

博士学位論文審査要旨

2018年2月2日

論文題目： 川端康成の代筆問題及び文体問題に関する計量的研究

学位申請者： 孫 昊

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 矢野 環

副 査： 文化情報学研究科 教授 浦部 治一郎

副 査： 文化情報学研究科 教授 山内 信幸

副 査： 創価大学大学院文学研究科 教授 山中 正樹

要 旨：

本論文は、川端康成の作品の代筆問題、文体問題について計量的手法により分析を行った。

論文の第1章では、研究背景として川端康成の生い立ちを概観した上で、研究の着眼点となる川端康成の代筆問題と文体問題を紹介した。第2章では、代筆問題と文体問題の解決に用いた計量的研究の学史をまとめ、研究の主要手法である著者識別に必要な文体特徴量と計量的手法を説明した。第3章から第5章では、『乙女の港』、『花日記』と『コスモスの友』の代筆問題を検証し、『乙女の港』と『花日記』は川端康成と中里恒子の共同執筆、『コスモスの友』は川端康成作の可能性が高いことを示した。第6章から第8章では、『古都』、『眠れる美女』と『山の音』の代筆問題の検証を行い、『古都』は川端康成作の可能性が高く、『眠れる美女』と『山の音』は三島由紀夫の代筆ではないことを示した。第9章から第11章では、川端康成の文体存在問題、文体変化問題と語彙問題を扱った。文体存在問題では、泉鏡花、徳田秋声と横光利一との比較・対照を通じて、川端康成に文体が存在することが確認された。文体変化問題では、1974年に完結した全集から選んだ90篇小説の文体特徴量を経時的に調べ、終戦の1945年を境に語彙の豊富さと機能語の助詞、副詞、接続詞の変化を明らかにした。第11章では、川端康成の語彙問題を検証した。泉鏡花、徳田秋声と横光利一と比べ、川端康成は平易な表現を用いる傾向があり、平仮名の使用率も高いことを示した。第12章では、本論文のまとめと展望を述べた。

本論文は、コーパスを作成し、計量的手法で川端康成の作品における問題点についてデータ分析を行った結果と見解をまとめたものである。本研究は、従来の文学研究および文体研究とは一線を画すものであり、川端康成の作品研究に一石を投じるものである。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

2018年2月2日

論文題目： 川端康成の代筆問題及び文体問題に関する計量的研究

学位申請者： 孫 昊

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 矢野 環

副 査： 文化情報学研究科 教授 浦部 治一郎

副 査： 文化情報学研究科 教授 山内 信幸

副 査： 創価大学大学院文学研究科 教授 山中 正樹

要 旨：

学位申請者は2014年度4月より本学大学院文化情報学研究科博士課程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った。その成果を国際学会やワークショップで4回以上(Japan-China Symposium on Theory and Application of Data Science 2015, International Workshop for JSCS 30th Anniversary, International Quantitative Linguistics Conference 2016, International Conference on Culture and Computing 2017 など)発表し、国際論文誌に査読付きの論文2本(Neuroscience and Biomedical Engineering, 4(3), 174-180; Journal of Mathematics and System Science, 7, 127-141)を掲載した。申請者は英語の語学試験に合格している。

2018年2月2日金曜日、夢告館312号室で10:40から11:50までの公聴会と引き続き行われた30分の審査会において、種々の質疑応答の結果により博士(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： 川端康成の代筆問題及び文体問題に関する計量的研究

氏名： 孫 昊

要旨：

川端康成研究の未解明問題として代筆問題と文体問題が挙げられる。代筆問題は川端康成名義で発表した作品に代筆者が存在することを指す。その代表的な作品は小説の『乙女の港』、『花日記』、『コスモスの友』、『古都』、『眠れる美女』と『山の音』などである。文体問題は主に川端康成作品における文体存在問題と文体変化問題を指す。文体存在問題は川端康成が自分の文体を持っているかの問題で、文体変化問題は川端康成の文体に途中で変化が生じたかの問題である。

川端康成の代筆問題についての先行研究は史料学と文学の分野に集中している。史料学の分野では主に川端康成と代筆者の間の書簡を代筆の証拠とし、その内容は川端康成からの原稿作成依頼や、代筆者への執筆指導などである。今までの史料学に基づいた川端康成の代筆問題研究では、書簡の内容を代筆事実の有無を判断する基準としている。その往復書簡には、川端康成は執筆指導を行ったときに代筆者の原稿に手を加えたことが示されている。この場合川端康成も作品にかかわっていたため、原稿に対する貢献度に基づいた代筆の判断が必要である。

川端康成の文体存在問題に関しては、三島（1956）は『永遠の旅人—川端康成氏の人と作品』で「たとえば川端さんが名文家であることは正に世評のとおりだが、川端さんがついに文体を持たぬ小説家であるというのは、私の意見である。」と述べ、川端康成の「文体不在」論を唱えていた。また、寺田（1949）、三田（1994）と臼井（1952）も川端康成の文体は存在しないと主張している。このような諸説は川端康成の「文体不在」論の根拠となっている。

川端康成の文体変化が起きたと言われる時期は終戦を境に川端康成の文章が「魔界」に入った頃である。第二次世界大戦の敗戦と親友の相次いだ死は川端康成に精神的打撃を与え、川端文学も敗戦を境にして大きく変貌を遂げた。山中（1999）は、そこで川端文学が大きく前期と後期で分かれているとしている。また、五味（1981）は「川端康成は漢字の代わりに平仮名を多用している」と指摘した。

本論文では上述の川端康成の代筆問題と文体問題を課題とし、計量的手法を用いてその問題解明を試みた。

本研究では、著者識別（authorship attribution）を始めとする計量的手法を用いて川端康成の代筆問題と文体問題を明らかにする。著者識別は匿名文章の著者を推定する手法で、一般的にコーパス作成、文体特徴量の抽出と著者識別モデルの適用という3つの手順からなる。コーパス作成の段階では、代筆疑惑作品と、原著者、代筆疑惑者の作品コーパスを作成する。文体特徴量抽出の段階では、コーパスから文章著者の特徴を表す文体特徴量（stylistic features）を抽出する。著者識別モデル適用の段階では、抽出された文体特徴量を著者識別モデルに適用し、匿名作品の文体特徴量は原著者と代筆疑惑者のどちらに似ているかの判別を行い、その結果を用いて匿名文

章の著者を推定する。

本研究の目的に応じて3つのコーパスを作成した。1つ目は川端康成の代筆問題を検証するためのコーパスである。このコーパスには川端康成の小説20篇と、川端康成の小説を代筆したとされる中里恒子、澤野久雄、北條誠と三島由紀夫の小説それぞれ20篇が含まれる。2つ目は川端康成の文体存在問題を検証するためのコーパスである。このコーパスには川端康成の小説20篇と、泉鏡花、徳田秋声と横光利一の小説それぞれ20篇が含まれる。3つ目は川端康成の文体変化問題を検証するためのコーパスである。このコーパスには1974年に完成した川端康成全集のなかから選ばれた90篇の小説が含まれる。

本研究では文字記号の **bigram**、内容語抜きタグ付き形態素と文節パターンを文体特徴量として用いることにした。文字記号の **bigram** は日本語文の文字、仮名と記号の隣接しているペアである。内容語抜きタグ付き形態素は、名詞、動詞と形容詞以外の形態素とその形態素に付くタグ（品詞情報）の組合せである。日本語の文章からタグ付き形態素を抽出するためには日本語文章の形態素解析が必要である。本論文では形態素解析器 **MeCab** (IPA 辞書) を用いて形態素解析を行い、日本語の文章を形態素に分割した。本論文で用いた文節パターンは、金 (2013) の提案した4つの文節パターンの中で文学作品の著者識別に最も有効なものである。この文節パターンは、文節内の助詞・記号を除いた形態素の第1層品詞情報と助詞、記号の原型を組み合わせた文節パターンである。日本語文を文節ごとに分割する作業は日本語係り受け解析器 **CaboCha** を用いた。

本研究で用いた計量的手法は、主に対応分析、クラスター分析、エイダースト (**Adaptive Boosting: AdaBoost**)、高次元判別分析 (**High-Dimensional Discriminant Analysis: HDDA**)、ロジスティックモデルツリー (**Logistic Model Tree: LMT**)、ランダムフォレスト (**Random Forest: RF**) とサポートベクターマシン (**Support Vector Machine: SVM**) などがある。また、語彙の豊富さには **s** 指標を用い、対照作家との比較で分散分析を用いた。

代筆問題においては、文字記号の **bigram**、タグ付き形態素と文節パターンの3つの文体特徴量と対応分析、クラスター分析、**AdaBoost**、**HDDA**、**LMT**、**RF** と **SVM** の7つの計量的手法を用いて分析を行った。『乙女の港』、『花日記』では、代筆疑惑作品の各章は川端康成と可能な代筆者である中里恒子の作品グループに分類され、この2篇はいずれも共同執筆であることが分かった。『コスモスの友』は川端康成に多く分類されたため、中里恒子の代筆ではないと結論付けた。

『古都』の大多数の章は川端康成になっているが、文節パターン特徴量においては北條誠と澤野久雄に分類されたものも存在するため、この小説は三島由紀夫の代筆ではないと判断した。『眠れる美女』と『山の音』はほぼすべての回(章)は川端康成に分類されたため、この2篇の小説が三島由紀夫による代筆の可能性はほぼないと考えられる。

文体存在問題においては、文字記号の **bigram**、タグ付き形態素と文節パターンを用い、川端康成の文体を対照作家である泉鏡花、徳田秋声と横光利一の文体との比較を行った。その結果いずれの比較でも川端康成と対照作家の作品が分かれ、対象作家と比較する場合、川端康成は自分の文体を持っていると結論付けた。

文体変化問題においては、文字記号の **bigram**、タグ付き形態素と文節パターンを用い、川端康成の 90 篇の小説を時系列で調べた。終戦の 1945 年を境に語彙の豊富さと機能語の助詞、副詞、接続詞に変化があり、文字記号 **bigram**、タグ付き形態素、文節パターンと内容語の名詞、動詞、形容詞に変化がなかった。

語彙問題においては、対照作家と比べて語彙の豊富さを表す **s** 値がやや小さい。その原因は川端康成が極力平易な表現を用いたからである。また、川端康成が比較的平仮名を多用していることも明らかになった。

本論文は 12 章からなり、各章の要約を次に示す。

第 1 章では、研究背景として川端康成の生い立ちを概観した上で、本研究の着眼点となる川端康成の代筆問題と文体問題を紹介した。また、代筆問題と文体問題に関する先行研究の問題点を洗い出し、その問題解決のために計量的手法を用いる必要性を述べた。

第 2 章では、代筆問題と文体問題解決に用いた計量的研究の学史をまとめ、本研究の主要手法である著者識別に必要な文体特徴量と計量的手法を紹介した。

第 3 章では、川端康成の少女小説である『乙女の港』の代筆問題検証を行い、『乙女の港』は川端康成と中里恒子の共同執筆であることを示した。

第 4 章では、川端康成の少女小説である『花日記』の代筆問題検証を行い、『花日記』は川端康成と中里恒子の共同執筆であることを示した。

第 5 章では、川端康成の少女小説である『コスモスの友』の代筆問題検証を行い、『コスモスの友』は川端康成作の可能性が高いことを示した。

第 6 章では、川端康成の睡眠薬中毒時期に発表された小説である『古都』の代筆問題検証を行い、『古都』は川端康成作の可能性が高いことを示し、北條誠の文体要素も入っていると示した。

第 7 章では、川端康成の睡眠薬中毒時期に発表された小説である『眠れる美女』の代筆問題検証を行い、『古都』は可能な代筆者である三島由紀夫の作品ではないと示した。

第 8 章では、川端康成の長編小説である『山の音』の代筆問題検証を行い、『山の音』は可能な代筆者である三島由紀夫の作品ではないと示した。

第 9 章では、川端康成の文体存在性を取り上げ、対照作家である泉鏡花、徳田秋声と横光利一と比べて川端康成の文体特徴が明確に現れ、川端康成の文体は存在すると結論付けた。

第 10 章では、川端康成の文体変化性を取り上げ、1974 年に完結した全集から選んだ 90 篇の小説の文体特徴量を時系列で調べ、本研究で取り上げた文体特徴量に限って言うと先行研究で指摘された『雪国』、敗戦と『山の音』のいずれの時点においても川端康成の文体は変化が生じていないことを示した。

第 11 章では、川端康成の語彙問題を検証した。対照作家である泉鏡花、徳田秋声と横光利一と比べて川端康成は平易な表現を用いる傾向にあり、平仮名の使用率が多いことを示した。

第 12 章では、本論文のまとめと展望を述べた。