



Title	The Stylistics of Alfred Tennyson' s Poetry: A Stylometric Approach
Author(s)	藤田, 郁
Citation	大阪大学, 2024, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/96018
rights	
Note	やむを得ない事由があると学位審査研究科が承認したため、全文に代えてその内容の要約を公開しています。全文のご利用をご希望の場合は、大阪大学の博士論文についてをご参照ください。

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

論 文 内 容 の 要 旨

氏 名 (藤 田 郁)	
論文題名	The Stylistics of Alfred Tennyson’s Poetry: A Stylometric Approach (Alfred Tennyson 韻文作品の文体：計量文体論手法を用いて)
論文内容の要旨 本博士論文は、19 世紀ヴィクトリア朝の詩人 Alfred, Lord Tennyson (1809–1892) の韻文作品に計量文体論 (Stylometry) の手法を適用し、探索的に分析・考察を施したものである。 デジタル・ヒューマニティーズ (人文科学におけるコンピュータ技術の研究利用) は近年増加の一途をたどっており、英文学においても、定量的・計量的手法を用いた研究の発展、増加が顕著である。しかし、韻文作品研究においては、計量的手法の応用は散文に比べ未だ少ない傾向にある。 Stylometry は、統計分析や機械学習等の自然言語処理技術を文体研究へ応用することの総称である。コンピュータや自然言語処理技術が感情や知識を伴って文学作品を人間と同等に解釈し、「読む」ことはできない。しかし、Stylometry の最大の利点は、一度に大量のデータを客観的な視点で分析できる点にある。Stylometry は、作品の分析の範囲を広げ、客観的なデータを提示し、俯瞰して「読む」ことを可能にする。つまり、一部の作品や表現に焦点を当て、主観的に論じられてきた従来の研究に、より広い「Tennyson の作品」というコンテキストにおいて言えること、客観的な視点を投じることができる。 本論文は、Stylometry の特性を活かし、Tennyson の 593 作品に機械学習手法の一つである潜在的ディリクレ配分法 (以後 LDA) トピックモデルをあてがい、従来の Tennyson 研究とは異なる切り口から作品を分析している。得られた結果に精読に基づく質的な考察を施すことにより、Tennyson 作品の文体に関連した主題の研究に新たな知見を得ることを目的としている。本論文の研究課題は大別して以下の2点である。 1. これまでの Tennyson 研究で指摘されてきた作品の主題と同様の主題を扱っているにもかかわらず、これまでの研究の俎上に上がってこなかった作品があるか。 2. これまでの Tennyson 研究で論じられてこなかった主題が LDA によって明らかになるか。 LDA は、確率論的計算を用いてコーパス内の単語の背後に隠された意味的つながりを識別する手法であり、上記研究課題を達成するために適切な手法である。LDA がコーパス内の語の意味的つながりを明らかにする手法である点と、上記研究課題の焦点が作品内に描かれている主題であることから、作品内の主題や意味を顕著に伝達する要素である名詞に本分析の対象を限定した。 LDA を英文学作品に応用した先行研究は、主に散文テキストに焦点を当てたものであった。韻文作品を対象とした先行研究が限られているだけでなく、大規模なコーパスやデータの分析を得意とする LDA を比較的語数の少ない韻文テキストに応用し、適切な結果を得るためには、方法論に関する課題も達成せねばならない。本論文の著者が構築したコーパスに含まれる Tennyson の詩のデータサイズは、以前の散文テキスト研究のコーパスほど大きくはないが、35万以上の語数を有しており、LDA を用いた分析に足るデータ量であると判断した。LDA を Tennyson の作品に応用するにあたり、コーパス内のデータをどれくらいのサイズの文書のまとまり (セグメント) に分割するかが最も重要な方法論における研究課題である。本博士論文では、上記研究課題1., 2. に取り組む前に、Tennyson 作品に LDA をあてがう際の最も適切なセグメントサイズについて論じ、その結果を Tennyson 作品全体の LDA 分析に応用することで、手法の有用性を実証するだけでなく、Tennyson の作品に描かれている主題を捉えることに成功した。LDA による分析結果をもとに、本論では先行研究の検討とともに韻文作品の精読を行い、量的、質的研究手法の両方面から Tennyson 作品の考察を試みた。本研究は、作品を質的手法による精読 (close reading) と量的手法による遠読 (distant reading) を組み合わせて考察、論じることで、Tennyson の文体の新たな側面に光を当てただけでなく、韻文作品研究への量的手法応用の可能性を拡げるものである。本博士論文の構成は以下の通りである。 第 2 章では、Tennyson の作品、文体に関する先行研究と、計量文体論手法を用いた英文学作品研究やコーパス言	

語学分野の先行研究,そして LDA を用いた先行研究について概観する。Tennyson と他の詩人の文体,構文,語彙の類似点や相違点に焦点を当て、Tennyson の作品を検証した研究は数多あるが、定量的な方法を採用した研究は少なく、ほとんどの研究は精読に基づいた定性的なアプローチをとっている。これまでの Tennyson に関する研究の大半は、Tennyson の特定作品や一部の作品中の言葉が何を表現し、何を意味しているかを主観的に検討したり、他の詩人や作家の作品と比較したりするものであることを示した。さらに、これまで管見のかぎり確認された、定量的アプローチをとった Tennyson 研究について詳述した。また、これまでの英文学及び言語学(主に英語学)の分野で扱われてきた定量的手法について概観したのち、LDA を応用した研究について述べ、LDA を Tennyson 作品にあてがう際の方法論における研究課題を示した。

第 3 章では、本博士論文で用いる Tennyson の韻文作品を集めたコーパス(以後 T-Corpus)について説明し、LDA を応用する手立てを示している。T-Corpus 内には全 609 編の韻文作品が収められているが、その内 10 編は方言で記述されたものであり、本論文の分析の対象から除外した。本章では、方言によって書かれた作品を除外するに至った経緯を詳述している。また、T-Corpus 内の作品から名詞のみを抽出し、分析するための方法、作品をセグメントに分割する手法について説明している。

第 4 章では、LDA による Tennyson 作品の分析に適切なセグメントサイズについて論じる。LDA は、その計算の際に与えられた各セグメントを *bag of words* として扱うため、セグメント内の語数のばらつきが LDA の確率の計算に影響を与える可能性が指摘されている。Tennyson の各詩の語数は 12 から 20,000 以上に及び、大きなばらつきがある。そこで第 4 章では、分析対象の作品を 1,000 語以上からなる長編詩 67 編に限定し、同一条件下での LDA の複数の結果を比較することで、LDA 実行の際にいかように Tennyson 作品をセグメントとして扱うことが適切であるかについて論じた。第 4 章の結果、本論で扱う Tennyson の作品コーパスに適切なセグメントサイズは、作品毎の語数の平均値である 592 であることが示唆された。

第 5 章では、まず方言詩を除いた T-Corpus 内にある、内容が重複している作品の扱いについて詳述し、599 編の作品からさらに 6 作品を分析対象から外すことを説明している。その上で、重複作品を除く Tennyson の 593 の詩を上記第 4 章で得られた適切なサイズである 592 語のセグメントに分割し、LDA を適用して得られた結果を提示した。LDA の結果には、18 のトピックを最も特徴づける上位 20 のキーワードや、各トピックが対象コーパスに出現する程度を示す α (アルファ) 値を提示した。LDA の結果と作品を読み込むことによって著者が各トピックに付与したラベル、及び各トピックが顕著に現れる上位 65 作品も提示している。

第 6 章では、第 5 章で提示された LDA の結果をもとに、18 のトピックが何を表し、LDA が Tennyson における主題や要素をどのように検出したかを、LDA の結果と作品の精読から得られた知見を合わせ、詳細に論じた。例えば、T-Corpus に最も多く現れる Topic 1 と Topic 6 が顕著に現れる作品では、永遠の命や永遠の若さを表す *ontological immortality* (存在論的不死性)ではなく、*epistemological immortality* (認識論的不死性)を扱っている作品が多いことが分かった。さらに、認識論的不死性は、Tennyson においては 2 種類に下位分類できることを述べた。自然に関するトピックのうち、Topic 2 の主な主題は、海であり、海が Tennyson 作品内における孤独、絶望、悲しみ、死と関連していることを示した。先行研究は、Tennyson の海について死との関連を強調して論じる傾向があったが、本研究の結果から Tennyson の作品における海と死の関連だけではなく、孤独や絶望、悲しみといった感情と共に描かれ、登場人物の感情を鮮やかに描写する効果があることが分かった。Tennyson の長編詩 *The Princess* における自立した女性像や、“*The Lady of Shalott*”等に現れる不幸な女性に関する先行研究がある中で、LDA の結果も同様に作品内での女性の描きわけを捉え、トピックとして分類した。Topic 7 は主にヴィクトリア朝時代の女性に対する社会的観点から逸脱した、自立的、能動的な女性に関するトピックである。対して Topic 13 と Topic 11 は、当時の社会的態度に対応した、従順で受動的な女性に関連するトピックである。LDA は、作品内の女性登場人物だけでなく、男性登場人物の描き分けも捉えた。Topic 4 や Topic 10, Topic 14 は、主に男性の登場人物に関するトピックであり、男性登場人物の置かれた状況や登場するシーンはトピックによって異なるものの、いずれのトピックにおいても男性中心の状況・社会が描かれている点が共通しており、Tennyson 作品の中での男性は、女性と異なり家庭の中でも社会においても居場所を持ち、社会や家庭の中心として描かれていることが明らかになった。

最後に、第 7 章では、本博士論文の結論を簡潔に述べる。本博士論文第 5 章および第 6 章の LDA を用いた分析と考察の結果、これまでの Tennyson 研究で指摘されてきた作品の主題と同様の主題及び管見の限りでは論じられてこなかった主題が明らかとなり、これまでの研究で論じられてきた主題には、例えば *immortality* や海が挙げられるが、これらの主題はいずれもこれまで論じられてきた作品以外の作品にも描かれていることが明らかになった。さらに、先行研究では論じられてこなかった *immortality* の種類や、海と死や登場人物の感情との関連も LDA の結果及びその考察によって示唆された。

本博士論文では、LDA を用いて Tennyson の詩の文体に関する知見を提示した。LDA は、これまで Tennyson 研究に用いられたことがない手法であり、本論文は手法の妥当性の面においても挑戦的な研究であった。本論における LDA 分析の結果、先行研究で既に論じられている主題が、これまで論じられてこなかったと考えられる要素と共に確認されたことで、本研究、及び LDA を Tennyson 作品に応用する方法論の妥当性が担保されたといえることができる。また、精読と量的分析をバランスよく組み合わせることで、Tennyson の 593 編の詩に明示的に現れている要素の背後に隠れた新たな側面に光を当てると共に、量的手法と質的手法が補完し合いながら新たな知見を見出していけることを実証した。

論文審査の結果の要旨及び担当者

氏 名 (藤 田 郁)			
	(職)	氏 名	
論文審査担当者	主 査	教授	田畑 智司
	副 査	教授	小口 一郎
	副 査	准教授	ホドシチェク ボル

論文審査の結果の要旨

藤田郁氏の博士学位請求論文は、19 世紀英国ヴィクトリア朝の桂冠詩人 Alfred, Lord Tennyson (1809–1892) の韻文作品に計量文体論 (Stylometry) の手法を適用し、探索的に分析・考察を施したものである。同氏は Stylometry の特性を活かし、Tennyson の 593 作品に潜在的ディリクレ配分法に基づくトピックモデル (以降 LDA と表記) を適用し、従来の Tennyson 研究とは異なる切り口から作品を分析している。量的解析法である LDA より得られた分析結果と、テキストの精読に基づく質的な解釈を有機的に組み合わせることにより、Tennyson 作品において展開される主題およびそれらを表象する言語表現について説得力のある客観的データを提示し、新たな知見を示した研究である。本論文は以下のような構成となっている。

第 1 章は、本論文の学術的背景、研究の着想に至った経緯について述べている。特に、関連する国内外のデジタルヒューマニティーズ研究の動向と本研究の位置づけを示す導入部である。

第 2 章では、Tennyson の作品、文体に関する先行研究と、Stylometry を用いた英文学作品研究やコーパス言語学分野の先行研究、そして LDA を用いた自然言語処理研究について概説している。これまで、Tennyson と他の詩人とを比較し、文体、構文、語彙の類似点や相違点に焦点を当てた研究は数多存在するものの、計量的分析法を採用した事例は極めて少なく、ほとんどは精読と研究者の主観的関心に基づいた定性的な研究であり、分析手法や知見の再現可能性、複製可能性、反証可能性を十分に担保できていないうらみがある。そのため、藤田氏は、LDA を用いた計量的アプローチによって従来の研究を補完するだけでなく、これまでほとんど光が当てられなかった主題を掘り起こし、Tennyson 研究の地平を広げる取り組みをしている。本章ではさらに、LDA を韻文作品に応用する際に直面する課題を示し、それらに対する答えを第 3 章～第 5 章でもとめている。

第 3 章は、本博士論文で用いる Tennyson の韻文作品を集めたコーパスについて詳述し、LDA を応用する手立てに言及している。コーパスには全 609 編の韻文作品が収められているが、うち 10 編の方言詩と重複作品 6 点を除く 593 作品が本論文の分析対象である。テキストに生起する名詞を対象として LDA を実行することで、作品の主題と密接に関連する言語項目に焦点を絞る必要があることも本章で唱えられている。

著者は第 4 章で、LDA を用いた文学作品の分析を行う際、テキストを適切なサイズに分割し、テキストの分量、長短の差が分析結果を歪めることがないように配慮を施す必要性があることを指摘している。LDA は、テキストデータをいわゆる“bag of words”に還元してトピックの確率計算を行うため、分析対象とするテキストサイズの分散が多くなると、適切なアウトプットの解釈が困難になる恐れがあるためである。Tennyson の作品は総語数 12 の短詩から 20,000 語以上の叙事詩に及び、分散が極めて大きい。そのため本章では、分析対象の作品を 1,000 語以上からなる長編詩 67 編に限定し、分割サイズを変えることで、出力がどのように変化するか実験を繰り返し、最も粒度の高いトピックモデルを構築可能な分割サイズを推定した結果、最適解として、コーパスを構成するテキストの平均値である 592 語に設定して分析を行うことが肝要であると論じている。

さらに、第 5 章は、均等なサイズに分割したテキストをもとに、Tennyson 作品に現れる主題をできるだけ高い解像度で表しうるトピック数の推定に取り組んでいる。抽出トピック数を変化させることで perplexity 値, coherence 値などの指標がいかに変動するかに加え、得られたトピックの質的解釈の可否を注意深く考察した結果、最適解として、トピック数 18 に設定した際に最も粒度の高い分析が可能になることを指摘している。加えて、各トピックの質的解釈 (ラベリング) が提示され、各トピックが顕著に出現する作品リストも示されており、抽出したトピックと

作品の系譜との関係性が浮き彫りにされている。

第6章は、LDAによる分析結果と作品の精読に基づく知見を組み合わせることで、得られた18点のトピックそれぞれが何を表しているか、そして、Tennyson詩における主題や要素がどのように検出されているかを詳細に論じている。たとえば、最も大きな α 値を示す（すなわち、最も普遍度が高い）Topic 1とTopic 6が顕著に現れる作品では、永遠の命や永遠の若さを表す ontological immortality（存在論的不死性）ではなく、epistemological immortality（認識論的不死性）を扱っている作品が多いことを明らかにしている。さらに、認識論的不死性は、Tennysonにおいては2種類に下位分類できることを示している。他方、自然に関するトピックのうち、Topic 2の主題は、seaであり、seaがTennyson作品内における孤独、絶望、悲しみ、死の表象と関連していることを明らかにしている。Tennyson作品におけるseaについて、先行研究では死との関連のみが強調される傾向が認められるが、本論文はseaのトピックは死との関連だけではなく、孤独や絶望、悲しみといった感情と共に描かれ、登場人物の感情を鮮やかに描写するのに効果的に用いられている点にも注目している。また、Tennysonの長編詩*The Princess*における自立した女性像や、“*The Lady of Shalott*”等に現れる不幸な女性像を取り上げた先行研究があるが、LDAの結果も同様に作品内での女性の描き分けを捉え、異なるトピックに分類している点は興味深い。たとえば、Topic 7は主にヴィクトリア朝時代の女性に対する社会的観点から逸脱した、自立的、能動的な女性に関するトピックである。対してTopic 13とTopic 11は、当時の社会的態度に対応した、従順で受動的な女性像を反映するトピックである。LDAは、作品内の女性登場人物だけでなく、男性登場人物の描き分けも捉えており、Topic 4やTopic 10, Topic 14は、主に男性の登場人物に関するトピックである。男性の置かれた状況や登場場面はトピックによって異なるものの、いずれのトピックにおいても男性中心の状況・社会が描かれている点が共通しており、Tennyson作品の中での男性は、女性と異なり家庭の中でも社会においても居場所を持ち、社会や家庭の中心として描かれていることを説得力のあるデータで示している。

最後に、第7章が本博士論文の結論である。本研究の分析結果は、これまでのTennyson研究で指摘されてきた知見を量的に確証したり、補完する一方、従来の研究ではほとんど着目されなかった主題と作品にも光を当てるものである。先行研究で頻繁に論じられてきた主題として、例えばimmortalityやseaがあるが、本論文は、こうした主題が、従来ほとんど批評の光が当てられてこなかった作品にも通底していることを詳らかにしている。そして、本研究が、Tennyson作品の新たな読みの視野を広げることを願って論を結んでいる。

本博士論文で、藤田郁氏が活用したLDAは大量のテキストデータから有用なトピックを掘り起こすことを支援し、分析者が巨視的視点で発見した主題や意味を微視的に深く考察する糸口を与えてくれる。この手法を英文学作品、とりわけ多数の詩作品の横断的研究において利活用し、学術的価値の高い知見を得るには、量的分析により作品コーパスの全体像を把握する一方、個々のテキストの特徴を具に捉え、マクロな視点で立てた仮説や読みをミクロな分析や個々のデータの質的解釈により検証し、それをまたマクロな分析にフィードバックし、さらに仮説や読みを精緻化していくという仮説発見と検証の循環を築くことが不可欠である。藤田氏の論考は、その点において、質的な精読(close reading)と量的手法による遠読(distant reading)を相補的に組み合わせて考察することにより、Tennysonの作品の新たな側面に光を当てるとともに、韻文作品研究への量的手法応用の可能性を拓けるものである。定量的手法と分析結果の質的解釈のバランスを取って、Tennyson作品の意味的・主題的特徴を詳細に記述しており、客観性、再現性や反証可能性を担保するための科学的な手順を踏まえた実証的研究となっている点で、本博士学位請求論文は、Tennyson詩の研究に新たな地平をもたらす価値ある研究であると評価できる。

もっとも、LDAによるトピックのラベリング、すなわちトピックの解釈は、キーワードの機能的解釈に依存しており、分析者による解釈の恣意性をすべて排除することができているわけではない。また、本研究の論点の多くが依拠しているヒートマップやネットワークグラフによるテキストと主題との関係性の視覚化法は、皮肉にもデータ項目数が膨大であるため、極めて細かな文字とセルで表示がなされている。それゆえ、本論文の読者がこれらの視覚化手法の恩恵を容易に実感しにくい点については、今後表示方法の工夫改善が求められる。さらに言うのであれば、第6章で提示した精読についても、批評的知見の理解を深めた上で、分析技法をより精緻化していく余地があると思われる。しかし、これらは本論文の学術的価値を損なうほどの問題点ではなく、全体として本論文は、デジタルヒューマニティーズの手法と作品の質的読みを適切に組み合わせ、Tennyson作品において展開される主題およびそれらを表象する言語表現について説得力のある客観的データを提示し、新たな知見を示した意義深い研究であり、その学術的貢献は明確である。

以上により、本論文を博士（言語文化学）の学位論文として価値のあるものと認める。