

川端康成「山の音」の代筆疑惑検証

—計量文体学の観点から—

孫昊† 金明哲‡

†同志社大学 文化情報学研究科

Email:{sonnkou1985@gmail.com, mjin@mail.doshisha.ac.jp}

0. はじめに

川端康成 (1899~1972) は日本初のノーベル文学賞受賞者であり、数多くの名作を後世に残した。しかし、その一部には代筆疑惑が取り上げられ、現在まで議論されている。従来の文学の観点における川端康成の代筆研究方法は主に 2 通りあり、一つは史料に基づいた研究、すなわち川端康成本人が書いた代筆依頼の書簡もしくは関係者の証言に基づいた研究である。もう一つは文学研究者の長年培ってきた文体に関する鋭い観察力に基づいた研究である。ところが、川端康成の代筆依頼書簡があったとしても、本人にその代筆の殆どを書き換えられ、代筆というより川端康成本人の作品と見なすべき可能性もある。なお、文学研究者の観察力に基づいた研究では、その結論は研究者によってまちまちで、曖昧な結論しか得られない場合もある。したがって、代筆問題の研究に関しては、このような従来の研究法は不十分で、文章の特徴から客観的に文体を分析するのが必要である。

1. 先行研究

日本語文体計量分析に有効な特徴量として、読点の打ち方 (Jin and Murakami 1993)、形態素タグの n-gram 分布(金 2004)、文節パターン (金 2013) などが提案されている。

川端康成は 1960~1961 年睡眠薬を常用し、彼の作品に関する代筆疑惑の議論も 1960 年代初頭から発表された「眠れる美女」や「古都」などの作品に集中した。板坂(1997)は「眠れる美女」の三島由紀夫代筆説を述べ、大石(2009)は読点の使い方を用いて「眠れる美女」は川端康成本人の作品と実証した。なお、川端康成の「古都」は北条誠と沢野久雄が書いたとある出版関係者の証言もある。金(2013)は文節パターン特徴量とクラスター分析法を用いて川端康成、北条誠と沢野久雄の作品コーパスで検証し、「古都」は川端康成本人の作品と見なすほうが妥当だと結論づけた。

読点の打ち方は文章の書き手の特徴を反映する有力な特徴量として既の実証された (Jin and Murakami 1993)。村上(2001)は読点の打ち方を用いて分析し、川端康成の作品は「山の音」から変化し始めたことがわかった。この「山の音」は三島由紀夫の代筆という説もあり、本研究では「山の音」の代筆疑惑を計量文体学の観点から検証する。

2. 「山の音」について

「山の音」は川端康成戦後初の小説である。当初「山の音」の 14 章はそれぞれ昭和 24 年 9 月から昭和 29 年 1 月にかけて複数の雑誌に発表された。各章の発表時期及び初出雑誌を表 1 に示す。

表 1 「山の音」各章初出雑誌表

発表時期	章	雑誌名
昭和 24.9	山の音	改造文芸
昭和 24.10	蟬の羽	群像
昭和 24.10	雲の炎	新潮
昭和 24.12	栗の実	世界春秋
昭和 25.4	島の夢	改造
昭和 25.5	冬の桜	新潮
昭和 26.10	朝の水	文学界
昭和 27.3	夜の声	群像
昭和 27.6	春の鐘	別冊文芸春秋
昭和 27.10	鳥の家	新潮
昭和 28.1	都の苑	新潮
昭和 28.1	傷の後	別冊文芸春秋
昭和 28.4	雨の中	改造
昭和 28.4	蚊の群	別冊文芸春秋
昭和 28.10	蛇の卵	別冊文芸春秋
昭和 29.4	秋の魚	オール読物

表 1 に示したように、「山の音」各章の発表時期の隔たりが非常に大きく、特に「冬の桜」と「朝の水」の間は 1 年 5 ヶ月も経っていることから、代筆の疑問が浮上している。

3. 用いるコーパス

本研究では、「山の音」と同時期（昭和 24 年～昭和 29 年）に発表されていた川端康成と三島由紀夫の作品それぞれ 10 篇を電子化し、コーパスを構築した。用いた文章リストを表 2 と表 3 に示す。

表 2 川端康成作品コーパス

発表時期	作品名
昭和 24.1	かけす
昭和 24.5	千羽鶴
昭和 24.1	夏と冬
昭和 24.5	雨の日
昭和 25.1	笹舟
昭和 25.5	卵
昭和 25.5	瀧
昭和 25.7	蛇
昭和 27.3	見ない人
昭和 29.1~12	みづうみ

表 3 三島由紀夫作品コーパス

発表時期	作品名
昭和 24.10	薔薇
昭和 25.1	果実
昭和 25.7	日曜日
昭和 25.8	遠乗会
昭和 25.8	孤閨悶々
昭和 25.9	日食
昭和 26.1	女流立志伝
昭和 27.12	美神
昭和 29.6	潮騒
昭和 29.8	詩を書く少年

4. 研究方法

本研究では、金（1993,2004,2013）が提案した読点の打ち方、タグの Bigram 及び文節パターンを特徴量として用いた。コーパスに対する計量分析を下記の手順で行った。

- ① 形態素解析器 MeCab と構文解析器 Cabocha を用いてそれぞれ形態素解析及び構文解析を行った。
- ② 集計済みのデータに対して、クラスター分析及び主成分分析を行って結果を分析した。クラスター分析では、ward 法及び下記の KLD 距離を用いた。

$$KLD(X, Y) =$$

$$\frac{1}{2} \left\{ \sum_{i=1}^n x_i \log \frac{2x_i}{x_i + y_i} + \sum_{i=1}^n y_i \log \frac{2y_i}{x_i + y_i} \right\}$$

5. 分析及び結果

5.1 「山の音」の代筆検証

5.1.1 読点の打ち方特徴量

表 2 と表 3 の文章から読点の打ち方を特徴量として抽出した。本研究で用いた読点の打ち方特徴量は文中の読点とその一つ前の成分の組み合わせを集計し、出現頻度の上位 20 位を抽出したものである。読点の打ち方を用いたクラスター分析の結果と主成分分析を図 1 と図 2 に示す。図 1 に示したように、川端康成の作品と三島由紀夫の作品はそれぞれ異なるクラスターを構成し、矢印で示した「山の音」は川端康成の作品クラスターの左端にある。図 2 から、三島由紀夫と川端康成はそれぞれ異なるグループを構成し、矢印で示した「山の音」は川端康成作品グループにあることがわかる。

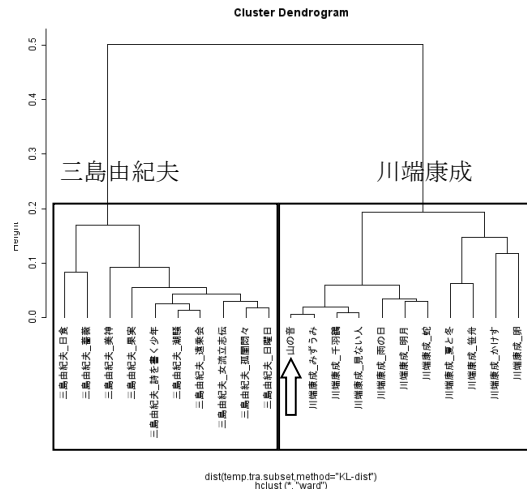


図 1 読点の打ち方のクラスター分析図 (Ward 法、KLD 距離)

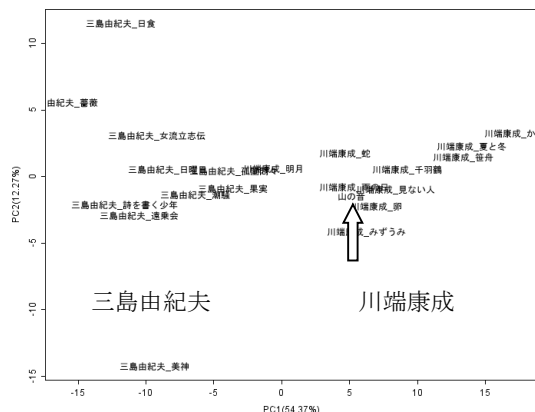


図 2 読点の打ち方の主成分分析図

5.1.2 タグの Bigram 特徴量

表 2 と表 3 の文章からタグの Bigram を特徴量として抽出した。本研究で用いたタグの Bigram 特徴量は形態素解析済みの隣接する二つのタグである。タグの Bigram 特徴量を用いたクラスターと主成分分析を図 3 と図 4 に示す。図 3 に示したように、読点の打ち方と同じく、矢印で示した「山の音」は川端康成の作品クラスターの左側にある。図 4 から、矢印で示した「山の音」は川端康成の作品グループにあることがわかる。

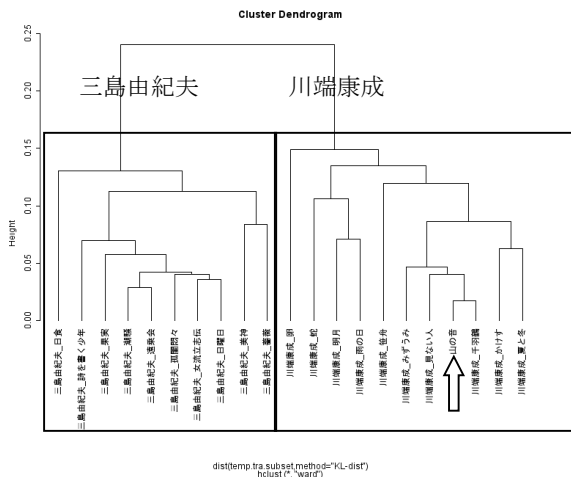


図 3 タグの Bigram のクラスター分析図 (Ward 法、KL 距離)

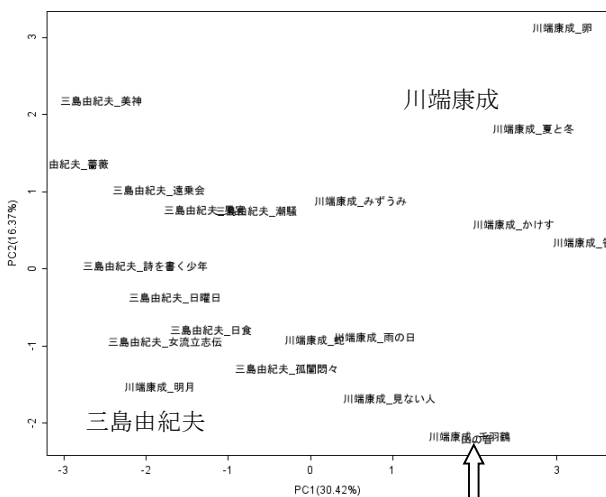


図 4 タグの Bigram の主成分分析図

5.1.3 文節パターン特徴量

表 2 と表 3 の文章から文節パターンを特徴量として抽出した。本研究で用いた文節パターンは、構文解析済みの第一層と第二層の情報の組み合わせ

わせである。文節パターン特徴量を用いたクラスター分析と主成分分析の結果を図 5 と図 6 に示す。図 5 に示したように、読点の打ち方及びタグの Bigram と同じく、矢印で示した「山の音」は川端康成の作品クラスターにある。図 6 から、矢印で示した「山の音」は川端康成の作品グループにあることがわかる。

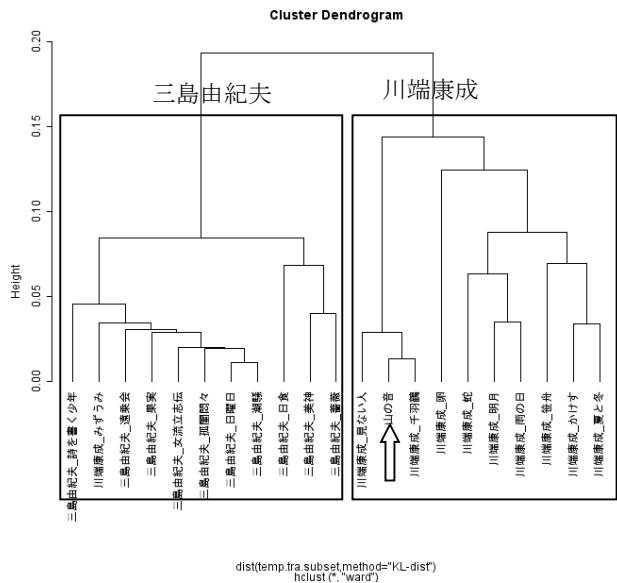


図 5 文節パターンのクラスター分析図 (Ward 法、KL 距離)

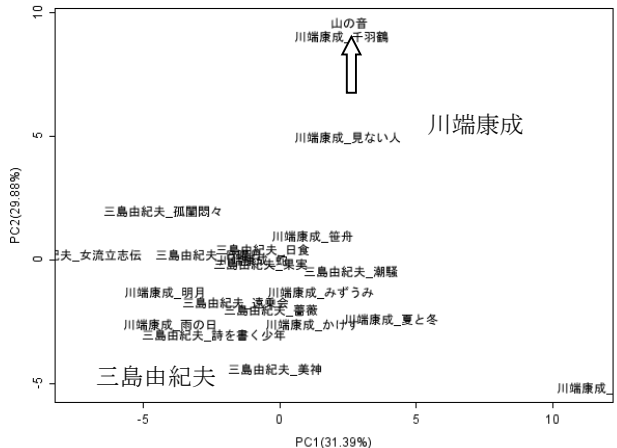


図 6 文節パターンの主成分分析図

以上のクラスター分析結果に示したように、読点の打ち方、タグの Bigram 及び文節パターン特徴量のどれを用いて分析しても川端康成の作品と三島由紀夫の作品は異なるクラスターを構成し、矢印で示した「山の音」は川端康成の作品ク

ラスターにある。主成分分析に関しても、川端康成と三島由紀夫の作品は分れており、矢印で示した「山の音」は川端康成の作品グループにある。

以上のクラスター分析と主成分分析での結果により、どの特徴量を用いても「山の音」は川端康成の作品に近いので、「山の音」は三島由紀夫の代筆だと言いがたく、川端康成本人の作品と見なしたほうが妥当である。

5.2 「山の音」各章の代筆検証

本論文の第2節にも言及したように、「山の音」各章の発表時期の隔たりが非常に大きく、その各章は三島由紀夫によって書かれた可能性もあるので、ここで「山の音」の14章を章ごとに検証する。

図7に示したように、読点の打ち方を用いてクラスター分析を行った結果、三島由紀夫の作品は一つのクラスターとなり、「山の音」の全14章及び川端康成のほかの作品は一つのクラスターとなる。したがって、「山の音」を章ごとに見ても川端康成本人の作品と見なしたほうが妥当である。

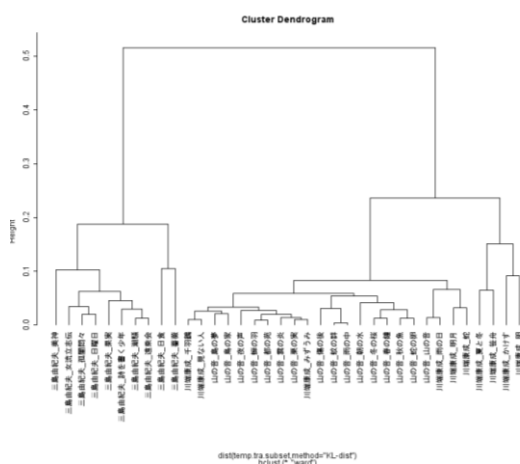


図7 各章読点の打ち方のクラスター分析図 (Ward法、KL距離)

6. 結論及び課題

本研究は、読点の打ち方、タグの Bigram 及び文節パターンを特徴量とし、クラスター分析と主成分分析を用いて「山の音」の代筆疑惑を検証した。「山の音」及びその各章は三島由紀夫の代筆と言いがたく、川端康成本人によるものと見なすほうが妥当だという結論に至った。

本研究は川端康成と三島由紀夫の作品それぞれ10篇という小規模のコーパスでしか検証でき

ていない。今後の課題としては、「山の音」の14章の発表期間内(昭和24年9月から昭和29年1月)における川端康成と三島由紀夫の全作品を含んだコーパス(川端康成:49篇、三島由紀夫:71篇)で検証する必要がある。

謝辞

本論文は同志社大学文化情報学研究科特別研究員として文化情報学研究科在籍期間中に行いました。特別研究員として採用して下さった同志社大学文化情報学研究科の諸先生に心より感謝します。

参考文献

- [1]Jin, M. and Murakami, M. (1993). Author's Features Writing Styles as Seen Through Their Use of Commas. *Behaviormetrica*,20,1,63-76.
- [2]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回全国大会講演論文集,384-385.
- [3]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.
- [4]板坂剛(1997).極説三島由紀夫一切腹とフラメンコ,夏目書房.
- [5]大石拓生(2009).川端康成にまつわる諸問題の計量的分析—代筆疑惑の検証と、文体変化時期の推定—,同志社大学文化情報学部卒業論文.
- [6]村上征勝(2001).川端康成の文体終戦直後に大きな変化—機械的に分析,朝日新聞.
- [7]羽鳥徹哉・原善(1998).川端康成全作品研究事典,勉誠出版.
- [8]板坂剛・鈴木邦男(2010).三島由紀夫と一九七〇年,鹿砦社.