

博士学位論文審査要旨

2022年1月25日

論文題目： 近代以降の日本小説の文体変化に関する計量的研究

学位申請者： 李 広微

審査委員：

主査：	文化情報学研究科	教授	金	明哲
副査：	文化情報学研究科	教授	田口	哲也
副査：	文化情報学研究科	教授	山内	信幸
副査：	文化情報学研究科	教授	沈	力
副査：	文化情報学研究科	教授	宿久	洋

要 旨：

本論文は、近代以降の日本語の社会的文体の自然的な変化と個人文体の意図的な変化に焦点を当て、自ら大規模のコーパスを作成し、統計的データ解析法および最新の機械学習法などを駆使し、計量分析を行ったうえで、考察を行った。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・目的・構成、第2章では研究に使われるコーパスの構築および分析手法について説明した。第3章から第5章は近代以降の小説の文体に関わる三つの要素の自然的な変化、第6章では意図的な変化についての計量分析及び考察を行った。第3章では助詞の使用に関して、第4章では文末表現の使用に関して、第5章では接続表現の使用に関して、統計的解析法による分析や機械学習(elastic net、ランダムフォレスト、構造的トピックモデルなど)によるモデリング手法を用いて、通時的な変化の特徴を分析した。第6章では作家個人による文体の意図的な変化に着目して顕著な変更要素を明らかにすると同時に、第3章から第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、近代以降の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、個人文体のどの要素を目的とした文体に変化させたかについて分析した。第7章では、本研究で得られた結果の要約及び今後の課題について述べた。

本研究は、文体及び関連の基礎的研究や贋作・模倣・成りすましの分析に関する実務に有益な情報を与えるものである。よって本論文は、博士(文化情報学)(同志社大学)の学位論文として十分な価値を有するものと認められる。

総合試験結果の要旨

2022年1月25日

論文題目： 近代以降の日本小説の文体変化に関する計量的研究

学位申請者： 李 広微

審査委員：

主査：	文化情報学研究科	教授	金 明哲
副査：	文化情報学研究科	教授	田口 哲也
副査：	文化情報学研究科	教授	山内 信幸
副査：	文化情報学研究科	教授	沈 力
副査：	文化情報学研究科	教授	宿久 洋

要 旨：

学位申請者は2017年度4月より本学大学院文化情報学研究科博士後期課程に在学しており、国内会議および国際会議での研究発表を通じて研究活動を積極的に行い、それらの成果を論文誌に3本公刊し、さらに1本採択されている。また、英語の語学試験にも合格していることから語学（英語）について十分な能力を有していると認定されている。

申請者の学位申請に関し、2022年1月25日火曜日10:00から約1時間10分の公聴会と15分の審査会において、種々の質疑応答を行った。申請者は研究内容及び関連する質問に対する確に対応したことで、学位論文審査委員会は申請者が博士（文化情報学）（同志社大学）の学位を授与するに十分な学力を有することを確認した。

よって、総合試験の結果は合格であると認める。

博士學位論文要旨

論文題目： 近代以降の日本小説の文体変化に関する計量的研究

氏名： 李広微

要旨：

本研究では、近現代日本語小説を分析対象とし、社会的文体の経時的変化と個人文体の意図的な変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分析を行った。本研究は主に2つの部分によって構成される。

- (1) 近現代小説の文体における経時的変化に焦点を当て、文体に関わる言語項目を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変化を考察した。
- (2) 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、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、考察を行った。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、水村の意図的な文体変化について分析した。

本研究は、7章より構成されている。第1章では、日本文学の計量研究の現状及び直面する課題、本研究の目的・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章では近現代小説の文体をめぐって、助詞、文末表現と接続表現のデータを利用し、その変化と構造をモデリングしたうえ、見出された変化と小説の文体の変遷との関わりを考察する。第6章では、第3～5章で得られた回帰モデルを利用して、水村の作品が近現代文学の時間軸上の位置を確認する。そして、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、水村の意図的な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。以下、各章の要点について述べる。

第1章では、日本文学に関する計量的研究において、著者推定をはじめとするテキスト識別や分類などのタスクが主流になっており、自然言語処理及び統計分析の発展及び成果が文学作品の研究に十分に活用されていないということを指摘する。さらに、文体の変化をめぐって先行研究で触れられることの少ない2つの問題を取り上げ、統計的分析法及び機械学習法のアプローチで考察することを目的として掲げる。1つは近現代小説の文体における経時的変化であり、もう1つ個人文体の意図的な変化である。この2種類の文体変化に関しては、計量的アプローチによる体系的な考察が十分に行われてこなかった。

第2章では、コーパス、分析項目、手法について記述した。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。本研究では研究目的に応じて、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、コーパスから助詞、文末表現、接続表現という3つの文体要素のデータを抽出し、回帰分析をはじめとするモデリング手法を利用して分析を行う。一方、個人文体の変化に関して、夏目漱石の長編小説『明

暗』、水村美苗が漱石の文体を模倣して書いた『続明暗』、水村自身のほかの作品、漱石と同時代の5名作家の作品を用いて、コーパスを作成する。文の長さ、漢字、形態素、品詞と符号のbigram、文節のパターンなど計量可能な文体項目を特徴量とし、階層的クラスター分析、比率差による特徴項目の抽出などの方法を利用して分析を行う。

第3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かった。抽出された助詞データ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。また、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4章では、文末表現の経時的変化について考察を行った。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に利用した。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様式が多用されるようになったことが明らかになった。

第5章では、接続表現の経時的変化について考察を行った。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に用いた。一文あたりの接続表現の使用率が減っていることがわかった。一方、K特性値を指標とした多様性分析の結果によると、時期の推移に伴い、接続表現が単一から多様化してきた傾向が見られる。これにより、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、夏目漱石の『明暗』と水村美苗の小説『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。また、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。

第7章では、本研究の議論を振り返り、主な達成点をまとめた。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変化などは、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。さらに、時期区切りの設定調整や、会話文の追加などにより、研究をより深く、また広く展開させることができると考えられる。