

近代文学作品における継時的な文体の変化についての 計量的な検討：自然主義文学について

土山 玄（同志社大学 研究開発推進機構）

本研究では近代の日本文学史における文体の変化について検討を加える。分析において採り上げる作家は20世紀の文豪の一人である夏目漱石、そして同時期の文壇の主流であった自然主義文学の作家とされる国木田独歩、田山花袋、島崎藤村、徳田秋声の小説である。自然主義とは修飾語句を避け無骨な文章が記述され、自然主義の作家にとって小説とは事実そのままの忠実な再生であり、真実を書くことを創作の目標とした。その一方、夏目漱石の文章、特に初期の作品は軽妙であったと言われる。そこで、夏目漱石の文学作品と自然主義文学との間に文体的特徴の相違が認められるのか統計的な手法を用いて検討を加える。分析の結果、自然主義作家と夏目漱石との間において、打消の助動詞を伴う文末の動詞に出現傾向の相違が認められた。

Quantitative Analysis of Chronological Changes in Writing Styles in Japanese Modern Literature:

A Case Study of The Naturalism Works

Gen Tsuchiyama (Organization for Research Initiative and Development,
Doshisha University)

In this study, we investigate the differences in the writing styles of Soseki Natsume and some writers of the Naturalism literary movement using quantitative analysis. Naturalism was the main literary movement in Japan at the beginning of the 20th century. Many literary scholars consider that Soseki, one of the most famous writers in Japan, was opposed to Naturalism.

The primary analysis parameter used in this study is the relative frequency with which each word appears in the target text. Word counts are estimated on the basis of parts of speech. Analytical results indicate that the relative frequency of the auxiliary verbs at the end of sentences between them.

1. はじめに

一般に、文学作品などの文献を採り上げ、計量的に文章の分析を行う学問は計量文献学と称される。文章の計量分析は、主に書き手の文体に関わる習慣的特徴を統計的に解析する。文体という概念は多様であるが、計量文献学において文体とは計数可能な記述形式のことであり、これは文字や語の頻度、語や文の長さなどの文章を構成する量的な要素である。計量文献学に類する研究では語の意味や記述内容は考慮されず、従って研究の対象とはならない。この点において人文学的な文献研究と計量文献学は立場を異にしている。

従来、計量文献学に関連する学問分野では著者の識別・推定、文献の成立年代や成立順序の推定などが研究されてきたが、本研究ではこれまで取り扱われることの少なかった日本文学史、とりわけ近代の日本文学史における文体の変化につい

て検討を加える。分析において採り上げるのは20世紀の文豪の一人である夏目漱石、そして同時期の文壇の主流であった自然主義文学の作家とされる国木田独歩、田山花袋、島崎藤村、徳田秋声の小説である。

自然主義とは、それ以前の尾崎紅葉が創立した硯友社の凝りに凝った技巧的な表現を好まず、修飾語句を避け無骨な文章が記述され、真実を書くことを目標とした文学的思潮である。作品の多くは小説の中に著者の個性確立を試みる私小説であり、作者自身やその家族、友人などをそのままに記述する文学作品である。その一方、夏目漱石は自然主義文学とは対立する余裕派と称された。夏目漱石の文章、特に初期の作品は軽妙であったと評価され、例えば『吾輩は猫である』の文章は自然主義作家からは不真面目で俳諧的なユーモアだけの遊び半分の文章と批判されている。

本研究ではこれら5人の作家の小説を採り上

げ、日本における自然主義文学と、それに対立する夏目漱石について、その文体における計量的な相違点について検討を加える。これによって作風あるいは文学的思潮の相違に起因すると考えられる文体の相違を明らかにする。

2. データセット

本研究では夏目漱石、国木田独歩、田山花袋、島崎藤村、徳田秋声の5人の文学作品を小説に限定し、テキストデータを青空文庫¹から取得し、これを用いた。分析に用いた作品は表1に示す通りであり、作家別の作品数は夏目漱石が23作品、国木田独歩は32作品、田山花袋は7作品、島崎藤村は21作品、徳田秋声は10作品である。これらのテキストデータを形態素解析し、単語に品詞のタグ付けを行った。形態素解析においてはフ

表1 分析に用いた作品一覧

夏目漱石	一夜、薙露行、草枕、虞美人草、行人、坑夫、こころ、琴のそら音、三四郎、趣味の遺伝、それから、それから、手紙、二百十日、野分、彼岸過迄、文鳥、坊っちゃん、幻影の盾、道草、明暗、門、夢十夜、吾輩は猫である
国木田独歩	あの時分、遺言、初孫、運命論者、画の悲み、置土産、おとずれ、河霧、窮死、牛肉と馬鈴薯、源おじ、恋を恋する人、号外、郊外、少年の悲哀、小春、鹿狩り、酒中日記、女難、たき火、竹の木戸、富岡先生、二少女、二老人、初恋、春の鳥、非凡なる凡人、疲労、星、まぼろし、わかれ、忘れえぬ人々
田山花袋	一兵卒、田舎教師、少女病、新茶のかおり、トコヨゴヨミ、ネギ一束、蒲団
島崎藤村	朝飯、嵐、ある女の生涯、家（上）、家（下）、岩石の間、旧主人、刺繍、食堂、新生、足袋、並木、伸び支度、二人の兄弟、船、分配、芽生、夜明け前第一部上、夜明け前第一部下、夜明け前第二部上、夜明け前第二部下、藁草履
徳田秋声	蒼白い月、足跡、あらくれ、新世帯、仮装人物、黴、軀、縮図、挿話、爛

¹ <http://www.aozora.gr.jp/>

リーソフトのMecab0.996を用いた。

3. 分析

3.1 分析に用いる変数

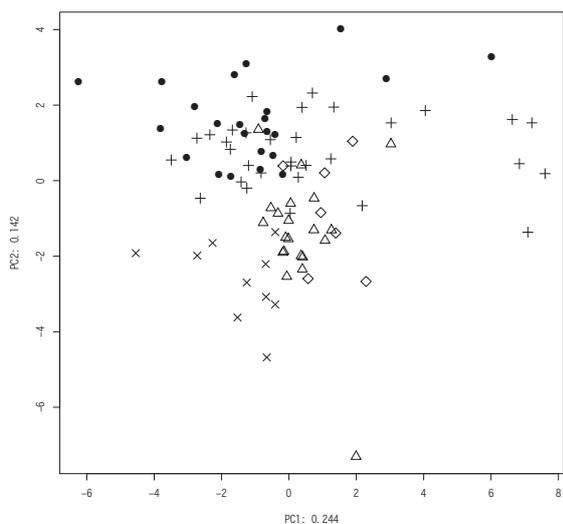
本研究では分析に単語のn-gramを用いる。単語のn-gramとは文中において隣接するn個の単語の組のことであり、n-gramモデルにおいて、n=1はunigram、n=2のときはbigram、n=3であればtrigramと称される。単語のn-gramの例として、「吾輩は猫である。名前はまだない。」という文における名詞のbigramは[猫 名前]、助詞と助動詞のbigramであれば[は だ]及び[だ は]となる。本研究では主にbigramまでを分析に用いた。

また、表1に示した93作品はそれぞれ延べ語数が同様ではないことから、単語のn-gramの出現頻度ではなく出現率を求め、これを分析に用いた。本研究において求めた単語のn-gramの出現率は各作品のn-gramの語数に対する割合である。次に、分析においては、名詞や動詞などの文中において語彙的意味を担う品詞は小説の内容が出現傾向に影響を与えることが予想されるため、原則として名詞のunigramや名詞のbigramのように単独で用いることはせず、機能語と称される助詞や助動詞のような文中において文法的機能を担う品詞と組み合わせ、分析に用いた。また、本研究において用いた分析手法は主成分分析であり、これは多次元データに対する次元縮約の手法であり、元のデータの変数より新たに合成変数を求めることで情報の縮約を行う統計手法である。なお、本研究では相関係数行列を用いた主成分分析を行った。主成分分析による分析結果は第1主成分と第2主成分の主成分得点を散布図として可視化した。本研究ではその散布図において、夏目漱石の小説と他の作家の小説が分離して付置した場合、両者の間に計量的な相違が認められたと考える。

3.2 分析結果と考察

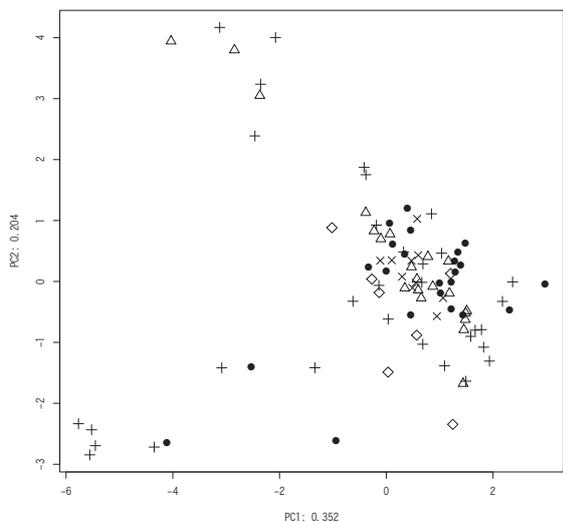
まず、単語のunigramについて検討を加えた。ここで、主成分分析を行った品詞は助詞及び助動詞である。

Mecabによって助詞と認定された単語は159語あるが、出現頻度上位25語までの累積頻度は助詞の総度数の95%を上回る。よって、本研究では出現頻度上位25語の助詞に主成分分析を行った。分析結果は図1に示す通りであり、夏目漱石の作品と自然主義の作家の作品は分離せず、両者の付置される領域は重複していると言える。なお、第1主成分の寄与率は24.4%、第2主成分の寄与率は14.3%である。また、分析に用いる単



● : 夏目漱石 + : 国木田独歩
 △ : 島崎藤村 × : 徳田秋声 ◇ : 田山花袋

図 1 助詞における出現頻度上位 25 語について主成分分析の結果



● : 夏目漱石 + : 国木田独歩
 △ : 島崎藤村 × : 徳田秋声 ◇ : 田山花袋

図 2 助動詞における出現頻度上位 10 語について主成分分析の結果

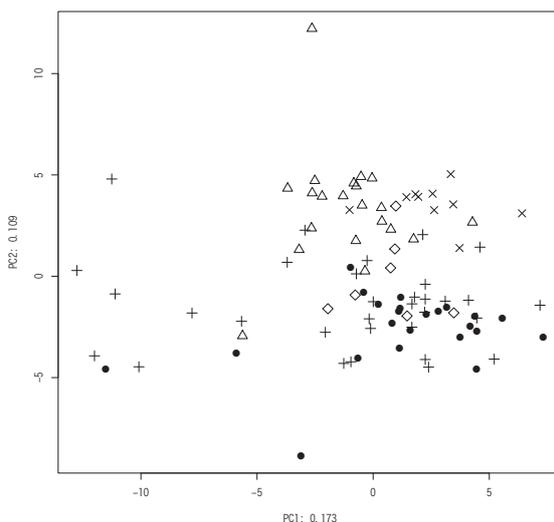
語数を増減させても分析結果は大きく変化しないことから、助詞の unigram の出現傾向は相違しないと考えられる。

次に、助動詞の unigram について分析を行った。Mecab によって助動詞とされた単語は 36 語あり、出現頻度上位 10 語の累積頻度が助動詞の総度数の 95% を上回る。よって、助詞に対する

分析と同様に、この 10 語について主成分分析を行った。分析結果は図 2 に示す通りである。図 1 と同様に夏目漱石の小説と自然主義の作家の小説は分離して位置することはない。なお、第 1 主成分の寄与率は 35.2%、第 2 主成分の寄与率は 20.4% である。また、図 2 において夏目漱石の小説の 3 作品が他の漱石作品から離れて位置しているが、これらは「一夜」、「薙露行」、「幻影の盾」である。分析に用いる変数を増減させても、得られる分析結果を大きく変わらない。

このように、これらの分析結果から助詞及び助動詞の unigram においては、自然主義文学と夏目漱石の作品との間における文体的特徴の顕著な相違は認められないと考えられる。

次に、単語の bigram について検討を加えた。単語の unigram について行った分析と同様に助詞及び助動詞について分析を行った。助詞の bigram の組み合わせは 5136 パターンあることから出現頻度上位 100 変数を用いて主成分分析を行った。主成分分析によって得られた結果は図 3 に示す通りであり、unigram についての分析結果と同様に、夏目漱石の小説と自然主義の作家の小説は分離して位置することはない。なお、分析に用いた 100 変数の累積頻度は総度数の 75.1% に該当し、第 1 主成分の寄与率は 17.3%、第 2 主成分の寄与率は 10.9% となる。次いで、助動詞の bigram についても同様に出現頻度上位 100 変数について分析を行った。この 100 変数の累積頻度は総度数に対する 95.0% にあたる。図 4



● : 夏目漱石 + : 国木田独歩
 △ : 島崎藤村 × : 徳田秋声 ◇ : 田山花袋

図 3 助詞の bigram における出現頻度上位 100 変数について主成分分析の結果

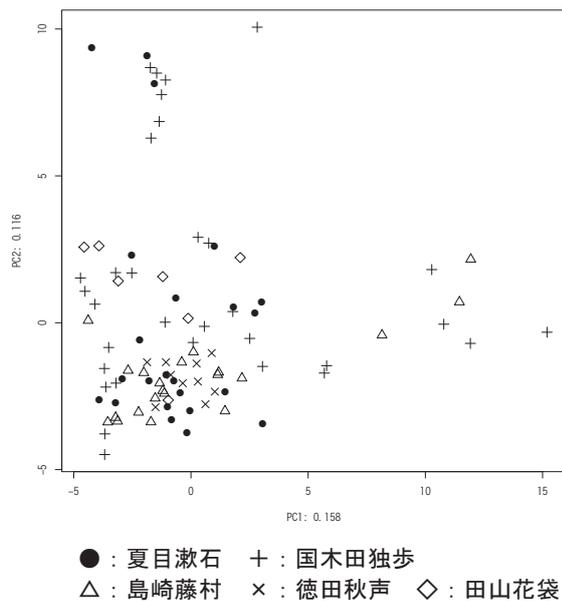


図 4 助動詞の bigram における出現頻度上位 100 変数について主成分分析の結果

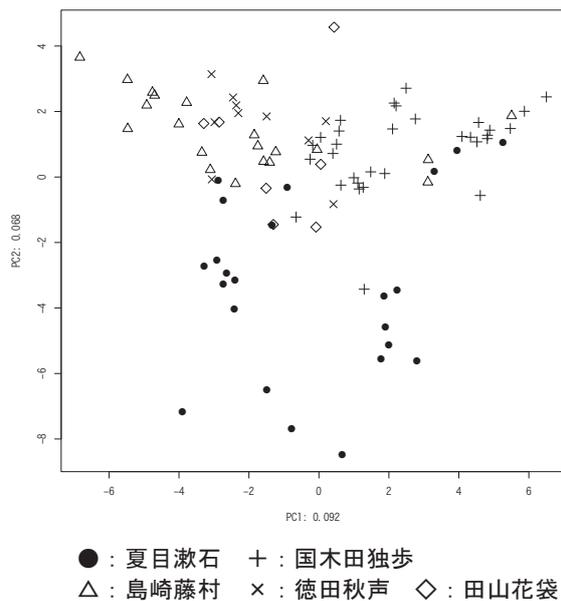


図 5 動詞と助動詞の bigram における出現頻度上位 100 変数について主成分分析の結果

はその分析結果である。助動詞の bigram の分析結果においても、夏目漱石の小説と自然主義の作家の小説は分離しない。

以上より、機能語である助詞及び助動詞を用いた分析からは自然主義文学と夏目漱石の作品との間において、計量的な相違は認められない。

そこで、本研究では助詞や助動詞と他の品詞を組み合わせた bigram を用いて主成分分析を行った。その結果、句読点の前に現れる動詞と助動詞の bigram において、自然主義文学と夏目漱石の作品との間に文体的特徴の相違が認められた。この動詞と助動詞の bigram は、動詞の次に助動詞が現れる組み合わせだけを計数している。図 5 はそのような動詞と助動詞の bigram について行った主成分分析の結果である。用いた変数は出現頻度の上位 100 変数である。この 100 変数の累積頻度は総度数の 55.6% に該当する。図 5 においてはおよそ夏目漱石の作品が自然主義文学の作品が付置している領域から分離して位置する。ただし、「一夜」、「薙露行」、「幻影の盾」、「道草」、「こころ」、「三四郎」、「手紙」の 7 作品が自然主義の位置する領域に混在している。

次いで、変数を増やし出現頻度の上位 200 変数について主成分分析を行った。その結果、図 6 に示すように夏目漱石の作品が自然主義文学の作品は分離して付置されると考えられる。図 6 において自然主義文学と混在している 3 作品は「一夜」、「薙露行」、「幻影の盾」である。なお、上位 200 変数の累積頻度は総度数の 63.4% である。

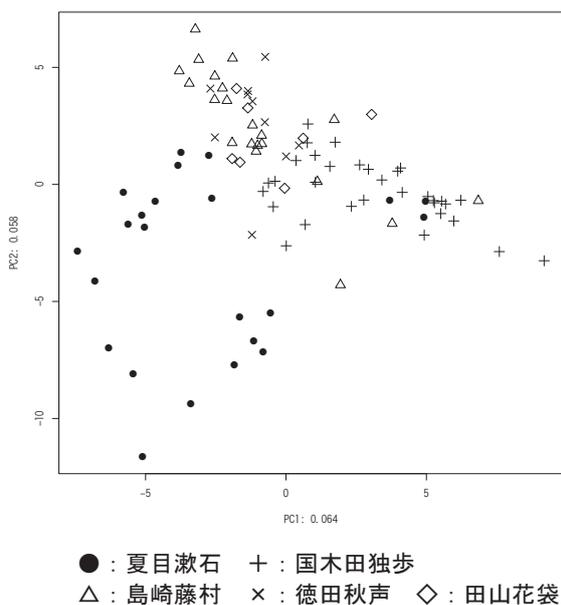


図 6 動詞と助動詞の bigram における出現頻度上位 200 変数について主成分分析の結果

このように、本研究の分析結果から、句読点の前に現れる動詞と助動詞の bigram に夏目漱石の小説と自然主義の作家の小説の間における文体的特徴の相違が認められると考えられる。

次に、どのような動詞と助動詞の bigram に顕著な相違が認められるのか明らかにするために分析を行った。用いた手法は独立サンプルの t 検

表 2 特徴的な句読点の前に現れる動詞と助動詞の bigram の一覧

	p_value	夏目漱石	自然主義	頻出
いる/動詞-た/助動詞-、/記号	5.25E-05	0.013555137	0.027743715	自然主義
来る/動詞-た/助動詞-、/記号	7.00E-05	0.006268794	0.012825298	自然主義
く/動詞-た/助動詞-。/記号	0.000104036	0.000593513	0.005781837	自然主義
言う/動詞-た/助動詞-。/記号	0.000145431	0.003907352	0.017131177	自然主義
云う/動詞-た/助動詞-。/記号	0.000184235	0.014887286	8.62E-05	夏目漱石
構う/動詞-ない/助動詞-。/記号	0.000353739	0.001244882	0	夏目漱石
分る/動詞-ない/助動詞-。/記号	0.000362231	0.003177718	0.000108153	夏目漱石
来る/動詞-た/助動詞-。/記号	0.000498425	0.013962771	0.02612438	自然主義
云う/動詞-だ/助動詞-、/記号	0.000548794	0.00226676	0.000204331	夏目漱石
云う/動詞-た/助動詞-、/記号	0.000849247	0.004558241	1.46E-05	夏目漱石
開く/動詞-た/助動詞-。/記号	0.000907363	0.001074074	0.000179217	夏目漱石
出す/動詞-た/助動詞-。/記号	0.000929745	0.009405036	0.003647441	夏目漱石
落ちる/動詞-た/助動詞-。/記号	0.000967913	0.002397386	0.00037311	夏目漱石
はじめる/動詞-た/助動詞-。/記号	0.001015259	0.000127176	0.001379638	自然主義
歩く/動詞-た/助動詞-。/記号	0.00103254	0.000869032	0.00282339	自然主義
分る/動詞-ない/助動詞-、/記号	0.001049251	0.001585971	4.03E-06	夏目漱石
答える/動詞-た/助動詞-。/記号	0.001077098	0.00712741	0.002364155	夏目漱石
問う/動詞-た/助動詞-。/記号	0.00118104	9.30E-05	0.001768437	自然主義
言い出す/動詞-た/助動詞-。/記号	0.001405648	1.43E-05	0.000794251	自然主義
気がつく/動詞-た/助動詞-。/記号	0.001545457	0.001879223	0.00018979	夏目漱石
やって来る/動詞-た/助動詞-、/記号	0.001651785	2.32E-05	0.000227358	自然主義
答える/動詞-た/助動詞-、/記号	0.001652447	0.001091286	0.000183602	夏目漱石
言う/動詞-ない/助動詞-、/記号	0.001852131	1.20E-05	0.001247688	自然主義
始める/動詞-た/助動詞-。/記号	0.001881692	0.003641179	0.001263413	夏目漱石
込む/動詞-だ/助動詞-、/記号	0.002024368	0.000399897	8.48E-06	夏目漱石
なる/動詞-た/助動詞-。/記号	0.002216927	0.019068107	0.010704367	夏目漱石
ある/動詞-た/助動詞-、/記号	0.002270618	0.002262436	0.005900462	自然主義
動く/動詞-ない/助動詞-。/記号	0.002493365	0.001017138	0	夏目漱石
知れる/動詞-ない/助動詞-。/記号	0.002595407	0.005169566	0.001509416	夏目漱石
送る/動詞-た/助動詞-。/記号	0.002987179	0.0001345	0.001036372	自然主義

定である。本研究では t 検定における p 値を特徴的な変数であることをあらかず特徴度とした。このような変数の抽出に際し、「夏目漱石作品」と「自然主義文学作品」という 2 つの群に分け、上述の動詞と助動詞の bigram の出現率について夏目漱石作品における平均値及び自然主義文学作品における平均値を求め、これについて検定を行った。これを本研究で用いた小説に現れるすべての単語について行い特徴度の推定、及び特徴的な変数を抽出した。表 2 は検定の結果であり、p 値

の小さい 30 変数を示した。表 2 において網がけされている変数は打ち消しの助動詞が含まれる変数であり、表 2 に示した結果からこれらは夏目漱石の作品に多く現れていることが推測される。特に、夏目漱石の小説においては句点の直前、すなわち文末に動詞と打ち消しの助動詞の bigram が特徴的に出現していると考えられる。

次に、文末に現れる動詞と助動詞の bigram において「ない」や「ぬ」といった打ち消しの助動詞が含まれるものについて調査した。図 7 は本

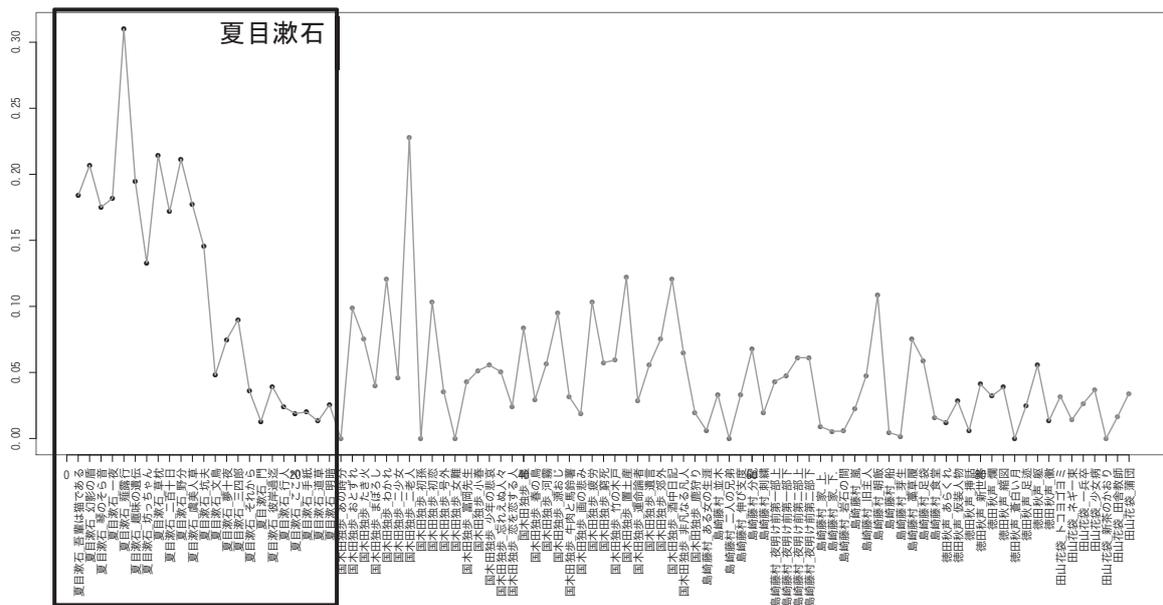


図 7 文末において打ち消しの助動詞を伴う動詞の出現率

研究で用いた 93 作品における出現頻度上位 50 語の動詞を対象とした、打ち消しの助動詞を伴う文末の動詞の出現状況を調査した結果である。図 7 における縦軸は、該当する bigram の累積頻度から求めた出現率である。図 7 より、夏目漱石作品において、特に夏目漱石の前期の作品において、打ち消しの助動詞を伴う文末の動詞は特徴的に出現していると考えられる。また、出現頻度上位 100 語に調査範囲を広げた場合においても図 7 と同様の結果が得られた。

このように、自然主義文学と漱石作品との間において、相違する文体的特徴は打消の助動詞を伴う文末の動詞の出現傾向であると考えられる。また、そのような動詞と助動詞の bigram の出現傾向は夏目漱石に高く、自然主義文学に低いと言える。ただし、図 7 に示したように夏目漱石の作品においても傾向が継時的に変化しており、『虞美人草』まで多く、以降は減少すると言える。

4. 結論

文末における動詞と助動詞の bigram についての主成分分析の結果から、自然主義文学と夏目漱石の作品との間における文体的相違が認められた。次いで、打ち消しの助動詞を伴う文末の動詞の出現傾向が顕著に相違する文体的特徴であると指摘できる。このような動詞と助動詞の bigram は夏目漱石の作品に多く、自然主義文学に少ないと考えられる。

今後、他の文体的特徴を用いた検討や、夏目漱石だけではなく森鷗外、谷崎潤一郎、川端康成の

小説を用いた分析が行われることによって自然主義文学という 20 世紀初頭の文壇を代表する文学的思潮における文体的量的な特徴がより多角的に明らかになると考えられる。

参考文献

- 1) ドナルド・キーン, 徳岡孝夫訳: 日本文学史近代・現代篇 二, 中央公論新社 (2011)
- 2) 金明哲: 読点の打ち方と文章の分類, 計量国語学, Vol. 19, No. 7, pp. 317-330 (1994).
- 3) 金明哲. (2002). 「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, pp. 225-140.
- 4) 金明哲: 文章の執筆時期の推定——芥川龍之介の作品を例として——, 行動計量学, 36(2), 89-103. (2009)