

日本の児童文学作品における文体変化についての 計量的な分析の試み

土山玄¹

概要: 文学作品を対象とした計量的な分析は広く行われているが、現役作家の文章を対象とした計量的な研究は十分に展開されているとは言えない。そこで、本発表では現役の児童文学作家であるイノウエミホコの著作を対象に、文筆活動を開始した時期から現在にいたるまで、どのように文体が継時的に変化したのか計量的な分析を行った。分析の結果、2010 年前後の作品においては助動詞の「た」や「らしい」という単語の出現率が相対的に高く、2020 年前後の作品においては「～したい」という表現の「たい」や「～っていう」という表現の「って」という助動詞の出現率が相対的に高くなることが明らかになった。計量的な判断に基づけば、読者に語りかけるような表現が多用されるようになったと考えられる。

キーワード: テキストアナリティクス, 児童文学, 特徴語, Boruta, t-SNE

Exploring Stylistic Changes in Japanese Children's Literature through Quantitative Analysis

GEN TSUCHIYAMA^{†1}

Abstract: Quantitative analyses of literary works have been widely conducted; however, relatively little research has focused on the quantitative analysis of the writings of contemporary authors. This study investigates the diachronic changes in the writing style of the contemporary Japanese children's author Mihoko Inoue by quantitatively analyzing her works from the beginning of her writing career to the present. The results indicate that works published around 2010 exhibit relatively high frequencies of the auxiliary forms "-ta" and "rashii", whereas works published around 2020 show relatively high frequencies of "-tai" and "-tte". From a quantitative perspective, these findings suggest that Inoue's writing style has gradually shifted toward a more conversational style that addresses readers more directly.

Keywords: Text analytics, Japanese children's literature, Features, Boruta, t-SNE

1. はじめに

文学作品を対象とした計量分析では、『源氏物語』をはじめとする古典文学作品や、著作権の保護期間が終了した近現代小説を対象とした研究が数多く報告されている[1][2][3]。一方、2026 年現在において、現役作家の作品を対象とした計量分析の事例は限られており、村上春樹の作品を対象とした研究などが散見される程度である[4][5][6]。この背景には、著作権によって保護されている作品の多くが電子テキストとして公開されておらず、分析に利用可能なテキストデータを容易に入手できないことに起因すると思われる。

そこで、本研究では児童文学作家であるイノウエミホコから作品原稿の Word ファイルの提供を受け、これを分析用データとして計量分析を実施した。著作権者本人から提供を受けた電子原稿を用いることで、従来は分析対象とすることが困難であった現役作家の作品を対象とした計量分

析を可能としている。なお、イノウエミホコは 2008 年に児童文学作家として活動を開始し、2026 年現在においても現役の作家である。

本研究において分析対象となる作品は表 1 の 8 作品 [7][8][9][10][11][12][13][14]である。ただし、これら 8 作品のうち『男子☆弁当部 弁当バトル! 野菜で勝負だ!!』[8]、『男子☆弁当部 オレらの初恋! ? ロールサンド弁当!!』[9]、『男子☆弁当部 オレらの青空おむすび大作戦!』[10]の 3 作品は提供されたデータに 2 つの原稿があることから、本研究における分析対象は 11 作品となる。

表 1 に示したように、本研究において対象とする「男子☆弁当部シリーズ」が 2010 年代に発表され、『ぼくたちはプライスレス!』および『ブルーなあたしとピンクなぼく』の 4 巻は 2020 年代に発表された作品である。本研究ではこれら 8 作品を対象として、現代の作家のデビューから約 10 年間にわたる文体的特徴の継時的な変化を明らかにする。

¹ お茶の水女子大学
Ochanomizu University

表 1 分析対象となる作品の一覧

作品名	出版社	出版年
男子☆弁当部 オレらの友情 てんこもり弁当	ポプラ社	2010 年 8 月
男子☆弁当部 弁当バトル！ 野菜で勝負だ！！	ポプラ社	2011 年 3 月
男子☆弁当部 オレらの初 恋！？ ロールサンド弁 当！！	ポプラ社	2011 年 8 月
男子☆弁当部 オレらの青空 おむすび大作戦！	ポプラ社	2012 年 2 月
ぼくたちはプライスレス！ (1) 中学生作家、取材を始め ます	KADOKAWA	2020 年 2 月
ぼくたちはプライスレス！ (2) 新キャラは、いじわる貴 公子!?	KADOKAWA	2020 年 6 月
ブルーなあたしとピンクな ぼく 1 2 人の入れ替わり大 作戦!?	KADOKAWA	2021 年 6 月
ブルーなあたしとピンクな ぼく 2 ドキドキいっぱい体 育祭!?	KADOKAWA	2021 年 10 月

2. データ

先にふれたように、本研究で採り上げる小説は 8 作品である。ファイルが 2 つある作品については、ファイルの作成された日付を参照し、「男子☆弁当部(2)_0607」,「男子☆弁当部(2)_1112」,「男子☆弁当部(3)_0306」,「男子☆弁当部(3)_0609」,「男子☆弁当部(4)_0815」,「男子☆弁当部(4)_1120」のようにラベリングを行い、ファイルの区別を行った。

これらの作品の電子テキストを Web 茶まめ[15]を使用して形態素解析を行った。形態素解析では 2025 年 12 月版の「現代語話し言葉」辞書を用い、品詞別に語彙素の集計した。本研究の分析では、先行研究[3][16]を踏襲し、文中において文法的機能を担う助詞および助動詞を分析に用いた。助詞の異なり語数は 67 語、助動詞の異なり語数は 34 語である。

なお、小説の分析では会話文を除外し、地の文のみを分析対象とすることが多い[3][16][17]。しかし、児童文学作品が計量分析の対象となることは少ないため、本研究においては全文を分析に用いた。

3. 分析と考察

3.1 分析手法

本研究では 8 作品のテキストデータの可視化に主成分分

析と t-SNE を用いた。t-SNE は高次元データを 2 次元や 3 次元へ非線形に次元削減し、データの構造を可視化する手法である。

主成分分析がデータ全体の分散をできるだけ保持するように線形変換を行うのに対し、t-SNE は個体間の局所的な近傍関係を保持するように個体の配置を決定する。そのため、主成分分析では明瞭に示されない個体間の近傍関係が、t-SNE では明示的に分離されることがある。その一方で、t-SNE では軸に意味がなく、データの全体構造がおおよそ反映されない。したがって、本研究では t-SNE を高次元データに潜むクラスタ構造や局所的な関係を可視化するために用いた。

次に、イノウエミホコの 2010 年代の作品と 2020 年代の作品の間で、出現傾向が顕著に相違する単語、すなわち特徴語を抽出するため、本研究では Boruta を用いた。Boruta はランダムフォレストを利用した特徴量選択手法であり、予測に有用な変数を抽出することを目的とした手法である。ランダムフォレストの変数重要度の推定と異なり、Boruta はランダムフォレストを何度も反復することで、各変数が「予測に有意に寄与しているか」について統計的に判断する。

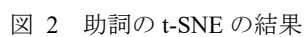
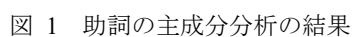
具体的には、元の変数を並べ替えて作成したシャドウ変数を基準とし、元の変数の重要度がシャドウ変数を有意に上回るかどうか検定する。その結果、有意とされた変数は Confirmed、不要な変数は Rejected、判定が保留された変数は Tentative と判定される。本研究では Confirmed と判定された単語を特徴語とした。

3.2 分析結果

まず、助詞の 67 語を対象に主成分分析と t-SNE を行った。図 1 が主成分分析の結果であり、第 1 主成分の寄与率は 30.21%、第 2 主成分の寄与率は 13.03% となる。図中において、2010 年代の作品は赤、2020 年代の作品は青となっているが、これは k-means 法によって求められたクラスタに基づいて各作品のカラーリングを行っている。

k-means 法はクラスタの中心となる初期値がランダムに付与されるため、クラスタリングを繰り返すと分析結果が異なることがあるが、本研究では k-means 法を 100 回反復し、クラスター内平方和が最小になる結果を最終的なクラスタリング結果として採用した。図 2 は t-SNE の分析結果である。t-SNE の方が主成分分析よりも顕著に 2 つのクラスタが分離しており、2010 年代と 2020 年代の小説は助詞の出現傾向が大きく異なることが推測される。

次いで、Boruta によって 2 つのクラスタ間における出現率が著しく異なる特徴語を抽出した。図 3 が Boruta の分析結果である。図 3 において、「ね」や「やら」といった図中における緑で示されている語りかけるような助詞が有意な特徴語として抽出されている。図 1 から、これらの助詞は 2020 年代に多用される単語である。



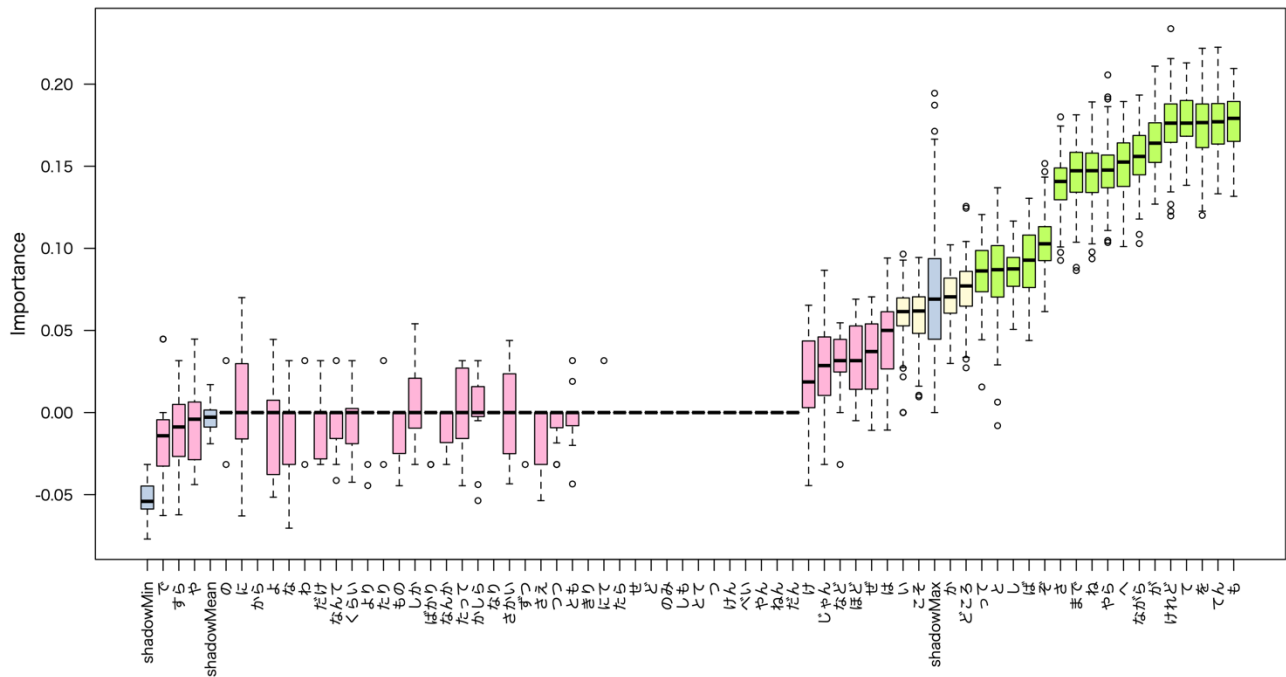


図 3 助詞の Boruta の結果

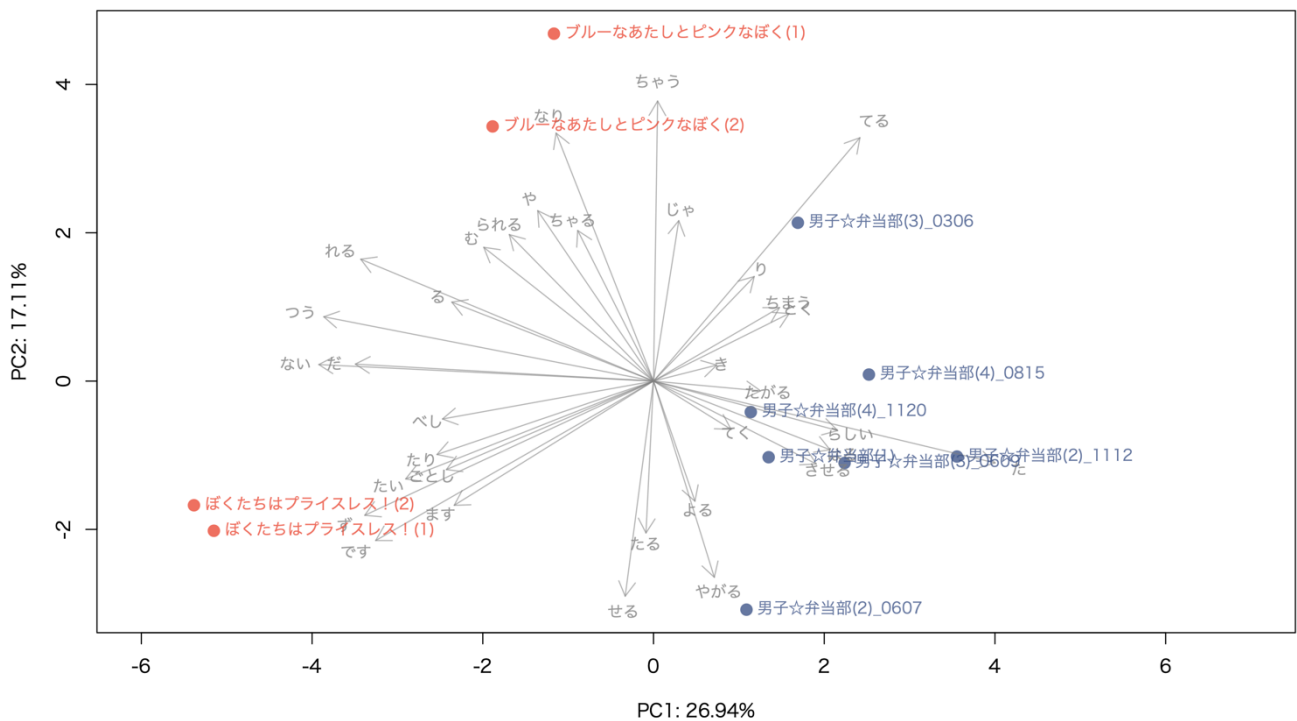


図 4 助動詞の主成分分析の結果

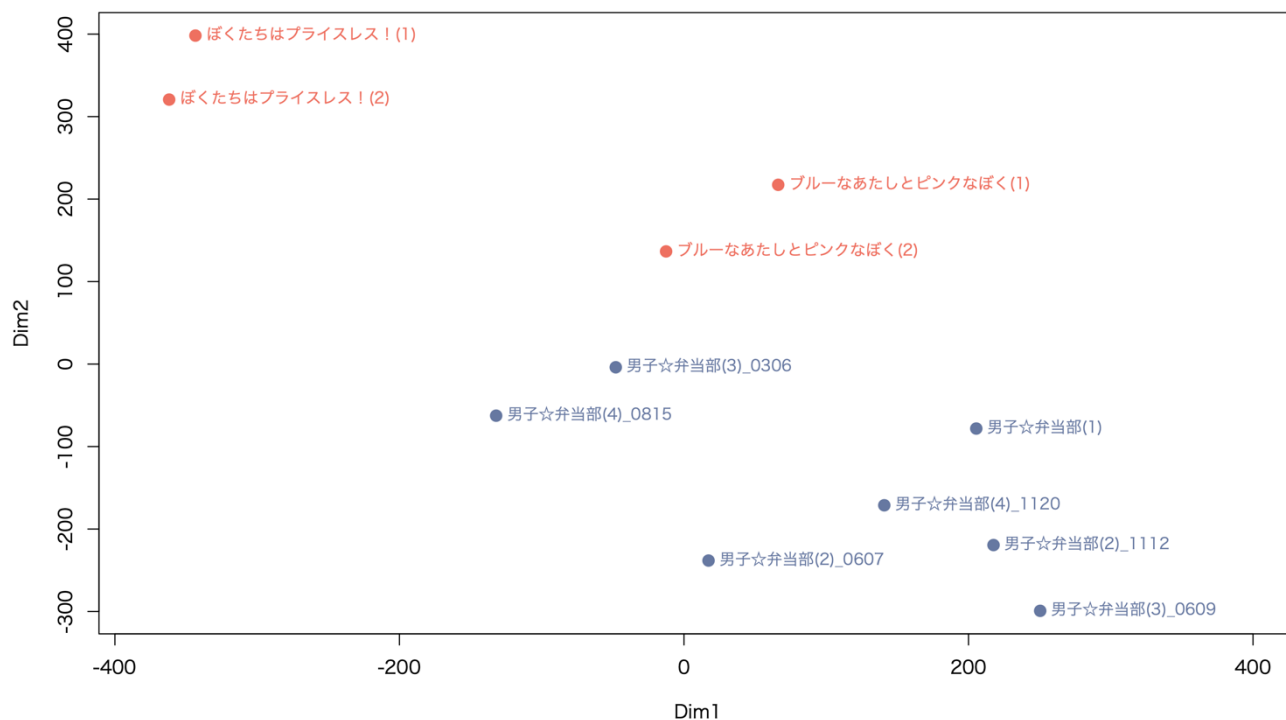


図 5 助動詞の t-SNE の結果

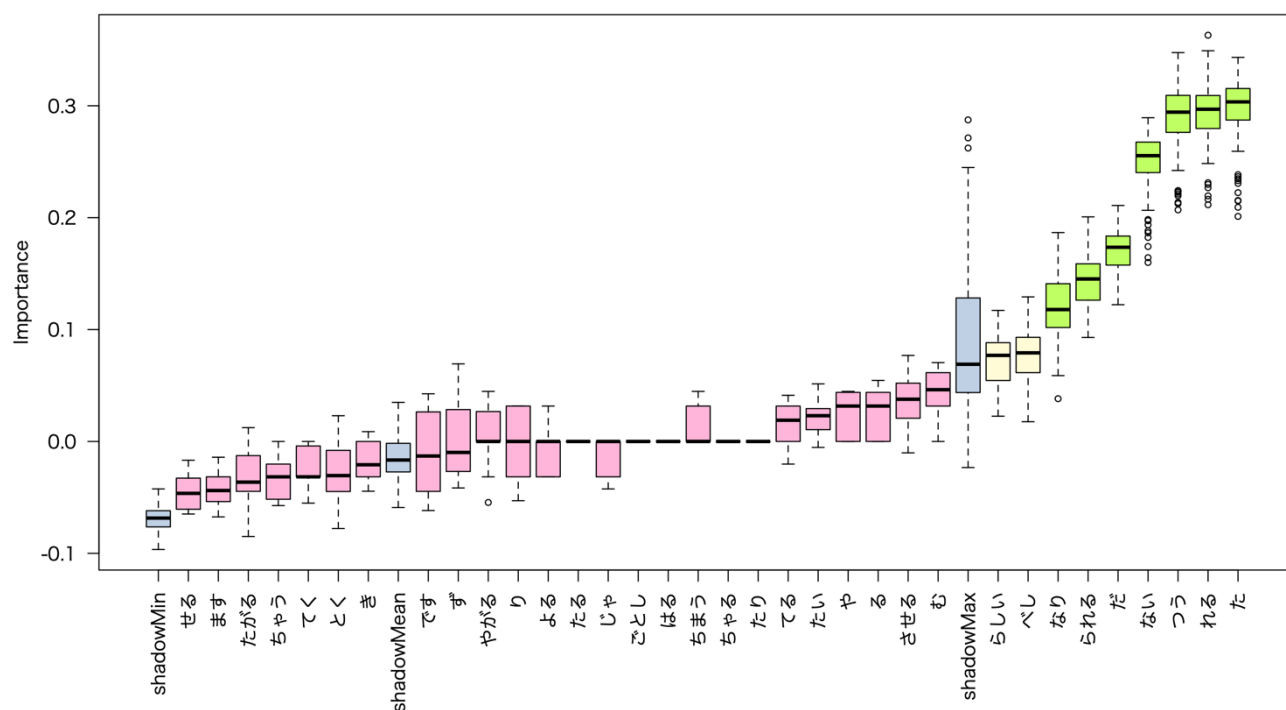


図 6 助動詞の Boruta の結果

次に、助動詞の 34 語対象に主成分分析と t-SNE を行った。図 4 が主成分分析の結果であり、第 1 主成分の寄与率は 26.94%，第 2 主成分の寄与率は 17.11%となる。助詞の分析結果と異なり、2020 年代の作品はシリーズごとに分離し、3 群に分かれると考えられる。ただし、第 1 主成分に注目すると、主成分得点が正の作品は 2010 年代の作品であり、負の値は 2020 年代の作品となる。寄与率の高い第 1 主成分で 2010 年代と 2020 年代の作品が分離しているため、ここに大きな相違があると考えられる。図 5 に示した t-SNE の分析結果も同様であり、主成分分析の結果を整合すると考えられる。2020 年代の 2 つのシリーズがそれぞれ離れて配置されることについては、より詳細な分析を行う必要がある。

また、図 6 のように Boruta によって 2 つのクラス間における特徴語を抽出した。図 4 および図 6 より「た」や「ない」といった単語が 2010 年代の作品を特徴付けている一方で、2020 年代の作品の特徴語として「つう」が抽出されている。

以上より、助詞と助動詞の分析から、計量的な判断に基づけば、デビューからおよそ 10 年が経過したことで、イノウエミホコの作品では読者に語りかけるような表現が多用されるようになったと考えられる。

謝辞 本研究では児童文学作家であるイノウエミホコ氏から、執筆時の貴重な原稿のデータをご提供いただきました。心より感謝申し上げます。

参考文献

- [1] 金明哲, 中村靖子, 上阪彩香, 土山玄, 孫昊, 劉雪琴, 李広微, 入江さやか: 文学と言語コーパスのマイニング, 岩波出版 (2021).
- [2] 村井源, 青山美月, 大田翔貴, 大場有紗, 福元隆希, 齊藤悠仁, 佐藤央一: 物語展開の共通因子とジャンルの関係性, 情報知識学会誌, Vol. 34, No. 3, pp. 256-270 (2024).
- [3] 土山玄: 夏目漱石の小説における文体の継時的変化について, 計量国語学, Vol. 33, No. 7, pp. 481-492 (2022).
- [4] 工藤彰, 村井源, 往住彰文: 村上春樹の初期三部作における構造解析, 情報知識学会誌, Vol. 19, No. 2, pp. 126-131 (2009).
- [5] 工藤彰, 村井源, 往住彰文: 計量分析による村上春樹文学の語彙構成と歴史的変遷, 情報知識学会誌, Vol. 20, No. 2, pp. 135-140 (2010).
- [6] 工藤彰, 村井源, 往住彰文: 計量分析による村上春樹長篇の関係性と歴史的変遷, 情報知識学会誌, Vol. 21, No. 1, pp. 18-36 (2011).
- [7] イノウエミホコ: 男子☆弁当部 オレらの友情でんこもり弁当, ポプラ社 (2010).
- [8] イノウエミホコ: 男子☆弁当部 弁当バトル! 野菜で勝負だ! !, ポプラ社 (2011).
- [9] イノウエミホコ: 男子☆弁当部 オレらの初恋! ? ロールサンド弁当! !, ポプラ社 (2011).
- [10] イノウエミホコ: 男子☆弁当部 オレらの青空おむすび大作戦! , ポプラ社 (2012).
- [11] イノウエミホコ: ぼくたちはブライズレス! (1) 中学生作家、取材を始めます, KADOKAWA (2020).
- [12] イノウエミホコ: ぼくたちはブライズレス! (2) 新キャラは、いじわる貴公子!?, KADOKAWA (2020).
- [13] イノウエミホコ: ブルーなあたしとピンクなぼく 1 2 人の入れ替わり大作戦!?, KADOKAWA (2021).
- [14] イノウエミホコ: ブルーなあたしとピンクなぼく 2 ドキドキいっぱい体育祭!?, KADOKAWA (2021).
- [15] Web 茶まめ <<https://chamame.ninjal.ac.jp/>> (参照 2026-07-06)
- [16] 金明哲: 文章の執筆時期の推定—芥川龍之介の作品を例として—, 行動計量学, Vol. 36, No. 2, pp. 89-103 (2009).
- [17] 土山玄: 夏目漱石の小説における文語表現について, じんもんこん 2018 論文集, Vol. 2018, pp. 269-276 (2018).