の特徴も有していると考えられる。

3.4.6. 補助動詞

補助動詞における主成分分析の結果は図 10 に示すように、源氏物語と宇津保物語という 2 群に分離せず、両作品の各帖はおよそ混在する。クラスター分析においても作品別のクラスターは形成されず、両作品は混在する。

ゆえに、補助動詞の語彙においては明確な使用傾向の相違を認めることはできない。



○：源氏物語 ●：宇津保物語
図 10 補助動詞の主成分分析

3.4.7. 助詞

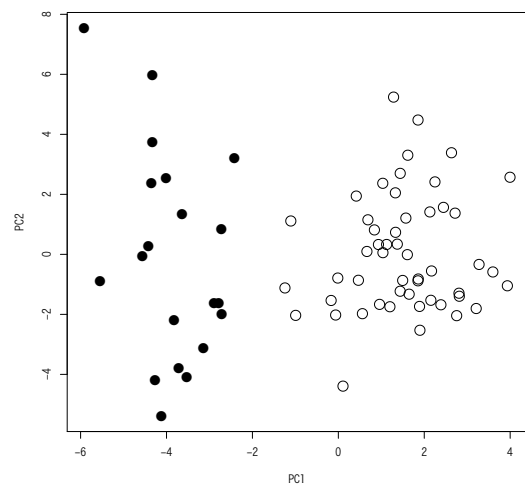
助詞の主成分分析の結果は図 11 に示すように、横軸の正負に作品別に分離して配置される。クラスター分析においても両作品の各帖はおよそ作品別に 2 つのクラスターに分けられる。したがって、源氏物語と宇津保物語の間に助詞についても、使用傾向に相違が認められる。

次にランダムフォレストより、両作品において顕著に使用傾向に相違が認められる語は「も」、「ど」、「して」である。

このうち「も」は漢文訓読文体において盛ん

に使用される語ではない（築島 1963）。ランダムフォレストより「も」は宇津保物語に比べ源氏物語に頻出する語であると言える。次いで、「ど」は宇津保物語に比べ源氏物語に多く用いられる接続助詞である。「ど」は接続助詞である「ども」と併用されるが、平安時代において、「ど」は和文体、「ども」は漢文訓読文体で用いられる語である（築島 1963）。

最後に、「して」は宇津保物語に頻出する。また、「して」が直前の語と接続した形式である「…くして」「…にして」「…として」「…ずして」は、平安時代において主に漢文訓読文体の形式であり（築島 1963）これらの 4 形式も宇津保物語に相対的に頻出する。



○：源氏物語 ●：宇津保物語
図 11 助詞の主成分分析

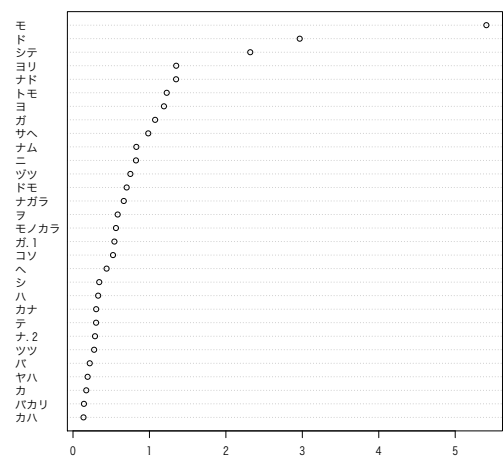


図 12 助詞のランダムフォレスト

源氏物語と宇津保物語はともに和文体の作品であることは既に述べたが、助詞の語彙の使用

傾向を見ると、源氏物語は宇津保物語に比べ、より和文体的であり、反対に宇津保物語は和文体の物語でありながら、漢文訓読文体の特徴も有すると指摘できる。

3.4.8. 助動詞

助動詞の主成分分析の結果は図 13 に示すように、源氏物語と宇津保物語の各帖は作品別に分離して配置される。クラスター分析においても概ね作品別に 2 つのクラスターを形成する。したがって、源氏物語と宇津保物語との間に、助動詞の語彙については使用傾向の相違が認められる。

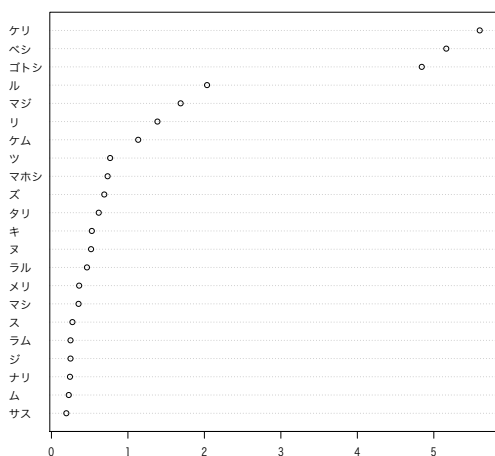
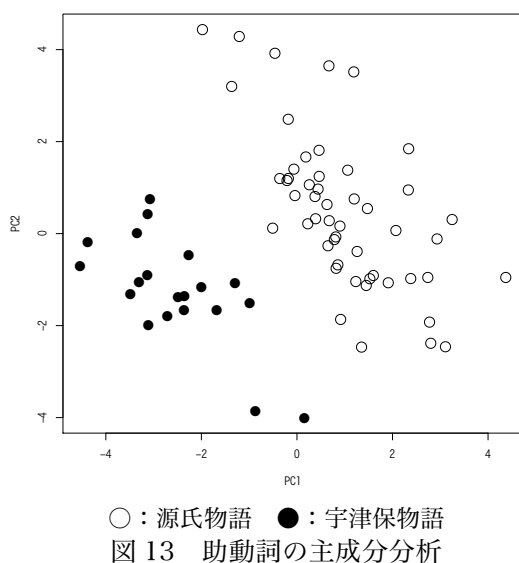


図 14 助動詞のランダムフォレスト

次に、ランダムフォレストより両作品において顕著に使用傾向に相違が認められる語は「けり」、「べし」、「ごとし」である。

「ごとし」は源氏物語に比べ宇津保物語に頻出する。一般に源氏物語と宇津保物語は和文体であるとされるが、「ごとし」は漢文訓読文体の直喩の表現形式である(山口 1984)。したがって、助動詞の分析より宇津保物語は源氏物語に比べ漢文訓読文体の特徴を有していると考えられる。

4. 考察

源氏物語と宇津保物語の文体を、その語彙に注目して品詞ごとに検討を加えるとき、いくつかの特徴を指摘することができる。まず、主成分分析およびクラスター分析により、名詞、動詞、形容詞、形容動詞、助詞、助動詞の語彙に使用傾向の明瞭な相違が認められた。一方、補助動詞については顕著な相違は見られなかった。

次いで、ランダムフォレストを用いた検討の結果、名詞では「けはひ」、「ありさま」など様相を意味する語の使用に、源氏物語の特徴を指摘した。動詞においては源氏物語が精神的行動を意味する語を多用し、宇津保物語が物理的行動を意味する語を多用する点にそれぞれ特徴を見いだした。形容詞では源氏物語において「こころぐるし」、「なつかし」など感情を指示する語の使用が認められた。形容動詞では「あはれなり」の多用が源氏物語を特徴づけることを指摘できる。副詞では「なかなか」など登場人物の期待に反する状況で用いられる語が源氏物語に多用されることが認められた。

また、源氏物語と宇津保物語はともに和文体であると考えられているが、助詞においては和文体に用いられることが多い「も」および「ど」が源氏物語に多用され、他方、主に漢文訓読の語であると考えられる「して」が「…くして」「…にして」、「…として」、「…ずして」の形式で宇津保物語に多用される。助動詞においても漢文訓読文体の直喩の表現形式である「ごとし」が宇津保物語に多用される。くわえて、宇津保物語に頻出する副詞「すなわち」も訓読語として使用される。ゆえに、宇津保物語は源氏物語に比べ漢文訓読文体の特徴も有しているといえる。

以上の検討を通じて、品詞ごとに語彙の使用傾向に相違を認めることができることを指摘したい。すなわち、名詞、動詞、形容詞、形容動詞に属する語彙は作品の内容の相違に関わり、他方、助詞、助動詞に属する語彙は作品の形式の相違、すなわち和文体であるか、あるいは和文体でありながら漢文訓読文体の特徴も有しているか、ということに関わると考えられる。そして副詞は作品の内容と形式の両方に亘る文体

的機能を有する要素であることを指摘することができる。

謝辞

本論文を投稿するにあたり、査読者からの建設的な御助言をいただき、誠に感謝申し上げます。

参考文献

- [1] 池田亀鑑：源氏物語大成 全14巻，中央公論社，1984
- [2] 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，勉誠出版，1994
- [3] 金明哲：助詞の分布に基づいた日記の書き手の識別，計量国語学，Vol.20，No.8，1997
- [4] 金明哲：助詞の n-gram モデルに基づいた書き手の識別，計量国語学，Vol.23，No.5，2002
- [5] 金明哲・村上征勝：ランダムフォレスト法による文章の書き手の同定，統計数理，Vol.55，No.2，2007
- [6] 築島裕：平安時代の漢文訓読語につきての研究，東京大学出版会，1963
- [7] 室城秀之：うつほ物語 全 改訂版，おうふう，1995
- [8] 山口仲美：平安文学の文体の研究，明治書院，1984