

博士論文

近代以降の日本小説の文体変化に関する
計量的研究

文化情報学研究科文化情報学専攻博士後期課程

1408171001

李広微

目次

第1章 研究の背景と目的	5
1.1 はじめに	5
1.2 研究の目的	6
1.3 社会的文体の経時的変化	7
1.4 個人文体の意図的变化	8
1.4.1 文体の意図的变化とそれに関する研究	8
1.4.2 水村美苗とその作品	9
1.5 論文の構成	11
第2章 分析対象作品・分析項目・分析方法	13
2.1 コーパスの構築	13
2.1.1 日本語文章の通時コーパスについて	13
2.1.2 近現代小説の通時コーパスの構築	14
2.1.3 個人文体の考察に関連するコーパスの構築	15
2.2 分析項目	16
2.3 分析手法と指標	18
2.3.1 系統樹分析	18
2.3.2 elastic net 回帰	19
2.3.3 ランダムフォレスト回帰	19
2.3.4 階層的クラスタ分析	20
2.3.5 構造的トピックモデル	20
2.3.6 主成分分析	22
2.3.7 ボックスプロット	22
2.3.8 K 特性値	22
2.3.9 比率差による特徴語の抽出	23
第3章 助詞の経時的変化に関するモデリングと分析	24
3.1 本章の背景と目的	24
3.2 助詞データの抽出	25
3.3 助詞の経時的変化についての分析	26
3.3.1 助詞の比率の経時的変化	26
3.3.2 助詞の系統樹分析	27
3.3.3 助詞のランダムフォレスト回帰分析	28

3.3.4	助詞の elastic net 回帰分析	30
3.4	考察	33
3.4.1	格助詞	34
3.4.2	そのほかの助詞	38
3.5	まとめ	39
第4章	文末表現の経時的変化に関するモデリングと分析	40
4.1	本章の背景と目的	40
4.2	文末表現に関する先行研究	41
4.3	文末表現のデータ	41
4.4	文末表現に関する分析	43
4.4.1	文末表現の多様性	43
4.4.2	文末表現の系統樹分析	44
4.4.3	文末表現のモデリング	45
4.5	考察	52
4.5.1	心理活動・思考体験の直接表出	53
4.5.2	時制変化及びその効果	53
4.5.3	未完結感を持たせる終わり方	55
4.6	まとめ	56
第5章	接続表現の経時的変化に関するモデリングと分析	57
5.1	接続表現の変化に関する先行研究	57
5.2	接続表現の集計	58
5.3	接続表現に関する分析	60
5.3.1	接続表現の使用率及び多様性の変化	60
5.3.2	接続表現の変化に関するモデリング	63
5.3.3	層別した接続表現についての多変量解析	67
5.4	考察	68
5.5	まとめ	69
第6章	個人文体の意図的变化に関する計量的分析	70
6.1	本章の背景と目的	70
6.2	回帰モデルによる水村作品の時期推定	71
6.3	『明暗』『続明暗』における言語表現の比較	72
6.3.1	文の長さ	73
6.3.2	漢字の出現率	74

6.3.3	タグ付き形態素.....	75
6.3.4	品詞及び記号の bigram.....	80
6.3.5	文節.....	83
6.4	『明暗』 『続明暗』におけるトピック構成の比較.....	86
6.4.1	トピックの数及びトピック別の遷移.....	87
6.4.2	トピックごとの特徴.....	90
6.5	まとめ.....	93
第7章	結論と今後の課題.....	95
参考文献	99
付録	105

第1章 研究の背景と目的

1.1 はじめに

文章の計量分析は、文章の用語的特徴及び構造的特徴を数値で把握・考察しようとする研究分野であり、計量文体学、コーパス言語学、計量文献学、テキストマイニング、スタイロメトリーなどの名称で呼ばれ、さまざまな方向で活用されている。文学作品の研究も例外ではない。

井口・往生・岩山（1996）は『文学を科学する』で工学（自然言語処理・人工知能・言語工学）と科学（認知科学・心理学）と文学（批評論）から文学作品にアプローチする理想を示し、当時までの成果をまとめた。この著書において、往生は小説の語りについて、表現技法、作家の個性的なものを表す特徴、文体論（stylistics）という3つの側面から分析することができる旨を指摘した。日本語の文学作品の文体計量は、早くから波多野（1935, 1965）、安本（1959）、樺島・寿岳（1965）などの研究が見られる。1935年の波多野の『文章心理学』が日本文学作品の統計的文体分析の嚆矢であり、数字を使うテキスト分析として当時「影響力はすさまじいものであった」（中村, 2016: 128）が、文の長さや品詞の使用率といったデータの記述的な分析であり、今から見ると「用いた統計手法は未熟なもの」（杉浦, 2017: 81）である。1960年代頃の文体計量分析には、推測統計や因子分析などの手法も用いられるようになったが、その時期はまだ自然言語処理の黎明期であり、コンピュータがまだ自然言語を自由に扱うことができなかった。テキストのサンプリング、言語項目のデータ集計などは、手作業に頼らざるを得なかった。

近年、自然言語処理及び人工知能技術の急速な展開に伴い、文章の計量分析が大きく発展し、質的にも大きな変化を遂げてきた。コンピュータを駆使してテキストの形態素解析や構文解析などを行い、専用ツールでデータを抽出するのが常識となっている。それと同時に、方法論に関してテキストデータを取り扱う手法が次々と開発されてくる。しかしながら、ビッグデータ時代の現在では、自然言語処理に関わる新しい技法やアルゴリズムなどは、文学作品でなく、一般的な文章に基づいて開発され、一般的な文章を対象とした分析にいち早く応用される傾向が見られる。

一方、日本語の文学作品を対象とした近年の計量的研究において、著者推定、真贋判定、作品成立過程に関する諸説の検証などが中核的な位置を占める。その代表的なものとして、語彙・語法の時間的偏りから『篁物語』の成立時代を10世紀後半と分析した安部（1996）、品詞構成率及び語彙頻度の計量的分析によって古典小説『源氏物語』の第三部の成立事情を考察した土山（2016）、江戸時代前期の井原西鶴の遺稿集の著者疑惑について計量分析を行った上阪（2016）などが挙げられる。そのほか、生活環境の変化や病気の影響で見られる作家の文体変化に関して、データ解析を通して変化時期や境界点を特定したりする研究（桑野・

金, 2008 ; 劉・金, 2017) が見られる。これらの研究では、変数とする言語項目間の相互関係からその背後に潜む文体的意味または内容や構造を解明するということより、テキストの識別や分類などのタスクに大きなウェイトが置かれている場合が多い。この点から見ると、文学の科学はまだ端緒に就いたばかりとも言える。

文学作品の分析にとって、種々の統計方法は「道具」である。中国の『論語』に「工欲善其事必先利其器」（工、其の事を善くせんと欲すれば、必ず先ず其の器を利にす）¹とあるように、強力な道具や適合性の良い手法を手に入れることが重要である。文章に向き合うための新しい統計手法やアルゴリズムを応用することで、文学作品の研究により豊かな発展を遂げることが可能である。杉浦（2017 : 94）は、「統計的に洗練された著者推定が主流になっているが、それはあくまでも著者が誰かという問題に注力するため、文学研究に利用するためにより多様な、そして文学的解釈に利用しやすい分析方法が望まれる」と指摘し、「デジタル・ヒューマニティーズや文学の統計分析は広大な未踏な領域を残している」と述べたが、実は、新しい分析手法の開発というよりもむしろ、自然言語処理及び人工知能の既存成果をいかに文学作品の特性に合わせて活用するかが当面の大きな課題と考えられる。

1.2 研究の目的

本論文では、近現代日本語小説を分析対象とし、文体の変化を中心にして、統計的分析法及び機械学習法のアプローチで考察を試みる。

文体は社会的文体と個人的文体に大別できる。社会的文体は、時代の変遷に伴って言語と共に変化している。また、国による国語政策、国民の文体に対する嗜好の時代性などが社会的文体の変化を後押ししている可能性もある。一方、個人的文体の変化には、年齢・環境・病気などに伴う自然的変化や、意図的变化などが挙げられる。個人文体の自然的変化に関する計量的研究として、芥川龍之介の作品における経年的変化の分析（金, 2009）と宇野浩二の文体変化の考察（劉・金, 2017）などがある。個人文体の意図的变化に関して、計量的手法で著者識別を行われる研究が存在しているが、作品の成立過程及び構造的特徴などに着目する計量分析が見当たらない。本論文は社会的文体の経時的変化と個人文体の意図的变化に焦点を当てる。

社会的文体の変化に関しては、口語体が確立されてから日本の近現代小説の文体にどのような変化が起こっているかに着目し、文体に関わる言語項目の使用様態及び変化をモデリング手法で捉えることを試みる。さらに、モデリングの結果を踏まえ、特徴的な変動要素を特定し、言語学や文体学の研究に新視点の発見に繋がるものを提示する。

¹ 訳語は『鑑賞中国の古典第2巻 論語』（加地伸行・宇佐美一博・湯浅邦弘、角川書店、1987 : 378）に拠る。

一方、個人文体に関しては、現代作家の水村美苗が近代文学への憧れを原点として創作した一連の作品は、近現代文学の変化の流れにおいてどのように位置づけられるかについて、本研究で得られた近現代小説の経時的変化に関するモデルを利用して考察を行う。さらに、水村が夏目漱石の文体と語りの特色に合わせて書いた『続明暗』に注目し、模倣作と原作における言語表現と物語構成の異同、水村の意図的文体変化について分析する。

1.3 社会的文体の経時的変化

文章の表現様式は時代によって異なっている。日本語の文章の文体は複雑な変遷を経てきた。古墳時代には、漢字漢文が大陸から日本にもたらされたとともに、漢字を用いて日本語の文章が書けるようになった。漢文による表現は日本語本来の表現との差異が大きくて理解されにくいいため、日本語の姿に近く変更されつつある。奈良時代に書かれた『古事記』は、変体漢文で書かれたものであり、漢文の構造が多く見られる。『古事記』に載る歌謡の部分は、日本語の音を漢字で表記した万葉仮名で書かれたのである。平安時代には、万葉仮名に続き、漢字を簡略化した表音的な片仮名、万葉仮名の草書体から生まれた平仮名が現れた。『古今和歌集』『土佐日記』『蜻蛉日記』『竹取物語』『源氏物語』など、仮名を主体とした仮名文による文学が花開くことになった。貴族の話し言葉に近い形で書かれた仮名文の発展につれて、変体漢文と仮名文の特徴を融合した和漢混淆文が成立した。中世の説話文学や近世の小説などを含め、中世以降の書き言葉には、当時の話し言葉から離れた和漢混淆文で記述されたものが多く存在する。室町時代になると、狂言の台本や抄物など、意識的に話し言葉に近い表現を取り入れた文献資料が見られる。その系譜は、江戸時代の近松門左衛門の浄瑠璃作品、洒落本・人情本・滑稽本などにまで及んでいる。明治時代に至って、言文一致活動が起こり、誰にでも読み書きできる口語体の文章形式が形成されるようになった。

社会的文体の経時的変化に関して、質的記述的手法に基づいた研究が昔から行われている。例えば、山本（1965）は、大量の文献を整理したうえ、言文一致体の歴史的な展開及び時期区分について分析を展開した。乾（2017）は、変体漢文から仮名の成立、和漢混淆文までの歴史を、万葉集や古事記を中心にして考察した。一方、電子コーパスの整備に伴い、データベースを利用した探索的な調査や分析が増えてきた。田中（2013）は、国立国語研究所が公開した『太陽コーパス』を用いて、文末語や助動詞の変化、語彙の数量増減及意味変化など、近代書き言葉の変遷をたどった。小説の文体の変化に特定した研究として、小西（2016）は、近代口語文翻訳小説を対象としてコーパスを構築し、基本統計量で文の長さ、文節の長さや品詞などのデータを算出し、明治中期の翻訳小説と近代口語文体確立期の創作小説との関係を分析した。また、宮島（2019）は、近代の小説が文語から口語へどのように変わっていつ

たかをめぐって、『明治文学全集』（筑摩書房）の小説をもとに、各時期の口語化の度合いを概観した。

時代を異にする文体の特徴に関する従来の研究は、明治期の言文一致運動までの変遷に着目している。言語一致運動から百数十年を経た現在に使われている書き言葉は、いまだ口語体の延長線に位置するが、この百数十年の期間において1つひとつの細かい変化が集積していると考えられる。しかし、口語体が確立されてから文体にどのような変化が起こっているのかに関して体系的な研究は見られていない。また、計量手法から見ると、先行研究では基本統計量を中心に用いるものが多かった。集合学習による回帰・分類などは、日本語文章の書き手の同定（金・村上, 2007）、文学作品における発話の教師なし話者の同定（遠田, 2019）などに用いられ、文体や文章様式の特徴を識別したり、テキストを分類したりするのに有効であることが実証されているが、日本語文章の社会的文体の構造や変化に関する分析への応用はほとんど見られていない。

1.4 個人文体の意図的变化

1.4.1 文体の意図的变化とそれに関する研究

個人文体の意図的变化は、大きく2つのケースが考えられる。1つは単に自分の文体を変えたいという意識で文体を変えるケースである。三島由紀夫の文体改造論はその例に挙げられる。三島（1956）は、「文体そのものが、私の意志や憧れや、自己改造の試みから出ている」と主張し（2016:203）、文体の自己改造によって作家としてのあり方を更新しようと試みていた。三島の文体改造に関して、西本（2010）は、作品のどのような部分に実質的に現れているかについて計量分析を行った。文体の意図的变化のもう1つは、他人の文体を真似して、自分の元の文体から対象となる文体に合わせて変化させることであり、つまり、文体模倣である。

模倣を文学ジャンルとして理論的に定義した研究（Genette, 1982）が既にあった。模倣を理論から考察したものとして、Hutcheon（1985）の *A Theory of Parody* などが見られる。また、小説の作法として、模倣を説明するものが存在する（大江, 1993；大塚, 2003）。そのほか、具体的な作品を取り上げて、間テキスト性などの視点から模倣作品と対象作品を比較する研究も少なくない。しかしながら、模倣作品の創作・受容過程及び構造特徴などに関して、未ださまざまな問題が十分に究明されていない。例えば、作品の模倣はどのように成立するのか、なぜ読者は似ている印象を受けたり、作品の雰囲気や形態の類似性を認識できたりするか、書き手はどのような要素に注目し、どこまで似せているか等々が挙げられる。

作品模倣に関する従来の先行研究は、主に、内省に基づく理論的・記述的観点に拠って分析を行った。情報技術が進んでいる現在、データをもとにした論理的、科学的な計量手法を

含め、複数のアプローチによる探索が実現可能となる。新しい分析手法を作品模倣の研究に導入することにより、文体のあり方への再認識を促し、文体・物語の研究、文章作成の実践及び教育などに有益な情報を提供することが期待できる。

例えば、文体の核を成すものは何であるか、個人的な文体がどのように形成され、顕現するのかについて異なる角度から認識することが可能である。小林(1968)によると、文体は、美的効果を発揮する言語要素が集まってできる構造であると定義された。しかし、美的効果を持つ言語要素は一体どういうものであるかは、明確にされていない。個性的な文体或いは美的効果のあり様を直接に浮かび上がらせることは困難であるが、観察或いは計量できるところに対して、データを作り、一定の方法で分析しながら、模倣作品と模倣対象作品との間の関係を探り、解明することで、ある程度に2者に共通する言語特徴を特定できる。個性的な文体を形成する言語要素を結び合わせることで、それらによって作りされた文体の全体像及び美的効果のあり様の理解を深めることができると考えられる。

また、模倣作についての研究から得られたデータまたは知見は、文体教育に利用することができるだけでなく、物語の自動生成及び創作支援などに関連するタスクに応用することが可能であると考えられる。近年、人工知能分野では、プロットの自動生成の研究(Pradyumna et al., 2018 ; 小方ら, 2018)が進められている。文章生成器に物語の基本的な構造を与えると、文章生成器が物語の生成文法に従って、自動的にストーリーを生成してくれるということが実現できるようになった。日本では、2016年に人工知能小説による小説が文学賞「星新一賞」の一次選考を通過したという報告(佐藤, 2016)もあった。しかし、現在に至っても、面白いとか、情緒的とか、村上春樹っぽいとかなどのような指令で文章生成のプログラムに小説を書かせることは実現されていない。人工知能の分野では、人の感情の理解と感情への働きかけなどの問題を如何に対処すべきかについて、文体模倣の行為及び模倣作品から示唆を得る可能性がある。

1.4.2 水村美苗とその作品

近現代小説の中では、他人の文体への模倣で好評された作品として、水村美苗の『続明暗』が代表的な例として挙げられる。

水村美苗(1951～)は小説家・評論家として90年代から活躍している。彼女は12歳に父親の仕事の関係でアメリカへ移住した。当初アメリカの生活に馴染めず、読書に没頭し、父親が日本から持ってきた改造社版『現代日本文学全集』(1926年版, 全63巻)を読むことによって、母国への恋しさ及び孤独を埋めていた。イエール大学の博士課程(仏文専攻)を修了した後、20年程ぶり(1980年代)に日本へ一時帰国した。その時、それまで読んできた日本の小説とは似ても似つかないものがもてはやされていることに衝撃を受け、日本語が日本

人の精神世界や情緒を豊かに表現する書き言葉としての輝きを失っていくという危機を感じとった。のちに再び渡米して、プリンストン大学で日本近代文学を教えながら、創作活動を行う。現在も小説の創作活動を続けている。

水村が最初に発表した小説『続明暗』は、夏目漱石の未完に終わった長編小説『明暗』の結末を描いたものである。夏目漱石（1867～1916）の遺作『明暗』は、1916年5月から同年12月まで『朝日新聞』に連載されたが、作者の病没によって第188回で中断され、未完のまま閉じられた。未完でありながら、三十余万字で漱石作品として最大の長編となり、「数少ない真の近代小説の1つである」（江藤，1974）とも評されている。そのあらすじはこうである。会社員の津田は上司の奥さん（吉川夫人）の紹介でお延と半年ほど前に結婚し、東京で暮らしている。二人の夫婦関係は目立った問題がなかったが、どこかぎくしゃくしている。実は、結婚以前、津田は吉川夫人の紹介で清子という女性と付き合っていたが、清子は突然津田の元を去り、津田の旧友に嫁いでしまった。そんな過去が津田の心にずっと引きずっており、結婚生活に影を落としている。吉川夫人は、痔の手術を受けた津田に、清子が温泉で療養していることを伝え、二人に会わせようとする。津田が夫人の勧めを受け入れ、妻に内緒で温泉場に向って、清子と再会したところで物語は途切れてしまった。柄谷（2013：678）が「これが未完結であることは、読むものを残念がらせ、その先を想像させずにおかない」と言ったように、温泉場での二人はどうなるか、妻のお延はどんな行動をとるか等々は、謎として残され、読者の関心・想像を掻き立てる。江藤（1968）、大岡（1988）、大江（1990）をはじめとする多くの作家、批評家は、その結末を推測し、エッセイや論文で発表している。

1990年に、当時新進の作家だった水村美苗は漱石の文体に合わせて、『明暗』の新聞連載番号を引き継いで、第189～288回からなる『続明暗』を書きあげ、物語を終わらせた。この続編は、漱石の独自の文体をよく真似ていることで高く評価され、1990年の芸術選奨新人賞の受賞作品に選ばれた。小説家の辻邦生は『続明暗』を「信じがたいロマネスクの離れ業と文体的パロディーの才で、書きついだもの」（辻，2005：25）と評した。また、文学評論家の安原顯は「（漱石独自の当て字や文体模写などの）制約のすべてをもの見事にクリアーし、ある面では漱石のオリジナルより感動的な小説にしてしまった」（安原，1990：224）と称賛し、国語学者の山田潔も「この作品を読了した時の衝撃はいまだに記憶に残る。その衝撃の内容は作品としての説得力もさることながら、文体が漱石のそれに酷似していることであった。それは、用字・語彙・語法をはじめ、漱石の語用論のすべてに及ぶ」（山田，2001：48）と評した。水村の『続明暗』は、あらゆる点で漱石の文体に酷似していると高い評価を受けているが、この2つの作品の文体の類似性と相違点などについて、従来の評論や鑑賞は印象批評に基づいて評価するに止まっており、客観的かつ体系的に捉える視点が欠落している。

『続明暗』に続き、水村は『私小説 from left to right』（1995）、『本格小説』（2002）及び『母の遺産』（2012）といった一連の長編小説を創作した。第2作の『私小説 from left to right』は、日本語に英語を交ぜた横書きの文体を用いて書かれた自伝風作品であり、野間文芸新人賞を受賞した。第3作の『本格小説』は、エミリー・ブロンテの『嵐が丘』を下敷に、戦後日本を舞台に書き換えた恋愛小説であり、読売文学賞を受賞した。また、実体験を下敷きにした『母の遺産』は大佛次郎賞を受賞した。

『私小説 from left to right』と『本格小説』はサブタイトルが「日本近代文学」である。これについて、水村（2009）は、もう一度漱石の時代に戻ってみようというメッセージが込められていると自らを説明している。一方、『母の遺産』には「新聞小説」というサブタイトルが付けられている。この小説自体が新聞に掲載されたものであると同時に、百余年前の尾崎紅葉の新聞連載小説『金色夜叉』が絶えず意識され、主人公たちの生き方にも『金色夜叉』的な要素が盛り込まれている。水村（2012）は『母の遺産』の創作について、「（新聞小説の）過去の栄光を確認し、その終わりを記すため“白鳥の歌”のようなものにしようと思いました」と述べている。近代の文学を理想的な国語文学としている水村が創作した作品は、近現代小説の変化の流れの中でどのように位置づけられるのか。また、社会的文体の変化の流れに逆行しようとする作家個人による文体づくりの試みはどこまで到達できるかとの疑問が浮上する。

1.5 論文の構成

本論文は7章から構成されている。第1章では、本研究の背景及び目的・構成について述べる。第2章では、研究に使われるコーパスの構築、分析項目及び分析手法について説明する。第3章から第5章は、近現代小説の文体に関わる、言語項目のデータを元に、その変化と構造をモデリングし、考察する。第3章では、助詞の使用に関する情報に基づいて、回帰モデルを通して、その変動状況を顕現させると同時に、特徴的な変動要素を特定し、小説の文体の変化との関わりを考察する。第4章では、文末表現の使用に関する情報に基づいて、同じく回帰モデリングを行い、その遷移の傾向から、小説の語り方や技法に関する変化を分析する。第5章では、接続表現の使用に関する情報に基づいて、その多様性の変化を分析したうえ、回帰モデル及びトピックモデルによるモデリングを行い、変化の特徴を考察する。第6章では、作家個人による文体の意図的变化に着目して分析を行う。まず、第3～5章で得られた回帰モデルを利用して、作家水村美苗が近代文学への憧れを原点として創作した『続明暗』等一連の作品が、近現代文学の変化の流れにおいてどのように位置づけられるかについて分析する。次に、水村が自分の文体を変化させ、夏目漱石の文体と語りの特色に合わせて書いた『続明暗』を取り上げ、模倣作と原作における言語表現及び物語展開のパターンの

異同，水村の意図的文体変化の特徴について分析する．第 7 章では，本研究で得られた結果及び今後の課題を要約する．

第2章 分析対象作品・分析項目・分析方法

本章では、分析に必要なコーパスの構築、分析項目を示し、全体の研究の流れ及び分析方法について説明する。

2.1 コーパスの構築

2.1.1 日本語文章の通時コーパスについて

文章の文体における経時的变化を捉えるには、通時コーパスが必要である。日本語文章の通時コーパスとして、国立国語研究所が日本語史研究の基礎資料として開発を進めている『日本語歴史コーパス』は代表的なものであり、奈良時代から明治・大正時代の広範な文献を納めている。表1に示されるように、明治・大正期間の文献は、2021年現在、雑誌・教科書・明治初期口語資料・近代小説（21編）という4つの部分から構成されており、小説の数は限られている。

表 2.1 『日本語歴史コーパス』の公開内容（2021年12月時点）

奈良時代編	『万葉集』全編，『続日本記』の62編の宣命，『延喜記』巻第八の27編の祝詞
平安時代編	『古今和歌集』『竹取物語』『枕草子』『源氏物語』などに代表される16編の仮名文学作品
鎌倉時代編	説話・随筆作品5作品，日記・紀行作品5作品
室町時代編	『虎明本狂言集』『天草版平家物語』『天草版伊曾保物語』
江戸時代編	洒落本30作品，人情本8作品，松の浄瑠璃24作品，芭蕉の紀行文6作品
明治・大正編	『明六雑誌』『太陽』『女学雑誌』などに代表される8種類の雑誌データ，小学校・高等小学校で使用された国語教科書のデータ，明治初期刊行の口語体資料11作品，明治中期から大正末期の小説21作品（21著者）
和歌編	平安時代から鎌倉時代にかけて編纂された，八代集と称される和歌集のデータ

同じく国立国語研究所が構築を進めている『現代日本語書き言葉均衡コーパス』(BCCWJ)は、書籍全般、雑誌全般、新聞、白書、ブログ、ネット掲示板、教科書、法律などのジャンルにまたがって書き言葉の資料を格納しているが、現時点ではその収録対象の刊行年代は、最

大 30 年間（1976～2005）であり、メインとなる書籍の場合は、1986 から 2005 年になっている。

そのほか、小説などの文学作品を中心に扱う電子コーパスはいくつか公開されている。そのなかで、青空文庫は規模が一番大きく、古今東西の 16815 編（2021 年 12 月時点）の作品を収録している。しかし、著作権保護の制限により、公開している電子コーパスには、作者の死後 50 年を経るなどして著作権の消滅した作品がほとんどである。

このように、20 世紀初期から近年までの日本小説を収録し、公開された通時コーパスは未だ見つからない。

2.1.2 近現代小説の通時コーパスの構築

山本（1965）では、言文一致運動の流れについて、慶応 2 年～昭和 21 年（1866～1946）までの時期を文学系統の資料により 7 期に分けている。そのなか、明治 33 年～明治 42 年（1900～1909）は口語文体の確立期、それ以降は成長・完成期と見なされる。本研究では口語体が確立された後の文章を中心に考察を行うため、1910 年以降の小説を主な分析対象とする。

1910 年から 2014 年までの間に出版した小説を分析対象として集めて、コーパスを作成した。1 年ごとに 5～6 作品をサンプリングし、作家個人の文体の影響を避けるため、1 人につき 1 作品を使用することにした。

作品を選定する際に、基本的には以下の 2 点を指標とし、そのいずれかに該当するものを採用する。1 つは、ある時期の代表的な小説として文学選集に収録されていることである。

『現代日本文学全集』（筑摩書房、1953～1959）、『日本短篇文学全集』（筑摩書房、1968～1970）、『昭和文学全集』（小学館、1986～1990）、『戦後短篇小説選』（岩波書店、2000）などの叢書を主に参考とした。1950 年代以降の作品に関してはこのような文学選集にカバーされるものが少ないため、文学賞の受賞作をもう 1 つの指標とする。主に、芥川龍之介文学賞や直木三十五賞の受賞作から選出して用いた。芥川賞や直木賞に限らず、ほかの文学賞の受賞作の一部も使用した。例えば、吉本ばなの「キッチン」（第 6 回海燕新人文学賞を受賞）は今や世界 35 ヶ国で翻訳されており、広く世界中に知られている。このような作品も近現代小説の代表作としてコーパスに取り入れた。一方、この 2 つの指標で 1 年に 5 作品を集めにくい場合があり、この時、芥川賞などの受賞作及び候補作の掲載歴が多い文芸誌（『文学界』『群像』『すばる』『新潮』）から作品を選んで使用した。

コーパス構築にあたって、紙媒体から光学読み取り機器で電子化作業を行った。作成したコーパスには 555 名の作家の作品が含まれている。本研究の分析では、時代の流れで見られ

る大きな傾向を掴むことを主眼とするため、個人の作品における差を考慮しないこととする。コーパスのリストは本論文の付録に示す。

テキストの長さのバランスを保つため、5千字を目安に長さを統一する。つまり、5千字以下の短編小説は全文を使用し、5千字を超えた場合、5千字を目安として、文が完了するところまで切り取って用いた。対象テキストに対し、括弧やダッシュなどの記号で標記された会話文を削除し、地の文を分析の対象とした。

2.1.3 個人文体の考察に関連するコーパスの構築

個人文体の意図的变化の考察に関しては、夏目漱石の長編小説『明暗』及び水村美苗が漱石の文体を模倣して書いた『続明暗』を中心に分析を行う。主要の分析対象である『明暗』『続明暗』を電子化し、分析に用いる。『明暗』は『漱石全集』（岩波書店、1966年）の第7巻に収録されたもの、『続明暗』は初版（筑摩書房、1990）を使用する。

『続明暗』の文体と『明暗』との類似度を顕在化させるため、ほかの5名の作家が書いた小説を比較作品として、コーパスに取り入れる。比較対象とする5名の作家は泉鏡花（1873～1939）、永井荷風（1879～1959）、森鷗外（1862～1922）、島崎藤村（1872～1943）、幸田露伴（1867～1947）にする。この5人を選択する主な理由は、夏目漱石とおおよそ同年代に活躍していた作家であること、水村の読書ないし文学創作に接点があることである。水村（1991）自身の話によれば、彼女は12歳の時にアメリカへ連れられた時、海外の生活に溶け込めず、読書に没頭し、父親が日本からアメリカに持っていた昭和二年発行の改造社版の『日本現代文学全集』（全63巻）ばかりを読んでいた。そして、『続明暗』を執筆する時、漱石の小説になかった場面を描く際、泉鏡花及び永井荷風などの作品を参考にしていただことである。今回の分析は、5名の近代作家による50作（5×10）の短編小説を比較作品として用いる。

『続明暗』は模倣を意識して書かれたものであるが、水村自身の文体の痕跡が残されている可能性があると考えられる。これを確認するため、水村が書いたほかの小説もコーパスに取り入れ、併せて考察する。水村が書いた小説は、『続明暗』のほか、『私小説 from left to right』（1995）、『本格小説』（2002）、『新聞小説 母の遺産』（2012）という3篇の長編小説がある。そのなかで、『私小説 from left to right』は日本語と英語の2言語で書かれたものなので、比較に用いないことにする。

対象テキストを全部新仮名・新字に統一し、会話文を削除し、地の文のみを使用する。コーパスのリストを表2.2に示す。

表 2.2 コーパスリスト

作者	作品
夏目漱石	明暗
水村美苗	続明暗, 母の遺産, 本格小説
森鷗外	じいさんばあさん, 寒山拾得, 阿部一族, 心中, 普請中, 最後の一句, 山椒大夫, 沈黙の塔, 木精, あそび
泉鏡花	露肆, 草迷宮, 卵塔場の天女, 瓜の涙, 妖術, 春昼, 歌行燈, 茸の舞姫, 伯爵の釵, 七宝の柱
島崎藤村	芽生, 岩石の間, 船, 分配, 三人, 刺繍, 食堂, 伸び支度, 並木, 熱海土産
永井荷風	すみだ川, 裸体, ひかげの花, 雨瀟瀟, 或夜, 老人, 寐顔, にぎり飯, 羊羹, 勲章
幸田露伴	観画談, 雁坂越, 名工出世譚, 雪たたき, 太郎坊, 魔法修行者, 平将門, 連環記, 蘆声, 鷺鳥

2.2 分析項目

社会的文体は、「超個人的」ものとして捉えられ、基本的に文章の外的な形態に関わるもののみが注目される。社会的文体の経時的変化に関する考察は、論文・著書ごとに異なった構想の下に記述されているが、基本的には、文字（表記法）、語彙、文法、音韻などの要素を取り上げて、比較・分析を行っている。文体はさまざまな言語項目に内蔵されており、その全貌を描き出すことは容易にできるものではなく、異なる構成要素を単位として文体の歴史を多層的に検討することが必要となる。本研究では、文体表現の経時的変化の考察にあたり、助詞、文末表現、接続表現を分析項目として取り上げる。これらの項目の選択理由及び背景について、後続の各章で述べることにする。

個人的文体は、作者の個性や人格を表し出すものであり、人によってそれぞれに異なるのである。個人的文体をめぐる有効な特徴項目として、小林（1975）は、自身の研究を踏まえ、テーマ・構成、構文法、語彙、品詞、リズム、テンポを挙げている。小林が挙げている項目を参考にし、『明暗』『続明暗』との比較分析において、文の長さ、漢字の出現率、形態素、品詞、文節パターンなどといった計量的に分析可能な文体項目を抽出して分析を行う。また、文体論には、語学的文体論と展開文章論が含まれている。後者の場合、「主題の細叙・布衍・変形を経て進行する作品の展開を辿る」（中村、1993:123）ことを課題とする。本研究では、個人的文体に関して、言語表現の特性および文章展開の様相から分析を進める。

① 文の長さ

文の長さは、1文に何文字を使用しているのかの数値であり、文体の特徴を示す指標の1つとしてよく使われている。波多野（1950）は文章心理学の観点から平均文長と作家の創作態度や性格との結びつきについて考察し、安本（1958）と佐々木（1976）は文の長さの分布の正規性について、樺島（1990）は同一文章における文の長さの変化について計量分析を行った。本研究では、文の長さについて、平均文長及び文の長さの分布を合わせて分析する。

平均文長は、テキストファイルごとに総文字数を文の数で割って平均文長を算出する。文の長さの分布については、テキストファイルごとに対し、文字数を5つずつ増やしながら、各長さの文の数を集計する。すなわち、文字数が1~5の文、文字数が6~10の文、といったように分けて数える。

② 漢字の出現率

漢字の出現率は、各テキストでの漢字の総数を総文字数で割ることによって計算する。

③ タグ付き形態素

形態素は、意味を持つ表現要素の最小単位である。「津田はよく眠れなかった」という文を例に、MeCab 形態素解析器で解析をすると、「津田_人名」「は_係助詞」「よく_副詞」「眠れ_動詞」「なかつ_助動詞」「た_助動詞」のような6個の形態素が抽出される。

③ 品詞・記号の bigram

形態素解析器を用いてテキストを単語・記号に区切り、品詞タグ及び記号タグを付与し、その bigram データを集計する。n-gram は、文字・音素・単語などを単位とし、隣接している n 個を1つの組としたものである(n=1 は unigram, n=2 は bigram, n=3 は trigram)。金（2004）は日本語の品詞の n-gram を用いて、書き手の識別を行い、その有効性を示した。

④ 文節パターン

構文について、従来の文体研究では、主辞の内頭と外頭、主辞・述辞の正常辞列（S・P型）と異常辞列（P・S型）などを取り上げ、分析するものが多く見られる。本研究では、テキスト計量分析でよく使われている「文節パターン」を利用し、構文の構成要素について分析を実施する。文節パターンとは、構文分析の基本単位となる文節を、形態素タグを通してモデル化したものである。内容との依存関係が大きく抑えられるため、文節パターンは書き手識別特徴量として有効であることが金（2013）の研究で実証された。文節パターンの抽出について、例文「彼はまた考えつづけた。」を CaboCha で文節に切り分けると、表 2.3 のような構文情報が得られる。構文情報はいくつかの層に分かれているが、第1層の情報のみを使えば、この例文から、「名詞_助詞」「副詞」「動詞_動詞_助動詞_記号」のような3つの文節パターンが抽出される。本研究では、文節内の助詞及び記号は原型を用い、それ以外は形態素のタグを用いる。

表 2.3 構文解析の例 (CaboCha)

*0 2D 0/1-2.242775			
彼	名詞	代名詞	一般
は	助詞	係助詞	*
*1 2D 0/0-2.242775			
また	副詞	助詞類接続	*
*2-1D 0/2 0.000000			
考え	動詞	自立	*
つづけ	動詞	非自立	*
た	助動詞	*	*
。	記号	句点	*

2.3 分析手法と指標

本研究で用いた分析手法は、系統樹分析、ランダムフォレスト回帰分析、elastic net 回帰分析、構造的トピックモデル、主成分分析、ボックスプロット、階層的クラスター分析である。そのほか、語彙の豊富さの指標である K 特性値なども分析に利用する。

2.3.1 系統樹分析

系統樹とは、言語や生物といった対象の集合の要素に対して特徴量のベクトルを抽出し、それを用いて、関連性を樹木の枝分岐の形式に示したものである。系統樹の生成手法は多く提案されており、距離行列法がその代表的な1つである。距離行列法とは、個体間の距離を何らかの形で定義して、その距離行列を基に系統樹を推定することである。

距離行列に基づいた系統樹の生成法は複数存在しているが、本研究では、最も広く利用されている近隣結合法 (neighbor joining, NJ法) を利用する。近隣結合法は、距離行列か複数の要素の近隣を段階的に見出し、最終的に全ての要素が連結された系統樹を得る方法である。距離行列の計算について、本研究では Jensen–Shannon Divergence (JSD) の平方根を用いる。文章 (テキスト) a で集計した語などのベクトルを $T_a = (x_{a1}, x_{a2}, \dots, x_{am})$ 、文章 (テキスト) b で集計した語などのベクトルを $T_b = (x_{b1}, x_{b2}, \dots, x_{bm})$ としたとき、JSD の平方根 $RJSD(T_a, T_b)$ は次のように定義される。ここでは m は変数の数、 $\sum_{i=1}^m x_{ai} = 1$ 、 $\sum_{i=1}^m x_{bi} = 1$ である。

$$RJSD(T_a, T_b) = \left\{ \frac{1}{2} \sum_{i=1}^m \left(x_{ai} \log \frac{2x_{ai}}{x_{ai} + x_{bi}} + x_{bi} \log \frac{2x_{bi}}{x_{ai} + x_{bi}} \right) \right\}^{\frac{1}{2}}$$

2.3.2 elastic net 回帰

回帰分析は、「現象の結果とそれに影響をおよぼすと考えられる複数の要因と結びつけるモデル」（小西，2019：2）である．要因となる変数を「説明変数」，結果となる変数を「被説明変数」とする．そして，説明変数が1つの場合は単回帰分析，説明変数が複数の場合は「重回帰分析」と呼ぶ．重回帰分析では，観測データを下記の式で表す．式の中で， x_q ($q=1, 2, 3, \dots$)は説明変数， β_0 は定数項， β_q は回帰係数， ε は誤差項を表す．

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_q x_q + \varepsilon$$

従来の回帰分析において，残差の二乗和の最小化または尤度最大化によって回帰変数を求める手法が適用されている．しかしながら，説明変数間に相関がある場合，過学習や多重共線性などの問題が起こりやすい．諸問題の対策として，正則化などの方法が提案されてきた．正則化には，変数選択と次元削減を特徴とした L1 正則化，係数の縮小による過学習抑制を特徴とした L2 正則化がある．

本研究では，線形回帰分析として，L1 正則化と L2 正則化を発展させた elastic net 回帰を利用する．elastic net 回帰は，Zou and Hastie（2005）によって提唱され，以下の式で与えられる．

$$L(\beta) = \frac{1}{2} \sum_{i=1}^n (y_i - \beta X_i)^2 + \lambda \sum_{j=1}^p \{\alpha \beta_j^2 + (1 - \alpha) |\beta_j|\}$$

式の中の λ は，罰則の強さを調整するパラメータである． λ については，交差検証（10-fold cross-validation）に基づき最適値を推定する．式の中の α ($0 \leq \alpha \leq 1$) は，L1 正則化と L2 正則化の混合比率を調整するパラメータである． α が 0 のときには ridge 回帰（L2 正則化）となり， α が 1 のときには lasso 回帰（L1 正則化）となる．

2.3.3 ランダムフォレスト回帰

回帰分析は線形回帰と非線形回帰に分けられる．2.3.2 の elastic net 回帰は線形回帰の 1 つである．非線形回帰としては多項式，指数関数，ランダムフォレストなどが挙げられる．本研究では非線形回帰としてランダムフォレストを用いる．ランダムフォレスト（random

forest, RF) は、大量の回帰木或いは決定木を用いたアンサンブル学習法の 1 つであり、回帰分析及び分類分析に応用されている (Breiman, 2001)。

ランダムフォレストを用いた回帰モデルの構築過程としては、まず、分析対象のデータセットから、ブートストラップサンプリング (復元抽出) の手法で B 組のサブデータセットを抽出する。各々のサブデータセットの個体数は、一般的に元データセットの個体数の $2/3$ に当たる。

次に、抽出されたサブデータセットの中から変数をサンプリングし、未剪定の回帰木を生成する。未剪定の回帰木の生成に使われる変数の数は、元データの変数の数の正の平方根にあたる数を取るのが一般的である。作成された回帰木の評価は、テストのためにとっておいた OOB (out-of-bag) データを用いる。このような処理を B 回繰り返すが、B は分析者によって決まる。本研究では 500 を用いた。

最後に、回帰問題では生成された B 個の回帰木の結果の平均を統合の結果とする。

ほかのアンサンブルメソッドと比較して、ランダムフォレストは多くの利点を持っている。特に、過剰適合が起こりにくいこと、且つ特徴量の重要度の信頼性が高いことで評価されている。

2.3.4 階層的クラスタ分析

階層的クラスタは、個体間の近さ (類似性または距離) を測り、得られた距離行列を用いて、似ているものを段階的にグルーピング (クラスタ化) する手法である。個体間の近さは、ユークリッド距離、マンハッタン距離、一致係数など、分析の目的やデータの種類によって複数提案されている。また、グルーピングの手法は、最近隣法、最遠隣法、群平均法、重心法、メディアン法、ウォード法など様々がある。

本研究では、個体間の距離は JSD の平方根によって求める。また、グルーピングする際に、グループ内の分散が小さく、かつグループ間の分散が大きい組み合わせでグループ分けをするウォード法を採用する。

2.3.5 構造的トピックモデル

トピックモデルとは、文書が生成される過程を確率的に表現したモデルである。トピックモデルでは、文書には潜在的なトピックがあり、単語はトピックごとに何らかの確率で分布されていると仮定する。トピックに集まる単語群は、実際に文単位の表層において必ずしも共起しているとは限らないが、推定可能の潜在的共起性が潜んでいると考えられる。トピックモデルではこのような単語の潜在的共起関係からトピックの所属確率を推定していく。そのため、推定された「トピック」は「潜在的意味のカテゴリー」と理解され、「潜在的トピック

ク」とも呼ばれる。トピックモデルのコンセプトを一言で述べるならば、文書と単語の関係を利用して、文書に内在する隠れた構造を抽出し、文書間の関係と特徴を考察するということである。一方、トピックモデルは最初に言語処理の数理モデルとして提案されたのであるが、文書に限らず、幅広いドメインで離散データに隠された潜在的なトピックを推定し、その傾向を探ることに応用できる。

トピックモデル分析を行うため、まず、文書に対し形態素解析処理を行い、単語に分割する。次に、個々の文書 (d) を行に、文書ごとに現れる単語 (w) の出現頻度を列に並べたベクトル (Bag of Words) を生成する。さらに、このベクトルに基づいて各文書が扱うトピックを析出するトピックモデルを構築する。トピックモデルにさまざまなヴァリエーションがあるが、本章では、時系列に沿ってトピックの変化を探ることができる構造的トピックモデル (STM: Structural Topic Model, Roberts et al., 2013) を用いる。ほかのトピックモデル、例えばトピックモデルの代表である LDA (Latent Dirichlet Allocation, Blei et al., 2003) に比べ、STM は基本的な考え方が同じであるが、2 つの顕著な特徴を持つ。1 つ目はトピック間の相関を認めること、2 つ目はトピックに関連する変数 (例えば、時間、属性、性別など) を共変量として導入することである。このため、STM はテキスト内のトピックの析出、トピックの経時的変化の考察及び説明に適したモデルと考えられる。

STM の生成過程は次の通りである (Roberts et al., 2014) 。

- ① 文書の共変量 X_d に基づいてロジスティック正規化された線形モデルを用いて、各文書におけるトピック比率を求める。

$$\vec{\theta}_d | X_d, \Sigma \sim \text{LogisticNormal}(\mu = X_d, \Sigma)$$

- ② トピック k を表す単語が個々の文書においてどのように分布しているかを求める。そのため、ベースライン単語分布 m 、トピック k における偏差 K_k 、共変量グループ g における偏差 K_g 、トピック間の相互作用 K_i が用いられる。

$$\beta_{d,k} \propto \exp(m + K_k + K_{g_d} + K_{i=(k,g_d)})$$

- ③ 文書の単語 $n = 1, 2, \dots, N_d$ に対して、当該文書のトピック比率に従って、単語ごとのトピック割り当てを求める。

$$Z_{d,n} | \vec{\theta}_d \sim \text{Multinomial}(\vec{\theta})$$

そして、選択した特定のトピックに対し、それに属する単語を抽出する。

$$W_{d,n}|z_{d,n}, \beta_{d,k=z} \sim \text{Multinomial}(\beta_{d,k=z})$$

推定されたトピックは、明確な名称が直接に得られないものであるが、研究者が各トピックを構成する単語の出現傾向から解釈して、トピックのラベルを付与することができる。

2.3.6 主成分分析

主成分分析は、多くの変数を持つデータに対する次元縮約の手法であり、データの情報の損失を抑えながら、もとのデータの変数を主成分と呼ばれる少数個の合成変数に集約する。本研究では、相関係数行列に基づく主成分分析を利用して、情報の縮約を行う。元の変数の情報を多く含む主成分から順に第1主成分、第2主成分などと呼ぶ。本研究では、第何成分まで考察するかは、平行分析を利用して、判断する。平行分析では、用いたデータと同じサイズの乱数のデータマトリックスを生成し、その相関行列の固有値を計算し、それを実データの固有値がどの程度意味を持つかを判断する指標とする。

2.3.7 ボックスプロット

ボックスプロットとは、グループ別にデータの分布状況を示すグラフ表現法であり、その形から箱ひげ図とも呼ばれる。ボックスプロットでは、数値データは四分位数に分割され、長方形(箱)の下側の辺は第1四分位数、上側の辺は第3四分位数を示し、箱の中央の線は中央値を示す。そして、ボックスの両側から伸びる破線(ひげ)の長さは、箱の長さの1.5倍以内の最小値と最大値を示す。なお、ひげの外にはみ出ているものは外れ値として印で示される。

2.3.8 K 特性値

K 特性値 (characteristic K) は、ユール (Yule, 1944) が提案した語彙の豊富さを示す指標である。K 特性値は、単語の出現頻度がポアソン分布に従うと仮定している。TTR などの指標と違って、K 特性値は延べ語数と異なり語数のみではなく、語彙の出現回数(頻度スペクトル)も考慮している。テキストにおける延べ語数が N 、異なり語数が V 、 m 回出現した単語数を $V(m, N)$ とした場合、K 特性値は下記の式で求められる。

$$K = 10^4 \frac{\sum_{all m} m^2 V(m, N) - N}{N^2}$$

式からわかるように、1000語からなるテキストで、全ての単語が異なる場合、 K 特性値は、ゼロになる。それに対し、仮に1000語が全て同じであるとしたら、 K 特性値は9990となる。つまり、 K 特性値が小さいほど使用語彙が豊富であると見なす。

2.3.9 比率差による特徴語の抽出

テキストに現れている頻度情報を用いて、特徴語句を抽出する方法は多く提案されている。本研究では、語句がテキストにおける比率の差を比較することによって、特徴的な語句を抽出する。例えば、テキストAとテキストBにおいて、ある項目の比率がそれぞれ \hat{p}_A 、 \hat{p}_B であり、テキスト内のすべての項目の合計が n_A 、 n_B であるとき、下記の式に従って z 値を求める。すべての項目に対し、求めた z 値の絶対値が大きい順に並べ替える。 $|z|$ 値が大きいほど当該項目に顕著な差があると考えられる。

$$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{p(1-p)\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$$

$$p = \frac{\hat{p}_A n_A + \hat{p}_B n_B}{n_A + n_B}$$

第3章 助詞の経時的変化に関するモデリングと分析

3.1 本章の背景と目的

日本語は、叙述すべき事柄を請け負う語彙の1つひとつに助詞の形で添えて確めながら先へ先へと進めていく表現方式を取っているという特徴を持つと言われる(森田, 2007: 18)。このような構文的性格のため、日本語文章の特性及びその変化を助詞から窺い知ることができる。山口(1982)は、いくつかの助詞と助動詞の使い方を手がかりに、古代語の再現がもたらした文体効果及び江戸文語文の書き手の尚古・擬古意識を考究した。柳父(2004)は、名詞中心の西洋式表現を倣って作られた「～ハ」構文が、どのように文の主語、句読点、文末語、漢字などに影響を与えているかを分析し、近代の「翻訳的」文体の成立過程を提示した。これらの研究で共通するのは、時枝(1949: 6)が強調した「沈潜の深さ、凝視の鋭さ」によって、一見何でもない平凡な言語現象の中に存する「変」のいくつかを嗅ぎ出すところである。見出された「変」が1つひとつの点として繋がれば、線が生まれ、面が形づくられ、そして基盤が形成された時、さらに何らかの新しい方向性が見えてくると考えられる。しかし、面に行くまで多数の「変」の点を見つけ出すことが必要であり、研究者の個人の感受力だけでは限界がある。また、数多くの変化する点のなか、文体様式に重要な役割を果たしているものもあれば、それほどでないものもあるため、特徴的な変化を見極めることが困難である。

近年では、コーパス言語分析の環境整備に伴い、データの計量分析を通して助詞の変化の傾向及び特徴を考察する研究が増えてきた。コーパスに基づいた近現代日本語の助詞についての先行研究は、大まかに2種類に分けられる。1つ目は、使用パターンの分類または語法記述・意味記述の精緻化を主な目的とする研究である(水谷, 1999; 荻野, 2006; 石川, 2017)。この類の研究では、助詞の文法特性・機能に注目することが1つの特徴である。2つ目は、異なるレジスターのテキストまたは発話における助詞の使用実態を比較する研究である。中西(2007)は、日本語学習者と日本語母語話者の書きことば・話しことばにおけるとりたて助詞の使用の違いを調査した。宮内(2012)は、12種類の接続助詞について分析を行い、これらが話し言葉的文体と書き言葉的な文体、フォーマルな文体などの違いによって使い分けられていると報告した。丸山(2015)は、格助詞について、書籍、雑誌、新聞、白書、Yahoo! ブログの6つの媒体に出現した頻度を用いて比較を行った。このように、カテゴリー別に集約した助詞の頻度を基に比較を行い、分析対象の特徴を記述する研究が行われているが、いくつかの課題が残されている。第1に、「書籍」や「雑誌」または「書き言葉」などといったカテゴリーの観点で分類しているが、そこにもさまざまな種類(若しくはジャンル)のテキストが集積されており、一様に、1つのパターンに括ることに問題がないとは言い切れない。この点は、小林・小木曾(2013)の研究に裏付けられている。小林・小木曾(2013)が

中古和文の物語文学と日記文学における助詞・助動詞の使用傾向を調査した結果によると、ジャンルによって大きな差が見られることがわかった。そのため、テキストのジャンルに応じて細分化して考察する必要があると考えられる。第 2 に、異なるカテゴリー間の比較はあったが、カテゴリー内の経時的な変化を全体的に把握する分析は少ない。

そこで、本章では、モデリングを通して助詞の経時的変化を捉え、小説の文体との関わりについて考察を試みる。計量分析には、まず、系統樹分析を通して、助詞の使用に明らかに変化が発生していることを確認したうえで、主な変動要素を特定するため、回帰分析を用いて、助詞に関するモデルを作成し、モデル構築に大きく寄与する助詞項目を抽出し、分析を試みる。

3.2 助詞データの抽出

コーパス内の作品に対し、形態素解析を行い、助詞データを抽出する。形態素解析のための辞書は何種類もあり、異なる特徴や利点を持つ（工藤，2018）。形態素解析器 MeCab（IPA 辞書）の場合、「にあたって」や「にしたがって」などの動詞を含む複合助詞について、まとまった形で抽出することができるので、ほかの辞書に比べて、より多様な形式の助詞情報が含まれている。その理由で、本章では、MeCab 形態素解析器を使用することにした。

IPA 辞書において、助詞が、第 2 層では、格助詞、係助詞、副助詞、接続助詞、終助詞、並列助詞、副詞化助詞、連体化助詞、特殊助詞という 9 つの種類に分類されている。本章では、第 2 層を用いた。また、データを集計する際、「ヲ_格助詞」「を_格助詞」のような異表記は同じ項目としてカウントした。一方、「たり_並列助詞」「だり_並列助詞」のような語彙の活用形の影響により清濁の違いが生じる項目については、異なる項目としてカウントした。その結果、コーパスから計 130 項目の助詞が抽出された。詳細項目を表 3.1 に示す。

表 3.1 コーパスより抽出された助詞項目

格助詞	に, を, が, と, で, から, の, へ, という, として, より, って, についで, にとって, によって, とともに, といった, にたいして, にたいする, にかけて, っと, ん, において, につれて, による, をもって, にしたがって, とかいう, にかんする, をとおして, にて, における, にあたる, により, にかんして, をつうじて, につれて, っていう, にわたって, につけ, にとり, にたいし, ていう, にわたる, をめぐって, にしたがい, にあたり
係助詞	は, も, さえ, しか, こそ, すら, や, ぞ
副助詞	まで, だけ, ほど, など, でも, ばかり, かも, くらい, じゃ, ぐらい, とも, なんて, ずつ, だって, のみ, なんか, なぞ, なり, しも, なんぞ, やら, ばっかり, だに
接続助詞	と, が, て, で, ながら, ば, から, ので, のに, し, けれども, けれど, けど, つつ, とも, ものの, ちゃ, なり, どころか, や, たって, じゃ, からは
終助詞	か, な, よ, ね, わ, の, け, べ, さ, ぞ, なあ, かしら, やら, ねえ, ぜ, かい, ねん, つけ, や
並列助詞	と, や, たり, とか, やら, だの
副詞化助詞	に, と
連体化助詞	の
特殊助詞	に

3.3 助詞の経時的変化についての分析

助詞使用の経時的変化を分析するに当たり、まず、すべての品詞項目における助詞の比率に変化があるかを確認する。その次に、集計した130項の助詞データに対し、系統樹を用いて、助詞内に変化が発生しているかについて、その概観を示す。次に、その結果を踏まえて、助詞のデータを説明変数、作品の初出年代を目的変数として、回帰モデルを構築し、助詞使用状況を回帰モデル化する。そして、モデリングに大きく寄与する説明変数を見つけ出し、考察する。

3.3.1 助詞の比率の経時的変化

MeCabの品詞体系では、品詞は第1層のタグによって、名詞・接頭詞・動詞・形容詞・副詞・連体詞・接続詞・助詞・助動詞・感動詞・記号・フィラー・その他に分けられる。各年度の小

説における助詞項目の比率を算出し、その経年変化を図3.1にプロットする。助詞の比率は約28%~32%であり、ほぼ一定していると考えられる。

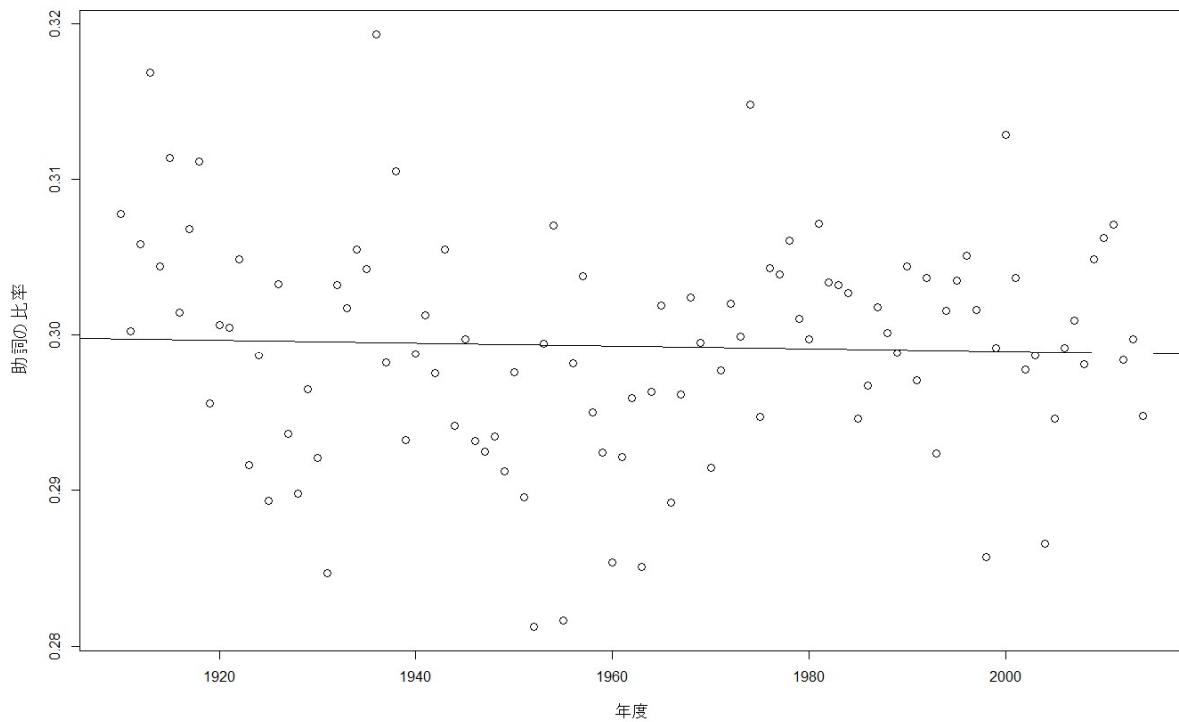


図3.1 助詞の比率の経時的変化

3.3.2 助詞の系統樹分析

助詞内における変化の状況を概観するため、集計した130項目の助詞データを相対頻度に変換したうえで、系統樹分析を行う。図3.2に助詞の使用率のデータを用いた系統樹分析の結果を示す。図3.2の点線の右側は1960年以前の作品が多く配置されているに対し、左側はほとんどが1960年以降の作品である。これにより、助詞の使用が時期の推移に伴って変化している傾向が窺える。

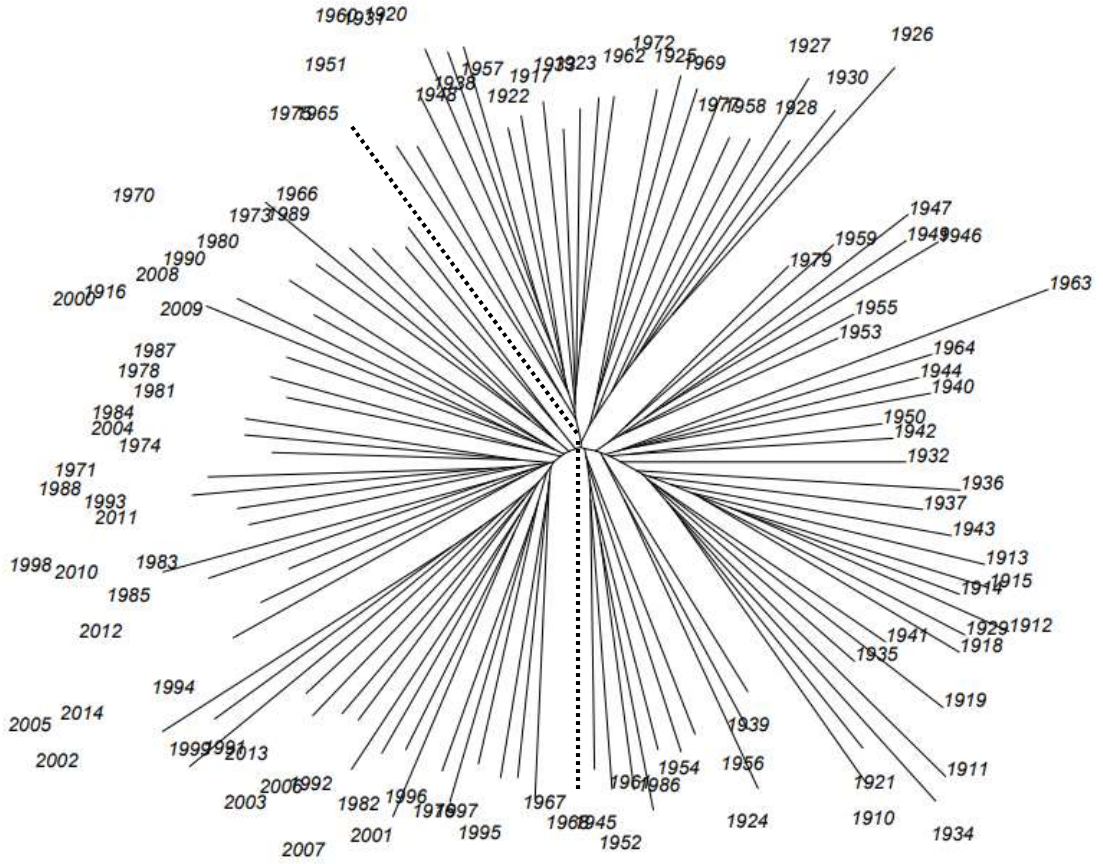


図 3.2 助詞を用いた系統樹分析の結果

3.3.3 助詞のランダムフォレスト回帰分析

続いて、時系列情報と助詞使用状況を表現する回帰モデルを構築し、モデリングに寄与する特徴的な助詞項目を見つけ出す。回帰モデルを構築するため、各年度の番号を表す1~105という数字列（y列と呼ぶ）を目的変数とし、130個の助詞項目を説明変数とする。本項では非線形的関数を用いたランダムフォレストを用いてモデルの構築を実施する。

回帰木を500個としたランダムフォレストを行ったエラー率の推移グラフを図3.3に示す。図3.3からわかるように、エラー率は木の数が200以上になると比較的安定している。回帰木を500個としたときのランダムフォレストの回帰モデルは、元データの分散を約79%説明している。

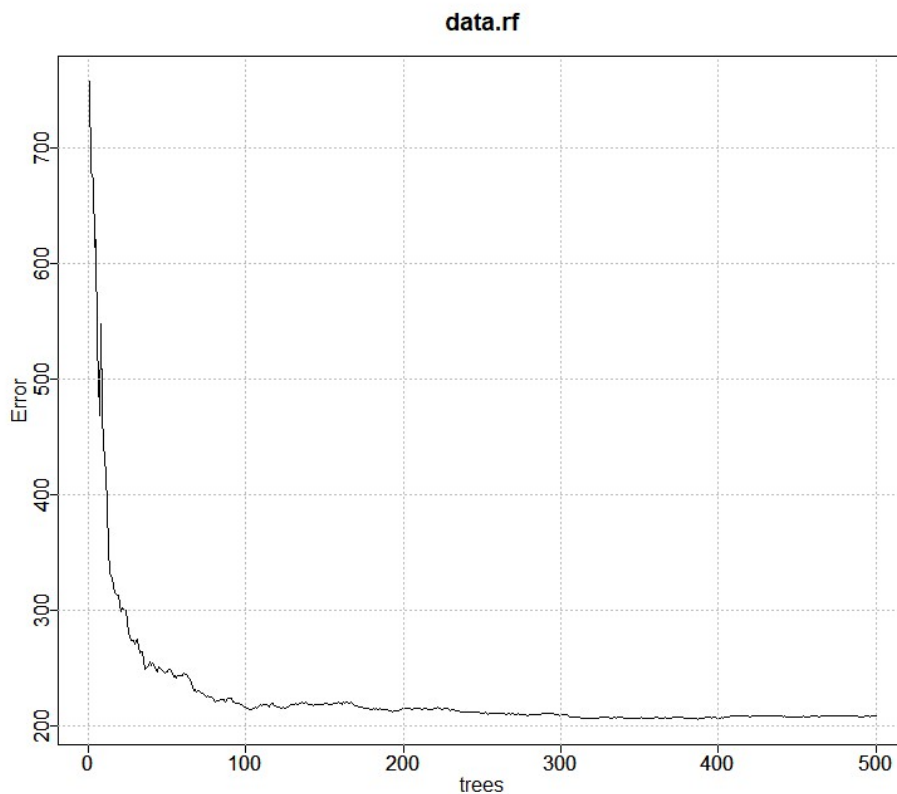


図 3.3 回帰木の数とエラー率

また、その残差（絶対値）の四分位数はそれぞれ 0.10 (Min) , 1.70 (IQ) , 3.20 (Median) , 6.30 (3Q) , 14.00 (Max) であり、平均値 (Mean) は 4.43 である。残差の最大値より、推定誤差が最も大きいものは約 14 年である。

ランダムフォレスト回帰では、モデルに用いた変数の重要度を求めることができる。回帰木は、ノード（変数）の不純度に基づいて分岐させている。ランダムフォレストでは、不純度は変数毎の不純度の合計をすべての変数の総不純度で割った値である（金, 2021: 207-209）。RFによって選択された上位15個の変数のドットプロットを図3.4に示す。図からわかるように、上位10項目以降では変数間の差が非常に小さい。図3.4に示された上位10個の変数で、元のデータをどれぐらい説明できるかを見るため、10変数を用いて再びランダムフォレストを行った結果、元データの分散を約82%説明できることがわかった。

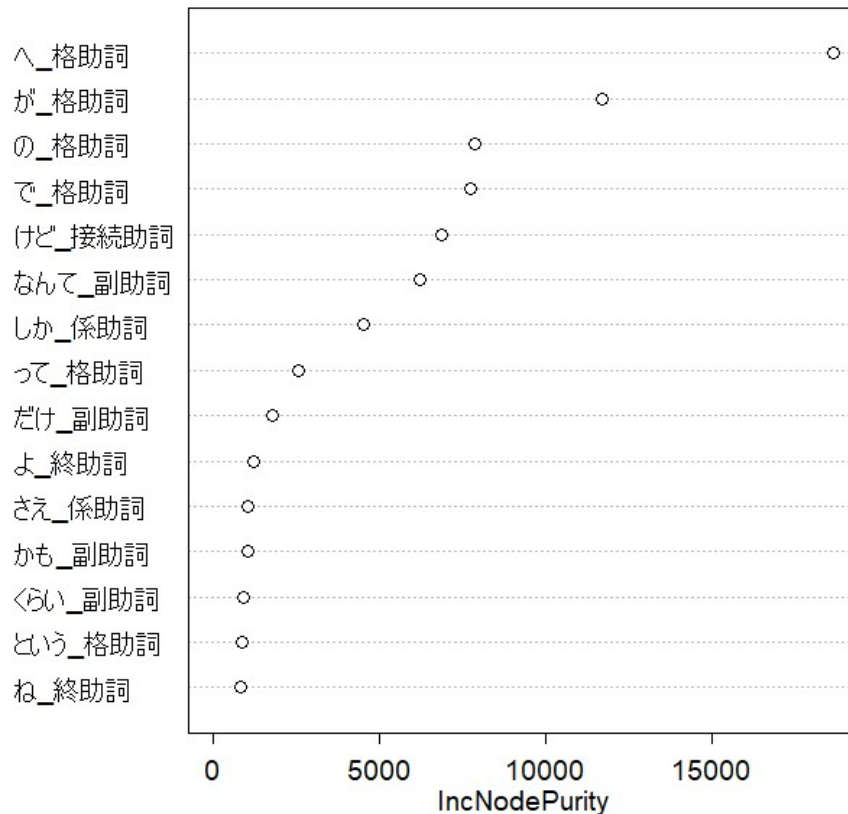


図3.4 ランダムフォレストによる変数の重要度

3.3.4 助詞の elastic net 回帰分析

前項で作成されたランダムフォレストモデルを確認するため、線形回帰 elastic net モデルで選択された変数で対照する。

elastic net 分析を行ったところ、自由度調整済み決定係数が 0.97 である回帰モデルが得られた。また、その残差(絶対値)の四分位数はそれぞれ 0.06 (Min), 1.67 (IQ), 3.20 (Median), 5.16 (3Q), 9.98 (Max) であり、平均値 (Mean) は 3.54 である。残差の最大値より、推定誤差が最も大きいものは約 10 年である。

elastic net回帰モデルの構築では、複数の相関が強い説明変数が存在し、多重共線性がある場合、相関が強い変数の中から代表となる変数が選択してモデルを構成している。今回の分析では8つの説明変数が選ばれた。この8つの説明変数とその係数を表3.2に示す。係数の絶対値が大きいほど回帰モデルでの影響が大きいと判断される。

表 3.2 elastic net に選ばれた説明変数の係数

説明変数	係数
1.の_格助詞	-2349.24
2.へ_格助詞	-2294.87
3.なんて_副助詞	1909.79
4.で_格助詞	1097.08
5.しか_係助詞	369.62
6.さえ_係助詞	-296.90
7.が_格助詞	196.26
8.よ_終助詞	87.81

この8個の変数には、7個がランダムフォレストモデルで選出した変数の上位10個に含まれており、残りの「さえ_係助詞」はランダムフォレストモデルで選出した第11番に相当する。これによって、2つの回帰モデルで得られた結果は高い一致度を示していることがわかる。どの結果が一番理想的であるかに関しては決定しがたいため、共通な項目に焦点を与えて考察を行う。2つの回帰モデルより選択された「へ_格助詞」「が_格助詞」「の_格助詞」「で_格助詞」「なんて_副助詞」「しか_係助詞」「さえ_係助詞」「よ_終助詞」という8つの変数が特徴的項目として言語表現の変化の大まかな流れを反映できているかを確かめるため、8つの助詞の使用率を用いて、改めて系統樹分析を行い、その結果を図3.5に示す。

図3.5から見ると、全体的には、下から上の方向へ、大まかに時期の流れが読み取れる。

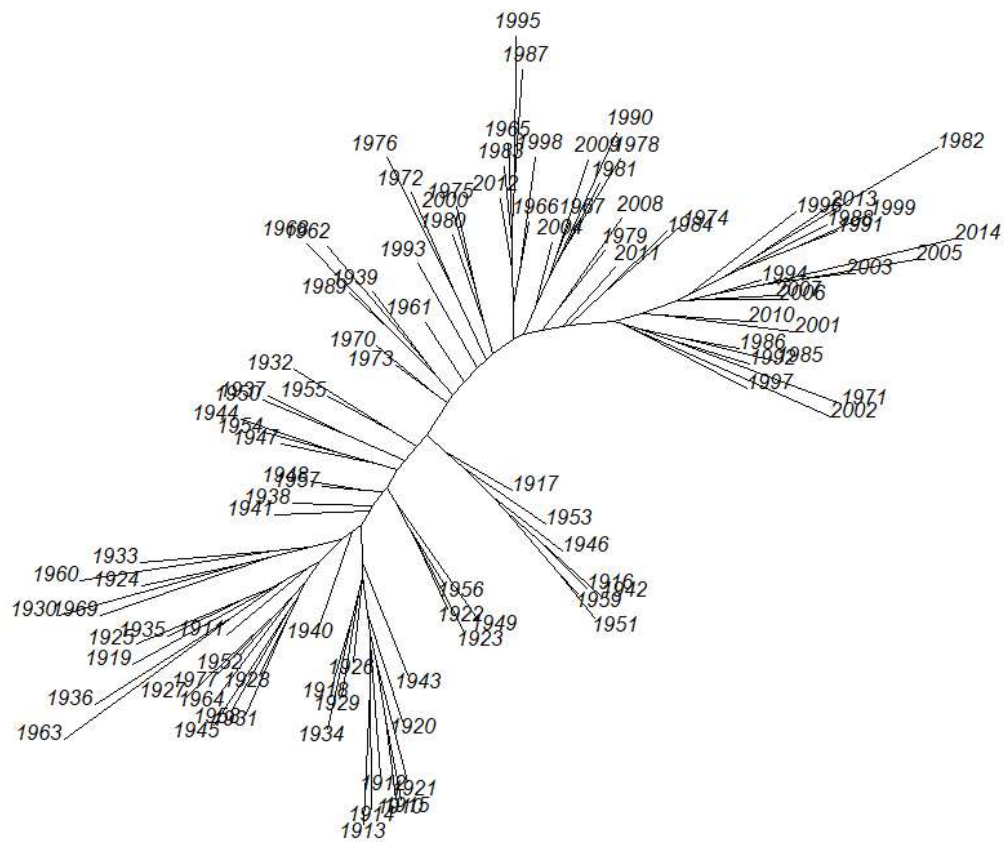


図 3.5 8つの助詞項目を用いた系統樹分析

構築した回帰モデルで得られた8つの助詞について、それぞれの使用率に回帰折れ線を加えたプロットを図3.6に示す。図3.6より、「へ_格助詞」「の_格助詞」「さえ_係助詞」が右肩下がりであり減り続けており、「が_格助詞」「で_格助詞」「しか_係助詞」「なんて_副助詞」「しか_係助詞」「よ_終助詞」が時期の推移に伴い増加の傾向を示すことがわかった。

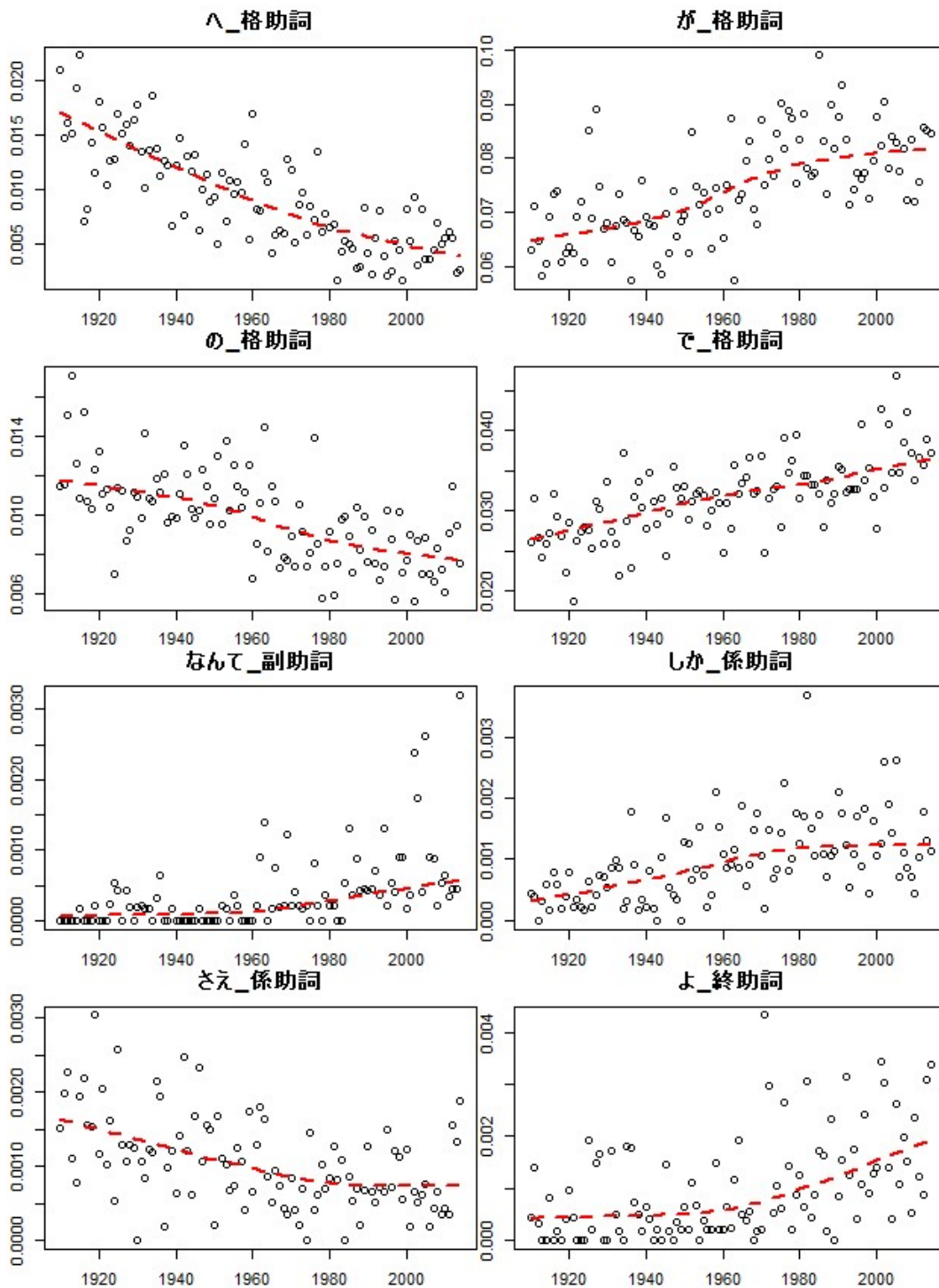


図3.6 8個の特徴的助詞の使用率の変化

3.4 考察

回帰モデルで選択された特徴的助詞項目には、格助詞「へ」「で」などのように、意味（語義）が多様であり、出現頻度が高いものもあれば、語義の種類が比較的になく、それほど

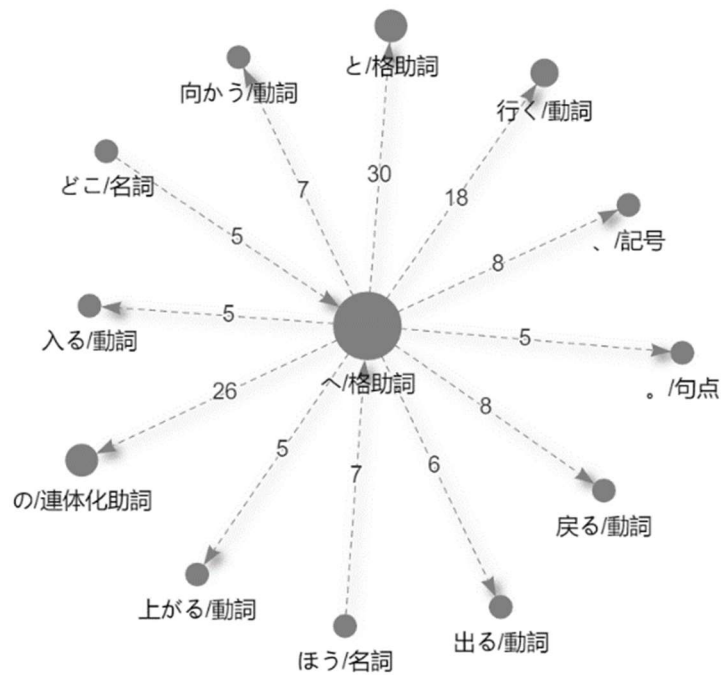


図3.8 頂点格助詞「へ」のネットワークグラフ (2005~2014年)

図 3.7, 3.8 から見ると、格助詞「へ」の出現頻度が減少するにつれ、格助詞「へ」に関わる bigram 項目が大幅に減っている。すべての bigram 項目において、「名詞+へ」及び「へ+動詞」といった共起が多数であるため、その減少は特に顕著に見える。動詞に着目すると、図 3.7 より「落ちる」「置く」「出す」「入れる」「着く」「かける」など、＜着点＞を表す動詞との共起は多く見られるが、図 3.8 ではほとんど見られない。それに対し、「行く」「戻る」「入る」「出る」「向かう」などのような＜方向＞を表す動詞は、図 3.7 と図 3.8 の両方に現れている。現在の言語感覚から見ると、＜方向＞を表す動詞と共起する「へ」は、その多くが格助詞「に」で代用することができ、格助詞「に」のほうがより自然に聞こえる場合もある。コーパス内の 1915 年の作品における 1 文を例として挙げる。

- (1) こうして夜明けになって電灯を消すと、小川と云う電気職工は或る遠い下町の学校へ通っている為め大急ぎで帰って行く。機関手の沢井と俊治だけ後へ残って機械の掃除やら日誌の記入やらを済ませてから、瓦斯機関の汽筒から湧いたお湯で風呂へ這入る。

(1915 年『地に頬つけて』)

例(1)において、「学校へ通っている」「後へ残って」「風呂へ這入る」などの類の「へ」は、時期の推移につれ、「に」に取って代わられてきた可能性があると考えられる。コーパス内の「へ」の数量は膨大であり、それをすべて調べて文脈と用法を判別するのは困難であるが、『地に頬つけて』という作品に対し、全文を確認した結果、79%の「へ」が「に」で代用することが可能であることがわかった。一方、格助詞「へ」と「に」に対し、相対頻度データのピアソン相関係数を確認した結果、 -0.42 が得られ、弱い負の相関が存在すると判断される。この点から見ると、格助詞「へ」の使用変化について、格助詞「に」との間の勢力消長・役割分担による影響が否めないが、すべて格助詞「に」の代用に帰するものではないと推察する。

図3.7にも図3.8にも存在する項目として、「へ+の_連体化助詞」「へ+と_格助詞」「へ+、_読点」がある。前の2つは、それぞれ「への+名詞」「へと+動詞」に等しい項目であり、その中の「へ」は、ほかの助詞によって代替されにくいと考えられる。

一方、図3.7にはなかったが、図3.8には現れた項目として、「へ+。_句点」が見られた。使われたコーパスを確認したところ、述語を省略して、格助詞「へ」で終える文は、1966年以前の作品に全くなく、コーパスの後半に存在している。そうした表現は、新聞記事の見出しや広告ヘッドラインなどによく使われると指摘されている(李, 2002)が、最近になって小説にも使われるようになったことが今回の分析で明らかになった。その数はまだ多くはないが、述語の省略により文内に余白・余韻が残されたり、文末表現に変化がもたらされたりするような効果がある。「へ+句点」の用例を例(2)と例(3)に示す。

(2) 紙川さんは神奈川県横浜市へ。私は埼玉県さいたま市へ。

(2009年『この世は2人組ではできあがらない』)

(3) 洗濯を終えて、公園へ。砂場の前にしゃがむ、美和さんの背中が見えた。

(2010年『うちに帰ろう』)

3.4.1.2 格助詞「で」「の」「が」

格助詞「で」「の」「が」に対し、同じく最初の10年間(1910~1919)及び最後の10年間(2005~2014)におけるbigram項目もネットワークで分析することが可能である。

格助詞「で」が文章に大量に使われており、「で」と1つ前/後の語彙による組合せが多い。ここでは、格助詞「で」と1つ前の語彙に限定し、出現頻度が5以上の共起項目を用いる。コーパス内の最初の10年間(1910~1919)及び最後の10年間(2005~2014)のデータから、格助詞「で」及びその前に現れる語彙項目をbigramの形式で抽出して、ネットワークグラフを描く。その結果を図3.9, 3.10に示す。

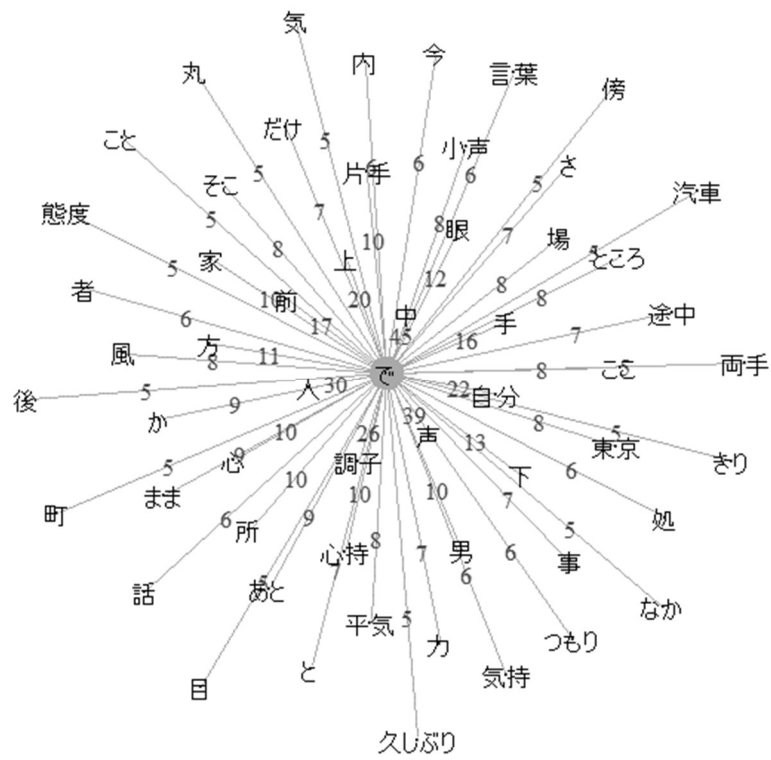


図3.9 頂点格助詞「で」のネットワークグラフ (1910~1919年)

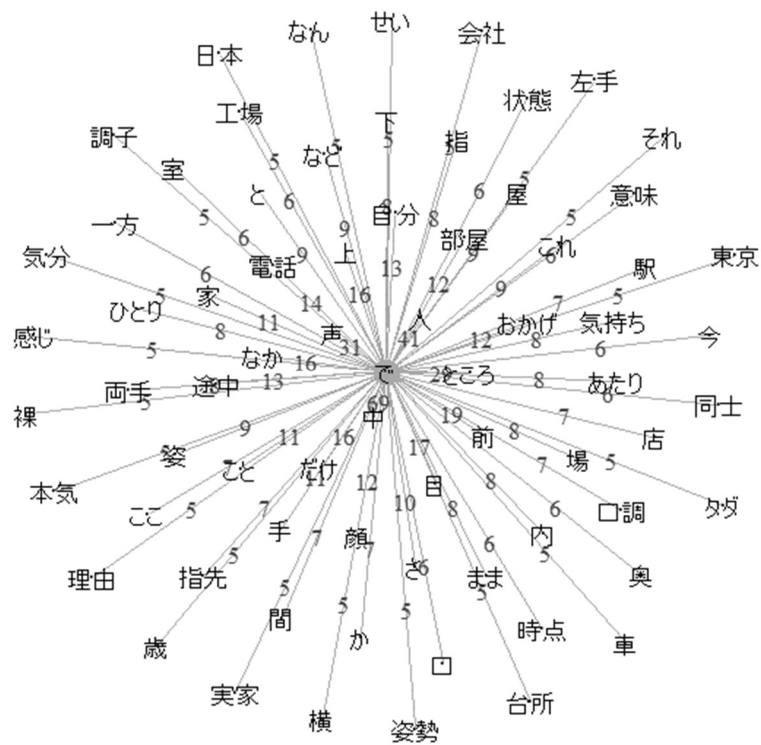


図3.10 頂点格助詞「で」のネットワークグラフ (2005~2014年)

2つの区間を比較した結果、格助詞「で」において、場所を表す名詞との共起が増えていることが確認された。「場」が明確に提示されることによって「コト的体験的把握」がよりしやすくなると尾野（2018）は述べている。つまり、より知覚的・感覚的に捉えやすくなり、そのことによって、「共感」の度合いが高まるのである。できごとの起きる場所や動きの空間範囲を明示することによって、読者に物語の内容をより感覚的に捉えさせ、共感を呼びやすくする効果が見込まれる。「で」格の増加には、空間変換による遠近感を現出させたり、読者に現実感覚を与えたりするような作家達の表現意識が働いていると推測される。

一方、格助詞「の」と「が」について、ネットワーク分析により、2つとも、主に、「名詞」の後と「動詞」「形容詞」の前に現れており、この使用傾向が変わっていないが、格助詞「の」は減少し、「が」は増えていることが確認された。『広辞苑』（第7版、新村編、2018:2275）によると、述語の表す内容をもたらした主体について、主文で「が」を使うのに対し、従属文では「の」を使うとする考えもあったが、現代語では、従属文でも「が」で表すことが多い。今回の結果はこの点を裏付けている。とは言え、格助詞「が」の増加量は格助詞「の」の減少量を大きく上回るが見逃せない。

3.4.2 そのほかの助詞

格助詞に比べ、同じく特徴的項目である副助詞「なんて」、係助詞「しか」、「さえ」及び終助詞「よ」は出現頻度が相対的に少なく、語彙内における語義差異もそれほど大きくない。

係助詞「しか」は使用率が上がる傾向を示している。「しか」は、「花子しかいない」のように否定述語と拘束関係を持っている。この「しか」の拘束関係が出現したのは近世初期以降であることが指摘されている（山口、1991）。そして、「しか」の類の助詞の成立は、否定表現の変革に大きく関わっており、その本質は日本語述語構造そのものの歴史的変革との指摘もある（宮地、2007）。「しか」を始めとする否定呼応表現は、既に研究者に注目されていたが、質的、記述的に言及されるにとどまっていた。本章の分析結果は、先行研究の見解の定量的裏付けになると同時に、その意義を改めて認識し、再評価するための契機ともなりうる。

終助詞「よ」も時期の推移に伴って増加している。コーパスを確認したところ、「よ」のみではなく、終助詞の全体使用率は2000年前後に増えてくる傾向を示した。コーパスクリーニングの際、括弧やダッシュなどの記号で標記された会話文を削除したため、使われたコーパスにおける終助詞は、括弧などの記号が付けられない引用文または独話や心内発話文などに現れることが多い。「よ」を始めとする終助詞の増加は、内的モノローグの叙法やくだけた言葉遣い、喋り口調という特徴を持つ小説が多くなっていることを示唆している。その原因

について深く検討する余地があるが、インターネットの普及に相まって発達してきたブログなど新メディアの表現による影響の可能性が考えられる。

3.5 まとめ

本章では、助詞の経時的変化について考察を行った。品詞の比率から見ると、助詞の比率は終始 28%~32%の間に維持されており、大きな変化が認められない。抽出された 130 項の助詞項目を用いて系統樹分析を行った結果、助詞内に経時的変化の傾向が観察された。ランダムフォレストによる非線形的回帰モデル及び *elastic net* による線形的回帰モデルを用いて、助詞の使用についてモデリングを行った。構築された 2 つのモデルに共通する特徴的な変化項目として、格助詞の「へ」「の」「で」「が」、副助詞の「なんて」、係助詞の「しか」「さえ」及び終助詞の「よ」が特徴的項目として特定された。見出された特徴的助詞項目から、文体学や語学の研究のヒントとなり得る可能性が示唆された。

第4章 文末表現の経時的変化に関するモデリングと分析

4.1 本章の背景と目的

明治期の言文一致における文体変革について、柄谷（2004）は「事実上語尾の問題に帰着する」と指摘した。当時、文末の処理をめぐって、「だ」調、「です」調、「であります」調、「である」調など様々な実験・模索が行われていた。この過程において、欧文翻訳の影響により、「た」は述語の過去形の機能を担わされ、「た」止めの文末形式が生み出された。藤井（2010）は、日本の古代物語の基調が非過去であるが、近代において過去時制を持つ文体が制覇をなし遂げたと指摘した。近代からタ形で文を終わらせる形式が一般的になってきた。文末におけるこのような変化は、語り構造の変動にもつながる。野口（1994）は、言文一致以前の物語には一人称的な語りしかなかったが、タ形の文末形式によって三人称客観描写が創出されたと述べている。この三人称客観描写は「近代小説を特徴づける」（柄谷，2004：18）作法とも評される。

一方、口語文の欠点は、文末表現にも顕著に現れる。主語-目的語-動詞といったSOV型の語順を取る日本語では、文末に動詞（+助動詞）がくることが多い。そのため、多くの作家（谷崎，1975；丸谷，1980；井上，1987）が指摘しているように、口語文は、大部分が「る」止めか、「た」止めか、「だ」止めであり、変化に乏しい傾向がある。特に、小説の場合、過去の出来事を取り扱う形で構成されるものが多いため、過去時制が連続しやすい。この点を意識して、文章の文末に変化を持たせるように工夫する作家がいる。例えば、三島由紀夫は自分が小説を執筆する際、書いた文章を読み返し、過去形の多いところを直すようにしていると述べた（三島，1973）。

「作品が最大の効果をもって読み手に伝わるような種々の言語装置を駆使する」（山本，2002：2）作家にとって、文末の形式も小説世界を巧みに構成するための言語装置の1つである。どのような形式の文末を取るかは、文体の硬軟、語句の選択、リズムだけではなく、物語の時間的展開や場面転換などにも影響を及ぼす。また、特定の時代や社会を生きた作家たちの取捨選択により、その時代の言語装置の特徴が形成される。口語文体が確立されてから現在に至り百数十年が経過し、この期間に文体は時の推移と共に変化しつつあり、統計的に見れば、小説の文末形式の分布に何らかの変化があると予想されるが、これを取りあげて論じた研究はまだ見当たらない。そこで、本章では、近現代小説における文末表現を分析対象とし、計量分析によって、文末表現の多様性及びその分布状況に変化が生じているか、または、どのような項目に顕著な変動が見られるかについて考察を行う。

4.2 文末表現に関する先行研究

現代書き言葉の文末表現に関して、特定のジャンルにおける使用状況または性別・ジャンル間の差異などに着目した研究事例は数多くある。高崎ら（2007）は、『文芸春秋』の巻頭随筆 500 編を対象に、最終述語に関わる部分を文末形式と見なして、随筆における文末の表現形式を概観した。また、山崎（2011）は、現代日本語書き言葉均衡コーパス（BCCWJ）に基づいて、出版書籍・雑誌・新聞・図書館書籍・白書・教科書・ベストセラー・Yahoo! 知恵袋・Yahoo! ブログ・国会会議録といった 10 種類の媒体に対し、上位の文末語の分布を調査すると同時に、「た」系列・非「た」系列に分けて、それらの分布状況をまとめた。丸山（2012a; 2012b）は、同じく BCCWJ を利用して、文末位置から逆向きに文字単位の n-gram を抽出し、異なるレジスターにおける文末 n-gram のバリエーションを比較した。そして、橋本（2014a）は、終助詞の使用を手掛かりとし、書き言葉と話し言葉の文末表現における男女の差異を分析した。

一方、通時的な分析は、思考動詞などの使用を手掛かりにした分析（渡邊，2016）が見られる。また、文末に現れた品詞の類型によって、新聞の文章を口語敬体・口語常体・文語体に分けて、その推移状況を考察した研究もある（ヤロシュ島田，2021）。これらの通時的分析に使用した資料は、いずれも 20 世紀前半以前のものである。近現代小説に関しては、作家個人の作品における文末形式の経時的変化についての分析（土山，2019）はあるが、口語文体の確立から現在までの期間において、文末表現の全体分布がどのような変化をしているかを考察する研究はまだ見当たらない。また、「地の文章の文末形式は、ふつうにはシタが基調になっている」（野村，2015：23）という認識を持つ研究者もいれば、「ごく最近の小説では、ル形を多く使用したり、体言止めを多く使用したりする小説も増えているように思われる」（橋本，2014b：194）と述べる研究者もいるが、実際のデータを提示した研究は行われていない。

4.3 文末表現のデータ

本研究では、形態素解析器を利用して、文ごとに最後の語を文末表現のデータとして抽出する。

本章の文末データ抽出には JUMAN 辞書を用いる。その理由は主に 2 つある。1 つ目は、ipadic や UniDic などの辞書に比べて、JUMAN 辞書は長い単位で単語を分割しているため、得られた文末データにより多くの情報が含まれるためである。2 つ目は、学校文法をベースとするほかの辞書と違って、JUMAN 辞書は益岡・田窪文法を採用しており、文末形式の類別には便利であるためである。例えば、JUMAN 辞書では判定詞が品詞の 1 つとして導入されたため、名詞述語の文ははっきりと識別できる。また、JUMAN 辞書では、学校文法で助

動詞として扱う単語の多くが、活用の一部として認定されているため、文末の単語が持つ本質的な意味を判断しやすい。表 4.1 には、JUMAN 辞書と ipadic 辞書を用いた結果の比較例を示す。

表 4.1 JUMAN 辞書と ipadic 辞書を用いた結果の比較例

JUMAN 辞書	ipadic 辞書
景色/ <u>が</u> /きれいだ <u>った</u>	景色/ <u>が</u> /きれいだ <u>っ</u> / <u>た</u>
彼/ <u>は</u> /家/ <u>を</u> /出 <u>た</u>	彼/ <u>は</u> /家/ <u>を</u> /出/ <u>た</u>
明日/ <u>は</u> /雨/ <u>だ</u> ろ <u>う</u>	明日/ <u>は</u> /雨/ <u>だ</u> ろ/ <u>う</u>

形態素解析器により得られた文末データについて、主に、その品詞タグを分析に利用する。また、品詞タグのほか、活用形式の情報や形態素を適宜に用いる。

品詞タグに関して、JUMAN では第 1 層の品詞タグとして、名詞、動詞、形容詞、判定詞、助動詞、副詞、助詞、接続詞、感動詞、接尾辞、接頭辞、連体指示語がある。接頭辞、連体指示語は、文末語としてコーパスに現れていないため、除外とする。本章の分析において、接尾辞と助詞はさらに細分類した第 2 層タグを使用し、ほかの項目は第 1 層の品詞タグを用いる。つまり、接尾辞は、形容詞性述語接尾辞、形容詞性名詞接尾辞、動詞性接尾辞、名詞性接尾辞といったタグを利用し、助詞は、格助詞、副助詞、接続助詞、終助詞といったタグを利用する。なお、細分類した助詞について、その形態素原形も併せて表示させる。

また、一部の品詞（動詞、形容詞、接尾辞、判定詞、助動詞）には活用があり、その活用形にヴォイス、アスペクト、テンス、モダリティなどの情報が多く含まれているので、活用情報も分析に活用する。動詞、形容詞、接尾辞に関しては、JUMAN 辞書で定義した活用関係情報²（基本形、タ形、推量形、命令形など）を付け加える。判定詞及び助動詞に関しては、活用関係の代わりに直接に形態素原形を表示する。

表 4.1 に挙げられた 3 つの文に関して、抽出された文末データの表示形式を表 4.2 に示す。

表 4.2 文末データの表示例

	JUMAN 辞書	品詞タグ	活用関係	データの表示
1	景色/ <u>が</u> /きれいだ <u>った</u>	形容詞	ダ列タ形	形容詞ダ列タ形
2	彼/ <u>は</u> /家/ <u>を</u> /出 <u>た</u>	動詞	タ形	動詞タ形
3	明日/ <u>は</u> /雨/ <u>だ</u> ろ <u>う</u>	判定詞	ダ列基本推量形	判定詞_だろ <u>う</u>

² 活用関係に関する定義及び分類の詳細は、JUMAN Ver.7.0 のマニュアルを参照されたい。 <https://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

コーパス内の作品を初出年月に従い、年度ごとに文末データを抽出し、分析に用いる。抽出した文末表現の項目は合計 168 個がある。

4.4 文末表現に関する分析

4.4.1 文末表現の多様性

まず、文末表現の多様性に変化が生じているかを考察するために、計量言語学でよく利用される語彙の豊富さを測る K 特性値を利用して分析する。

各年度の文末データに基づいて算出した K 特性値のプロットを図 4.1 に示す。図 4.1 の点線は、年度と K 特性値の中心的な分布傾向を表す回帰直線であり、右下がりとなる傾向が認められた。つまり、文末の形式が、時期の推移に伴い、多様になっていることを示唆している。

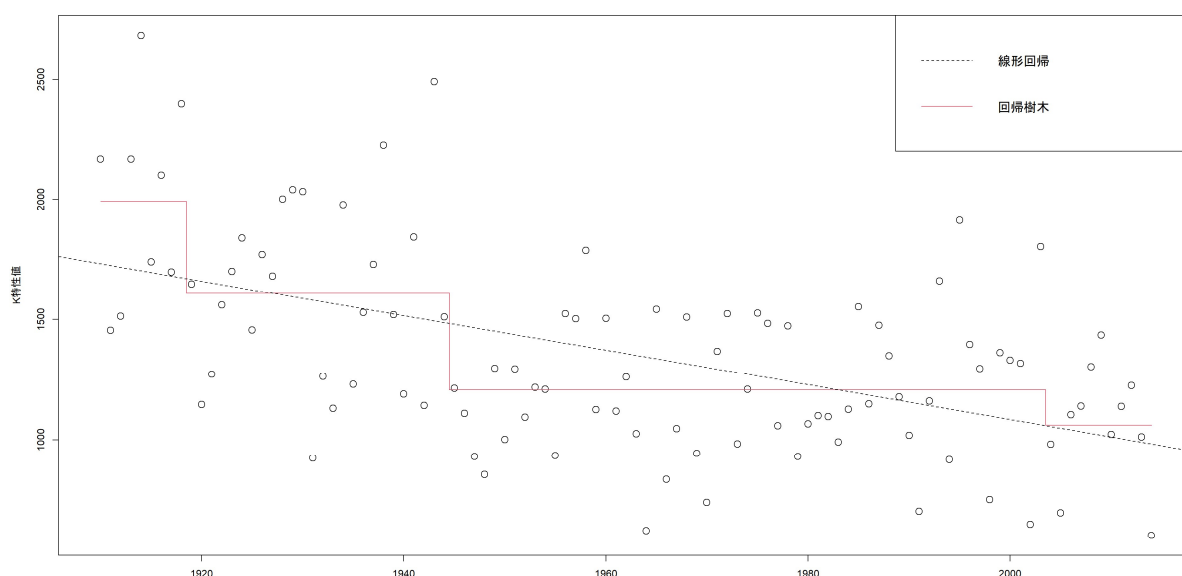


図 4.1 文末語の K 特性値の推移

変化の状況をさらに詳しく確認するため、 K 特性値と年度との関係を回帰樹木で推定する。回帰樹木では、説明変数の K 特性値を分岐基準に基づいて 2 進分岐させ、木を成長させる。樹木回帰関数としてのあてはめ結果を図 4.1 に実線で示し、回帰樹木の構造を図 4.2 に表示する。図 4.2 から見ると、 K 特性値が年度 1944.5 未満/以上で 2 つの部分に分かれている。次に 1944.5 未満の個体からなる部分集団は、1918.5 未満/以上で 2 分割され、他方、1944.5 以上の個体の部分集団は、2003.5 未満/以上で 2 分割された。また、図 4.1 に示された回帰折れ線の形状からもわかるように、1945 年頃までの期間は変化が相対的に激しく、1945 年頃

以降は比較的穏やかになってきた。そして、前後の二段階はそれぞれ 1919 年，2004 年前後を境界に細かく層別できることが読み取れる。

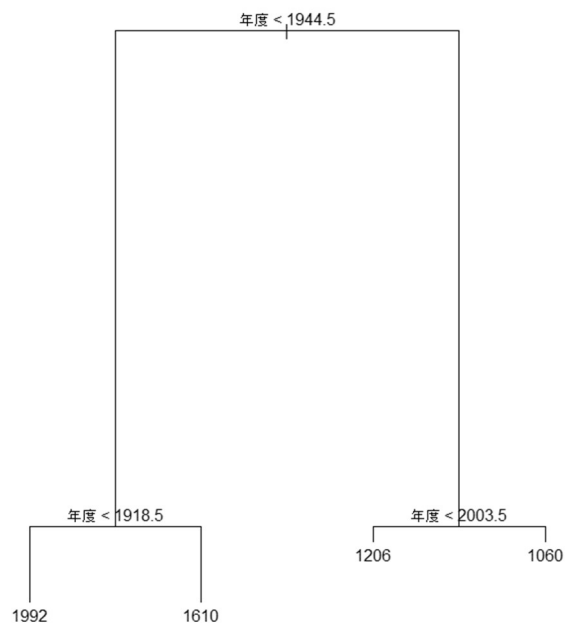


図 4.2 回帰樹木の構造

以上の分析により，文末の表現形式は多様化しつつあり，1945 年を境に大きく 2 グループに分類され，1945 年以降の変化が比較的緩やかになったことがわかった。また，コーパス内の作品に対し，1945 年前後に分けて比較してみたところ，「助動詞_のに」「終助詞_ぜ」「終助詞_さ」「助動詞_みたい」「助動詞_みたいだった」などの項目は，1945 年以前は見られなかったが，1945 年以降の作品によく使用されることがわかった。

4.4.2 文末表現の系統樹分析

文末表現における変化の状況を概観するため，集計した 168 項目のデータを相対頻度に変換したうえ，系統樹分析を行う。図 4.3 に文末表現のデータを用いた系統樹分析の結果を示す。図 4.3 を大まかに上・下という 2 つの部分に分けることができる。下側には 1970 年以前の作品が多く配置されているに対し，上側には 1970 以降の作品が多く分布されている。これにより，文末表現の使用が，時期の推移に伴って，変化している傾向が窺える。

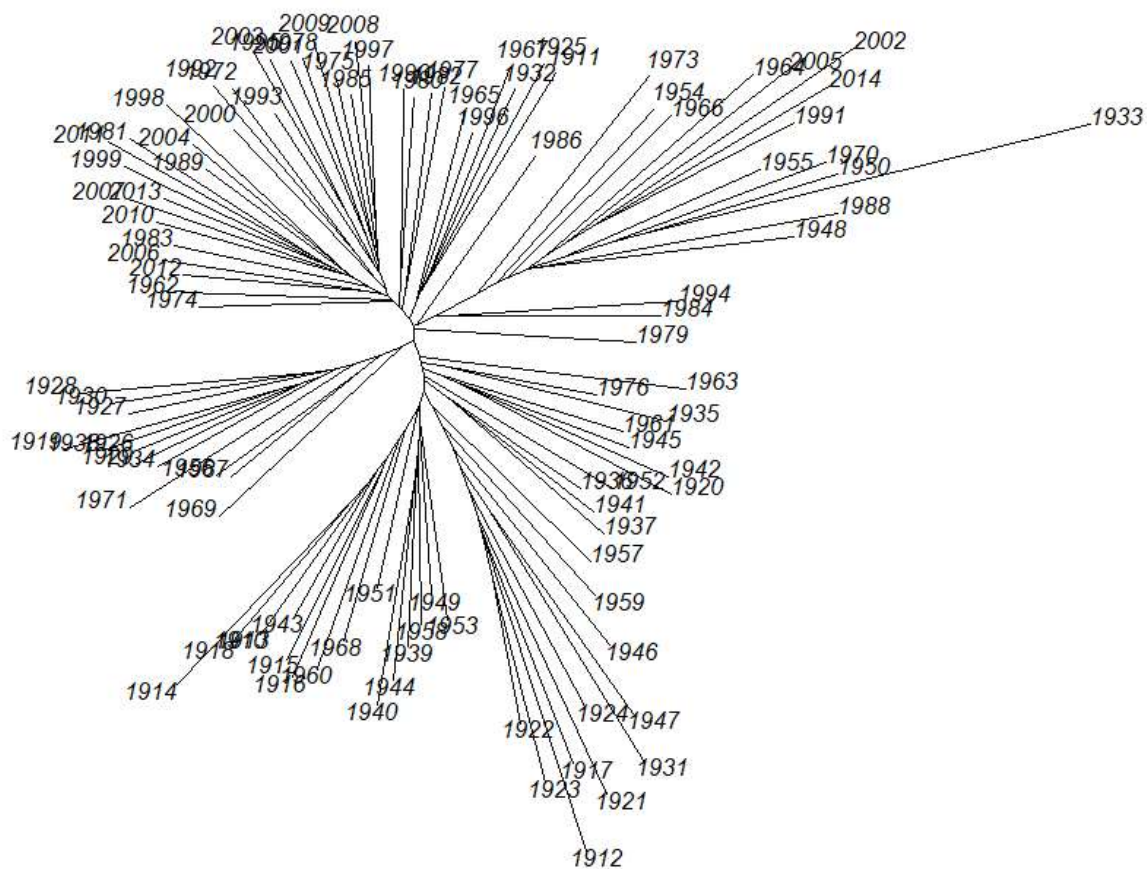


図 4.3 文末表現を用いた系統樹分析の結果

4.4.3 文末表現のモデリング

時間とともに変化する文末の要素を特定するため，抽出した文末表現の相対頻度データを用いて，作品の年度を目的変数とした回帰モデルを構築し，分析を行う．モデルの構築は，前章と同じくランダムフォレスト及び elastic net を利用する．

4.4.3.1 文末表現のランダムフォレスト回帰分析

回帰木を 500 個としたランダムフォレストを行ったエラー率の推移グラフを図 4.4 に示す．図 4.4 からわかるように，エラー率は，木の数が 200 以上になると，比較的安定している．回帰木を 500 個としたときのランダムフォレストの回帰モデルは，元データの分散を約 72% 説明している．その残差プロットを図 4.5 に示す．図 4.5 から残差の多くは，おおむね正負 20 の区間内に配置されていることがわかる．

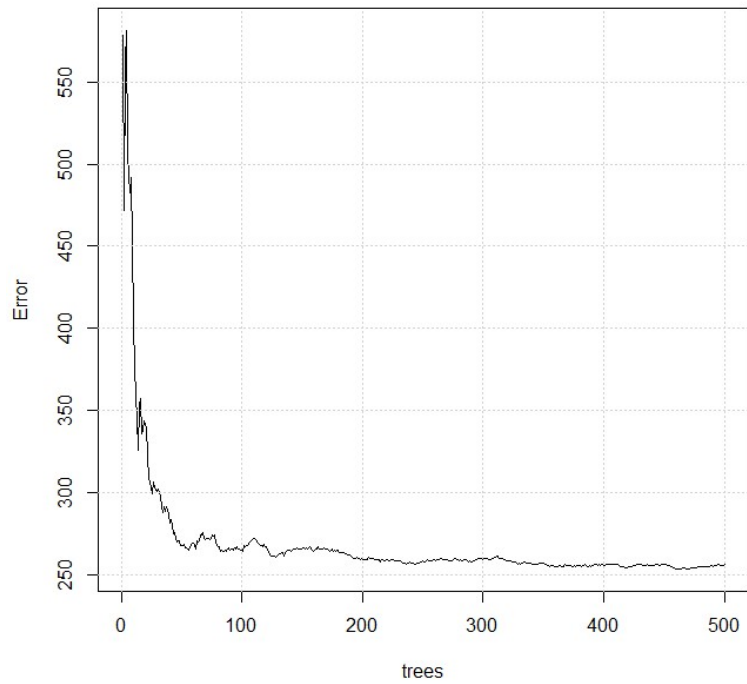


図 4.4 回帰木の数とエラー率

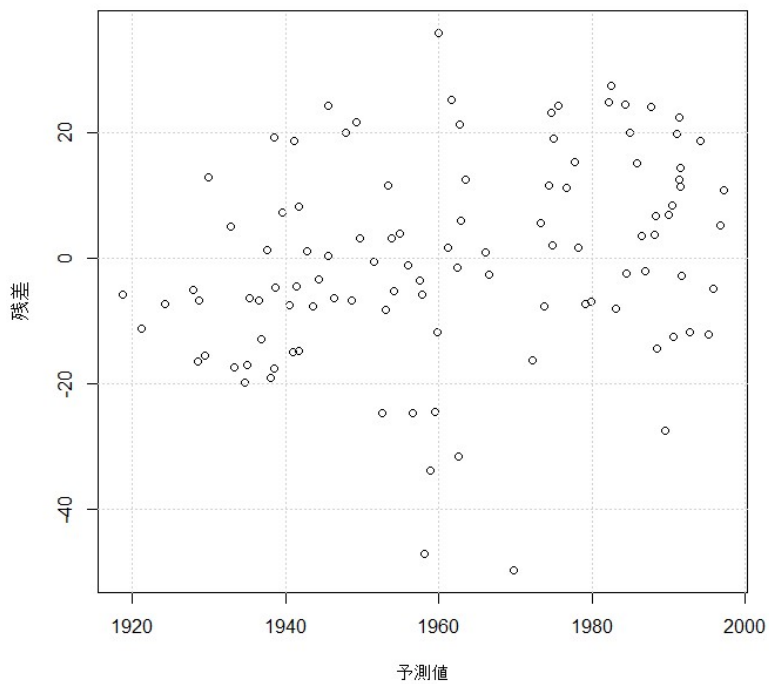


図 4.5 モデルの残差プロット

作成されたランダムフォレストに採用された変数の重要度に関して、上位 20 項目を図 4.6 に示す。図 4.6 からわかるように、上位 10 項目以降では変数間の差が小さくなっている。

図 4.6 に示された 20 個の変数で、元のデータをどれぐらい説明できるかを見るため、20 変数を用いて、再びランダムフォレストを行った結果、元データの分散を約 72%説明できることがわかった。また、この 20 変数を用いて重回帰分析を行うと、自由度調整済みの決定係数は 0.71 である。よって、これらの 20 変数（項目）を用いて分析することには意味があると判断する。

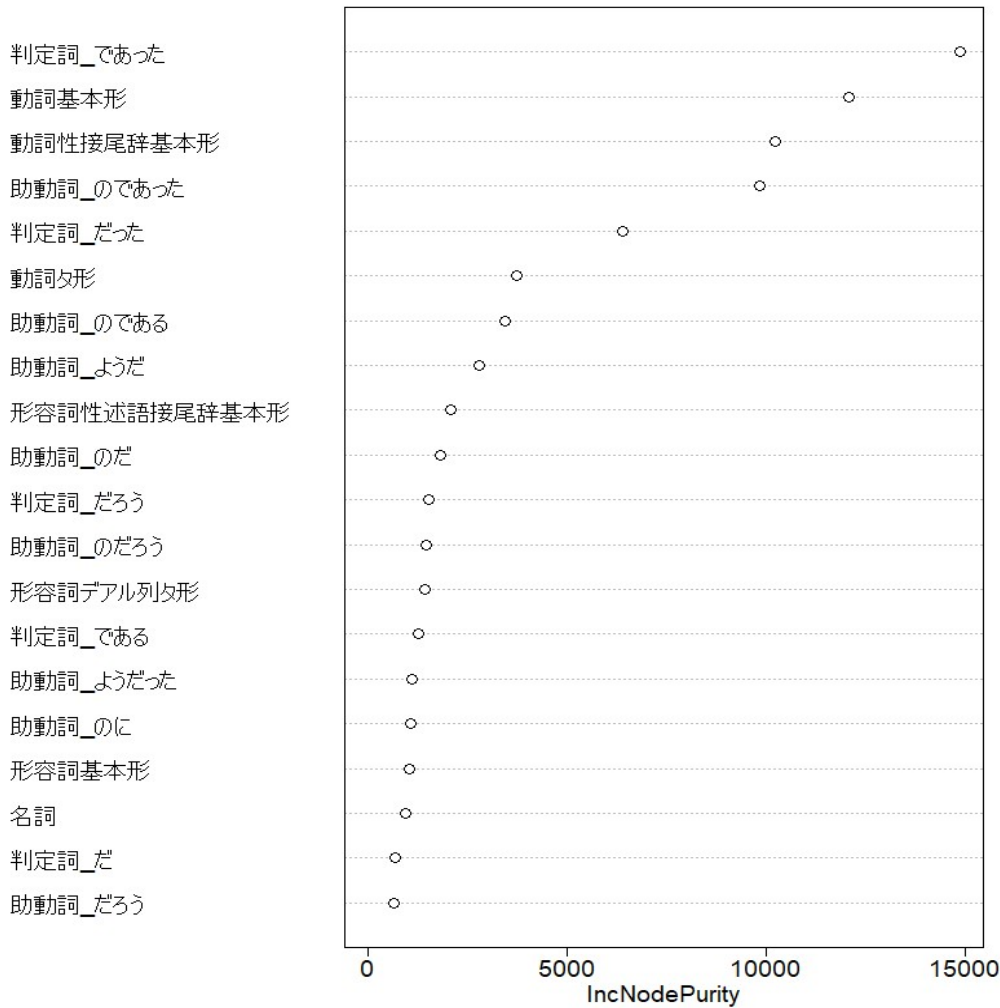


図 4.6 ランダムフォレストによる上位 20 個の変数

ランダムフォレストで選択された上位 20 個の変数が言語表現の変化の大まかな流れを反映できているかを確かめるため、20 個変数の使用率を用いて、改めて系統樹分析を行い、その結果を図 4.7 に示す。図 4.7 において、全体的には、下から上の方向へ、大まかに時期の流れが読み取れる。

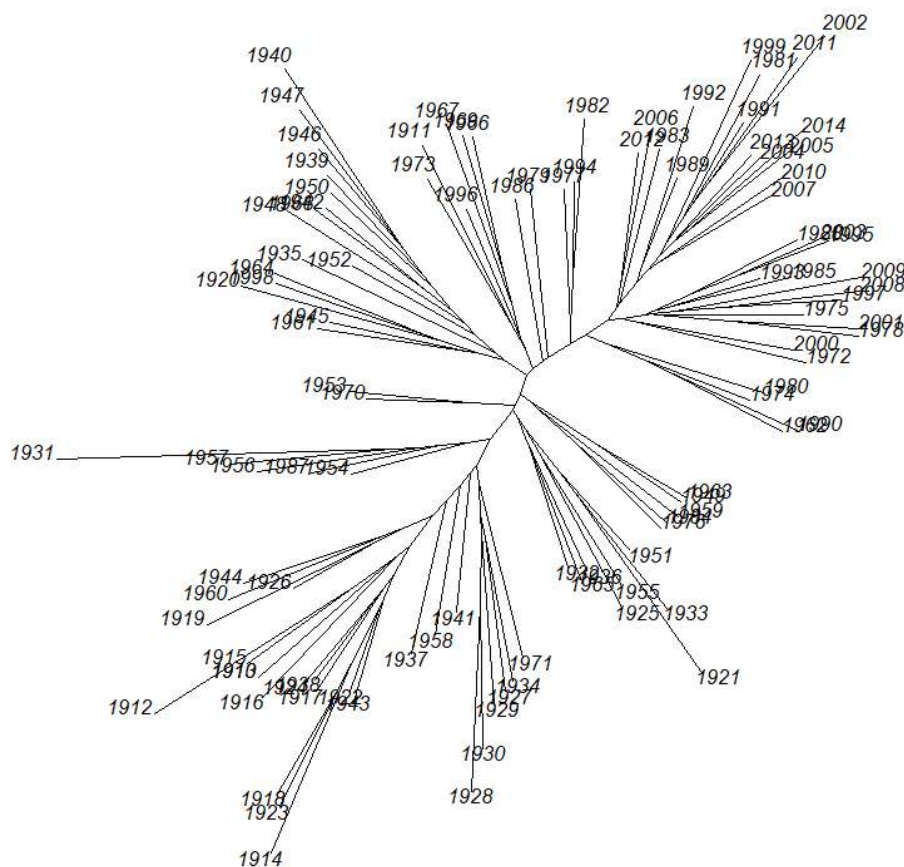


図 4.7 ランダムフォレストで選択した上位 20 個の変数の系統樹

4.4.3.2 文末表現の elastic net 回帰分析

非線形回帰のランダムフォレストに続き、線形回帰の elastic net モデルで選択された変数で対照する。elastic net 分析を行った結果、自由度調整済み決定係数が 0.55 である回帰モデルが得られた。elastic net 回帰モデルの構築では、168 個の説明変数から 52 個が選ばれた。ランダムフォレストと同じように、重要度が高い上位 20 個を図 4.8 に示す。

図 4.6 と図 4.8 を見比べたところ、ランダムフォレストで選択された変数と elastic net で選ばれた変数とは相当異なることがわかった。elastic net で選ばれた変数は、出現回数が極めて少なく、排他性が強い項目が多い。このような特徴を持つ上位 20 個の変数を用いて、系統樹による分析結果を図 4.9 に示す。図 4.9 では、多くの年度が 1 つの塊を形成し、図の下側に位置しており、残りは散在的になっているので、時間的変化する流れが確認しにくい。

ロバスト安定性に合わせて考慮した結果、非線形回帰のランダムフォレストによって選択した変数を中心に考察を進めることにする。

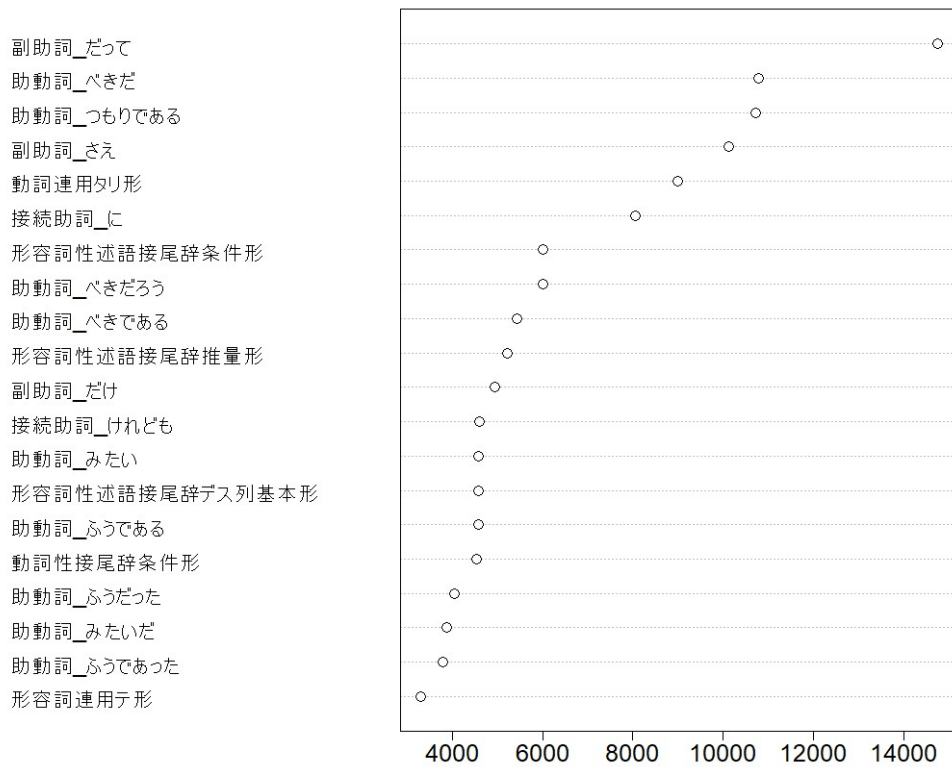


図 4.8 elastic net による上位 20 個の変数

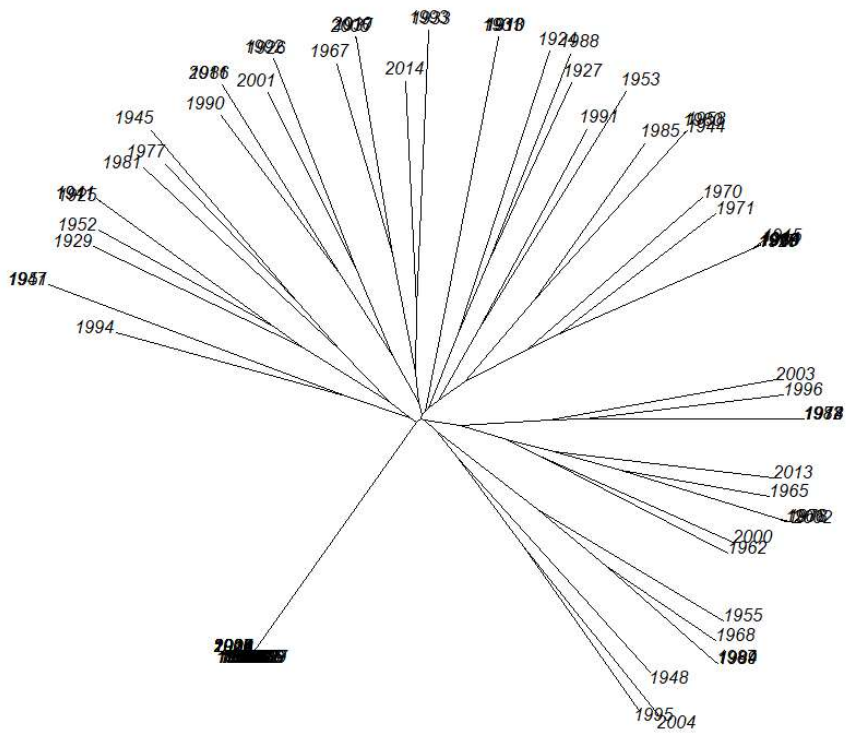


図 4.9 elastic net で選択した上位 20 個の変数の系統樹

4.4.3.3 重要な変数及びその変化

ランダムフォレスト回帰モデルによって選択された 20 個の変数は「判定詞_であった」「動詞基本形」「動詞性接尾辞基本形」「助動詞_のであった」「判定詞_だった」「動詞タ形」「助動詞_のである」「助動詞_ようだ」「形容詞性述語接尾辞基本形」「助動詞_のだ」「判定詞_だろう」「助動詞_のだろう」「形容詞デアル列タ形」「判定詞_である」「助動詞_ようだった」「助動詞_のに」「形容詞基本形」「名詞」「判定詞_だ」「助動詞_だろう」である。これらの 20 項目が経時的にどのように変化しているかを見るため、それぞれの比率の推移の散布図を図 4.10 に示す。

図 4.10 に示された変数及びその推移により、以下のいくつかの変化が見られたと考えられる。

1 つ目は、「である」体と「だ」体の表現における変化である。選択された 20 項目の最上位は「判定詞_であった」であり、「助動詞_のであった」「助動詞_のである」「判定詞_である」「形容詞デアル列タ形」と共に「である」体と見なすことができる。これらの項目はすべて減少する傾向にある。それに対して、「判定詞_だった」「助動詞_ようだ」「助動詞_のだ」「助動詞_のだろう」「助動詞_ようだった」「判定詞_だろう」「判定詞_だ」などの変数項目は「だ」体を特徴づける項目と見なすことができ、いずれも増加の傾向を見せている。「である」体、「だ」体は同じく常体として扱われ、小説では混在使用が可能であるが、分析により、異なる推移傾向が見られた。

2 つ目は、時制の変化である。「動詞基本形」「動詞性接尾辞基本形」「動詞タ形」「形容詞性述語接尾辞基本形」「形容詞基本形」は、動詞または形容詞が述語となる文末項目である。そのなかで、「動詞タ形」のみが過去形であり、減少の傾向を示しているが、ほかの項目は現在形と見なすことができ、増加の傾向が見られる。この結果は、文末形式の時制に変化が生じていることを示唆する。この点を確認するため、時制に関わる全ての文末項目に対し、過去形と現在形に分けて集計し、それぞれの使用率の推移を図 4.11 にプロットする。図 4.11 では、過去形の文末表現は下がる傾向を示している。

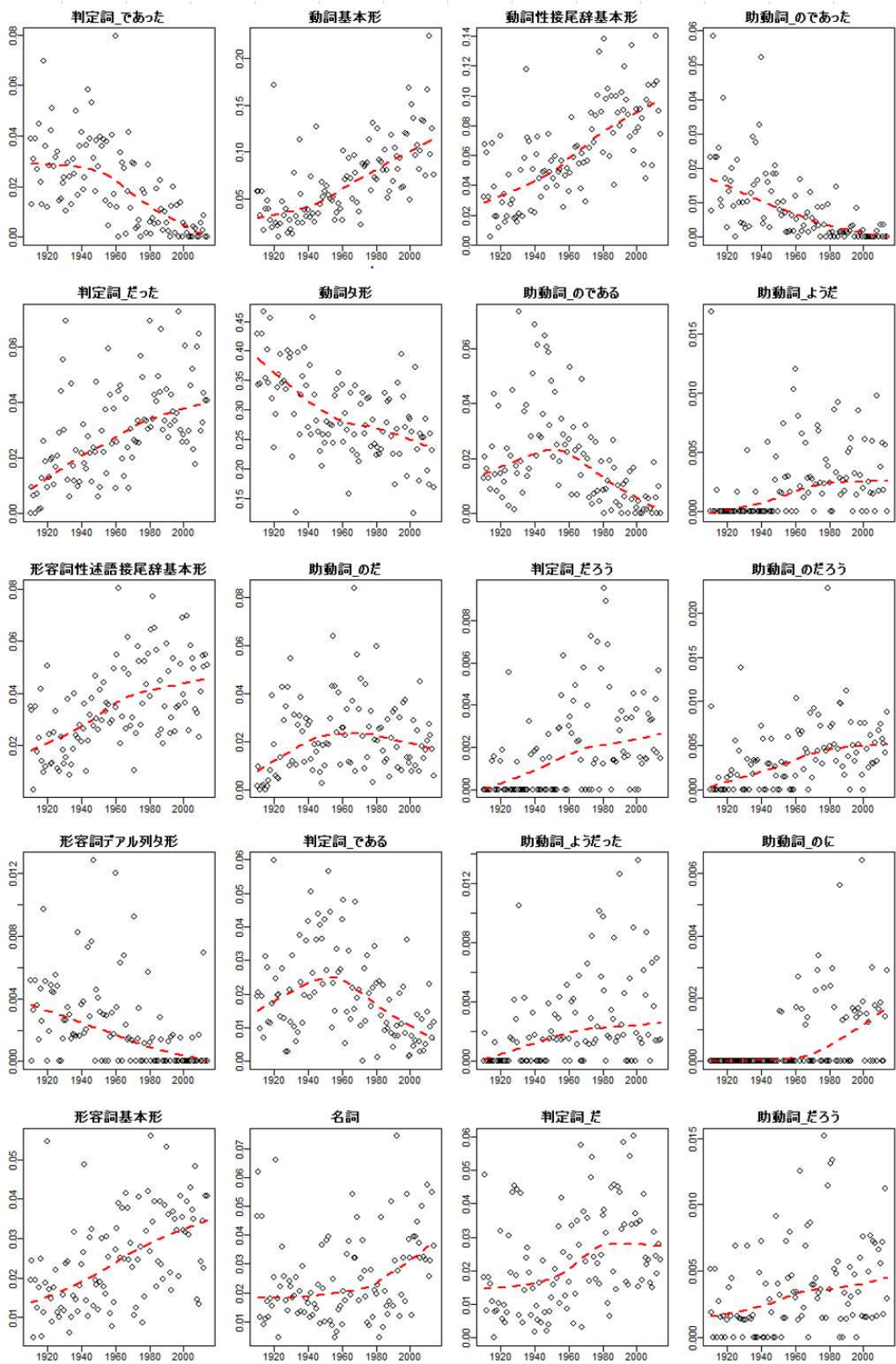


図 4.10 上位 20 項目の使用率の推移

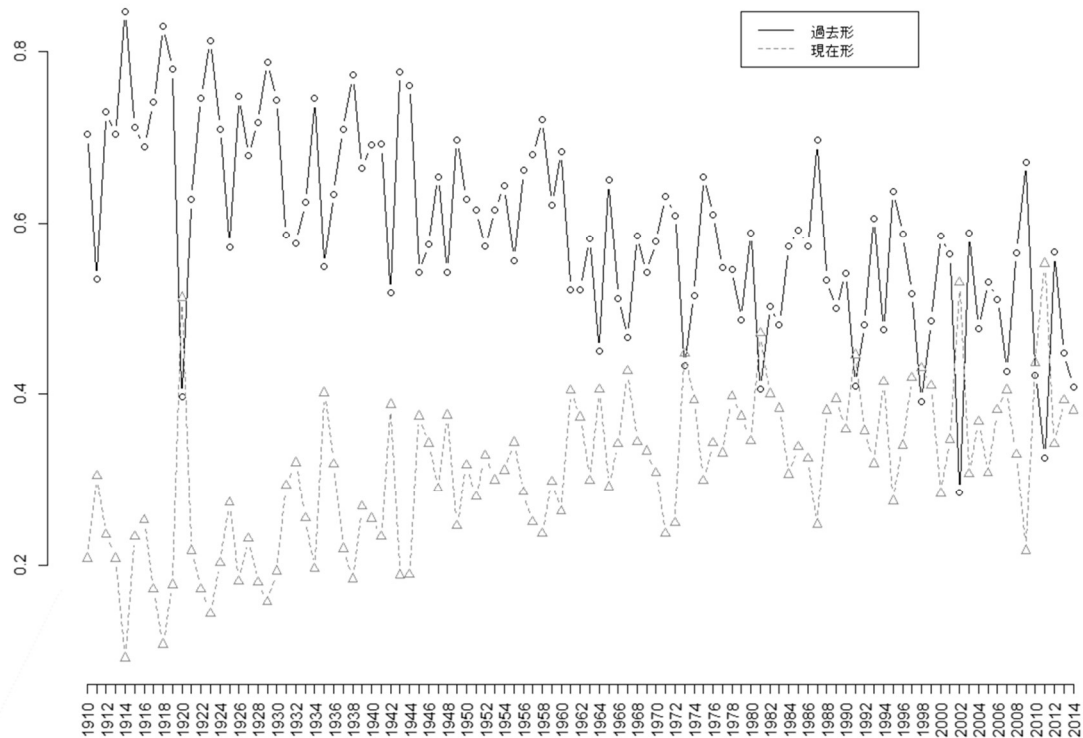


図 4.11 過去形と現在形の文末項目の使用率

3つ目は、「名詞」及び「助動詞_のに」の増加である。図 4.10 から見ると、体言止めの文末表現は 1980 年頃以降から増える傾向を示している。「助動詞_のに」で文を終わらせる表現は、全体の比率が多くないものの、同じく増加の傾向が見られた。

また、上位 20 項目のなかで、「助動詞_のだろう」「判定詞_だろう」及び「助動詞_だろう」は推量の意味を含んでおり、増加の傾向を示している。同じく増加の傾向を示した「助動詞_ようだ」「助動詞_ようだった」は、大まかに推量・推定と比況・比喩という 2 つの用法に分けられる。コーパス内の使用文をすべて抽出して、人手で分類した結果、9 割以上が推量・推定と判定された。この結果を踏まえて、上位 20 項目以外の項目に対しても、活用形のある項目及び助動詞を中心に、推量の意味を含む項目を調査した。その結果、「助動詞_らしい」は増加しており、「助動詞_らしかった」「助動詞_みたいだ」「助動詞_みたいだった」「動詞推量形」などの項目はやや増加傾向にあることが読み取れた。これにより、推量形式の文末表現の使用が増加していると推察できる。

4.5 考察

前節において、文末表現の多様性の変化を分析したうえ、回帰モデリングを行い、1910 から 2014 までの期間における文末表現の特徴的变化項目を特定した。そして、ランダムフォレ

ストフォレストによるモデルの構築に大きく寄与した項目を手がかりにして、見出された変化傾向をまとめた。続いて、観察された文末の変化は、何を意味するか、または、小説の文体の変化にどのように関わっているかについて分析する。

4.5.1 心理活動・思考体験の直接表出

小説の地の文では、推量の文末はよく作中人物或いは語り手の内的思考・意識の直接表現に使用する。そのなかで、「だろう」は人物・語り手の立場での思考・判断、「ようだ」「らしい」などは、事態・状況への認識などの表現に用いる。この類の文末形式の多用は、作中人物・語り手の知覚・判断をそのまま語ったりする傾向が強まっていると推察できる。

作中人物の心の内の直接表現に当たって、独白・呟きのような文体がよく用いられる。その叙法として、内的独白、体験話法などの諸形式が存在する。現代の小説は、「物語に描かれる現実の諸相のなかでも特に意識描写に格別の関心を向けてきたのであって、それがまた結果としてすこぶる精妙な表現手段を開発することにもなったのである」（シュタンツェル 1989: 123）と言われている。一方、作中人物・語り手の内的独白は、多くの場合、現在形が用いられる。山本（2016）は、小説の地の文末辞であるタ形とル形の使い分けについて、作中人物の知覚体験・内的独白・語り手を視座にした事態描写にル形を使用し、語り世界の客観的事実を外視点で提示する場合タ形を用いると指摘した。

「だ」体と「である」体の使い分けについて、於（2010）によると、「だ」体は特定の人に親しく話しかけるような叙述を書くとき、または、話し手の独話や自問自答などの心理活動を表すときに使われやすいのに対して、「である」体は客観性・恒常性を持つ事実の伝達或いは多数者の意見や社会的常識の記述に使われやすい傾向がある。

「だろう」、ル形、「だ」体の文末項目は、相対的に（ル形とタ形、「だ」体と「である」体と比較した場合）知覚体験や心理活動に使われやすいという共通する特徴がある。これらの項目はいずれも増加の傾向を示している。これにより、小説における心理活動・思考体験の直接表出が増加している可能性があると考えられ、今後、実験的研究を通して、更に調査するに値する。

4.5.2 時制変化及びその効果

波多野（1988）は、自然主義派作家の作品では圧倒的に過去時制が愛用されていることを指摘し、「自然主義は無技巧を標ぼうしただけに、特に技巧的に考えられた現在終止などは使おうとしても使えなかったものに相違ない」と解釈した。技巧を排した客観的な表現の方法とされているが、「単純過去形」で実現した客観描写の流行は一時的なものであった。4.4.3.3の分析により、20世紀前期の小説において過去時制が基調になっていたが、非過去形が時間

の推移につれてどんどん増えてくることがわかった。時制表現の変化は、小説における時間構造や叙述技巧の変化を示唆する。

4.5.2.1 過去形と現在形の交錯

時制使用の変化は、具体的な出来事の描写・叙述において、過去形と現在形との併用が好まれる傾向に関係していると考えられる。寺村（1984）は、過去形と現在形が入り混ざる叙述が、「読者を過去のある時に、つまりある舞台の前に、引き込む」という効果を持つと説明した。時制の交替は、語られる物語内部の時間的展開だけではなく、語りの視点変化にも関わっている。

- (1) 仁子さんが井を持ってきた。飯の上に庖丁で叩いた鱈と鰯、葱が盛られ、しょう油の香りがする。茄子の漬物が二切れ、端の方に載っている。箸で混ぜ、掻き込む。仁子さんは魚のあらを川に捨て、俎板と庖丁を洗い終ると義手を外し、柱にねじ込んである鉤にぶら下げ、奥の座敷に上がった。やがてそちらから流れてきた煙草の煙が、店の中をかすかに濁らせた。

(2011年『共喰い』)

例文(1)では、第1文「持ってきた」は、出来事の継起性を表している。それに続く「する」「載っている」「掻き込む」は非過去形で提示されている。これにより、出来事の現場から直接に知覚される光景が、実況報告として伝達されている。また、後ろに見る通りに、物語内部の出来事間の時間関係に戻り、「上がった」「濁らせた」を通して、時間の流れの前進性を提示する。一方、視点から見ると、冒頭の過去形の語りで超越的なポジションから場面を提示する。その後、視点を登場人物に合致させ、現場の光景を現在形でそのまま伝える。続いて、過去形に戻り、もとの俯瞰的な視点位置に戻って、ストーリーを展開させる。このように、時の関係は文の前後関係に従って自由に流れ、語りのポジションも常に変化することが感じられる。

地の文における時制が、20世紀初期の過去形中心から過去・非過去の併用へと移り変わったことは、作家たちが多様性を持つ表現を創り出し、修辭的な効果を得るために意識的に変化を求めてきた可能性が高いと推察する。

4.5.2.2 「カメラアイ」技法

動詞述語や形容詞述語の現在形で文を終わらせる表現の増加は、「カメラアイ」という技法の多用にも関係していると考えられる。「カメラアイ」は元々映像芸術的概念である。20

世紀に映画が飛躍的な発展を遂げ、言語表現のあり方にも影響を及ぼした。それによって、「カメラアイ」が物語技法的な術語としても用いられるようになった。

特に、情景描写の場合、語り手が知覚する状態や周りの光景を映像を展示するような形で描き出すには、現在形の述語が用いられることが多い。これにより、読者は語り手の視点のある場所へと連れ出され、あたかもその出来事が眼前で繰り広げられているかのような臨場感を体験できる。

(2) 床に野菜くずが散らかっていて、スリッパの裏が真っ黒になるくらい汚いそこは、異様に広いといい。ひと冬軽く越せるような食料が並ぶ巨大な冷蔵庫がそびえ立ち、その銀の扉に私にもたれかかる。油が飛び散ったガス台や、さびのついた包丁からふと目を上げると、窓の外には淋しく星が光る。

(1988年『キッチン』)

また、例(2)では、語り手＝主人公であり、視点を担っている「私」自身からは自分が見えないので、本来は、「その銀の扉に私にもたれかかった」と過去形のほうが自然である。しかし、原文では「その銀の扉に私にもたれかかる」と現在形の述語が使われており、レンズ越しに映る「私」を読者に見せているように見受けられる。現在形述語の増加は、このような表現技法の応用による影響もあると考えられる。

4.5.3 未完結感を持たせる終わり方

4.4.3.3より、名詞及び「助動詞_のに」を用いて文を終える表現が増加していることがわかった。体言止めは和文古来の表現技法である。『枕草子』の冒頭の「春はあけぼの」はその代表的例である。中村(2005:156)は、体言止めは「本来は名詞の後に何らかの記述が続くものを、あえて省略し、感極まったという気持ちを投影させる」ので、読者に余韻、余情を感じさせる効果があると指摘する。省略によって、ある種の未完結感が文章の空隙に醸し出される。「のに」で終わる文にも同じような未完結感が存在している。本来「のに」に続く記述は、「予測可能」あるいは「文脈から自明」(家田, 2005)といった原因により、省略されたり、落ちたりしたと理解できる。

第3章の助詞に関する分析により、述語を省略し、格助詞「へ」で文を終える表現が増加していることがわかった。文末における格助詞「へ」の増加は、本章の考察に見出された文末における名詞と「のに」の多用に呼応していると考えられる。このような形式の文末表現の増加は、余白・余韻を持たせたり、文末部に変化を持たせたりするという作家たちの意識的な配慮に関係していると推察できる。

4.6 まとめ

本章では、文末表現の経時的変化について考察を行った。 K 特性値を指標として、文末表現の多様性について測定した結果より、時期の推移に伴い、文末表現はより多様になってきていることがわかった。また、ランダムフォレストによるモデリングによって特徴的な変数項目を特定した。同じく、常体として扱われる「である」体と「だ」体の表現が異なる推移傾向を見せており、「である」体が減少しているのに対し、「だ」体のほうが増えていることがわかった。さらに、時制に関して、過去時制の表現が減少し、非過去時制の表現が増加していることがわかった。また、体言止めなどの文末表現様式が多用されるようになったことが明らかになった。特徴的な変化傾向から、小説の語り方や技法に関する変化の一端を窺い知ることができた。

第5章 接続表現の経時的変化に関するモデリングと分析

隣り合う2つの文は、通常、意味的につながっていると理解される。このつながりまたは関係性は、暗黙のままということもあるが、接続詞や接続詞と同様の機能を果たす連語といった接続表現によって示されることが多い。接続表現が、文章展開において、文脈の流れを明示し、文章の推移に論理的展開を与えるという役割を担っている。本章において、近現代小説における接続表現の使用変化について考察を行う。

5.1 接続表現の変化に関する先行研究

先行研究（塚原，1958；京極・松井，1973）では、日本語における接続詞は、はやくから発達したのではなく、ほかの品詞の派生により成立したとの見解が一般的である。中古の和文においては、接続詞たる表現の数量は少ない。例えば、国立国語研究所による「日本語歴史コーパス（平安時代編）」のなかで、接続詞とされる語彙は7語である。中世に入って、漢文訓読語の接続詞が日本語の文章に取り入れられるようになった。また、近代に至って、接続詞の使用は更に著しく発達した。近代以前の日本語の文章では、文の切れ目が曖昧ということがあがるが、翻訳を通じて摂取された西欧語の文体の影響により、日本語の文章に句読点が整備されると同時に、文と文を結ぶ働きをする接続表現が更に多く使われるようになった。

市川（1966：40）は「各時代の表現における接続の様相を分析することによって、日本語の表現における発想法の変遷、あるいは、日本語がどのように論理性を獲得してきたかという過程などを解明することもできる」と指摘した。接続表現の経時的変化及び異なるカテゴリーの文章における差異などに関しては、量的調査による手法がよく用いられる。例えば、江湖山（1981）は、数編の散文・小説及び数冊の和歌集を対象とし、古代和歌と近代和歌、古代の散文文芸作品と近代の散文文芸作品における主語・述語及び接続詞の調査を行った。その分析によると、接続詞の点では、散文文芸が近代作品では種類・数量・使用率のどの面でも、古典に対して飛躍的に多く、西洋文章から論理性を導入するという明治以後の文章表現の特色を表している。それに比べ、定型と抒情詩であるという和歌の本質に規制され、古典和歌と近代和歌との間にほとんど差がなく、接続詞が極めて少ないと報告されている。また、木坂（1987）は、明治37年～昭和33年の小説20作品の接続語を調査し、私小説が接続語の多様化への傾向を促すこと、昭和期以降の作品において接続詞の個性的な用法が増え、新鮮な文体効果をもたらしていることを指摘した。

現代小説の接続表現の使用について、石黒（2009）で、『新潮文庫の100冊』（1995年、CD-ROM版）及び『新潮文庫の絶版100冊』（2000年、CD-ROM版）に基づいて、137種

の接続表現について、社説・コラム・論文・エッセイ・小説・シナリオなどのジャンルごとに文頭における出現頻度を集計し、上位項目を比較した。このうち、小説では、「しかし」「だが」「でも」「ただ」の使用頻度が高かったと示された。一方、村田（2014）では、ブログ・新聞社説・論文・文学作品（小説）に対し、選定した35語の接続表現について、1文あたりの使用頻度を調べた結果、文学作品では「それで」「すると」「それとも」などの使用率が高かったと報告されている。これらの研究では、使用頻度の集計によって、現代小説における接続表現の使用特徴をある程度提示することができた。しかしながら、接続語句は時期と共に変化しつつあるため、たとえ同じく20世紀の小説にしても、20世紀初期と末期において異なる傾向が見られる可能性がある。このような変化の過程及び特徴を表示する研究はまだ見当たらない。そこで、本章では、近現代小説における接続表現の使用をマクロの視点から考察し、大きな傾向を捉えることを試みる。

5.2 接続表現のデータ

接続表現の範疇及び認定の基準は、諸学説によってさまざまであり、いまだ完全に統一されていない。佐久間（1992：9）は「接続表現」について、以下のように述べている。

「接続表現」とは、文章論における「接続語句」に当たるものであるが、品詞論における「接続詞」や「接続助詞」や構文論における「接続語」「接続句」に対する用語であって、その取り扱う範囲のやや広いものである。

本研究も基本的にこの定義を認めるが、本章では、文章内の文と文をつなぐ形式、つまり、「文の接続」に着目して分析を行うため、文と文を結ぶ接続詞・接続連語及びそれに準ずるものに限定する。

接続表現の抽出について、形態素解析器を利用して、品詞体系の「接続詞」にあたるものを抽出することが一案として考えられるが、形態素解析は品詞体系に基づいて行われるものなので、抽出される「接続詞」は限られている。例えば、短単位で設計された UniDic 辞書では、接続詞は現代語において意味の持つ最小単位で分割されている。そのため、「しかし」、「そして」、「ただし」などの短単位の語彙は接続詞として認定されるが、「だから」「しかしながら」などは「だ+から」「しかし+ながら」のように分割される。一方、UniDic 辞書より分割単位が長い JUMAN 辞書では、「だから」「しかしながら」は接続詞としてカウントされているが、「そうすると」などのような長い文字列からなる接続連語は、まとまった接続詞として取り扱われていない。また、「つまり」「例えば」「一方」などのような接続の

機能を担っているが、品詞体系上副詞、名詞などに分類される語彙も多く存在する。このように、形態素解析により得られた接続詞は、接続表現の一部分に限られる。

そこで、本研究では、国立国語研究所（2004）が公表した分類語彙表に収録されている接続表現の範囲で調査することにした。分類語彙表には 209 項目の接続語が収録され、累加・展開・換言・補充・転換・理由・選択の 7 つカテゴリーに分けられている。本研究では、文頭で用いられているもののみを分析対象としている。同語異表記の項目を 1 つとして集計する。例えば、「それ故」で言えば、「それ故」のほかに、「それゆえ」「其れ故」も含めて同じ項目として数える。接続表現として認定しがたいものが混入する可能性を減らすため、集計結果を目視で確認し、不適切なものを削除する。出現頻度が 0 である項目を取り除いた後、表 5.1 に示される 131 項目のデータが得られた。

表 5.1 接続表現リスト

かつ	また	そうして	そして	しこうして
それから	それに	そのうえ	おまけに	ただでさえ
のみならず	それどころか	そればかりか	しかも	まして
いわんや	で	ですから	だから	されば
だからこそ	さればこそ	そこで	それで	それゆえ
そのため	ゆえに	したがって	よって	それかあらぬか
こうして	かくて	かくして	すると	では
じゃあ	じゃ	それじゃ	そうしたら	それなら
そんなら	とすれば	ならば	しからば	そうすると
して見ると	というと	ついては	つきましては	つぎに
しかし	しかしながら	けれども	けれど	だけれども
だけれども	だけど	けど	が	だが
ところが	しかるに	でも	それでも	さりとて
そのくせ	くせに	それでいて	にもかかわらず	なのに
それなのに	とはいえ	とはいっもの	そのかわり	かえって
ひるがえって	いちめん	いっぽう	はんたい	ぎやくに
それにたいして	あるいは	または	さもなければ	それとも
すなわち	そく	いわば	いいかえれば	いってみれば
はやいはなしが	つまり	しょせん	けっきよのところ	つまるところ
ようするに	いずれにせよ	どっちみち	なんといつても	たとえば
なかんずく	なかでも	ただ	なお	ただし

もっとも	ちなみに	ねんのため	そも	そもそも
さて	ところで	それにしても	なにはともあれ	かんわきゆうだい
それでは	おもえば	おもうに	ときに	ここに
なぜ	なにゆえ	なんで	なんだって	なんとして
どうして	どうしたことか	なんぞ	なんとなれば	なぜなら
だって				

5.3 接続表現に関する分析

接続表現に関する先行研究について、馬場(2020)の「接続詞関係研究文献一覧」には1000を超える文献がまとめられている。それらの文献から見ると、接続表現の研究において、内省の研究手法がまだまだ主流であることがわかった。近年、コーパスに基づいた量的データを使用した分析が増えてきたが、データの解析に関して、接続表現の使用頻度または使用率の集計にとどまっているものが多い。研究の客観的妥当性を高め、内省分析及び単なるデータ集計だけでは発見できなかった新たな事実を見つけ出すために、計量分析の新手法を併用し、多角度からアプローチする必要がある。

本章の分析では、まず、接続表現の推移状況を大まかに把握するために、使用率及び語彙の豊富さという2つの指標の変化を計算し、プロットで示す。次に、接続表現の使用について、ランダムフォレストによる回帰分析及びトピックモデルを利用して、モデリングを行う。接続表現の使用に関するモデルを構築したうえ、重要な変数を特定し、異なる時期に見られる接続表現の特徴を明らかにする。最後に、接続表現のカテゴリーの変化を把握するため、層別した接続表現のデータを用いて主成分分析を行う。

5.3.1 接続表現の使用率及び多様性の変化

本項では、接続表現の使用率及び多様性の変化について検討する。

まず、使用率について、各年度の作品から得られた接続表現の出現頻度を文の数で正規化して、その結果を接続表現の使用率とする。年度毎の接続表現の使用率のプロットを図5.1に示す。

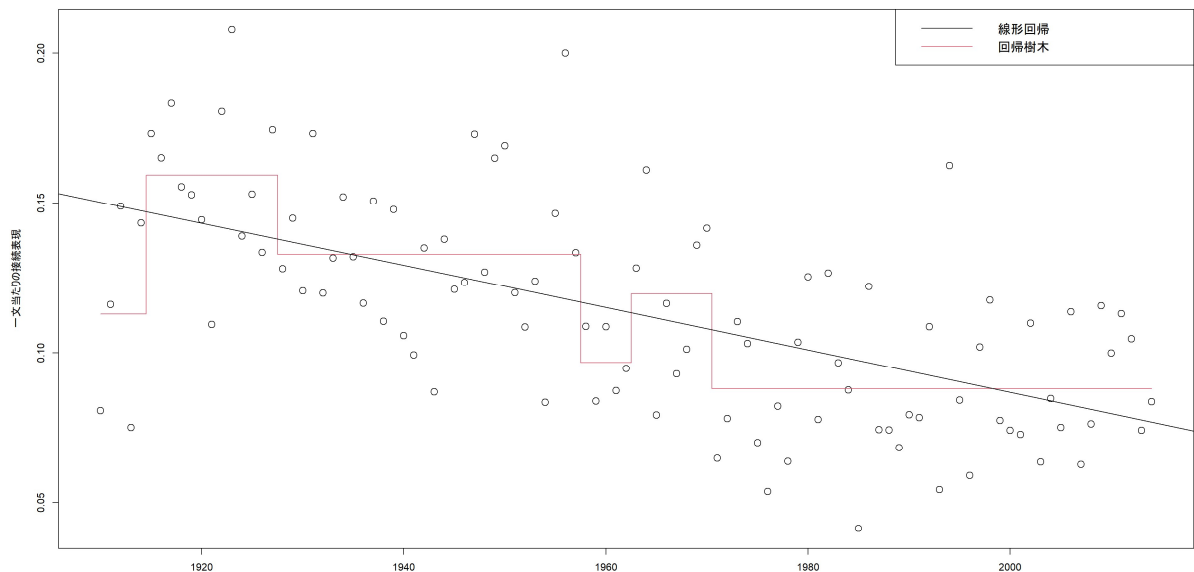


図 5.1 一文あたりの接続語の推移

先行研究の木坂（1987）は、明治期の小説に比べ、昭和期の作品において、接続詞の個人的な用法が増え、新鮮な文体効果をもたらしていると指摘した。本研究で構築した通時コーパスに基づいた調査から見ると、昭和前期において、接続表現多く使用される傾向が見られるが、それが長く続かなかった。図 5.1 により、1920 年代及び 1930 年代は比較的に高い値を示すが、全体的から見ると、1930 年代以降は接続表現の使用率が減少しつつ、1980 年代以降は相対的に安定していることが窺えた。

接続表現の使用が減少する原因、または、それが小説の文体効果との関わりについて、複合的な視点から考察する必要があるが、文学者及び文芸評論家または研究者の文章論から多少示唆が得られると考える。

「さて」とか「ところで」とか「実は」とか「なんといっても」とか「とは言うものの」とか、そういう言葉を節の初めに使った文章は、如何にも説話体的な親しみを増しますが、文章の格調を失わせます。（三島，1973：145）

接続詞を多く用いると、論理的にととられるが、その反面、描写性が薄れて説明調になったり、簡潔さが失われたりする場合もある。（市川，1978：78）

安易な接続語の使用は、一文をしっかりと書こうとする努力を失わせる。接続語が、前後の文をぼかしてしまう。（馬場，1983：104）

昭和後期の小説家や評論家による文章論では、文章の筋をはっきりさせることができる半面、深みや余韻などを失わせる恐れがあるので、その多用を慎むべきという意見が多く見られる。図 5.1 の分析により、接続表現の使用を控える傾向が強まっていることがわかったが、接続表現の使用率は延べ語数のみを考慮した指標であるから、すべての接続表現が一様に減らされているかどうかは使用率だけでは判断できない。

続いては、延べ語数及び異なり語数両方とも考慮した語彙の豊富さの指標を用いて、接続表現の多様性の変化を検討する。K 特性値の推移を図 5.2 に示す。図 5.2 から見ると、時期の推移につれ、接続表現がより多様化してきた傾向が見られる。以上の分析により、接続表現の出現頻度が減少するものの、接続表現の項目間のバランスがよりよくなったと判断できる。

変化の状況をさらに詳しく確認するため、K 特性値と年度との関係を回帰樹木で推定する。樹木回帰関数としてのあてはめ結果を図 5.2 に赤線で示し、回帰樹木の構造を図 5.3 に表示する。図 5.3 では、回帰樹木は、まず、年度 1982.5 未満/以上で 2 つの部分に分かれている。そして、1982.5 未満の部分は、1918.5 及び 1941.5 を境界に分割されている。

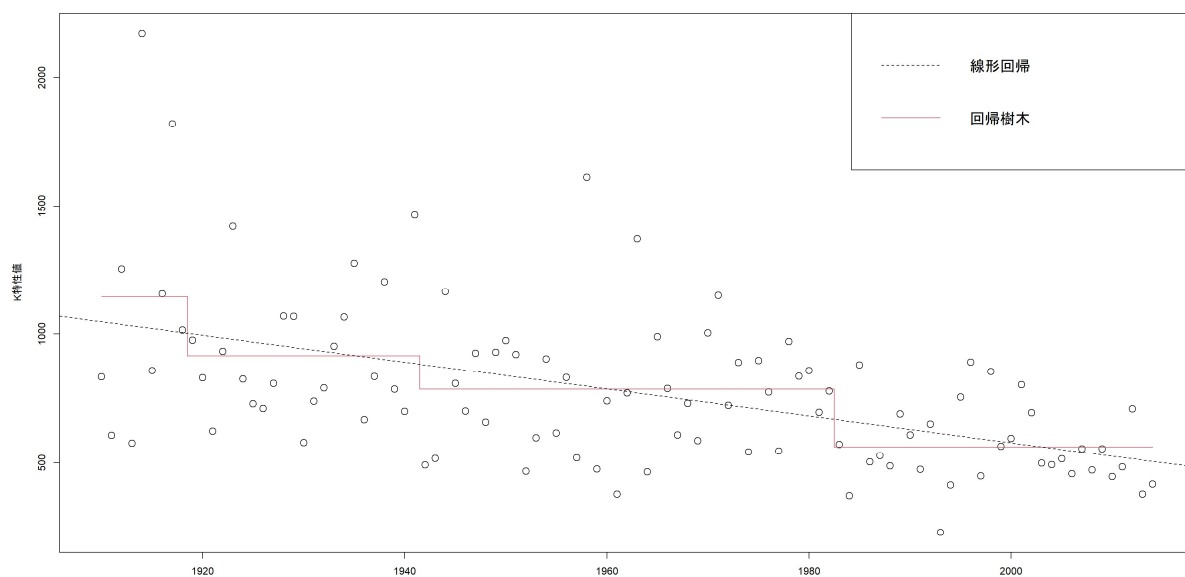


図 5.2 接続表現の K 特性値の推移

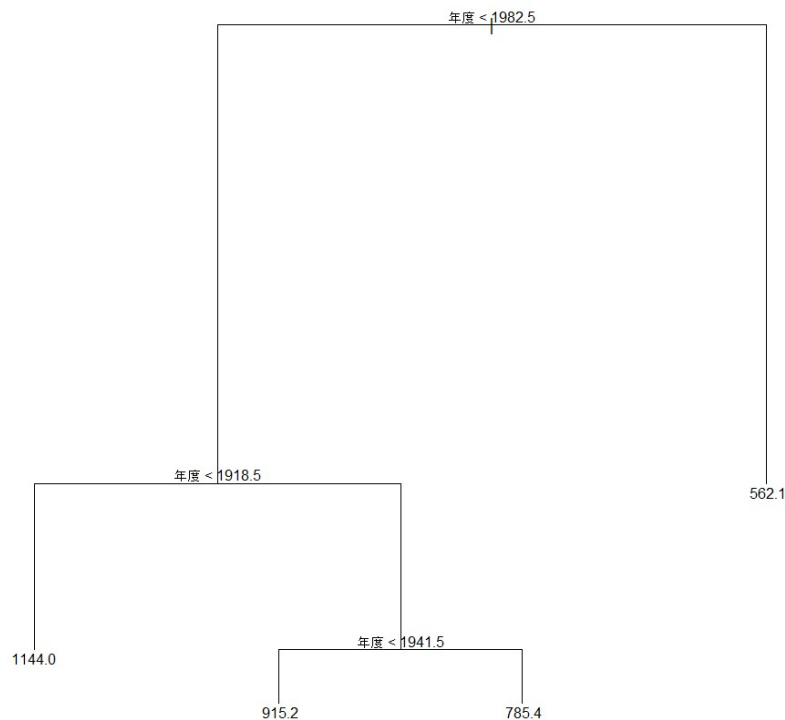


図 5.3 回帰樹木の構造

5.3.2 接続表現の変化に関するモデリング

本項では、接続表現の使用に関して、モデルを構築する。構築したモデルを踏まえて、異なる時期に見られる接続表現の特徴を明らかにする。

5.3.2.1 ランダムフォレストによるモデリング

コーパス内の作品の初出年度を目的変数、131 項目の接続表現を説明変数として、ランダムフォレストを利用して、回帰モデルを構築する。そして、構築した回帰モデルに寄与する上位の変数を重要な変動要素として抽出する。ランダムフォレストの回帰モデルは、元データの分散を約 65%説明している。また、その残差（絶対値）の四分位数は、それぞれ 0.10 (Min), 2.40 (IQ), 4.60 (Median), 8.20 (3Q), 21.90 (Max) であり、平均値 (Mean) は 5.57 である。

回帰木のノードの純度の増分に基づいて計算した変数の重要度の結果を図 5.4 に示す。図から見ると、「そして」「すると」「そうして」「けれども」「でも」「が」「しかし」「それから」「また」「だから」「もともと」「それでも」「そこで」「だが」「それで」が上位の 15 項目となる。

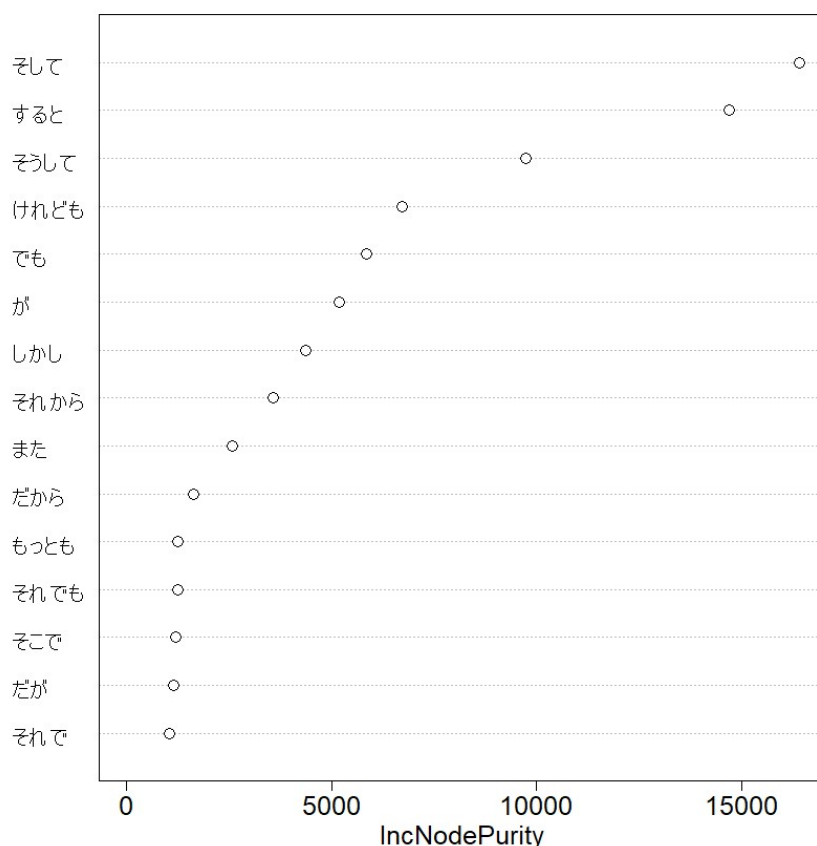


図 5.4 回帰モデリングに寄与する変数（上位の 15 項目）

5.3.2.2 構造的トピックモデルによるモデリング

ランダムフォレストによる回帰モデリングの結果を対照・比較するため、接続表現のデータセットに構造的トピックモデルを適用し、得られたトピックの経年変化及びトピックを構成する重要語を考察する。

トピックモデルは、トピック数を事前に指定する必要がある。トピック数の決定について、現在はこれといった決まりがなく、研究者が研究目的及びデータの実際状況に合わせて調整することが一般的である。データ駆動型のアプローチを使用して、トピックの数を推定する方法も提案されている。例えば、期待対数尤度 (held out likelihood) , 残差分散(residual dispersion), 意味的コヒーレンス(semantic coherence), lower bound (変分下限) などを計算し、その結果を参考にして、それに基づいて、指定した範囲内のトピックの数を比較し、最適な数を決める方法がある。

トピック数の範囲を 2 から 4 までに設定し、関数 `searchK` でトピックを推定した計算結果のプロットを図 5.5 に示す。モデルの尤度を最大とする Held-out Likelihood 及び意味情報

の一貫性に関する Semantic Coherence は、値が最大となるトピック数を推奨する。それに対し、モデルの推測残差を最小とする Residuals, 変分下限を示す Lower Bound は、値が最小となるトピック数を推奨する。この 4 つの指標を総合的に評価した結果、理想的なトピック数が 2 に該当すると判断される。

トピック別の遷移傾向を図 5.6 に示す。トピック 1 は時期の推移につれ減少するのに対し、トピック 2 は増加しつつある。

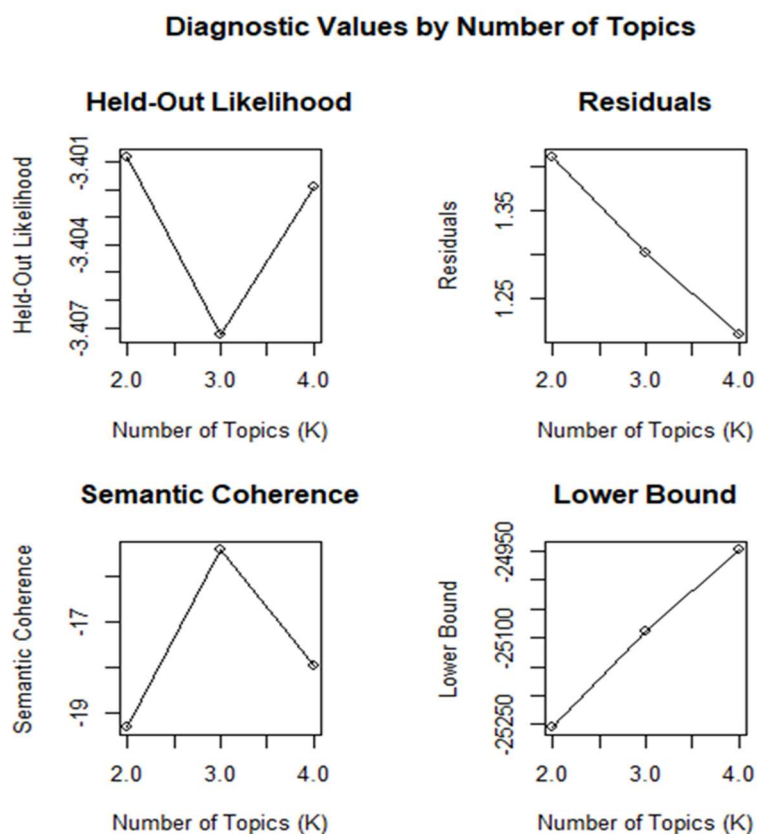


図 5.5 トピック数の推定結果

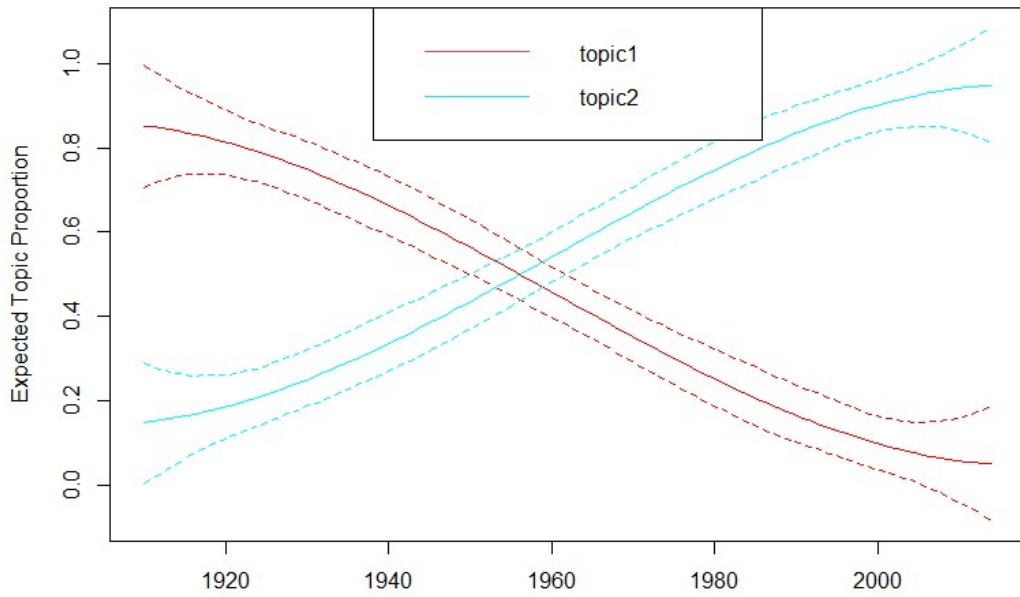


図 5.6 トピック別の遷移状況

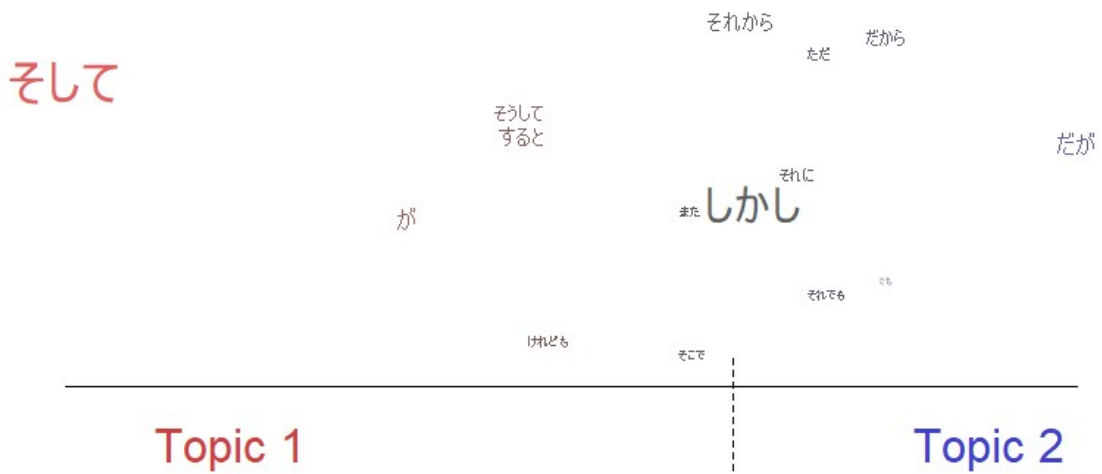


図 5.7 推測されたトピック別の重要語

2つのトピック間の重要語を比較するため、各項目がトピック1とトピック2における出現確率を計算し、差が大きい項目をトピック別の重要語として図5.7にプロットする。前節の回帰モデリングで選出した上位15項目の変数と対照して見ると、図5.4に示された第11番目の「もっとも」と第15番目の「それで」を除き、ほかの13項目はすべてトピックモデ

リングによる結果と同じである。これにより、回帰モデリングとトピックモデリングの結果はほぼ一致すると判断される。

また、図 5.7 において出現確率が高い語が突出して大きく示される。トピック 1 において順接・累加を表す「そして」が一番顕著な重要語であるのに対して、トピック 2 において逆接・反対を表す「しかし」は最も特徴的な項目として表示される。また、トピック 1 を特徴づける「そして」がほかの重要語と離れているところに位置しているが、トピック 2 を特徴づける「しかし」の周りに多くの重要語が分布している。

5.3.3 層別した接続表現についての多変量解析

前項では、モデリングにより、特徴的な変化項目を特定できた。続いて、本項では、カテゴリ一別から接続表現の経時的变化の特徴が見られるかどうかを分析する。分類語彙表には、接続語が累加・展開・換言・補充・転換・理由・選択という 7 つのカテゴリに分けられている。集計した接続表現データを 7 つのカテゴリにまとめ、相関係数を用いた主成分分析を行う。

考察には第何主成分までを使うかを定めるため、平行分析を行った。その結果により、3 つ目の主成分以降は固有値の変化がほとんどないので、2 つ目までの主成分を採用すると判断した。第 1 主成分と第 2 主成分の主成分得点と負荷量の散布図をそれぞれ図 5.8 と図 5.9 に示す。



図 5.8 主成分得点の散布図

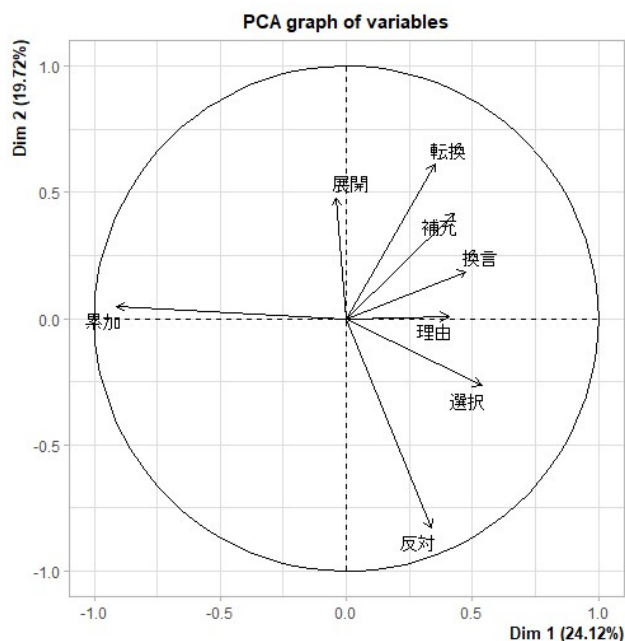


図 5.9 主成分負荷量の散布図

図 5.8 を見ると、第 1 主成分（横軸）で、左側に 1950 年代以前の作品が多く集まっているのに対し、右側に 1960 年代以降の作品が多く見られる。そして、図 5.9 より、左側の作品群に「累加」の接続表現が多く、右側の作品群に「反対」「選択」「理由」「換言」「補充」「転換」の接続表現が多く使用されているということがわかる。

5.4 考察

前節の分析より、全体的には、接続表現の使用頻度は減少していると同時に、接続表現の分布はより均衡がとれていることが観察された。その原因について、以下の 2 つの可能性が考えられる。

まず、接続表現の使用は、文章の展開や語調をなめらかにすることができる一方、文と文との間の余白を狭く、浅くする。それに対し、接続表現なしに文と文の関係づけを明確に示さない場合、文章が簡潔になると同時に、文間の余白も生まれるため、読み手に自分で余白を埋めることが求められる。接続表現を少なくすることは、深く広い文間の余白を作り出し、小説の叙事性や物語性を強化することにつながっている可能性があると考えられる。

一方、接続表現は、文と文をつなぐ道具の 1 つであるが、必ずしも唯一のものではない。文の接続の方法は、ほかにも多数存在する。例えば、代名詞指示や同一語句を利用して、前文の内容を後文に持ち込んで、2 つの文を内容的に関連付ける。また、文末に「のだ」や「からだ」などを使用するによって、接続詞なしで文の関係を表示することができる。市川(1978)

は、2文以上からなる文章の統合方法は、線条的直接的統合、線状の間接的統合、非線状の直接的統合、非線状の間接的統合に分けられ、接続詞による文の接続は線状の直接的統合であると指摘した。接続表現を用いることなく、ほかの語彙的手段や方法で2つの文のつながりを示すことが可能なので、接続表現の減少は、文脈のあり方の多様化に関係している可能性も考えられる。

5.3の分析により、「そして」は20世紀前半の小説を一番特徴づける項目であるが、「そして」を始めとする累加型の接続表現は、20世紀後半以降の小説には減少する一方であることがわかった。累加型の接続表現は、前の事柄に、同類の事柄を付け加えて述べたり、前後の事柄を並べて述べたりするために使われる。国語学者の阪倉（1983）は、かつて『源氏物語』を例として、和文の「延伸」という展開の特質を指摘した。要するに、古典日本語は、次から次へと延ばしつぎ足して行くというかたちの表現が基本になっている。節と節、文と文は、ぎくしゃくした連なりがなく、ただ平面的に並べられ、後ろへ後ろへと続いていくのである。20世紀前半の日本語文章には、「そして」を始めとする接続表現が多用されたのは、2つの事象が同時的なものか継起的なものかということだけを表示する平板的な思考の影響によるものと考えられる。哲学者及びフランス文学者である森有正（1972）は、ヨーロッパ文章は肯定、否定、疑問、懐疑、推測などを駆使して、論理の結構を組み立てることに比べれば、日本語は主体と対象を区別せず、修飾的副文章をひたすら主体側に結びつけて、読者に疊感的な感覚を与えていることを指摘している。この点は、「累加」カテゴリーに属する接続表現の多用にも反映されている。一方、20世紀後半及び21世紀のはじめの文章には、「しかし」など、前の事柄に反する事柄を述べる反対型の接続表現のほか、選択型、理由型、換言型、補充型、転換型などが併用されるという特徴が見られる。

5.5 まとめ

本章では、接続表現の経時的変化について考察を行った。1文あたりの接続表現の使用率から見ると、接続表現の使用率が減少している傾向を示していることがわかった。一方、 K 特性値の指標を用いた分析では、接続表現の使用率が減少しているが、接続表現の項目間のバランスがよりよくなったと判断された。一方、モデリング及び多変量解析の結果より、「そして」は20世紀前半の小説を一番特徴づける項目であることがわかった。それに対し、20世紀後半及び21世紀のはじめの文章において、「しかし」など前の事柄に反する事柄を述べる反対型の接続表現が特徴的な項目になると同時に、接続表現の分布もより均衡がとれている。

第6章 個人文体の意図的变化に関する計量的分析

6.1 本章の背景と目的

第3章～第5章では、近現代日本語小説の通時コーパスに基づいて、助詞・文末表現・接続表現といった文体要素の使用様態について分析を行った。その結果、小説の文体には明らかな経時的变化が認められた。本章では、個人よる文体の意図的变化を中心に分析を行う。

現代作家の水村美苗は海外で成長期を過ごしていたが、日本近代文学を読み耽りつつ育った。20年を超える海外生活を経験した後、日本へ戻ってきたら、日本で使われている日本語と近代日本文学を通じて自分が慣れ親しんできた日本語のレベルの格差を強く感じた。その後、彼女は小説家としての道を歩み始め、10代からくり返し読んできた日本近代文学の蓄積をバネに、夏目漱石の遺作を書き継いだ『続明暗』をはじめとする一連の作品を発表した。水村（2020）は、自分が日本の近代文学の歴史を振り返りつつ、日本近代文学の最後に来た者の一人として書いていると述べており、自身のホームページ³に「水村美苗・日本語で日本近代文学を書く小説家」という自己紹介文を掲げている。

本章では、水村の『続明暗』をはじめとする作品を分析対象として、文体の意図的变化を中心に分析を展開する。主に、以下のことを明らかにする。

まず、水村は社会的文体の自然変化を意識しながら、時代の流れに流されないように意図的な文体づくりを行っているが、その試みはどこまで到達できるかについて、第3章～第5章に用いられたコーパス及び得られた分析結果を利用して、分析・検証を行う。

また、創作時期から見ると、『明暗』と『続明暗』2作品には、74年の隔たりがある。2人の作者の生きた時代環境が異なり、個人経歴と知識素養や個性も大きく異なるため、文体模倣に不可避な言語制限や限界が存在することが考えられる。この点から考えると、『続明暗』と『明暗』の文体は、似ていながらもどこか異なる部分があると推測できるが、『続明暗』の文体が何をどこまで似せているが、水村はどのように自分の文体を変えて漱石の文体に似せるのか、または、どのような差異があるかという問題は明らかにされていない。なお、『明暗』と『続明暗』のストーリーの展開や構成にどんな変化が見られるかに関して、伝統のテキスト読解による内容分析（関，2010）はあるが、計量的な研究は見当たらない。『明暗』『続明暗』のような何十万文字の長編小説について、その語り内容及び構成における関係が長い時間軸の中にどのような変動を見せるか、テキストを構成する言葉の数値の変動をもとに把握できれば、物語構造に関する作品の分析に新たな可能性を切り開く可能性がある。

本章の構成としては、まず、水村美苗の『続明暗』とその模倣対象作品『明暗』、水村の他の2作品（『本格小説』『母の遺産』）に対し、第3章から第5章で得られた助詞・文末表

³ <https://www.mizumuraminae.com/about-japanese> (2021年12月12日アクセス)

現・接続表現の使用状況に関する回帰モデルを利用して、作品の発表年度を推定し、水村作品が近現代文学の時間軸上の位置を確認する。

その次に、テキストから文体特徴に関わる言語パターンを抽出し、計量分析を通して、『明暗』『続明暗』における言語表現の類似度及び差異、『続明暗』と水村のほかの作品との関係を把握し、テキスト間の相互関係を視覚化する。最後に、『明暗』『続明暗』2作品におけるトピックの時系列の変化を捉え、トピック構成の仕方の差異を比較する。

6.2 回帰モデルによる水村作品の時期推定

データマイニングでは、既知のデータに関する回帰モデルを作成した後、構築したモデルを用いて、未知のデータの特徴を新たに予測することができる。例えば、過去のデータを分析して、将来の株価や経済指標などを予測する研究が多く見られる。文学作品の場合、作品の執筆時期の推定問題などに応用されている。本節では、通時コーパスを元に構築した回帰モデルを用いて、作家個人の作品を時系列における位置（時期）を推定し、その結果と実際と比較する。

夏目漱石の『明暗』と水村美苗の『続明暗』『本格小説』『母の遺産』に対し、第3章～第5章と同じようなやり方で、4編の対象作品から、それぞれ助詞・文末表現・接続表現のデータを抽出する。その後、第3章～第5章の助詞・文末表現・接続表現の分析で構築した回帰モデルを用いて、『明暗』『続明暗』『本格小説』『母の遺産』に対し、発表年度の推測値を求める。得られた推定結果を表6.1に示す。

表6.1から見ると、夏目漱石の『明暗』（1916年）は、発表年度が20年代～30年代初期と推定された。第3章～第5章におけるモデルの残差評価の結果により、残差の多くはおおむね正負20の区間内に配置されており、その結果は本節の『明暗』の年度推定の結果とほぼ一致している。一方、水村の『続明暗』（1990）は40年代～50年代、『本格小説』（2002）は40年代～80年代、『母の遺産』（2012）は40年代～70年代の作品と推定された。この3作品の推定年度は、いずれも実際の発表年度より早かった。夏目漱石の『明暗』と水村の『続明暗』の推定年度には約20年の開きがある。20年は、第3章と第5章におけるモデルの残差とほぼ一致するので、これが推定残差と見なすことも可能である。一方、水村が自分の文体を意図的に夏目漱石の『明暗』の文体に変化させようと努力したにもかかわらず、このような要素においては、配慮または力が及ばなかったのではないかとすることも考えられる。

また、通時コーパスの作品群を用いて対照した場合、水村の作品は、近現代文学の時間軸において、実際の発表時期より前の作品であるように見えると考えられる。そのなかで、意識的に漱石の文体を真似て書いた『続明暗』は、一番古めかしく見えることは、興味深い事実として、指摘できる。

表 6.1 回帰モデルによる時期推定の結果

作品名	発表年度	文体要素	推定結果
明暗	1916 年	助詞	1931.7
		文末表現	1922.8
		接続表現	1932.0
		助詞+文末表現+接続表現	1930.5
続明暗	1990 年	助詞	1954.4
		文末表現	1933.7
		接続表現	1940.4
		助詞+文末表現+接続表現	1947.7
本格小説	2002 年	助詞	1983.9
		文末表現	1940.7
		接続表現	1976.3
		助詞+文末表現+接続表現	1966.1
母の遺産	2012 年	助詞	1974.3
		文末表現	1945.3
		接続表現	1964.3
		助詞+文末表現+接続表現	1963.1

6.3 『明暗』『続明暗』における言語表現の比較

本節では、『続明暗』の文体が『明暗』とどこまで似ているか、また、どのような差異があるかを調べるため、文体に関わる言語項目をもとに分析する。第 2 章の 2.1.3 項で説明したように、『続明暗』の文体と『明暗』との類似度を顕在化させるため、漱石とおよそ同年代に活躍していた 5 名の近代作家の作品 (5 人×10 編) を比較として用いる。また、『続明暗』と水村のほかの作品との関係を把握するため、水村の『本格小説』『母の遺産』(以下水村 A, 水村 B と略称する) も比較作品として取り入れる。

作品の特徴をより明確に表出させるため、長編小説である『明暗』『続明暗』『本格小説』『母の遺産』という 4 つの作品に対し、それぞれ文字数がほぼ均等になる 10 個のファイルに分割する。5 名近代作家による 50 編の作品は短編小説であるため、分割せず作品ごとに独立したデータとする。

文の長さ、漢字の出現率、タグ付き形態素、品詞・記号、文節パターンといった項目についての分析結果を以下に示す。

6.3.1 文の長さ

『明暗』，『続明暗』，水村のほかの2作品，泉鏡花などの等5名の作家の作品群に対して，ファイルごとに総文字数を文の数で割って平均文長を算出する．図6.1は，平均文長をボックスプロットで示したものである．『続明暗』の平均文長は，『明暗』とほぼ同じ水準で，水村の作品Aと作品Bより低い．

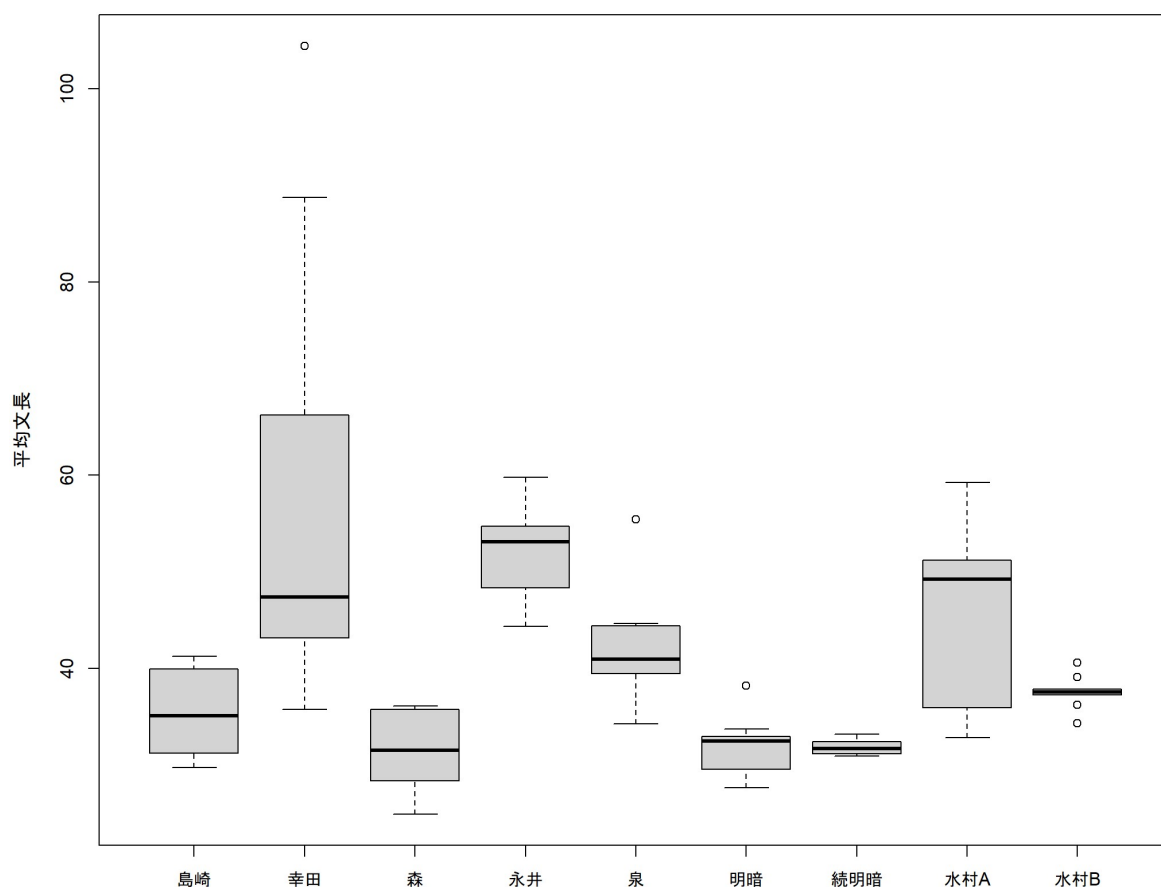


図6.1 平均文長のボックスプロット

『明暗』『続明暗』の平均文長が相似していることを確めたあと，続いて，この『明暗』『続明暗』及び水村の作品Aと作品Bを中心に，文の長さの分布状況を観察する．文字数を5つずつ増やししながら，各長さの文の数を集計する．すなわち，文字数が1~5の文，文字数が6~10の文といったように分けて数える．なお，文字数が100を超える文をまとめて1つの項目として数える．集計したデータを作品ごとの文の総数で割り，相対度数データに変換し，図6.2の折れ線グラフで示す．

図 6.2 に示された曲線の曲り具合から見ると、『続明暗』が『明暗』にもっと近似していることが読み取れる。水村の作品 A と作品 B における文の長さの分布は完全一致しているわけではないが、水村が『続明暗』を書く時、文の長さに傾注を配分し、意識的に『明暗』に合わせていることが推察できる。

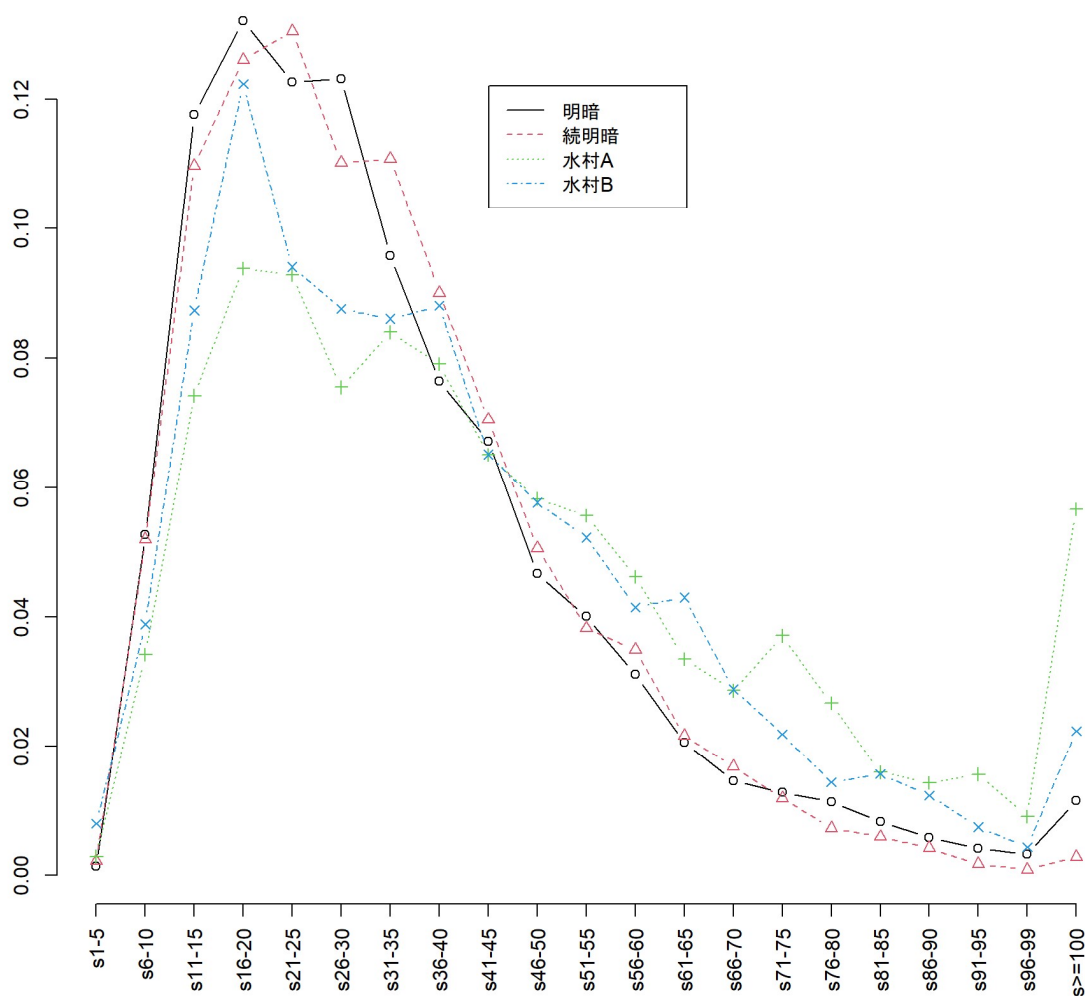


図 6.2 文の長さの分布

6.3.2 漢字の出現率

現代の日本語の文字体系は、主に、漢字、仮名（平仮名・カタカナ）、ローマ字という 3 種が用いられる。漢字が文の中で占める割合は、「文章表記の様相を手軽に示す指標」（佐竹，1987:109）とされる。一般的には、漢文の素養が高く古典に詳しい人は、漢字を多く使用すると言われているが、文学を書く際にあえて仮名書きにして柔らかい雰囲気を演出したり、特殊なニュアンスを伝えたりするようにする場合もあるから、作家または作品ごとの差が存

在すると考えられる。一方、安本（1963）は、1900年から1954年までの計100篇の小説における漢字の出現率を調査した。その結果、漢字の使用が減少していることが明らかになり、この傾向が続くとするなら、2191年に漢字含有率はゼロになると予測された。これにより、明治・大正期の作家の文章に比べて、全体的には若い世代の作家の文章は漢字が少ないことは推察できる。

本項では、『明暗』、『続明暗』、水村のほかの2作品、泉鏡花などの5名の作家の作品群に対して、ファイルごとに漢字の総数を総文字数で割って算出し、図6.3のボックスプロットに示す。図6.3から、水村の作品A、Bは、漢字の出現率が明らかに『続明暗』とほかの近代作家の作品より低いことがわかる。水村が『続明暗』を執筆した際、漢字を多く使用するよう配慮していたことが推察できる。

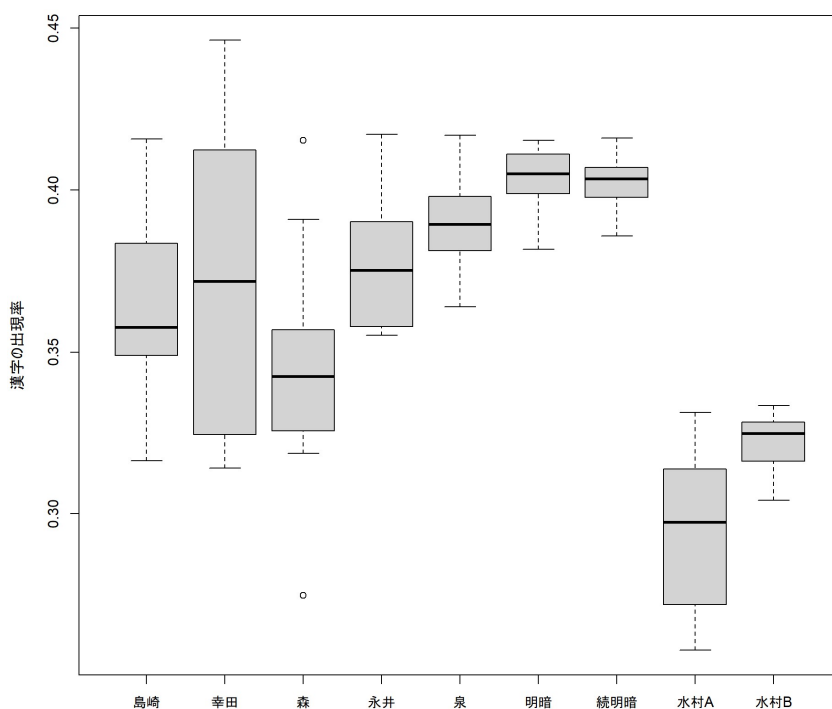


図 6.3 漢字の出現率のボックスプロット

6.3.3 タグ付き形態素

文の長さにつき、形態素の頻度について、語彙レベルの計量分析を加える。本項では、物事の形態や性質を表す「形容詞」「形容動詞語幹」「自立動詞」「非自立動詞」「サ変接続」「副詞」といった類の形態素項目を中心に、階層的クラスター分析を行う。本項では、単語長と曖昧性解消との観点から考慮し、形態素解析器 MeCab（デフォルトの IPA 辞書）を用いた。

全ての形態素項目を集計すると、項目数が多くなるだけではなく、データ行列の中にテキストの個性づくりにほとんど効かない項目が多く含まれる。そこで、出現頻度が計 45 回以下の形態素項目を 1 つの項目 (others) にまとめ、計 565 個の変数を使って、分析を実施する。階層的クラスタ分析に使われたデータは、全部相対頻度に変換されたものである。クラスタ分析では、最も似ているグループが先ず小さな階層を作り、その階層を含む形で別のグループとの階層が定まる。このようにして各組合せが順番に括られ、最後に、データ全体を包括するクラスタが階層的に形成されることになる。

タグ付き形態素のデータで行った階層的クラスタ分析の樹形図を図 6.4 に示す。樹形図の横軸 (上側) の目盛は、クラスタ間が結合した際の距離を表している。図 6.4 を見ると、水村の作品 A と作品 B は、目盛数が約 0.6 の位置で結合し、1 つのクラスタを構成した。一方、『続明暗』と『明暗』は、目盛数が約 1.05 の位置で、1 つのクラスタを構成した。そして、水村の作品 A・作品 B クラスタは、『明暗』『続明暗』を含むクラスタと目盛数が約 1.7 の位置でつながり、最終的に 1 つの木になった。図 6.4 により、『明暗』と『続明暗』の距離は、水村のほかの作品と『続明暗』の距離と比較して、圧倒的に近いことがわかった。これにより、水村が自身の他作品と『続明暗』では大きく文体を変化させていると推察できる。また、『明暗』と『続明暗』は、それぞれ小さなクラスタを成してから、目盛数が約 1.05 の位置で結合しているので、『明暗』の状態や性質を表す語彙が『続明暗』にある程度活かされているものの、2 者の間に相違が存在すると考えられる。

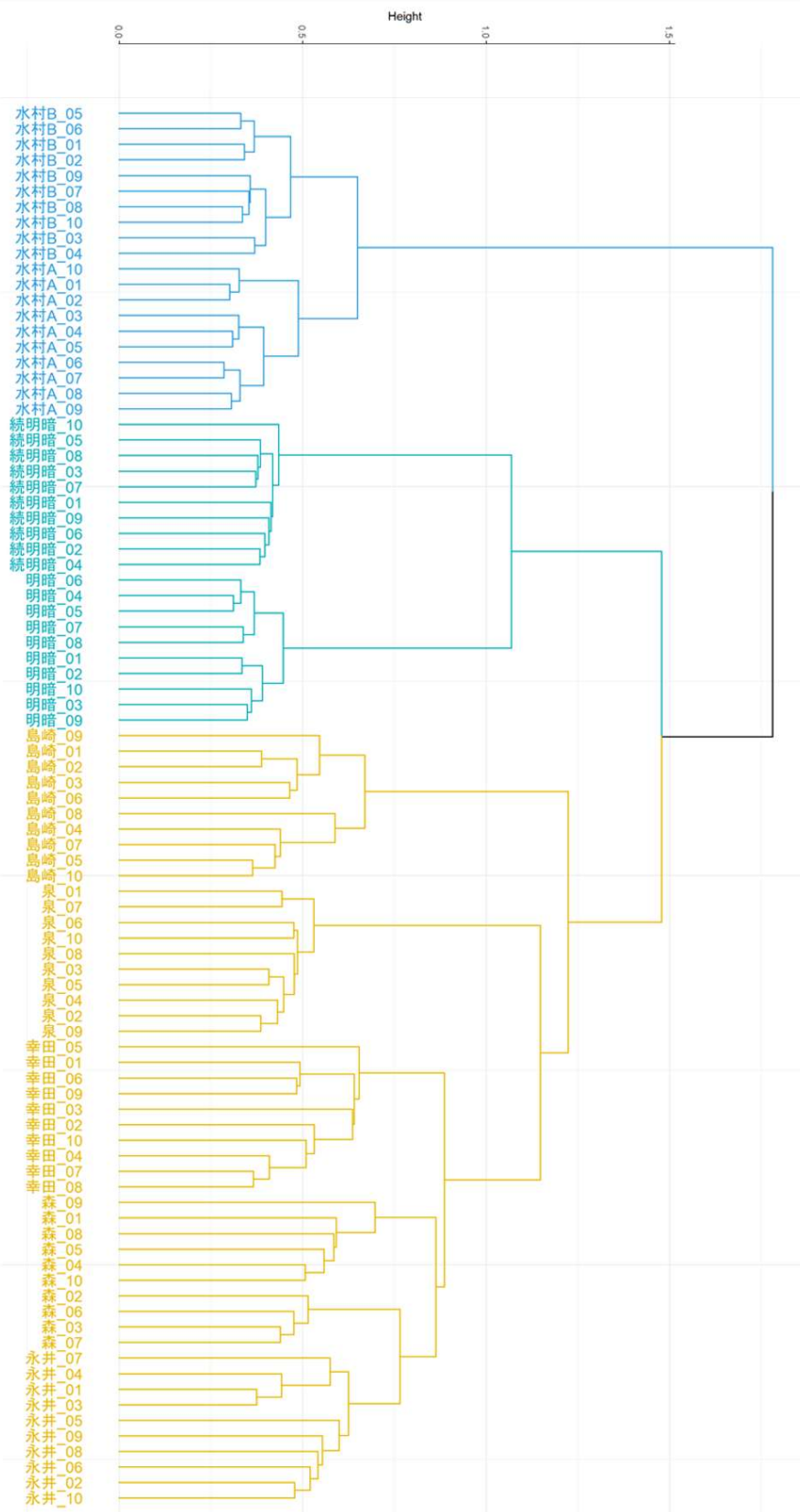


図 6.4 タグ付き形態素のクラスター分析

『明暗』『続明暗』における共通する特徴項目を確認するため、『明暗』『続明暗』によるクラスター及び他作品によるクラスターに対し、比率差を通して特徴的形態素項目を抽出する。そのうちの上位 10 のタグ付き形態素を表 6.2 に示す。統計の結果により、ほかの作品に比べ、『明暗』『続明暗』によるクラスターでは、「なかつ_形容詞」「すぐ_副詞」「答え_自立動詞」「云わ_自立動詞」「なら_非自立動詞」「解ら_自立動詞」「急_形容動詞語幹」「突然_副詞」などが多く出現することがわかった。

表 6.2 『明暗』『続明暗』と他作品におけるタグ付き形態素（上位 10）

	タグ付き形態素	『明暗』 『続明暗』	他作品	z 値
1	なかつ_形容詞	579	360	20.68
2	すぐ_副詞	368	145	20.23
3	言っ_自立動詞	2	898	19.84
4	答え_自立動詞	128	34	13.37
5	云わ_自立動詞	162	67	13.36
6	なら_非自立動詞	265	184	13.07
8	解ら_自立動詞	81	7	12.50
9	き_非自立動詞	8	347	11.61
8	急_形容動詞語幹	171	103	11.42
10	突然_副詞	138	76	10.93

「なかつ_形容詞」は、『明暗』の形容詞・形容動詞・動詞・副詞の形態素項目における出現率が 0.968%、『続明暗』における出現率が 1.1%である。「なかつ_形容詞」が特徴的形態素項目になったことは、2 作品で否定文が多く使われることを示唆し、ほかの特徴項目「なら_非自立動詞」「云わ_自立動詞」「解ら_自立動詞」の頻出にも繋がっている。否定表現は、実現しなかった事態を提示する形式であり、それが多く使われていることは、語り手が恒常的に物語の現実と異なる可能性を考えながら叙述を展開していくことをある程度暗示している。この点に関連して、高木（1966）は『明暗』の叙述にきわめて否定文が多いと指摘している。否定叙述の多用という特徴は水村の『続明暗』にも現れており、水村が『明暗』の叙述特徴を体得して、『続明暗』に活かしたと推定できる。

また、「突然_副詞」「急_形容動詞語幹」というような驚きの表現も 2 作品の特徴的形態素項目になり、この 2 項目が『明暗』における出現率が 0.65%、『続明暗』における出現率が 0.45%である。清水（1984）は、「『明暗』キー・ワード考——<突然>をめぐる」で、『明暗』において「突然」という言葉は筋の展開の中に頻出し、その使用はさまざまなヴァリエーションを伴って、『明暗』の世界に陰翳のある動勢を与えていると指摘した。例えば、以下の用例（1）と（2）は『明暗』から抜粋したものであり、用例（3）と（4）は『続明暗』から取り出した文である。

(1) すると其時お秀の口から最後の砲撃のように出たという句が、突然彼女の心を震わせた。

『明暗』

(2) すると突然津田の心が揺いた。何人とも答えなかった手代は、かえって弁解がましい返事をした。

『明暗』

(3) 突然清子の云わんとしている事が意味をなし、津田の頬にも勢い能く血がのぼった。

『続明暗』

(4) 突然頭に上ったその言葉には、無信心な津田の胸にも、無条件に生を受けた罪業の念を喚び起すような一種の響きがあった。

『続明暗』

作品中の人物が予想外の事態の出現によって、瞬間的なパニックまたは驚きに陥り、その心に動揺が与えられる。秋山（1987）は、「心理叙述の細密化」が『明暗』の特色の1つであると指摘している。「突然」「急」などの語彙は、他者及び環境の突然の変化に伴う人物の心理の動き及び感情の起伏、曲折を感じさせ、心理変化の表れと連動している。『明暗』『続明暗』に対し、「突然」を含む文を全部抽出して確認した結果、「突然疑惑の焰が彼女の胸に燃え上った」「突然一種の恐怖が湧いた」「突然ある疑がお延の心に芽生えた」などのように、「突然」を含む文では、「突然」と心理・感情を表す語句が共起している文は約40%を占めている。また、「突然」が感情・心理を表現する語句と直接共起せずに、他者及び環境の変化を表す文も存在している。そのなかで、下記の例（5）のように、外部環境の変化にもたらされた感情の起伏、曲折を描写する文が「突然」を含む文に続いていくケースがある。この類は、「突然」を含む文の中で約12%を占めている。

(5) 婦人が突然訊いた。我に返ったお延は返答に窮した自分を見出した。同時に、自分が返答に窮するのを承知で其様な質問を問い掛ける夫人の悪意を臆気ながら感じた。お延はもう身体の中に残っていない力を無理に奮い立たせ、夫人の追及に備えた。

『続明暗』

また、「すると連想が急に彼の胸を不安にした」「急に一種の淋しさが彼女の胸を打った」「急に悲しい気分囚えられた」のように、「急」を含む文では、「急」が心理・感情の動きを表現する語句と共起する文は約23%を占めている。以上の分析により、水村が『続明暗』を書いたとき、細微な心理現象まで細かく描き尽くすようにしていると推察できる。

一方、『明暗』と『続明暗』における差異を考察するため、2作品に出現する形態素に対し、同じく、比率差によって特徴的項目を抽出し、|z|値の大きい上位10個を表6.3に示す。表6.3から見ると、『明暗』では「なら_非自立動詞」「すぐ_副詞」「出来_自立動詞」「又_副

詞」などの項目が特徴項目になる。一方、『続明暗』では「もう一度_副詞」「やがて_副詞」「向かっ_自立動詞」「仕舞っ_自立動詞」「続_自立動詞」「再び_副詞」などが特徴項目である。抽出された特徴項目により、2 作品の形態素項目における差異の多くは似た意味を持つ形態素項目にあることがわかる。例えば、『明暗』では「すぐ_副詞」が多用されることに対し、『続明暗』ではその類義語の「やがて_副詞」のほうがよく使われている。これと同じように、「又_副詞」が『明暗』の特徴項目であるに対し、似た意味を示す「再び_副詞」「もう一度_副詞」が『続明暗』で頻出する。また、『明暗』の特徴項目である「出来_自立動詞」は、『続明暗』においてそれほど頻繁に現れなかったが、検索したところ、『続明暗』では「出来_自立動詞」の代わりに同じ意味を表せる「可能_形容動詞語幹」が多く使われている。さらに、少数とは言え、表記の違いによる差異も見られる。例えば、漱石の小説では仮名の「しまう」と漢字の「仕舞」との2種類の表記が混じて使用されているが、『続明暗』では全部「仕舞」と表記されているので、「仕舞っ_自立動詞」は『続明暗』の特徴項目になった。このように、『明暗』『続明暗』の形態素項目における差異は、類義語の使用・選択や表記のような微細な違いに集中していると考えられる。

表 6.3 『明暗』と『続明暗』におけるタグ付き形態素（上位 10）

	タグ付き形態素	『明暗』	『続明暗』	z 値
1	なら_非自立動詞	239	26	10.68
2	すぐ_副詞	312	56	7.95
3	出来_自立動詞	163	31	7.39
4	又_副詞	122	18	7.02
5	もう一度_副詞	0	34	6.83
6	やがて_副詞	15	70	6.52
7	向かっ_自立動詞	0	30	6.50
8	仕舞っ_自立動詞	10	48	6.27
9	続け_自立動詞	4	37	6.24
10	再び_副詞	21	63	6.02

6.3.4 品詞及び記号の bigram

本項では、抽出された品詞及び記号の bigram のデータを使って階層的クラスタリングを行った。出現頻度が計 45 回以下の項目を 1 つの項目 (others) にまとめ、計 597 個の変数を使用した。『続明暗』と『明暗』は、目盛数が約 0.6 の位置で結合し、1 つのクラスターを構成した。そのクラスターは、水村の作品 A と作品 B を含むクラスターと目盛数が約 1.05 の位置で結合した。つまり、『続明暗』と水村の他作品との距離に比べて、『続明暗』と『明暗』との距離は近いことがわかる。

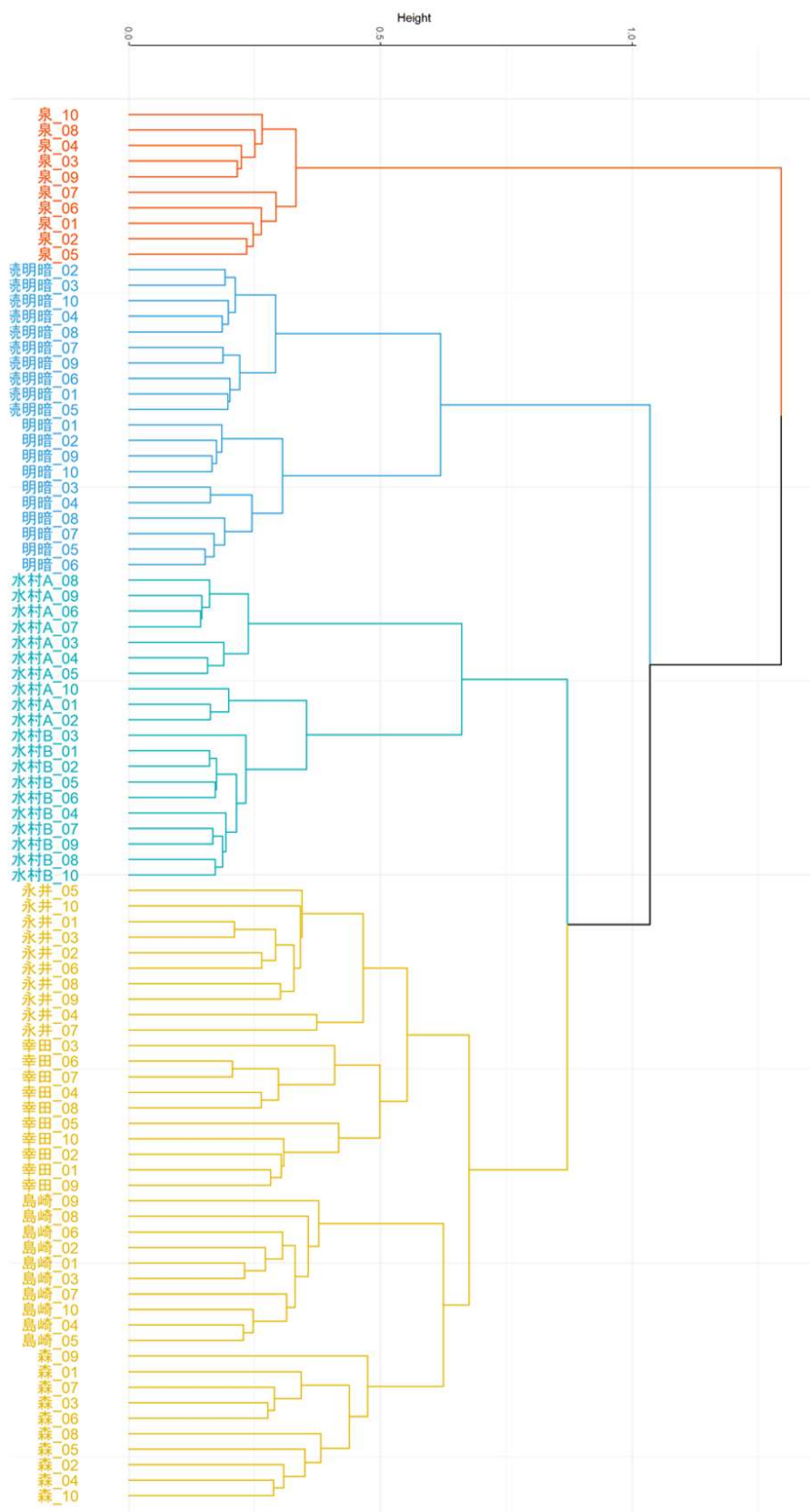


図 6.5 品詞及び記号の bigram による階層的クラスター分析

ほかの作品に比べ、同じクラスターを成している『明暗』『続明暗』にどんな項目が多く出現するかを考察するため、『明暗』『続明暗』を1つのグループ、他の作品群を1つのグループとして、2グループからそれぞれの特徴的形態素項目を抽出し、上位の10項目を表6.4に示す。その結果、『明暗』『続明暗』において「接頭詞_一般名詞」「助動詞_句点」「句点_接頭詞」「代名詞_係助詞」「自立動詞_助動詞」などが多く出現し、2作品の共通する特徴項目になることがわかった。「接頭詞_一般名詞」が特徴項目となったのは、2作品に共通する名詞が多いからである。また、コーパスを調査して見ると、『明暗』『続明暗』2作品における「助動詞_句点」は98.2%が「た。」や「だ。」であり、「自立動詞_助動詞」は99.3%が「動詞_た」に等しいことがわかった。佐藤(1991)は、夏目漱石の小説では、タ形の文末と非タ形の文末の割合が作品及び時期によって多様であり、初期小説『吾輩は猫である』では、文末部が現在形で結ばれているのが大きな特徴であるのに対して、遺作の『明暗』では、「～タ」止めが著しいと指摘した。この特徴が『続明暗』に受け継がれていることが明らかにされた。

表 6.4 『明暗』『続明暗』と他作品における品詞・記号の bigram (上位 10)

	品詞及び記号の bigram	『明暗』 『続明暗』	他作品	z 値
1	接頭詞_一般名詞	2731	3148	39.68
2	助動詞_句点	10389	20528	38.66
3	句点_接頭詞	935	513	35.73
4	読点_一般名詞	2268	14877	34.15
5	人名_接尾名詞	50	3665	32.70
6	代名詞_係助詞	1713	1901	32.40
7	接続助詞_読点	2152	12926	28.98
8	自立動詞_助動詞	12406	28621	28.54
9	代名詞_連体化助詞	914	831	27.34
10	人名_係助詞	1756	2462	26.26

一方、『明暗』『続明暗』は共に1つのクラスターを構成しているが、そのクラスターの中で、2作品のファイルはまたそれぞれ小さなクラスターを形成しているため、2者には差異があることがわかった。この差異を確認するため、2作品それぞれの特徴項目を抽出し、上位10項目を表6.5に示す。『続明暗』に対し、『明暗』では、「代名詞_格助詞」「代名詞_係助詞」「代名詞_連体化助詞」「句点_代名詞」「格助詞_読点」「読点_代名詞」「読点_副詞」「助動詞_代名詞」などが特徴項目となる。一方、『続明暗』では、「自立動詞_読点」「一般名詞_格助詞」「格助詞_一般名詞」などが特徴項目である。読点に係る項目が多く見られ、2作品における読点の打つ位置の違いを示している。文章を綴る際に、読点によってテンポを意識的に整えることができるが、どの文字・単語の後ろに打つかについては定則がないため、

そこに書き手の癖が表れやすいと考えられる。『明暗』の特徴項目「格助詞_読点」「読点_副詞」を例に挙げると、その読点は、打たなくても文法や文意に大きな影響がない場合が多く、打つかどうかは個人によって異なる。2 作品の特徴項目により、『続明暗』の文体模倣は読点の打つ位置まで行き届いていないと判断できる。

表 6.5 『明暗』と『続明暗』における品詞・記号の bigram (上位 10)

	品詞及び記号の bigram	『明暗』	『続明暗』	z 値
1	代名詞_格助詞	1489	288	22.50
2	代名詞_係助詞	1400	313	20.35
3	代名詞_連体化助詞	785	129	17.39
4	自立動詞_読点	60	331	16.88
5	句点_代名詞	1087	287	16.24
6	一般名詞_格助詞	9036	8442	16.15
7	格助詞_読点	1463	487	15.71
8	読点_代名詞	730	173	14.21
9	助動詞_代名詞	351	30	13.68
10	格助詞_一般名詞	2150	2317	12.80

6.3.5 文節

出現頻度が計 45 回以下の文節パターン項目を 1 つの項目 (others) として集計し、計 451 個の変数を用いて行った階層的クラスター分析の結果は図 6.6 に示す通りである。文節パターンによるクラスターのデンドログラムは、6.3.4 の品詞・記号の bigram による結果とほぼ同じである。『続明暗』と『明暗』は、先に目盛数が約 0.7 の位置で結合し、1 つのクラスターを構成してから、水村の作品 A と作品 B を含むクラスターと目盛数がより高い位置で結合した。つまり、水村の他作品に比べ、『明暗』『続明暗』における文節パターンの出現傾向の共通性がより高いと判断できる。

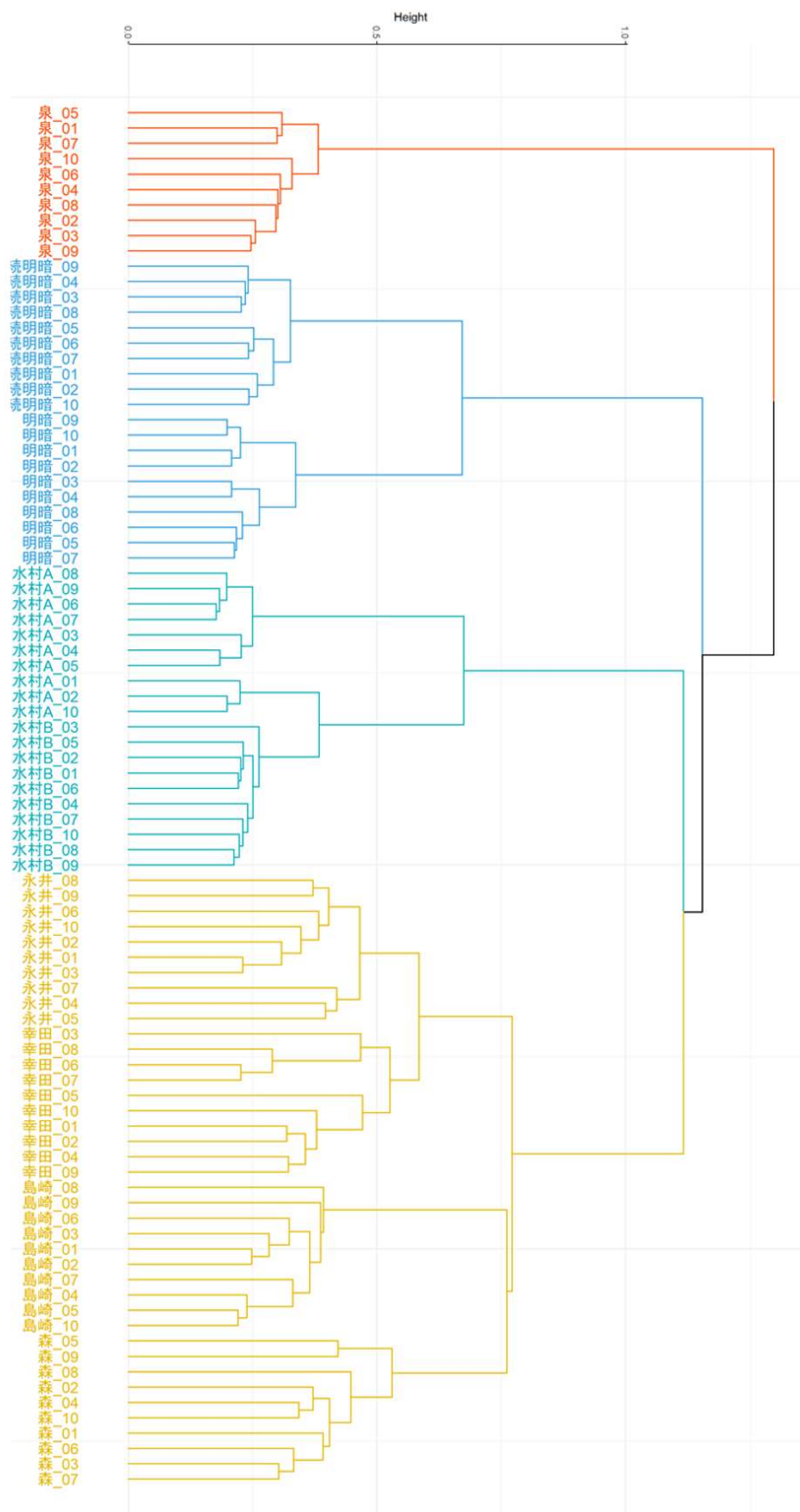


図 6.6 文節パターンによる階層的クラスター分析

『明暗』『続明暗』のクラスターにおける文節パターンの特徴項目を抽出して、上位10項目を表6.6に示す。表6.6から見ると、「動詞_助動詞_。」「接頭詞_名詞_が」「接頭詞_名詞_の」「名詞_は」「接頭詞_名詞_は」「名詞_動詞_助動詞_。」「名詞_から」「動詞_て_動詞_助動詞_。」などが見られた。なかでも最も特徴のある項目は「動詞_助動詞_。」である。テキストを確認して見ると、「動詞_助動詞_。」は、「動詞_た_。」に等しいことがわかった。『明暗』『続明暗』は「動詞_た_。」という文末表現によって特徴づけられていると言える。

工藤(1995)は、小説の地の文において、スル・シタ・シテイル・シテイタというアスペクト・テンス形式は、物語世界内部の出来事の時間関係提示のため総動員され、過去形と非過去形の競合は、「視点の相違」あるいは「アクチュアル性の明示の有無」という対立に結びつくと言っている。「動詞_助動詞_。」を始めとする文節パターンが、『明暗』『続明暗』における視点人物の認識及び叙述の構成に連鎖し、個性的な文体効果に寄与するのではないかと思われる。

表 6.6 『明暗』『続明暗』と他作品における文節パターン (上位 10)

	文節パターン	『明暗』 『続明暗』	他作品	z 値
1	接頭詞_名詞_は	819	382	34.45
2	動詞_助動詞_。	4413	7272	31.94
3	接頭詞_名詞_の	781	649	25.74
4	動詞_。	63	1810	21.78
5	動詞_，	319	3056	21.10
6	名詞_助動詞_助動詞_。	723	760	21.00
7	名詞_は	4761	9932	20.69
8	接続詞	1595	2501	20.46
9	名詞_名詞_の	1050	5928	19.61
10	名詞_名詞_は	243	2440	19.27

一方、『明暗』『続明暗』2作品の文節パターンにはどのような差異があるかについて検討するため、この2作品のみを用いて、比率差の比較によって特徴的な項目を抽出し、上位10項目を表6.7に示す。特徴抽出の結果によると、『明暗』では、「副詞」「接続詞」「連体詞」「名詞_へ」などが多く用いられているに対し、『続明暗』では「動詞_。」「名詞_助動詞_助動詞_。」「名詞_が」「動詞_と」「名詞_は、」「動詞_て_動詞_。」などが特徴的文

節パターンであることがわかった。テキストに戻って確認すると、『続明暗』の特徴項目の中の「動詞_。」「名詞_助動詞_助動詞_。」はそれぞれ動詞（辞書形）の文末表現、名詞述語（過去形）の文末表現に等しいことがわかる。「動詞_。」を更に詳しく見ると、その中で「見える。」が一番多くて、約 22.2%を占めている。「見える／見えた」は、主に、視覚的経験または視覚的経験による判断・思考を記述する動詞であり、それを使った描出表現の文において、観察の動作より観察される事態が中心である。『明暗』の場合、「失望の色が見えた。」「彼らは楽しそうに見えた。」といったように、視覚的経験などを記述した際、文末に「見える。」を使ったことは 1 回もなく、全て「見えた。」を使用した。それに対し、『続明暗』において、「伸び縮みするように見える。」という表現もあるし、「岩がぬるぬる光るのが見えた。」のように「見えた。」が使われる文もあって、半々になる。「見える。」の使用が水村自身の文体の痕跡と判断される。これと同じように、「名詞_だった。」という文節は、『明暗』には滅多になく、水村の表現特色であることがわかった。

表 6.7 『明暗』と『続明暗』における文節パターンの（上位 10）

	文節パターン	『明暗』	『続明暗』	z 値
1	動詞_。	29	290	17.59
2	名詞_助動詞_助動詞_。	17	190	14.42
3	副詞	3712	1919	12.89
4	接続詞	1166	429	12.57
5	名詞_が	1720	1822	11.19
6	動詞_と	89	234	10.98
7	名詞_は_,	611	187	10.84
8	連体詞	1639	774	10.31
9	動詞_て_動詞_。	8	90	9.93
10	名詞_へ	650	230	9.76

6.4 『明暗』『続明暗』におけるトピック構成の比較

6.3 では、言語表現に関わる項目について考察を試みた。続いて、テーマ・構成に着目し、『明暗』『続明暗』における内容・話題の変動様態を俯瞰的に捉える。本節では、構造的トピックモデルを通して、2 作品からトピックを抽出し、その推移・変化の状況を考察する。

6.4.1 トピックの数及びトピック別の遷移

本項のトピック分析では、時系列と内容構成の2つを共変量として使う。前節に分割された『明暗』『続明暗』に対し、文章の順序に従い、時間軸の情報として1~20の順番を付ける。形態素解析器 MeCab を利用して、内容構成の考察に最も情報量が高い単語と考えられる名詞・形容詞・形容動詞・動詞を、表記の基本形という形で各ファイルから抽出し、分析に使う。ただし、内容考察との結びつきの程度を考慮して、「数」「接尾」「副詞可能」類の名詞、「非自立」類の動詞、「自立」動詞に属する「する」「ある」「いる」「なる」を除外する。計9382項目の単語が抽出された。名詞、動詞、形容詞などの内容語に注目してモデルを構築しているため、そこから得られたトピックは、同じ事柄・話題について述べられている文書群が属するクラスと理解できる。

本項の分析では、トピック数の範囲を6までに設定し、チューニングを行った。その結果を図6.7に示す。トピック数が2であるとき、held-out likelihood と semantic coherence が最大、residual dispersion と lower bound が最小になっているため、理想的なトピック数は2に該当すると判断される。

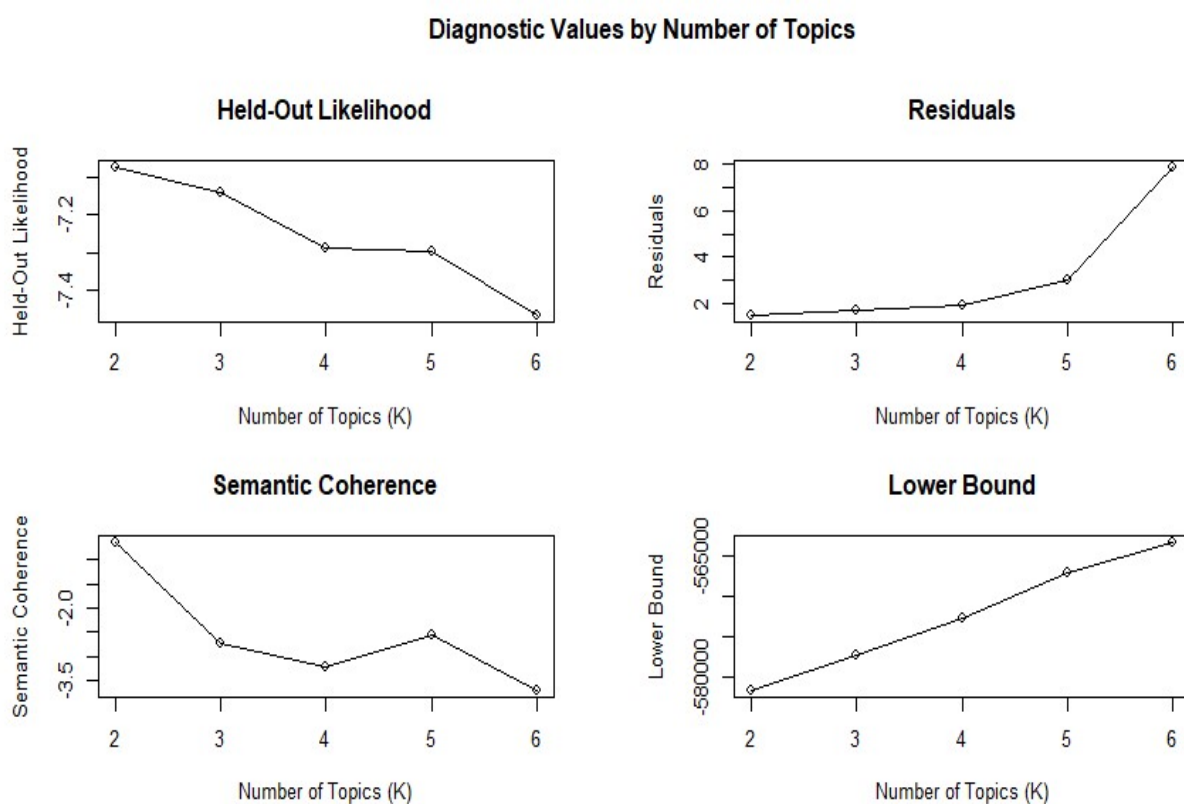


図 6.7 最適なトピック数についての分析図

トピック数を 2 としたトピックの時系列変化の非線形モデルの推測結果を図 6.8 に示す。横軸が時系列にあたり、曲線で 2 つのトピックの推定確率の変化が示されている。時系列はテキストの内容の順序に従うものなので、1～10 は『明暗』にあたり、11～20 は『続明暗』にあたっている。図 6.8 では、左から右へ見ていくことで、トピックの時間的推移が認められた。トピック 1 はピークが時系列の前半にあるのに対し、トピック 2 は時系列の後半でピークを成している。これにより、『続明暗』では『明暗』と異なる展開が描かれていることが推察される。

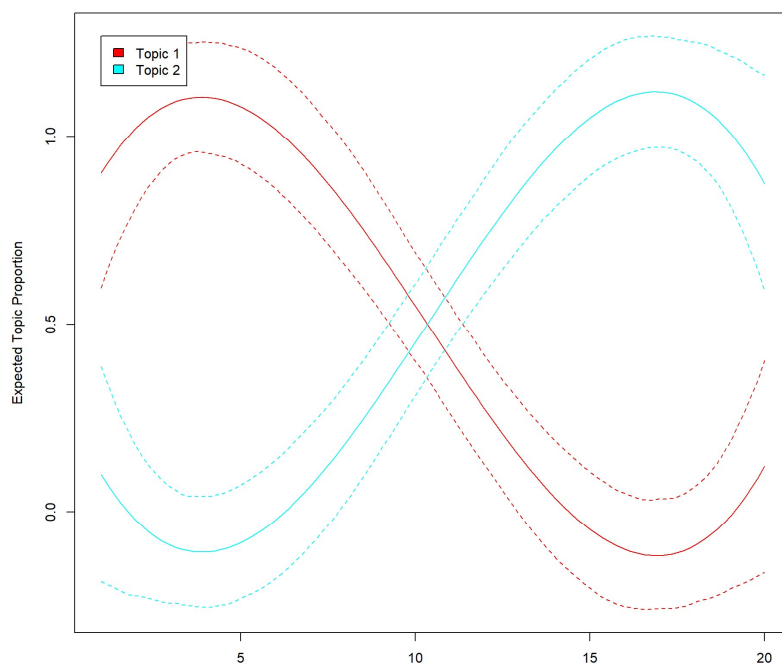


図 6.8 トピックの時系列変化（『明暗』『続明暗』）

続いて、2 作品におけるトピックの構成パターンにどんな差異があるのかを比較するため、2 作品に対し、それぞれトピックモデルを適用した。図 6.9 と図 6.10 は、トピック数を同じく 2 に設定した場合の結果である。

『明暗』のトピックを示す図 6.9 では、トピック 1 は、最初の頃がピークで、徐々に下がっていき、最後のところに改めて取り上げられるという形になっている。トピック 2 は、トピック 1 と真逆の形を取っている。一方、『続明暗』のトピックを示す図 6.10 では、2 つのトピックがほぼ並列的に進行しているように見て取れる。これにより、『明暗』『続明暗』2 作品におけるプロット構成の仕方が異なることがわかる。

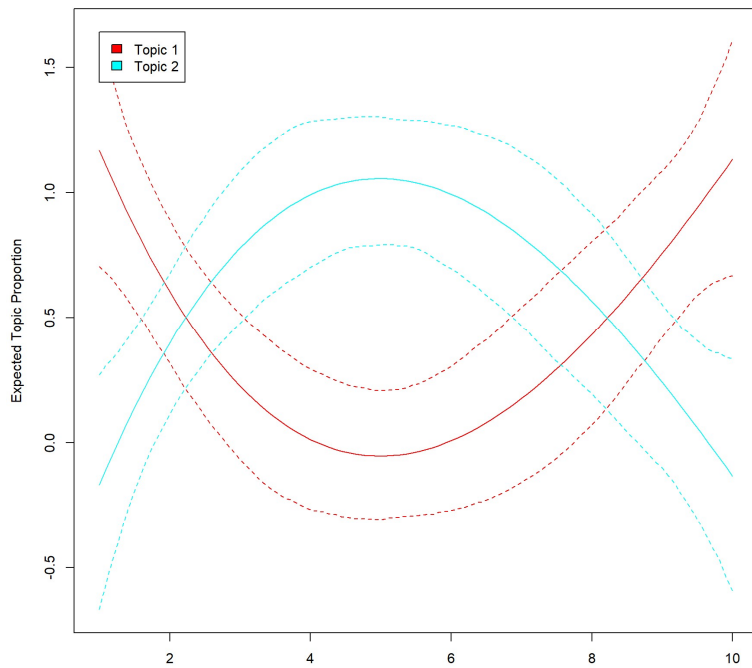


図 6.9 トピックの時系列変化 (『明暗』)

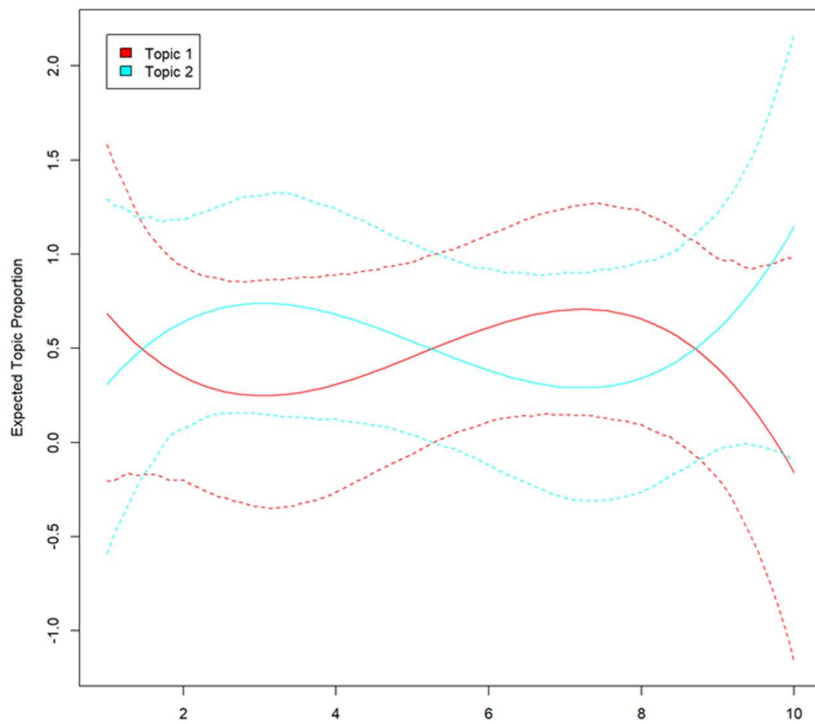


図 6.10 トピックの時系列変化 (『続明暗』)

6.4.2 トピックごとの特徴

構築された STM モデルからトピックに強く関連している語群を抽出し、トピックごとの中身及び特徴を考察することができる。単語の重み付け手法によって、4 種類の特徴語彙を取り出すことが可能である。本項では、図 6.8 に示されている、『明暗』に対応するトピック 1 と『続明暗』に対応するトピック 2 を比較してみる。

表 6.8 には、トピック 1 とトピック 2 から抽出した関連性の高い語群（上位 15 項目）を示している。そのなか、**Highest Prob** は出現頻度が高い単語を示す。**FREX** の単語は全体的な単語頻度の加重平均及びトピックに対する単語の排他性によって計算されたものである。**Lift** は、ほかのトピックにおける単語の使用頻度を利用して、トピックに固有の単語を選出する。**Score** は、頻度情報の TF-IDF (TF: Term Frequency, 単語の出現頻度 ; IDF: Inverse Document Frequency, 逆文書頻度) に近い指標で、すべてのトピックにおける分布を考慮された上位の単語を示す。

表 6.8 2つのトピックに強く関連する語群

Topic1	
Highest Prob	彼, 津田, 延, 彼女, ない, 自分, それ, 云う, 見る, 小林, 秀, 夫人, 眼, 斯, 其処
FREX	看護, 不愉快, 喜, 弱点, 封筒, 見合わせる, 小切手, 自白, 波瀾, 握る, 非常, 片付く, 働, 空想, 気が付く
Lift	浮べる, 良心, ひろげる, 藪, 有利, 晴々, 和解, 区切, あてる, アルバム, いい, いたずら, ウーロン茶, かく, かたちづくる
Score	彼, 看護, 彼女, 其処, 働, 気が付く, お世話, それ, 不愉快, 弱点, 見合わせる, 叔父, 喜, 自白, 刺戟
Topic2	
Highest Prob	津田, 延, 清子, ない, 云う, 眼, 顔, 自分, 夫人, 見る, 小林, 女, 彼, 付く, 出る
FREX	安永, 貞子, 燈, 向かう, 起こす, 驚く, 寐, 変わる, 烟, 起こる, 烟草, 顎, 脊中, 益, 呉服
Lift	わに, 囲む, 意図, 益, 押し寄せる, 億劫, 介す, 回す, 広がる, 掛声, 岩肌, 寄越す, 極め, 呉服, 交わす
Score	安永, 貞子, 燈, 清子, 向かう, 起こす, 寐, 驚く, 変わる, 滝, 烟, いちじるしい, 脊中, 烟草, 起こる

Highest Prob の単語について、2つのトピックから抽出した30個の Highest Prob 単語は異なり語数が20であり、共通する頻出語が多く見られる。これら20個の単語の相対頻度に基づいて、単語とテキストファイルの対応関係を階層的クラスタのヒートマップ（図6.11）で示す。従来のクラスタ分析では、グルーピングされている各クラスタの特徴を直接に読み取ることができないが、今回のように変数の数が少ない場合、ヒートマップによって個体の樹形図と変数の樹形図とを同時に示し、クラスタの特徴を直観的に考察することができる。変数間の距離については、 m 個の文章における変数 k （語など）のベクトルを $\mathbf{x}_k = (x_{k1}, x_{k2}, \dots, x_{kn})$, 変数 l のベクトルを $\mathbf{x}_l = (x_{l1}, x_{l2}, \dots, x_{ln})$ とした時、次に示すキャンベラ距離、クラスタリング法としてはワード法を用いた。式の中の n は用いた文章の数である。

$$d_{can}(\mathbf{x}_k, \mathbf{x}_l) = \sum_{i=1}^n \frac{|x_{ki} - x_{li}|}{|x_{ki} + x_{li}|}$$

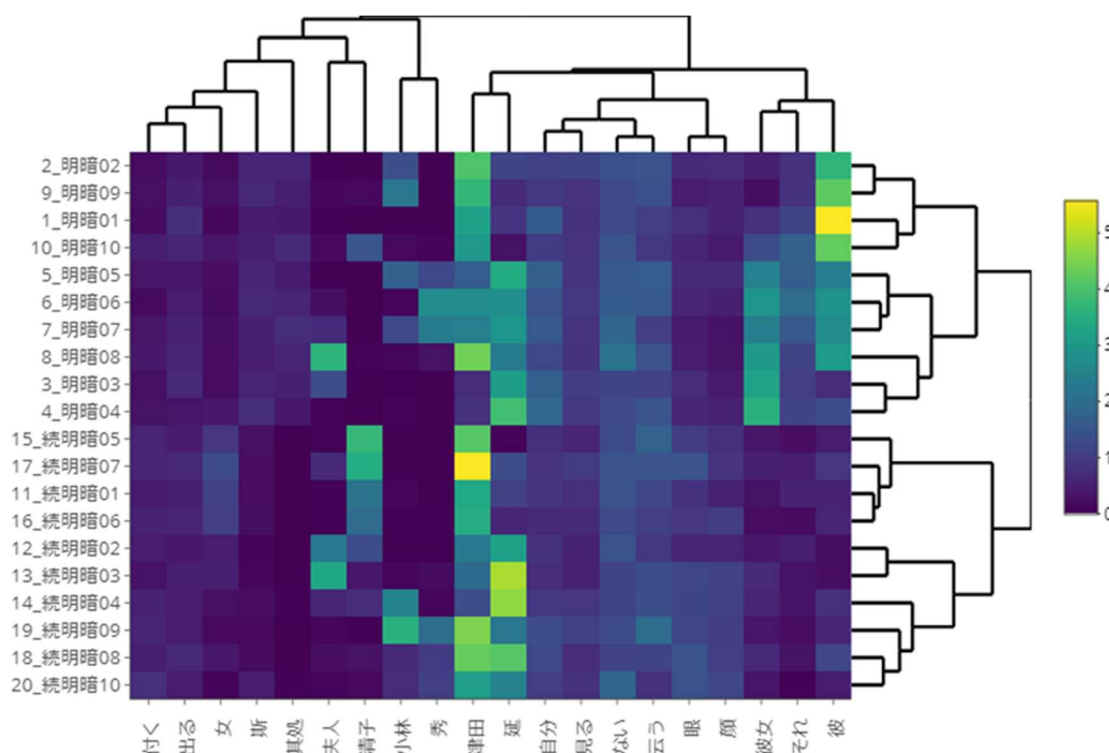


図 6.11 Highest Prob 単語を用いたクラスタリングのヒートマップ

ヒートマップでは、値の大小が色彩の色と濃淡によって示される。図6.11から見ると、20個のテキストファイルが『明暗』『続明暗』という2つの大きなクラスタに分かれている。また、単語の分布状況を見ると、主人公の「津田」の出現頻度が一番高く、終始高い出現頻度を保っている。妻の「延」が「津田」に続く2位となる。友人の「小林」、若夫婦の家庭

生活に風波を立てた「夫人」，妹の「秀」などの登場人物についても，ヒートマップを通して小説のどのあたりで活躍しているかを確認できる．また，津田の昔の恋人である「清子」について，『明暗』の 01～09 番目のテキストファイルにおける値が大きくなかったが，10 番目のファイルにおいて顕著に見える．そして，それに続く『続明暗』において頻繁に出ている．また，人名や代名詞のほか，2 つのトピックに共通する頻出語として，否定を表す「ない」，動作を表す「見る」「云う」などが見られる．これらの共通項目は，2 作品における主役の一貫性及び文体の共通性をある程度反映している．

FREX (frequency-exclusivity) 単語について，同じく階層的クラスタのヒートマップを作成し，図 6.12 に示す．共通項目の多い Highest Prob 単語と違って，FREX 単語はトピックに対する単語の排他性を考慮して計算されるものであるため，2 つのトピックの特徴を異なる角度から提示している．トピック 2 において，新しく登場する人物「安永」「貞子」のほか，「呉服」「益」「顎」「背中」「起こす」「燈」「起こる」「変わる」「驚く」「寐」「向かう」「煙草」「煙」などの語彙は，物語の筋を示唆する特徴的な項目である．それに対し，トピック 1 では，「空想」「見合わせる」「握る」「不愉快」「小切手」「非常」「封筒」「片付く」「働」「看護」「気が付く」「喜」「波瀾」「自白」「弱点」は特徴的な項目である．

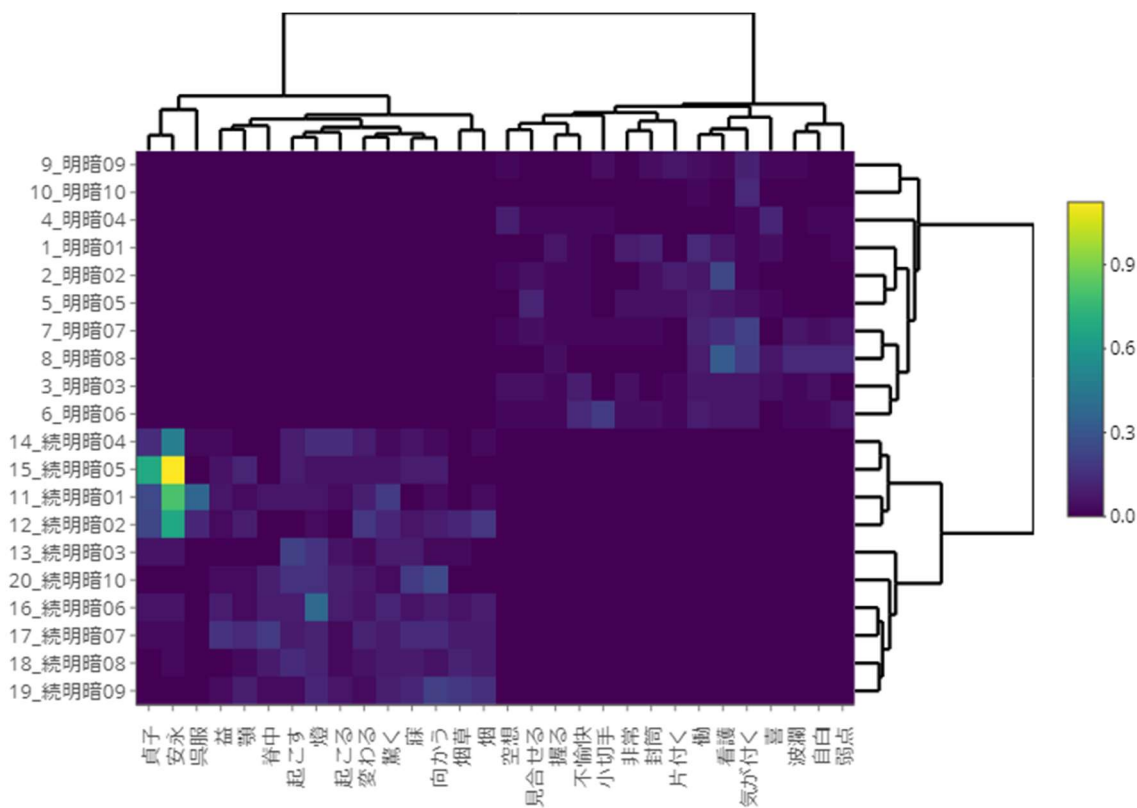


図 6.12 FREX 単語を用いたクラスタリングのヒートマップ

Bischof *et al.* (2012) は、排他性 (exclusivity) を有する単語がトピックの要約語に適することを指摘している。FREX 単語を中心に、Highest Prob, Lift, Score の語群も併せて考察してみると、『明暗』『続明暗』における 2 つのトピックのアウトラインが見えてくるようになった。トピック 1 では、努力して夫の愛を獲得しようとするお延、虚栄心高くて体面を保つため父に仕送りを送ってもらっている津田が中心人物である。津田の友人小林、妹のお秀及び仲人の吉川夫人も重要な登場人物である。津田が手術に必要なお金を手紙で父に頼もうとしたが、返済の約束を守らなかった津田に怒った父は、送金を見合わせにする。妹のお秀からの見舞金にお延の義理の叔父から小切手を与えられた御蔭で、津田の入院資金が集まってきた。トピック 1 の単語群は、津田の入院のいきさつ、悩みや秘密を抱える夫婦の日常生活、主人公の性格・弱点の解剖などの話題を示唆しているように思える。一方、津田が吉川夫人の勧めに従って昔の恋人清子が滞在する温泉場へ向かって、清子と再会したため、トピック 2 では清子についての描写が増えてきた。そして、宿で出会った温泉客の安永、貞子が脇役として新しく登場する。水煙る朝、津田は清子と温泉場の近くの滝を見に行き、そこで清子の脊中を眺めながら、ずっと知りたかった捨てられた理由を訊ねる。夫を追って来たお延が、語り合う夫と清子の姿を見て衝撃を受け、雨に長く当たったこともあり、風邪をひいて床に臥せるが、その夜は一晩中眠れない。翌日、夫婦の信頼関係が崩れたことに絶望したお延が宿を出て滝壺で死のうとする。トピック 2 の単語群は物語の後半のあらすじに密に結びついていると考えられる。

コーパスから抽出したトピック 1 とトピック 2 がそれぞれ『明暗』『続明暗』の内容にマッチしている。構築したトピックモデルを通して、『明暗』とその続篇の内容・主題の変動様態が直観的に示されている。トピックモデルより、水村は『続明暗』において、『明暗』に出てきた話題・要素をうまく利用して、内容をつなげながら、新しい筋を展開させることが確認できた。

6.5 まとめ

本章は、水村の作品を中心に文体の分析を行った。まず、水村の小説『続明暗』『本格小説』『母の遺産』に対し、通時コーパスから得られた回帰モデルを用いて、発表年度を推定した。その結果により、水村の文体が同時代の作品から離れていることがわかった。また、3 つの作品の中で、夏目漱石の文体に合わせて書いた『続明暗』は、水村が理想的な国語文学としている近代の文学に一番近いことがわかった。

『続明暗』に対し、水村自身の作品及び漱石と同年代の作家達による作品を用いて比較分析した結果、『続明暗』は文の長さ、漢字の出現率、形態素、品詞と符号の bigram、文節のパターンにおいて、『明暗』と近似した傾向を示している。それにより、水村が『続明暗』を執筆した際、文章のリズム、語彙の選択と構文の組み立てなどに工夫を凝らしたと推察さ

れた。そして、2 作品より抽出した共通の特徴要素から見ると、叙述視点や細部の心理表現に関わる模倣が目立っていることがわかった。一方、『続明暗』では、漱石の筆致が注意深く模倣されているが、水村の痕跡が残されていることが分析で明らかにされた。さらに、トピックについて、明らかな時系列変化が観察された。水村が文体を『明暗』に合わせているものの、ストーリーの新たな展開に工夫していると推察された。

また、水村が『続明暗』のため、自身の文体を変えて『明暗』に合わせてようとする試みは、分析に取り上げた言語項目によく現れている。特に、表記の場合、『続明暗』における漢字の出現率が水村のほかの作品とは顕著に異なり、漢字と仮名とのバランスに注意を払っていたことが読み取れる。また、文の長さは作品の特性に合わせて調整しやすい言語項目と考えられ、水村のほかの2 作品において文の長さの分布は多少差異が存在しているのに対し、『続明暗』と『明暗』はほぼ一致している傾向を示している。一方、形態素、品詞と符号の **bigram**、文節のパターンを用いたクラスター分析の結果から、『続明暗』と水村自身の他作品との類似度は、『続明暗』と『明暗』との類似度に比べると、はるかに及ばないことがわかった。これにより、水村が『続明暗』において、個人の特徴を意識的に薄れさせていることが推察できる。

本章では、水村の文体模倣の出来具合及びその成立に関する特徴をめぐって、計量分析により、その実態の一部を解明することができた。また、本章の分析より、統計的分析法が文学作品の模倣に関する研究に応用する可能性を示すことができた。

第7章 結論と今後の課題

本研究では、近現代日本語小説を分析対象とし、社会的な自然変化と個人の意図的变化をめぐって統計的分析法及び機械学習法のアプローチで計量的分析を行った。社会的文体の変化に関しては、近現代小説の文体における経時的变化に焦点を当て、文体に関わる言語項目の使用様態及び変化をモデリングし、小説の文体及び表現の経時的变化を考察した。一方、個人文体に関しては、水村美苗が近代文学への憧れを原点として創作した『続明暗』などの一連の作品は、近現代文学の変化の流れにおいて、どのように位置づけられるかについて、本研究で得られた近現代小説の経時的变化に関するモデルを利用して考察を行った。そのうえで、水村が夏目漱石の文体と語りの特色を模倣して書いた『続明暗』を中心に、模倣作と原作における異同、ストーリー展開の変化及び水村の意図的文体変化について分析した。

近現代小説における文体の経時的变化に関して、1910年から2014年に出版された555編の小説を用いて小説の通時コーパスを作成し、助詞、文末表現、接続表現という3つの側面に着目して、それらの項目の変化構造をモデリングし、文章表現に関わる言語項目の変化を特定し、その要因や影響について分析した。

助詞について、系統樹分析より経時的变化の傾向が観察された。ランダムフォレストによる非線形的回帰モデル及びelastic netによる線形的回帰モデルを通して、格助詞の「へ」「の」「で」「が」、副助詞の「なんて」、係助詞の「しか」「さえ」及び終助詞の「よ」が特徴的項目として特定された。第3章では、それらの項目の変化と小説の表現との関わりについて論じた。

文末表現について、第4章では、形態素解析器を利用して、文ごとに最後の語を文末表現のデータとして抽出して分析を行った。 K 特性値を指標とした多様性分析の結果、時期の推移に伴い、文末表現は多様化してきたことがわかった。また、ランダムフォレストによるモデリングによって、特徴的な変数項目を特定することができた。同じく常体として扱われる「である」体と「だ」体の表現が異なる推移傾向を見せており、「である」体が減少しているのに対し、「だ」体のほうが増えていた。そして、時制に関して、過去時制の表現が減少し、非過去時制の表現が増加していた。また、体言止めなどの文末表現が多用されるようになったことが明らかになった。分析により、それらの特徴的な変化傾向は、小説の語り方や技法に関する変化にもつながっていることがわかった。

接続表現について、第5章では、国立国語研究所が公表した分類語彙表を利用して、通時コーパスから接続表現を抽出し、分析を行った。接続表現の使用頻度は全体的には減少しているが、その分布はより均衡がとれていることが観察された。また、モデリング及び多変量解析の結果より、20世紀前半の小説において、累加型の「そして」が一番よく使われる接続表現であるのに対し、20世紀後半及び21世紀のはじめの文章において、反対型のほか、選

択型、理由型、換言型、補充型、転換型などが併用され、文脈のあり方は多様化したことが推察された。

第6章では、水村美苗の小説『続明暗』『本格小説』『母の遺産』に対し、通時コーパスから得られた回帰モデルを用いて発表年度を推定した。その結果により、水村の文体が同時代の作品から離れていることがわかった。そして、水村が意図的に自身の文体を変えて、漱石の『明暗』の文体と語りに合わせて書いた『続明暗』を中心に、2作品における言語表現の異同、ストーリー展開の変化および水村の模倣の特徴を考察した。その結果より、水村が『続明暗』を執筆した際、文章のリズム、語彙の選択と構文の組み立てなどに工夫を凝らしており、叙述視点や細部の心理表現に関わる模倣が目立っていることがわかった。一方、『続明暗』では漱石の筆致が注意深く模倣されているが、水村の痕跡が残されていることが分析で明らかにされた。その結果として、トピックの時系列変化が観察され、水村がストーリーの新たな展開に工夫していることがわかった。

20世紀初期から近年までの日本語の小説における変化をモデリングの手法で捉える研究が行われていないため、本研究は全くの新しい試みであった。また、通時コーパスをもとに構築したモデルを利用して、個人文体の特徴を捉えることも、日本文学の研究において初めてのチャレンジであり、文学の計量分析に新しい可能性を示した。そのほか、数理モデルを適用し、物語におけるトピック・テーマの構成やプロセスを分析することも、自然言語処理の新技术を文学作品の特性に活用する事例になる。また、本論文に観察された助詞・文末表現・接続表現に関する特徴的变化などは、今後の言語学や文体学の研究にとって有効な情報になると考えられる。

一方、本研究では以下の課題が残されている。まず、出現頻度データを元に統計分析を行ったが、出現頻度は特に増減していないにもかかわらず、係り受け語が変化したパターンが存在する可能性がある。このような変化を視野に入れて考察することも検討に値する。また、本研究では、会話文を除外し、地の文のみを対象に分析したが、直接話法と間接話法を混交させた文体が生まれている現代では、会話文の分析は今後重要な課題になると考えられる。さらに、時期の区切りの設定調整などにより、結果に違いが生じるかどうかに関して、比較検証を行うことが必要と考えられる。いずれも今後の課題として取り組んでいきたい。また、本論文では、特定の文体要素を使用しているが、それ以外にも数量化できる言語項目が多く存在しているので、分析の対象項目を増やし、異なる角度から分析を展開することによって、より立体的な認識につながる可能性がある。

謝辞

本論文を結ぶにあたり、指導教員の同志社大学大学院文化情報学研究科金明哲教授に深く感謝申し上げます。金先生には、統計学の基礎にはじまり、研究に取り組む姿勢や研究の方向性などあらゆることを丁寧に指導していただきました。長期間に亘り、ご指導・励ましをいただき、温かく見守ってくださり、本当にありがとうございました。

同専攻教授宿久洋先生、田口哲也先生、山内信幸先生、沈力先生には、本論文の構成から細部にわたり、貴重なご意見とご助言をいただきました。厚く御礼を申し上げます。

博士後期課程に入学以来、授業や研究科のシンポジウムなどの場でいつも同志社大学文化情報学研究科の諸先生方から貴重なご教示・ご指導を頂いております。ここに深い感謝の意を表します。

研究会などの場においては、名古屋大学大学院人文学研究科の中村靖子先生をはじめとする多くの先生方に貴重なご教示をいただきました。心より御礼申し上げます。

また、多大なサポートをいただきましたデータサイエンス研究室の皆様に御礼を申し上げます。皆様のご協力がなければ、私1人で通時コーパスの構築を完成することができません。いつも励ましてくださり、助けてくださり、ありがとうございました。

本論文を構成する関連論文と学会発表

第3章は研究論文[1]に基づいて書かれたものである。

- [1] 李広微・金明哲 (2021) 「モデリングから見る小説における助詞の経時的変化」『情報知識学会誌』31(3), 371-383. 査読付き

第4章は論文[2]に基づいて書かれたものである。

- [2] 李広微・金明哲 (2022) 「現代小説の文末表現における通時変化の統計モデリングと分析」『計量国語学』33(5), 309-324. 査読付き

第5章は国際学会の口頭発表[3]に基づいて書かれたものである。

- [3] G. Li and M. Jin (2020) Diachronic changes of sentence-final expression in modern Japanese novels, International Conference on Language and Literature 2020, Jan.15-17, International Islamic University Malaysia.

第6章は研究論文[4]及び研究ノート[5]に基づいて書かれたものである。

- [4] 李広微・金明哲 (2019) 「統計解析からみた水村美苗著『続明暗』の文体模倣」『計量国語学』32(1), 19-32. 査読付き
- [5] 李広微・金明哲 (2020) 「『明暗』と『続明暗』のトピック変化の計量分析」『計量国語学』32(8), 496-505. 査読付き

参考文献

- [1] Bischof, J. & Airoldi, E. (2012). Summarizing Topical Content with Word Frequency and Exclusivity. *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*.
- [2] Blei, D.M. & Ng, A.Y. & Jordan, M.I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- [3] Breiman, L. (2001). Random Forests, *Machine Learning*. 45(1), 5-32.
- [4] Genette, G. (1982). Palimpsestes: La littérature au second degré. *Éditions du Seuil*.
- [5] Hutcheon, L. (1985). A Theory of Parody: The Teachings of Twentieth-Century Art Forms. *University of Illinois Press*.
- [6] Pradyumna, T. & Murtaza, D. & Lara J. M. & Animesh M. & Brent H. & Mark O. R. (2018). Controllable Neural Story Plot Generation via Reinforcement Learning, arXiv:1809.10739.
- [7] Roberts, M. & Stewart, B. & Tingley, D. & Airoldi, E. (2013). The Structural Topic Model and Applied Social Science, *NIPS 2013 Workshop on Topic Models: Computation, Application, and Evaluation*.
- [8] Roberts, M. & Stewart, B. & Tingley, D. (2014). stm: R package for structural topic models, *Journal of Statistical Software*, 10(2), 1-40.
- [9] Yule, G.U. (1944). The Statistical Study of Literacy Vocabulary, *Cambridge University Press*.
- [10] Zou, H. & Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *JR Statist Soc B*, 67(2), 301-320.
- [11] 秋山公男(1987) 『漱石文学論考—後期作品の方法と構造』桜楓社.
- [12] 安部清哉(1996) 「語彙・語法史から見る資料—『篁物語』の成立時期をめぐるて」『国語学』184, 14-27.
- [13] 井上ひさし(1987) 『文章読本(自家製)』新潮文庫.
- [14] 家田章子(2005) 「文末の『ノニ』に関する考察」『ことばの科学』18, 201-218.
- [15] 井口時男・往住彰文・岩山真(1996) 『文学を科学する』朝倉書店.
- [16] 石川慎一郎(2017) 「現代日本語における『デ』格の意味役割の再考：コーパス頻度調査に基づく用法記述の精密化と認知的意味拡張モデルの検証」『計量国語学』31(2), 99-115.
- [17] 石黒圭・阿保きみ枝・佐川祥予・中村紗弥子・劉洋(2009) 「接続表現のジャンル別出現頻度について」『一橋大学留学生センター紀要』12, 73-85.
- [18] 市川孝(1966) 「文体論と文章論」日本文体論協会編『文体論研究 8』32-40, 三省堂.

- [19] 市川孝(1978) 『文章論概説』 教育出版.
- [20] 乾善彦(2017) 『日本語書記用文体の成立基盤』 塙書房.
- [21] 江湖山恒明(1981) 『国語表現論の構想』 明治書院.
- [22] 江藤淳(1968) 『夏目漱石』 角川文庫.
- [23] 江藤淳(1974) 『決定版 夏目漱石』 新潮社.
- [24] 遠田哲史(2019) 「文学作品における発話の教師なし話者同定」 東京大学大学院情報理工学系研究科修士論文.
- [25] 大江健三郎(1990) 「解説」 夏目漱石著『明暗』 595-607, 岩波文庫.
- [26] 大江健三郎(1993) 『小説の方法』 岩波書店.
- [27] 大岡昇平(1988) 『小説家夏目漱石』 筑摩書房.
- [28] 大隈秀夫(1975) 『文章の実習』 日本エディターズスクール出版部.
- [29] 大塚英志(2003) 『物語の体操—みるみる小説が書ける 6つのレッスン』 朝日新聞社.
- [30] 大塚誠一(2007) 『小説道場』 小学館.
- [31] 小方孝・川村洋次・金井明人(2018) 『情報物語論』 白桃書房.
- [32] 荻野孝野(2006) 「日本語動詞の結合価の格助詞パターンと意味マーカに関する研究」 神戸大学大学院自然科学研究科博士論文.
- [33] 尾野治彦(2018) 『「視点」の違いから見る日英語の表現と文化の比較』 開拓社.
- [34] 樺島忠夫(1990) 『日本語のスタイルブック』 大修館書店.
- [35] 樺島忠夫・寿岳章子(1965) 『文体の科学』 総芸舎.
- [36] 柄谷行人(2004) 『日本近代文学の起源』 岩波書店.
- [37] 柄谷行人(2013) 「解説」 夏目漱石著『明暗』 678-685, 新潮文庫.
- [38] 木坂基(1987) 「近代文学の接続語」 『日本語学』 6-9, 65-83.
- [39] 京極興一・松井栄一(1973) 「接続詞の変遷」 鈴木一彦・林巨樹編『接続詞・感動詞』 90-136, ひつじ書房.
- [40] 金明哲(2004) 「品詞のマルコフ遷移の情報を用いた書き手の同定」 『日本行動計量学会第32回大会講演論文集』 384-385.
- [41] 金明哲(2009) 「文章の執筆時期の推定——芥川龍之介の作品を例として」 『計量行動学』 36(2), 89-103.
- [42] 金明哲(2013) 「文節パターンに基づいた文章の書き手の識別」 『行動計量学』 40(1), 17-28.
- [43] 金明哲(2018) 『テキストアナリティクス』 共立出版.
- [44] 金明哲(2021) 『テキストアナリティクスの基礎と実践』 岩波書店.
- [45] 金明哲・村上征勝(2007) 「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』 55(2), 255-268.

- [46] 工藤拓(2018) 『形態素解析の理論と実装』 近代科学社.
- [47] 工藤真由美(1995) 『アスペクト・テンス体系とテキスト：現代日本語の時間表現』 ひつじ書房.
- [48] 桑野麻友子・金明哲(2008) 「小倉左遷前後における森鷗外の文体変化」 『日本行動計量学会第36回大会講演論文集』 44-47.
- [49] 国立国語研究所(2004) 『分類語彙表増補改訂版』 大日本図書.
- [50] 小西貞則(2019) 『多変量解析入門』 岩波書店.
- [51] 小西光(2016) 「近代口語文翻訳小説コーパス構築の概要と計量的分析」 『国立国語研究所論集』 11, 37-61.
- [52] 小林英夫(1968) 『美学的文体論』 広済堂.
- [53] 小林英夫(1975) 『文体論の建設』 みすず書房.
- [54] 小林雄一郎・小木曾智信(2013) 「中古和文における個人文体とジャンル文体—多変量解析による歴史的資料の文体研究—」 『国立国語研究所論集』 6, 29-43.
- [55] 阪倉篤義(1983) 「日本的知性と日本語」 相良亨・尾藤正英・秋山虔編『知性』 3-38, 東京大学出版社.
- [56] 佐久間まゆみ(1992) 「接続表現の文脈展開機能」 『日本女子大学文学部紀要』 41, 9-22.
- [57] 佐々木和枝(1976) 「文の長さの分布型」 『計量国語学』 78, 13-22.
- [58] 佐竹秀雄(1987) 「語と表記の関わり」 水谷静夫編『文字・表記と語構成』 109-120, 朝倉書店.
- [59] 佐藤理史(2016) 『コンピュータが小説を書く日：AI作家に「賞」はとれるか』 日本経済新聞出版社.
- [60] 佐藤武義(1991) 「漱石の文体—文末表現を中心として—」 佐藤喜代治編『文章研究の新視点』 131-147, 明治書院.
- [61] 清水孝純(1984) 「『明暗』 キー・ワード考—〈突然〉をめぐって—」 『文学論輯』 30, 1-39.
- [62] シュタンツェル(1989) 前田彰一訳『物語の構造—「語り」の理論とテキスト分析』 岩波書店.
- [63] 新村出編(2018) 『広辞苑』 (第7版) 岩波書店.
- [64] 杉浦清人(2017) 「文学研究におけるデジタル・ヒューマニティーズの可能性—文章心理学・計量文献学・マクロ分析—」 『現代文芸論研究室論集』 7, 80-96.
- [65] 関恵実(2010) 『続・漱石—漱石作品のパロディと続編』 専修大学出版局.
- [66] 高木文雄(1966) 『漱石の道程』 審美社.
- [67] 高崎みどり・新屋映子・立川和美(2007) 『日本語随筆テキストの諸相』 ひつじ書房.

- [68] 高橋源一郎・水村美苗(1990) 「『続明暗』という小説の行為」 『すばる』 12(12), 224-241.
- [69] 田中牧郎(2013) 『近代書き言葉はこうしてできた』 岩波書店.
- [70] 谷崎潤一郎(1975) 『文章読本(完)』 中央公論社版.
- [71] 塚原鉄雄(1958) 『文法各論編』 明治書院.
- [72] 辻邦生(2005) 『手紙, 葉を添えて』 新潮社.
- [73] 土山玄(2019) 「文末表現の計量分析に基づく夏目漱石の小説の分類」 『研究報告人文科学とコンピュータ』 6, 1-4.
- [74] 寺村秀夫(1984) 『日本語のシンタクスと意味Ⅱ』 くろしお出版.
- [75] 時枝誠紀(1949) 「国語に於ける変の現象について」 『国語学』 2, 1-16.
- [76] 中西久美子(2007) 「書きことば・話しことばのコーパスにおけるとりたて助詞の使用の実態」 『京都外国語大学研究論叢』 69, 171-185.
- [77] 中村明(1993) 『日本語の文体—文芸表現の表現をめぐって』 岩波書店.
- [78] 中村明(2005) 『センスをみかく文章上達事典』 東京堂出版.
- [79] 中村明(2016) 『日本語文体論』 岩波書店.
- [80] 夏目漱石(1966) 『明暗』 岩波書店.
- [81] 西本薫(2010) 「三島由紀夫作品における文体改造に関する計量的研究」 同志社大学文化情報学研究科修士論文.
- [82] 野口武彦(1994) 『三人称の発見まで』 筑摩書房.
- [83] 野村剛史(2015) 「物語・小説のテンス・アスペクト形式」 『言語・情報・テキスト』 22, 23-36.
- [84] 橋本梨沙(2014a) 「書き言葉と話し言葉の文末表現—媒体・男女間の差異—」 『日本文学』 110, 217-232.
- [85] 橋本陽介(2014b) 『物語における時間と語法の比較詩学』 水声社.
- [86] 波多野完治(1935) 『文章心理学』 三省堂.
- [87] 波多野完治(1950) 『文章心理学』 新潮社.
- [88] 波多野完治(1965) 『最近の文章心理学』 大日本図書.
- [89] 波多野完治(1988) 『文章心理学入門』 小学館.
- [90] 馬場博治(1983) 『文章教室 20 講』 大阪書籍.
- [91] 藤井貞和(2010) 『日本語と時間—<時の文法>をたどる—』 岩波書店.
- [92] 丸谷才一(1980) 『文章読本』 中公文庫.
- [93] 丸山岳彦(2012a) 「『現代日本語書き言葉均衡コーパス』を用いた文末表現のバリエーションの分析」 『言語処理学会第 18 回年次大会発表論文集』, 591-594.

- [94] 丸山岳彦(2012b) 「『現代日本語書き言葉均衡コーパス』を用いた文末表現のバリエーションの分析(2)」 『第2回コーパス日本語学ワークショップ予稿集』, 207-214.
- [95] 丸山直子(2015) 「コーパスにおける格助詞の使用実態：BCCWJ・CSJ にみる分布」 『計量国語学』 30(3), 127-145.
- [96] 三島由紀夫(1973) 『文章読本』 中公文庫.
- [97] 三島由紀夫(2016) 『小説読本』 中公文庫.
- [98] 水谷静夫(1999) 「戦後小説での格結合型」 『計量国語学』 21(8), 345-360.
- [99] 水村美苗(1990) 『続明暗』 筑摩書房.
- [100] 水村美苗(2012) 「ひと 水村美苗」 『ずばる』 34(6), 220-223.
- [101] 水村美苗・石原千秋(1991) 「水村美苗氏に聞く—『続明暗』から『明暗』へ—」 『文学』 2(1), 80-94.
- [102] 水村美苗・田中広美(2020) 「私は近代日本文学の最後に来た者」 『公研』 58(12), 18-36.
- [103] 水村美苗・梅田望夫(2009) 「特別対談 日本語の危機とウェブ進化」 『新潮』 106(1), 336-355.
- [104] 宮内佐夜香(2012) 「接続助詞とジャンル別文体的特徴の関連について—『現代日本語書き言葉均衡コーパス』を資料として」 『国立国語研究所論集』 3, 39-52.
- [105] 宮地朝子(2007) 『日本語助詞シカに関わる構文構造史的研究—文法史構築の一試論』 ひつじ書房.
- [106] 宮島達夫(2019) 『言語史の計量的研究』 笠間書院.
- [107] 村上征勝・金明哲・土山玄・上阪彩香(2016) 『計量文献学の射程』 勉誠出版.
- [108] 村田年・ロマン ロッサ(2014) 「異なる文章ジャンルの判別可能性に関する調査:ブログ本文, 新聞社説, 文学作品, 論文を対象として」 『日本語と日本語教育』 42, 125-135.
- [109] 森有正(1972) 『木々は光を浴びて』 筑摩書房.
- [110] 森田良行(2007) 『助詞・助動詞の辞典』 東京堂.
- [111] 安原顯(1990) 「驚くべき傑作, 水村美苗の『続明暗』」 『現代詩手帖』 33(11), 224-227.
- [112] 安本美典(1958) 「文の長さの分布型について」 『計量国語学』 4, 20-24.
- [113] 安本美典(1959) 「文章の性格学への基礎研究—因子分析法による現代作家の分類」 『国語国文』 6, 19-41.
- [114] 安本美典(1963) 「漢字の将来—漢字の余命はあと二百三十年か」 『言語生活』 137, 46-54.
- [115] 柳父章(2004) 『近代日本語の思想：翻訳文体成立事情』 法政大学出版局.

- [116] 山口明穂(1982) 「尚古・擬古意識と文体史へのその反映」 森岡健二編『文体史 I』 246-264, 明治書院.
- [117] 山口堯二(1991) 「副詞『しか』の源流—その他を否定する表現法の広がり—」 日本語語源探求委員会編『語源探求 3』 34-48, 明治書院.
- [118] 山崎誠(2011) 「文末表現の分布と文体—『現代日本語書き言葉均衡コーパス』を利用して」 『Japio year book 2011』 280-283, 日本特許情報機構.
- [119] 山田潔(2001) 「水村美苗『続明暗』語用論」 『解釈』 47(5・6), 48-52.
- [120] 山本雅子(2002) 「小説文末辞/-ta//~ru/の意味—認知的観点から—」 『言語と文化』 7, 1-21.
- [121] 山本雅子(2016) 「語りの語用論」 山梨正明・吉村公宏・堀江薫・糸山洋介編『認知語用論』 175-254, くろしお出版.
- [122] 山本正秀(1965) 『近代文体発生の史的研究』 岩波書店.
- [123] ヤロシユ島田 むつみ(2021) 「明治大正期『読売新聞』文末表現の文体推移—文章の種類と文の種類に注目して」 『国際日本学研究論集』 13, 51-71.
- [124] 於日平(2010) 「日語文章内容的陳述方式和不同文体使用的關係」 『日語學習与研究』 6, 1-9.
- [125] 李欣怡(2002) 「格助詞で終わる広告ヘッドラインに隠されたもの:『文の述べ方』という視点から」 『ことばの科学』 15, 5-22.
- [126] 劉雪琴・金明哲(2017) 「宇野浩二の病氣前後の文体変化に関する計量的分析」 『計量国語学』 31(2), 128-143.
- [127] 渡邊由貴(2016) 「思考動詞による文末表現の史的研究」 早稲田大学博士論文甲第 4972 号.

関連 Web サイト

馬場俊臣(2020) 「接続詞関係研究文献一覧」

<http://www.sap.hokkyodai.ac.jp/baba/home/setuzokusiitiran.htm> (2021 年 12 月 12 日確認)

付録

通時コーパスのリスト

1910	長塚節: 太十と其犬, 「ホトトギス」2月; 志賀直哉: 網走まで, 「白樺」4月; 小寺菊子: 父の帰宅, 「文章世界」4月; 森鷗外: 普請中, 「三田文学」6月; 永代美知代: ある女の手紙, 「すばる」9月
1911	水野葉舟: 『壁畫』, 春陽堂, 4月; 水上瀧太郎: 山の手の子, 「三田文学」7月; 物集和子: 七夕の夜, 「青鞥」9月; 森志げ: 死の家, 「青鞥」9月; 関根黙庵: 枯尾花, 「新小説」12月
1912	青木健作: お絹, 「読売新聞」3月から連載; 長田幹彦: 零落, 「中央公論」4月; 加藤みどり: 執着, 「青鞥」4月; 葛西善蔵: 哀しき父, 「奇蹟」8月; 木内錠: 老師, 「青鞥」9月; 邦枝完二: 廓の子, 「三田文学」9月
1913	高浜虚子: 杏の落ちる音, 「ホトトギス」1月; 島崎藤村: 突貫, 「太陽」1月; 田村俊子: 木乃伊の口紅, 「中央公論」4月; 与謝野晶子: 明るみへ, 「東京朝日新聞」6月~9月; 中勘助: 銀の匙, 「東京朝日新聞」4月~6月; 平出修: 逆徒, 「太陽」9月
1914	上司小剣: 鱧の皮, 「ホトトギス」1月; 中村星湖: 女のなか, 「早稲田文学」1月; 斎賀琴: 夜汽車, 「青鞥」4月; 岩野泡鳴: 毒薬を飲む女, 「中央公論」6月; 相馬泰三: 田舎医師の子, 「早稲田文学」7月; 木下杢太郎: 穀倉, 「三田文学」10月
1915	正宗白鳥: 入江のほとり, 「太陽」4月; 田口掬汀: ふたおもて, 「大阪朝日新聞」5月~12月; 夏目漱石: 道草, 「朝日新聞」6月~9月; 谷崎精二: 地に頬つけて, 「早稲田文学」10月; 芥川龍之介: 羅生門, 「帝国文学」11月
1916	宮嶋資夫: 『坑夫』, 近代思想社, 1月; 有島生馬: 暴君へ, 「太陽」1月; 田山花袋: 黄い小さな花, 「中央公論」6月; 久米正雄: 競漕, 「新思潮」6月; 宮本百合子: 貧しき人々の群, 「中央公論」9月; 松岡譲: 赤頭巾, 「新思潮」11月
1917	江馬修: 飛驒街道, 「文章世界」3月; 折口信夫: 身毒丸, 「みづほ」6月; 里見淳: 銀二郎の片腕, 「新小説」6月; 有島武郎: カインの末裔, 「新小説」7月; 広津和郎: 神経病時代, 「中央公論」10月
1918	水野仙子: 嘘をつく日, 「文章世界」2月; 吉田絃二郎: 清作の妻, 「太陽」4月; 江口渙: 労働者誘拐, 「雄辯」4月; 宮地嘉六: 煤煙の匂い, 「中外」7月; 佐藤春夫: 田園の憂鬱, 「中外」9月; 加能作次郎: 世の中へ, 「読売新聞」10月
1919	南部修太郎: 猫又先生, 「三田文学」4月; 中戸川吉二: イボタの虫, 「新小説」5月; 加藤武雄: 嗚咽, 「文章世界」6月; 室生犀星: 性に目覚める頃, 「中央公論」10月; 水守亀之助: 帰れる父, 「文章世界」11月
1920	島田清次郎: 2人の男, 「新潮」1月; 武林無想庵: ピルロニストのやうに, 「改造」3月; 田中純: 妻, 「人間」5月; 吉屋信子: 地の果まで, 「大阪朝日新聞」; 賀川豊彦: 『地殻を破って』, 福永書店, 12月
1921	内田百閒: 花火, 「新小説」1月; 長谷川如是閑: 象やの糸さん, 「中央公論」1月; 永井荷風: 雨瀟瀟, 「新小説」3月; 佐々木茂索: ある死, 次の死, 「新潮」5月; 新井紀一: 怒れる高村軍曹, 「早稲田文学」8月; 前田河広一郎: 三等船客, 「中外」8月
1922	近松秋江: 黒髪, 「改造」1月; 寺田寅彦: 亮の追憶, 「明星」5月; 野上弥生子: 海神丸, 「中央公論」9月; 稲垣足穂: 星を造る人, 「婦人公論」10月; 中西伊之助: 死刑囚とその裁判長, 「早稲田文学」10月; 十一谷義三郎: 静物, 「東京朝日」11月
1923	長与善郎: 青銅の基督, 「改造」1月; 宇野浩二: 子を貸し屋, 「太陽」3月; 豊島与志雄: 電車停留場, 「女性」10月; 金子洋文: 犬喧嘩, 「我観」11月; 中河与一: 或る心中の話, 「文芸春秋」11月
1924	牧野信一: 父を売る子, 「新潮」5月; 白井喬二: 富士に立つ影, 「報知新聞」7月から連載; 小山内薫: 『足拍子』, プラトン社, 8月; 富ノ澤麟太郎: 流星, 「改造」10月; 菅忠雄: 銅鑼, 「文藝時代」10月; 細田民樹: 多忙な初年兵, 「改造」12月

1925	梶井基次郎:檸檬,「青空」1月;今東光:瘦せた花嫁,「婦人公論」1月;細田源吉:寡婦とその子達,「早稲田文学」5月;里村欣三:苦力頭の表情,「文芸戦線」6月;小島勲:地平に現れるもの,「早稲田文学」9月;石濱金作:ある死ある生,「改造」12月
1926	川端康成:伊豆の踊子,「文藝時代」1月;横光利一:ナポレオンと田虫,「文藝時代」1月;松永延造:アリア人の孤独,「不同調」2月;林房雄:林檎,「文芸戦線」2月;尾崎士郎:秋風と母,「女性」8月
1927	石坂洋次郎:海を見に行く,「三田文学」2月;葉山嘉樹:死屍を食う男,「新青年」4月;黒島伝治:櫓,「文芸戦線」9月;池谷信三郎:橋,「改造」9月;久野豊彦:ボール紙の皇帝萬歳,「文藝公論」8月
1928	野溝七生子:灰色の扉,「近代風景」1月;佐多稲子:キャラメル工場から,「プロレタリア芸術」2月;龍胆寺雄:放浪時代,「改造」4月;立野信之:軍隊病,「戦旗」5月;山内謙吾:三つの棺,「文芸戦線」6月;山本有三:波,「朝日新聞」7月~11月
1929	片岡鉄兵:菜の花月夜,『新選片岡鉄兵集』,改造社,3月;佐左木俊郎:熊の出る開墾地,「文章倶楽部」4月;小林多喜二:蟹工船,「戦旗」5月;江戸川乱歩:押絵と旅する男,「新青年」6月;渡辺温:アンドロギュノスの裔,「新青年」8月;森田草平:四十八人目,「改造」10月
1930	徳永直:眼,「能率委員会」1月;吉行エイスケ:女百貨店,「近代生活」2月;丸岡明:マダム・マルタンの涙,「三田文学」2月;下村千秋:天国の記録,「中央公論」7月;木村良夫:嵐に抗して,「ナッパ」10月
1931	坂口安吾:風博士,「青い馬」6月;須井一:綿,「ナッパ」8月;金田千鶴:霜,「つばさ」9月;浅原六朗:混血児ジョオヂ,「中央公論」11月;若杉鳥子:母親,「批判」11月
1932	伊藤整:生物祭,「新文芸時代」1月;嘉村礒多:七月二十二日の夜,「新潮」1月;田畑修一郎:鳥羽家の子供,「文学手帖」3月;南洋一郎:吠える密林,「少年倶楽部」4月~12月;小林秀雄:Xへの手紙,「中央公論」9月;深田久弥:あすならう,「改造」11月
1933	徳田秋声:和解,「新潮」6月;谷崎潤一郎:春琴抄,「中央公論」6月;海野十三:崩れる鬼影,「科学の日本」7月~12月;福田清人:『河童の巣』,金星堂,7月;堀辰雄:美しい村,「改造」10月
1934	藤澤桓夫:大阪の話,「中央公論」1月;和田伝:村の次男,「改造」3月;村山知義:白夜,「中央公論」5月;坪田譲治:善太の四季,「文学界」6月;川口松太郎:鶴八鶴次郎,「オール讀物」10月;林芙美子:泣虫小僧,「東京朝日新聞」10月~11月
1935	大谷藤子:須崎屋,「改造」1月;石川達三:蒼氓,「星座」4月;平林彪吾:鶏飼いのコムニスト,「文藝」7月;北条民雄:間木老人,「文学界」11月;武田麟太郎:一の酉,「改造」12月
1936	阿部知二:冬の宿,「文学界」1月;矢田津世子:神楽坂,「人民文庫」3月;小田嶽夫:城外,「文学生活」6月;十和田操:判任官の子,「文学生活」7月;富澤有爲男:地中海,「東陽」8月;鶴田知也:コシヤマイン記,「文芸春秋」9月
1937	大鹿卓:福壽草,「東陽」2月;倉田百三:光り合ういのち,「いのち」2月から連載;中村地平:土竜どんもぼっくり,「日本浪漫派」5月;寺島証史:怪奇人造島,「日本少年」8月;火野葦平:河豚,「文学會議」8月;橘外男:生不動,「新青年」年10月
1938	壺井栄:海の音,「自由」2月;中山義秀:厚物咲,「文学界」4月;原民喜:迷路,「三田文学」4月;滝井孝作:父祖の形見,「文芸春秋」6月;岡本かの子:老妓抄,「中央公論」11月
1939	北原武夫:雨,「文藝」3月;岩倉政治:稻熱病,「知性」2月;保高德藏:或る死,或る生,「改造」12月;長谷健:あさくさの子供,「虚実」4月;寒川光太郎:密獵者,「創作」7月;半田義之:鶏騒動,「文芸首都」6月

1940	上林暁:『野』, 河出書房, 9月; 網野菊:風呂敷, 「文藝」9月; 牛島春子:祝といふ男, 「満州新聞」9月~10月; 桜田常久:平賀源内, 「作家精神」10月; 舟橋聖一:『川音』, 実業之日本社
1941	多田裕計:長江デルタ, 「大陸往来」3月; 木村莊十:雲南守備兵, 「新青年」4月; 藤島まき:あめつち, 「文學界」4月; 上田広:或る日の水間部隊長, 「中央公論」7月; 芝木好子:青果の市, 「文藝首都」10月
1942	榊山潤:『街の物語』, 実業之日本社, 2月; 石塚友二:松風, 「文學界」2月; 森山啓:遠方の人, 「文學界」5月; 中島敦:幸福, 「南島譚」11月; 倉光俊夫:連絡員, 「正統」11月
1943	井上友一郎:竹夫人, 「日本評論」1月; 石塚喜久三:纏足の頃, 「蒙疆文學」1月; 永井龍男:手袋のかたっぽ, 「文學界」3月; 東野辺薫:和紙, 「東北文學」4月; 鷹野つぎ:草藪, 『限りなき美』, 立誠社, 11月
1944	小尾十三:登攀, 「國民文學」2月; 太宰治:散華, 「新若人」3月; 八木義徳:劉廣福, 「日本文學者」4月; 清水基吉:雁立, 「日本文學者」10月; 島村利正:仙醉島, 「新潮」10月
1945	織田作之助:髪, 「富士」1945年1月~4月; 三島由紀夫:エスガイの狩, 「文藝」1945年5月; 山本周五郎:『菊屋敷』, 大日本雄辯會講談社, 10月; 島木健作:黒猫, 「新潮」11月; 北島八穂:自在人, 「新潮」12月
1946	中里恒子:まりあぬ物語, 「人間」2月; 大佛次郎:地霊, 「朝日評論」3月~9月; 神西清:雪の宿り, 「文藝」3月; 高見順:妖怪, 「世界」5月; 田村泰次郎:肉体の悪魔, 「世界文化」9月; 石川淳:焼跡のイエス, 「新潮」10月
1947	丹羽文雄:厭がらせの年齢, 「改造」2月; 野間宏:顔の中の赤い月, 「総合文化」8月; 浅見淵:夏目抄, 「文壇」8月; 獅子文六:無頼の英霊, 「オール讀物」10月; 西尾正:墓場, 「真珠」11月
1948	木山捷平:春雨, 「早稲田文学」3月; 檀一雄:終りの火, 「人間」2月; 山代巴:蔭のとう, 「大衆クラブ」3月; 平林たい子:人生実験, 「世界」6月; 駒田信二:脱出, 「人間」7月; 藤原審爾:紅顔, 「八雲」9月
1949	小谷剛:確證, 「作家」2月; 外村繁:夢幻泡影, 「文藝春秋」4月; 田中英光:野狐, 「知識人」年5月; 井上靖:闘牛, 「文學界」12月; 前田純敬:夏草, 「群像」12月
1950	辻亮一:異邦人, 「新小説」2月; 金達寿:矢の津峠, 「世界」4月; 今日出海:天皇の帽子, 「オール讀物」4月; 田宮虎彦:幼女の声, 「文藝」8月; 大岡昇平:歩哨の眼について, 「文藝」11月
1951	武田繁太郎:風潮, 「文學者」5月; 三浦朱門:冥府山水図, 「展望」9月; 武田泰淳:巨人, 「世界」9月; 堀田善衛:広場の孤独, 「中央公論文藝特集」9月; 安部公房:詩人の生涯, 「文藝」10月
1952	小山清:小さな町, 「文學界」2月; 戸川幸夫:高安犬物語, 「旅」3月; 柴田錬三郎:河内山宗俊, 「オール讀物」7月; 小島信夫:小銃, 「新潮」12月; 五味康祐:喪神, 「新潮」12月
1953	杉森久英:猿, 「中央公論」2月; 大原富枝:小さい喪失者, 「新日本文学」2月; 真杉静枝:或る女の生立ち, 「新潮」5月; 松本清張:菊枕, 「文芸春秋」8月; 小山いと子:壁の中の風景, 「新潮」8月; 円地文子:ひもじい月日, 「中央公論」12月
1954	吉行淳之介:驟雨, 「文學界」2月; 大田洋子:残酷点々, 「群像」3月; 曾野綾子:遠来の客たち, 「三田文学」4月; 幸田文:黒い裾, 「新潮」7月; 由起しげ子:女中ッ子, 「小説新潮」12月; 庄野潤三:プールサイド小景, 「群像」12月
1955	川崎長太郎:入り海, 「新潮」2月; 石上玄一郎:空笑, 「世界」4月; 原田康子:挽歌, 「北海文学」6月から連載; 石原慎太郎:太陽の季節, 「文學界」7月; 藤枝静男:瘡我慢の説, 「近代文学」11月; 三原誠:白い鯉, 「季節風」12月
1956	中村真一郎:扇, 「文藝」1月; 多岐川恭:落ちる, 「増刊宝石」1月; 福永武彦:死神の馭者, 「群像」2月; 近藤啓太郎:海人舟, 「文學界」2月; 椎名麟三:門のある家, 「世界」8月

1957	小林勝:太白山脈,「文學界」5月;菊村到:硫黄島,「文學界」6月;長谷川四郎:家常茶飯,「世界」6月;伊藤人誉:穴の底,「文学四季」8月;開高健:裸の王様,「文學界」12月
1958	寺崎浩:傲慢,「群像」1月;大江健三郎:人間の羊,「新潮」2月;城山三郎:総会屋錦城,「別冊文藝春秋」10月;西野辰吉:骨のある土地,「群像」11月;阿川弘之:告別交響曲,「群像」11月
1959	山村正夫:断頭台,「宝石」2月;澤野久雄:風と木の対話,「早稲田文学」2月;斯波四郎:山塔,「早稲田文学」5月;三好徹:遠い聲,「文學界」5月;川上宗薫:シルエット,「文學界」7月
1960	有馬頼義:第三の現場,「宝石」2月;池波正太郎:錯乱,「オール讀物」4月;花田清輝:群猿図,「群像」6月;三浦哲郎:忍ぶ川,「新潮」10月;黒岩重吾:背徳のメス,「中央公論」11月
1961	有吉佐和子:三婆,「新潮」2月;宇能鴻一郎:鯨神,「文學界」7月;尾崎一雄:まぼろしの記,「群像」8月;山口瞳:江分利満氏の優雅な生活,「婦人画報」10月から連載;河野多恵子:幼児狩り,「新潮」12月
1962	川村晃:美談の出発,「文学街」3月;山川方夫:夏の葬列,「ヒッチコック・マガジン」8月;島尾敏雄:出発は遂に訪れず,「群像」9月;中野重治:貼り紙,「群像」10月;後藤紀一:少年の橋,「山形文学」11月
1963	田辺聖子:感傷旅行,「航路」8月;野坂昭如:エロ事師たち,「中央公論」11月;柴田翔:されどわれらが日々——,「象」11月;遠藤周作:札の辻,「新潮」11月;植谷雄高:闇のなかの黒い馬,「文藝」12月
1964	北杜夫:死,「世界」3月;中谷孝雄:のどかな戦場,「群像」4月;古田芳生:よしこの,「群像」4月;杉浦明平:壁の耳,「世界」6月;富士正晴:『帝国軍隊に於ける学習・序』,未来社,9月
1965	きだみのる:マイ・フェア・レディ,「群像」3月;津村節子:玩具,「文學界」5月;立原正秋:剣ヶ崎,「新潮」4月;梅崎春生:幻化,「新潮」6月;高井有一:北の河,「犀」8月
1966	門脇照男:風呂場の話,『花火』,文芸広場社,1月;なだいなだ:しおれし花飾りのごとく,「文學界」4月;野島勝彦:胎,「文學界」5月;加賀乙彦:フランドルの冬,「展望」8月;泉大八:ある朝のいつもの時間,「群像」9月;五木寛之:蒼ざめた馬を見よ,「文藝春秋」12月
1967	大城立裕:カクテル・パーティー,「新沖縄文学」2月;柏原兵三:徳山道助の帰郷,「新潮」7月;金井美恵子:愛の生活,「展望」8月;井上光晴:階級,「群像」10月;佐木隆三:奇蹟の市,「文藝」12月
1968	船山馨:薄野心中,「小説新潮」2月;丸谷才一:年の残り,「文學界」3月;大庭みな子:三匹の蟹,「群像」6月;小沼丹:懐中時計,「群像」6月;倉橋由美子:ヴァージニア,「群像」12月
1969	黒井千次:時間,「文藝」2月;庄司薫:赤頭巾ちゃん気をつけて,「中央公論」5月;吉村昭:青い街,「世界」6月;田久保英夫:深い河,「新潮」6月;深沢七郎:無妙記,「文藝」11月;清岡卓行:アカシヤの大連,「群像」12月
1970	渡辺淳一:『光と影』,文芸春秋,3月;吉田知子:無明長夜,「新潮」4月;古山高麗雄:プレオー8の夜明け,「文藝」4月;源氏鶏太:『口紅と鏡』,新潮社,7月;吉田健一:瓦礫の中,「文藝」7月
1971	及川和男:雛人形,「民主文学」5月;李恢成:砧をうつ女,「季刊藝術」7月;新田次郎:『八甲田山死の彷徨』,新潮社,9月;加藤富夫:玩具の兵隊,「文學界」9月;東峰夫:オキナワの少年,「文學界」12月
1972	畑山博:いつか汽笛を鳴らして,「文學界」4月;宮原昭夫:誰かが触った,「文藝」4月;後藤みな子:炭塵のふる町,「文藝」8月;山本道子:ベティさんの庭,「新潮」11月;郷静子:れくいえむ,「文學界」12月;三木卓:鴉,「すばる」年12月

1973	青木八束:蛇いちごの周囲,「文學界」6月;森敦:月山,「季刊藝術」7月;宮尾登美子:權,「展望」7月;後藤明生:『挟み撃ち』,河出書房新社,10月;野呂邦暢:草のつるぎ,「文學界」12月
1974	太田道子:微熱のとき,「新潮」2月;山本孝夫:笑い声,「群像」3月;福沢英敏:アイの問題,「文學界」3月;日野啓三:あの夕陽,「新潮」9月;阪田寛夫:土の器,「文學界」10月
1975	三浦綾子:尾灯,「小説宝石」5月;林京子:祭りの場,「群像」6月;高橋昌男:藁のぬくもり,「新潮」7月;中上健次:岬,「文學界」10月;岡松和夫:志賀島,「文學界」11月
1976	村上龍:限りなく透明に近いブルー,「群像」6月;水上勉:棗,「世界」7月;萩原葉子:葎麻の家,「新潮」7月;山田智彦:マンハッタン島の女,「海」10月;富岡多恵子:立切れ,「群像」11月
1977	和田芳恵:雪女,「文學界」2月;三田誠広:僕って何,「文藝」5月;松代達生:子殺し,「文學界」6月;高城修三:櫃の木祭り,「新潮」8月;宮本輝:螢川,「文芸展望」10月;田中小実昌:ポロポロ,「海」12月
1978	高橋三千綱:九月の空,「文藝」1月;結城信一:空の細道,「文藝」5月;高橋揆一郎:伸予,「文藝」6月;古井由吉:親坂,「世界」7月;中沢けい:海を感じる時,「群像」10月
1979	重兼芳子:やまあいの煙,「文學界」3月;安岡章太郎:離島にて,「世界」3月;広津桃子:石露の花,「群像」5月;森禮子:モッキングバードのいる町,「新潮」8月;野口富士男:散るを別れと,「文藝」10月
1980	佐々木基一:奇妙な家族,「群像」1月;竹西寛子:兵隊宿,「海」3月;北澤三保:狩人たちの祝宴,「文學界」5月;青山健司:囚人のうた,「文藝」12月;丸元淑生:遠い朝,「文學界」12月;尾辻克彦:父が消えた,「文學界」12月
1981	吉行理恵:小さな貴婦人,「新潮」2月;丸山健二:夜釣り,「世界」3月;長谷川卓:百舌が啼いてから,「群像」5月;峰原緑子:風のけはい,「文學界」6月;阿刀田高:靴の行方,「小説現代」6月
1982	草鹿外吉:灰色の海,「民主文学」2月;津島佑子:黙市,「海」8月;加藤幸子:夢の壁,「新潮」9月;色川武大:『百』,新潮社,10月;唐十郎:佐川君からの手紙,「文藝」11月
1983	中藪英助:髭のガートフ,「世界」5月;冥王まさ子:白馬,「群像」8月;笠原淳:杳二の世界,「海燕」11月;干刈あがた:樹下の家族,「海燕」11月;高樹のぶ子:光抱く友よ,「新潮」12月
1984	高橋たか子:恋う,「新潮」1月;辻井喬:亡妻の昼前,「文藝」4月;青野聡:朔行する星からの便り,「世界」6月;木辺弘児:月の踏み跡,「文學界」9月;木崎さと子:青桐,「文學界」11月
1985	海辺鷹彦:黄色い斥候,「文學界」2月;佐藤泰志:オーバー・フェンス,「文學界」5月;土居良一:土と雨,「群像」6月;米谷ふみ子:過越しの祭,「新潮」7月;村山精二:風の鳥瞰,「中央文学」8月
1986	奥泉光:地の鳥天の魚群,「すばる」1月;藤本恵子:比叡を仰ぐ,「文學界」6月;小川国夫:逸民,「新潮」6月;芹沢光治良:『神の微笑』,新潮社,7月;森田誠吾:『魚河岸ものがたり』,新潮社,9月
1987	村田喜代子:鍋の中,「文學界」5月;上田三四二:祝婚,「新潮」8月;三浦清宏:長男の出家,「海燕」9月;池澤夏樹:スティル・ライフ,「中央公論」10月;吉本ばなな:キッチン,「海燕」11月
1988	筒井康隆:ヨッパ谷への降下,「新潮」1月;伊藤桂一:少年の庭,「群像」2月;山田詠美:風葬の教室,「新潮」2月;新井満:尋ね人の時間,「文學界」6月;南木佳士:ダイヤモンドダスト,「文學界」9月

1989	伊井直行:さして重要でない一日,「群像」4月;崎山多美:水上往還,「文學界」4月;中村隆資:流離譚,「文學界」12月;瀧澤美恵子:ネコババのいる町で,「文學界」12月;大岡玲:表層生活,「文學界」12月
1990	島田雅彦:SORAMIMI,「世界」3月;小川洋子:妊娠カレンダー,「文學界」9月;泡坂妻夫:蔭桔梗,「群像」10月;岩阪恵子:淀川にちかい町から,「群像」10月;増田みず子:街塵,「群像」10月
1991	辺見庸:自動起床装置,「文學界」5月;高田宏:『山へ帰った猫』,PHP研究所,5月;江國香織:『きらきらひかる』,新潮社,5月;荻野アンナ:背負い水,「文學界」6月;松村栄子:至高聖所,「海燕」10月
1992	司修:影について,「新潮」2月;鷺沢萌:ほんとうの夏,「新潮」4月;藤原智美:運転士,「群像」5月;中村正軌:アリスの消えた日,「文芸春秋」9月;多和田葉子:犬婿入り,「群像」12月
1993	吉目木晴彦:寂寥郊野,「群像」1月;藤堂志津子:青い扉,「雑誌」3月;華城文子:アンモナイトさんの家,「群像」9月;長堂英吉:役者たち,「群像」10月;大西巨人:牛返せ,「群像」10月;矢代静一:生日快樂,「群像」10月
1994	室井光広:おどるでく,「群像」4月;笙野頼子:タイムスリップ・コンビナート,「文學界」6月;小浜清志:後生橋,「文學界」6月;眞鍋呉夫:雀隠れ,「群像」11月;中村邦生:ドッグ・ウォーカー,「文學界」12月
1995	保坂和志:この人の関,「新潮」3月;村松友視:鎌倉のおばさん,「新潮」6月;三浦俊彦:エクリチュール元年,「文學界」7月;又吉栄喜:豚の報い,「文學界」11月;浅田次郎:鉄道員,「すばる」11月
1996	塩野米松:ペーパーノーチラス,「文學界」2月;川上弘美:蛇を踏む,「文學界」3月;乃南アサ:凍える牙,「新潮ミステリー倶楽部」4月;伊達一行:夜の落とし子,「文學界」7月;小田実:「アボジ」を踏む,「群像」10月;柳美里:家族シネマ,「群像」12月
1997	岡田睦:一月十日,「群像」2月;鈴木隆之:インスタント・ロマンス,「群像」2月;目取真俊:水滴,「文學界」4月;吉田修一:最後の息子,「文學界」6月;坂上弘:台所,「群像」10月
1998	辻章:青山,「新潮」3月;大塚銀悦:濁世,「文學界」5月;若合春侑:脳病院へまゐります,「文學界」6月;花村萬月:ゲルマニウムの夜,「文學界」6月;平野啓一郎:日蝕,「新潮」8月
1999	宮沢章夫:サーチエンジン・システムクラッシュ,「文學界」10月;楠見朋彦:零歳の詩人,「すばる」11月;玄月:蔭の棲みか,「文學界」11月;藤野千夜:夏の約束,「群像」12月;赤坂真理:ミューズ,「文學界」12月
2000	町田康:きれぎれ,「文學界」5月;松浦寿輝:花腐し,「群像」5月;中井佑治:フリースタイルのいろんな話,「群像」6月;堀江敏幸:熊の敷石,「群像」12月;青来有一:聖水,「文學界」12月
2001	横田創:似ていない人,「群像」3月;玄侑宗久:中陰の花,「文學界」5月;石黒達昌:真夜中の方へ,「文學界」10月;長嶋有:猛スピードで母は,「文學界」11月;鈴木弘樹:グラウンド,「新潮」11月
2002	青山光二:吾妹子哀し,「新潮」8月;朔立木:スターバート・マーテル,『深層』,光文社,9月;村上春樹:『海辺のカフカ』,新潮社,9月;和田ゆりえ:鏡の森,「文學界」12月;大道珠貴:しょっぱいドライブ,「文學界」12月
2003	吉村萬壺:ハリガネムシ,「文學界」5月;栗田有起:お縫い子テルミー,「すばる」6月;綿矢りさ:蹴りたい背中,「文藝」8月;金原ひとみ:蛇にピアス,「すばる」11月;中村航:ぐるぐるまわるすべり台,「文學界」12月
2004	モブ・ノリオ:介護入門,「文學界」6月;宮下奈都:静かな雨,「文學界」6月;辻原登:枯葉の中の青い炎,「新潮」8月;岡崎祥久:ナラズモノの唄,「群像」12月;阿部和重:グランド・フィナーレ,「群像」12月

2005	中村文則:土の中の子供,「新潮」4月;松尾スズキ:クワイエットルームにようこそ,「文學界」7月;脇坂綾:青い玉,「群像」7月;絲山秋子:沖で待つ,「文學界」9月;清水博子:vanity,「新潮」10月
2006	島本理生:大きな熊が来る前に,おやすみ,「新潮」1月;中原昌也:点滅……,「新潮」2月;伊藤たかみ:八月の路上に捨てる,「文學界」6月;青山七恵:ひとり日和,「文藝」8月;佐川光晴:家族の肖像,「文學界」12月
2007	松井雪子:アウラアウラ,「文學界」3月;諏訪哲史:アサッテの人,「群像」6月;中山智幸:空で歌う,「群像」8月;朝比奈あすか:ちいさな甲羅,「群像」8月;川上未映子:乳と卵,「文學界」12月
2008	角田光代:福袋,「文藝」2月;木村紅美:月食の日,「文學界」5月;津村記久子:ポトスライムの舟,「群像」11月;吉原清隆:不正な処理,「すばる」12月;墨谷渉:潰玉,「文學界」12月
2009	戌井昭人:まずいスープ,「新潮」3月;磯崎憲一郎:終の住処,「新潮」6月;藤代泉:ボーダー&レス,「文藝」11月;木村友祐:海猫ツリーハウス,「すばる」11月;山崎ナオコーラ:この世は2人組ではできあがらない,「新潮」12月
2010	広小路尚祈:うちに帰ろう,「文學界」4月;今村夏子:こちらあみ子,『太宰治賞2010』,筑摩書房,6月;原田マハ:ジヴェルニーの食卓,「すばる」9月;朝吹真理子:きことわ,「新潮」9月;西村賢太:苦役列車,「新潮」12月
2011	水原涼:甘露,「文學界」6月;藤沢周:案山子,「新潮」6月;円城塔:道化師の蝶,「群像」7月;星野智幸:夜は終わらない,「群像」9月;田中慎弥:共喰い,「すばる」10月;澤西祐典:フラミンゴの村,「すばる」11月
2012	鹿島田真希:冥土めぐり,「文藝」2月;いしいしんじ:ある一日,『いしいしんじ』,新潮社,2月;水村美苗:『母の遺産』,中央公論新社,3月;黒田夏子:a b さんご,「早稲田文学」9月;新庄耕:狭小邸宅,「すばる」11月
2013	藤野可織:爪と目,「新潮」4月;鶴川健吉:すなまわり,「文學界」6月;松波太郎:LIFE,「群像」7月;浅川継太:ある日の結婚,「群像」7月;小山田浩子:穴,「新潮」9月
2014	柚木麻子:『本屋さんのダイアナ』,新潮社,4月;小林エリカ:マダム・キュリーと朝食を,「すばる」4月;柴崎友香:春の庭,「文學界」6月;小野正嗣:九年前の祈り,「群像」9月;岡本学:Identity Provider,「群像」10月;足立陽:島と人類,「すばる」11月