

日本語テキストを対象とした計量的研究の現状

上阪彩香†

日本語テキストに関する計量的研究は、日本語の単語認識の困難さ等が原因で欧米に遅れをとり 20 世紀にはじまった。しかしながら、情報処理技術の発展を受け、電子テキストを簡易に操作するためのツールの開発が急速に進められている。本報告では、日本語テキストを対象とした計量的研究の歴史および現状について述べる。

An introduction of the quantitative study of Japanese text

AYAKA UESAKA†

This paper reviews the history and recent research on the quantitative study of Japanese text. The quantitative analysis of Japanese text has lagged behind. There are several reasons for this. For instance, unlike English and other European languages, Japanese text does not have spaces between words. To perform quantitative analysis of Japanese text, spaces need to be added between the words in all sentences. However, the potential of quantitative analysis of textual data has dramatically advanced. Based on these developments, quantitative analysis has become a powerful tool in textual scholarship. In this paper, we discuss some software for Japanese text.

1. はじめに

テキストを対象とした学問分野のひとつに計量文献学がある。村上（1996）は、計量文献学を単語の使用率、品詞の構成比、文長・単語長の平均値や分布、語彙量等の文章の数量的性質に注目し、その統計分析によって、文献の著者の推定、真贋判定、成立年の推定、成立順序の推定等の解明を試みる研究分野としている。

歴史的資料として受け継がれてきた文学作品、宗教書、思想書のなかには、記述内容の検討や筆跡鑑定、化学的鑑定等の従来の検討方法だけでは解決できず問題が残されたままとなっているものも多い。このような問題に対しては、伝統的な文献研究とは異なる視点からの研究が必要である。

文章に執筆者の特徴があらわれ、文章を対象とした研究に計量的視点を取り入れることで、有用な情報が得られるということは、これまで指摘されてきている。フランスの博物学者 Georges - Louis Leclerc, Comte de Buffon (1707～1788) は、1753 年のアカデミー・フランセーズの入会演説で『文体論』について論じる中で、「Le style est l'homme même. (文体は人間そのものである)」と述べた (P. Guiraud、佐藤信夫訳、1959)。この一句は、現在は「文は人なり (The style is the man.)」という格言となっている。

P. Guiraud(佐藤信夫訳、1959) は「統計学とはまさしく

逸脱の科学〔偏差の研究〕である。また逸脱を観察し、測定し、解釈するための方法である。だからそれは、文体の研究にとってもっとも有効な道具のひとつとして受け入れられてしかるべきものだったのだ。」と述べている。

さらに、国語学者の波多野（1935）は、「作家が如何なる表現手段を多く使用するかわからないとき、我々は統計を使用するものである。作家は多く反無意識に自己の言語を駆使している。従って、どんな言語手段によって自分がある表現価値を創造し得たかは作家に聞いてもわからない場合がある。このようなときに手掛かりになるものが統計的調査なのである。」と文章の統計分析の有用性について指摘している。

2. 欧米における研究

2.1 欧米における早期の研究

文章の数量的な特徴に着目し、統計的手法を用いた本格的な研究が欧米で始まったのは 19 世紀である。1851 年に数学者 Augustes de Morgan (1806～1871) は、新約聖書の the Epistle to the Hebrews (ヘブライ人への手紙) が聖パウロの書いたものであるのかを判定するために聖パウロの手紙とヘブライ人への手紙に用いられている単語の長さの平均値を比較検討することを提案し、単語の長さの平均値を分析することで偽物の著作を見破ることができると

† 同志社大学 研究開発推進機構
Organization for Research Initiatives and Development, Doshisha University

述べた (1882)。この考えに触発され、T.C.Mendenhall は 1887 年、*Science* に単語の長さの分布に関する研究を発表し、1901 年にこの手法を William Shakespeare (1564～1616) 作品の著者の検討に適用した。Mendenhall は Shakespeare 作とされる戯曲が哲学者の Francis Bacon (1561～1626) の著作ではないかという疑問について、Shakespeare 作品 40 万語と Bacon 作品 20 万語の単語の長さの分布を用いて検討した。その結果、Shakespeare 作品では 4 文字からなる単語が最も多く用いられているのに対して、Bacon の著作では 3 文字からなる単語を最も多く用いており、2 人が用いた単語の長さの分布のモード (最頻値) が異なることから、Shakespeare 作品の著者は Bacon でないと結論付けた。Mendenhall の研究は、文章に統計的手法を用いた初期の研究として広く知られている。

Shakespeare 作品の著者問題に対しては、Williams (1975) が、Shakespeare、Bacon と同時代の作家で散文と韻文の文章を残している Sir Philip Sidney (1554～1593) を比較対象とし、この問題に取り組んだ。Shakespeare の韻文、Sidney の韻文、Sidney の散文、Bacon の散文の単語の長さを比較した結果、Sidney の韻文は Sidney の散文よりも Shakespeare の韻文と類似しており、Sidney の散文は Sidney の韻文よりも Bacon の散文に似ているということを示した。Bacon の作品が散文、Shakespeare の作品は韻文であることから、Mendenhall が見出した分布のモードの違いは作者の違いではなく、散文と韻文の違いではないかと指摘している。

執筆年代の推定を行った初期の研究としては、L. Cambell (1867) が Plato の 30 余りの対話篇の執筆年代の推定を語の出現頻度を用いた計量的研究が挙げられる。

2.2 具体的な著者問題の解明を目指した先行研究

Mosteller・Wallace (1963) は 1787 年から 1788 年にニューヨークの新聞で、Alexander Hamilton (1755～1804、初代アメリカ合衆国財務長官)、John Jay (1745～1829、初代連邦最高裁判所長官)、James Madison (1751～1836、第 4 代アメリカ合衆国大統領) の 3 人がアメリカ合衆国憲法の批准を促進するために作成した 77 編の論説 *the Federalist papers* (のちに 8 編追加して書籍として刊行された) の著者問題に取り組んだ。77 編の論説のうち、Jay の 5 編、Hamilton の 43 編、Madison の 14 編、Hamilton と Madison の共著の 3 編は著者が明らかであったが、Hamilton か Madison のどちらが書いたのか不明のものが

12 編存在した。この 12 編の著者に関して、機能語と文章の内容に影響を受けにくい単語の使用率を用いて、判別分析やバイズの手法等で分析した結果、12 編すべてが Madison の手によると結論づけた。Holmes・Forsyth (1995) もこの問題に対し、語彙の豊富さ、機能語を用いたクラスター分析と主成分分析等での検討を試みている。

1861 年に *New Orleans Daily Crescent* 紙で Quintus Curtius Snodgrass と署名され、10 回にわたり掲載された *Quintus Curtius Snodgrass letters* は、“*The Adventures of Tom Sawyer*” (1876) の著者として知られ、アメリカ近代文学の父ともいわれる Mark Twain (1835～1910) の著作であるという説が出された。これらの *letters* で描写された冒険は、アメリカ合衆国下院で南軍の脱走兵という非難を受けた Twain の南北戦争での行動を伝える重要な資料であるとされているが、Brinegar (1963) はこの問題に対し、単語の長さの分布の比較、カイ二乗適合度検定、二標本 *t* 検定を用いた分析で、Twain が著者であるという説を否定した。

Binongo (2003) は、L.Frank Baum の死後発表された “*The Royal Book of Oz*” の著者について検討した。“*The Royal Book of Oz*” は Baum が残した原稿を用いて Thompson が増補・編集したとされてきたが、現在では Thompson のオリジナルであると考えられている。この問題に対し、Binongo は出現率の高い上位 50 の機能語を主成分分析で検討し、Baum と Thompson の文章の特徴を明らかにしたうえで “*The Royal Book of Oz*” が Thompson の作品であると結論づけている。

Kestemont・Moens・Deploige (2013) は中世の有力な女性作家のひとり Hildegard von Bingen (1098～1179) の “*the Visio ad Guibertum missa*” および “*Visio de Sancto Martino*” の著者について検討した。Hildegard は秘書の Guibert of Gembloux に文章の文法訂正と文章の改変を許可しており、“*the Visio ad Guibertum missa*” と “*Visio de Sancto Martino*” はどちらも Guibert が Hildegard の秘書であったときに書かれた共著作品である。Kestemont らは、主成分分析等の手法を用い、Hildegard、Guibert 等の文章を分析し、共著作品は、共著者の文章とは明らかに異なる文体を示すと指摘した。

文献の著者に関する研究では文の長さ、単語の長さ、機能語の使用率等の様々な変数を用いて検討がなされてきた。Grieve (2007) は、提案された 39 種類の変数を比較分析

し、Word and punctuation mark profile、2-gram profile、3-gram profile 等が著者の分類に有効であることを示した。

3. 日本における研究例

3.1 日本における早期の研究

日本語における計量的な初期の研究として、波多野 (1935) の『文章心理学』が挙げられる。波多野は谷崎潤一郎 (1886~1965) と志賀直哉 (1883~1971) の文章を文の長さ、句読点、品詞の頻度等を項目ごとに比較検討し、志賀の文章を体言型の文章、谷崎の文章を用言型の文章であるとした。

大野 (1956) は『万葉集』、『土佐日記』、『竹取物語』、『枕草子』、『源氏物語』、『紫式部日記』、『讃岐典侍日記』、『方丈記』、『徒然草』の 9 作品を調査し、動詞、形容詞、形容動詞の構成比が、名詞の構成比の大小と逆順で並ぶ傾向があることを示し、水谷 (1965) はこの大野の語彙法則を定式化した。

さらに文学における文体論で、統計的手法を用い、指標モデルを考案した初期の研究として、樺島・寿岳 (1965) が挙げられる。

3.2 具体的な著者問題の解明を目指した先行研究

日本語の文献の著者問題に計量分析を試みた研究の嚆矢として、安本 (1957) が挙げられる。安本は、紫式部が著したとされる『源氏物語』の「宇治十帖」の作者が源氏物語の他の 44 帖の作者と同一人物であるかを、名詞、用言、助詞、助動詞、品詞数を数え、統計的仮説検定法を用いて検討した。その結果、宇治十帖は紫式部の作品ではないと結論付けた。

葦沢 (1965) は、『由良物語』の著者が加茂真淵と建部綾足にどちらであるのかを助詞、助動詞、接続詞の一部に判別関数を用いて検討した。その結果、加茂真淵というよりは建部綾足によって書かれたのではないかと結論付けている。

伊藤・村上 (1992) は日蓮著作の 24 編、日蓮偽作の 16 編、日蓮門下の日順と日興の著作 5 編、真偽未決の文献 5 編の計 50 編の文献に、文の構造に関する情報と言葉の出現率に関する情報を用いて、クラスター分析で分析し、真贋論争の続いてきた『三大秘法稟承事』の著者について検討した。その結果、『三大秘法稟承事』は日蓮の真作であると結論付けている。

金・樺島・村上 (1993) では、文章の中の読点がどの文字の後ろにつけられているのかの情報に基づいて、クラス

ター分析、主成分分析、判別分析で文学作品 (井上靖・中島敦・長谷川海太郎) と科学技術論文の著者別の分類を試みた。文章の中の読点に関する情報は著者によって異なり、さらに同じ著者のものは比較的安定していることを明らかにし、読点が著者の分類に有効なことを示した。

村上・今西 (1999) は、源氏物語を対象とし、巻の成立順序と「宇治十帖」他作者説について助動詞を用いた数量化三類で検討した。その結果、「宇治十帖」の異質性を裏付けた。

文章の特徴には、時と共に変化するものと変化しないものがある。金 (2009) は、芥川龍之介の作品の経年変化について統計的手法を用いて検討している。また村上・岸野・伊藤 (1990) では、日蓮遺文 23 編を対象に経年変化について調べ、動詞と副詞は使用率には変化が見られないが、普通名詞、固有名詞、形式名詞、代名詞、数詞、形容動詞、助詞、接頭語、接尾語、形容詞、助動詞、連体詞、接続詞の 13 品詞では、1265~1270 年の間に、使用率に変化が生じていると指摘した。

村上・岸野 (1990) では、漢文体と和文体について検討している。日蓮遺文を中心とした鎌倉期の文献 50 編を用いて検討した結果、和文体の文献では動詞、形容詞、助動詞といった活用のある語が多く用いられ、漢文訓読体では普通名詞、代名詞、副詞、接続詞といった活用のない語が多く用いられているということを明らかにした。

計量的に文章の執筆者を分析することは、広く認められるようになり、文学作品の真贋研究に限らず、現実社会の問題への適応も検討されている。

村上 (2004) には、保険金目当てのひき逃げ事件での容疑者の逮捕に、文章の計量分析が有力な情報として使用されたことが紹介されている。本分析事例では、容疑者の警察への上申書および保険会社への回答書、事件発生後に警察署に届いた目撃者の手紙および犯人の告白書・遺書、10 人の学生の文章、容疑者と同年齢の 10 人の文章、インターネットで公開されている 2 人の裁判所への上申書を対象に、どのような助詞がどの程度用いられているかに関する頻度情報、どの助詞のあとにどの助詞が出現するかに関する頻度情報、どの文字のあとに読点が付けられているかに関する頻度情報の 3 種類をクラスター分析、主成分分析を用いて検討した。その結果、容疑者の上申書、回答書、目撃者の手紙、犯人の告白書・遺書が多くの場合まとまりをもち、10 人の学生の文章、容疑者と同年齢の 10 人の文章、

2人の裁判所への上申書とは異なるという結果を得た。このような分析結果をもとに取り調べを行い、容疑者が犯行を認め、事件が解決したとされている。

財津・金(2015)は、少年Aによる「神戸連続児童殺傷事件」の犯行声明文、少年Aが作成したとされる犯行告白文4点および作文1点、「三菱重工爆破事件」の犯行声明文、無関係の5名の中学生の作文5点、「赤報隊事件」、「東京・埼玉連続幼女誘拐殺人事件」、「パソコン遠隔操作事件」、「黒子のバスケ脅迫事件」の4つの事件に関与した犯行声明文を対象に多次元尺度構成法およびクラスター分析を用いて、犯罪にかかわる文章における著者識別の有効性の検討を行った。文字の bigram、品詞の bigram、文の長さ、漢字・仮名の使用率に着目し、検討を行った結果、著者判別は可能であり、科学鑑定として有用であるとともに、法科学で扱うことができる分野の幅を広げる新しい手法であると述べている。

神田(2013)は、裁判員裁判導入前後の刑事事件の判決文(量刑理由)を対象とし、多次元尺度構成法等の手法を用いて、裁判官のみによる裁判と裁判員制度による裁判における法的判断の差異の可視化を試み、判決文に計量的研究を活用する有用性を指摘した。

4. 情報処理技術の発達

欧米に比べて、日本語の文献に関する数量的研究は遅れ、研究が行われるようになったのは、20世紀の後半である。研究が遅れた原因としては、日本語そのものの問題として日本語が分かち書きされていないことが挙げられる。近年、情報処理技術の発展を受け、電子テキストを簡易に操作するためのツールの開発が急速に進められている。日本語を対象とした研究のなかでも、現代文を対象とした研究に関しては、比較的容易におこなう環境が整いつつある。本章では、文章の計量分析に用いられているソフトウェアについて述べる。

4.1 電子化されたテキストデータ

文章をコンピュータを用いて検討するには、電子化されたテキストデータが必要となる。青空文庫[27]には、著作権の消滅した作品が無料で公開されており、テキスト形式で電子化されたテキストデータを入手することが可能である。

4.2 形態素解析ソフトウェア

日本語は分かち書きされていないため、文章の数量分析を行う際には、単語の認定作業が必要であるが、文章を自

動的に単語に分割し、品詞情報を付与するツールも公開されている。形態素解析を行うソフトウェアとしては JUMAN、Mecab、ChaSen が広く知られている。

日本語は時代によって変遷するが、この変遷に合わせた形態素解析を行うための辞書の開発も進められている。Web茶まめ[28]では、Unidic辞書を使用したMecabでの形態素解析をWEB環境で行うことが可能である。現代語、近代文語、中世文語等の幅広い時代の文章に対応している。

中国語は日本語と同様に、分かち書きされていないが、NLPIR (ICTCLAS) [29]やStanford POS tagger[30]を使用することで、形態素解析が可能である。Stanford POS taggerは、英語、ドイツ語、フランス語、アラビア語、スペイン語にも対応している。

4.3 分析用ツール

分析ツールも開発されており、日本語のテキストに対応しているテキスト分析ツールとして、以下の三種類があり、フリーソフトウェアとして公開されている。

(1) MTMineR

MTMineRは、Windows用のツールであり、主に日本語、中国語、韓国語、英語に対応している。長さの分布、n-gram、共起、KWIC (Keyword In Context) 検索等の機能があり、情報の集計およびテキストの整形をおこなう。統計分析処理言語Rも設定を行うことによって、MTMineRで使用することが可能である[31]。

(2) KH Coder

Windows、Linux、Mac OS Xでの使用が可能で、日本語、英語、中国語、韓国語、ロシア語、カタロニア語等の言語に対応している。抽出語リストの作成・検索、KWIC検索、Rでの分析結果をKH Coderで視覚化する可能である[34]。

(3) CasualConc

Mac OS X用のツールである。日本語、韓国語、中国語、英語等の言語に対応している。KWIC検索、単語クラスター検索、コロケーション検索、単語リスト・n-gramの作成の機能がある。Rをインストールすることで、CasualConcで作成した頻度表等の視覚化をおこなう[35]。

5. おわりに

最後に、研究の限界と展望について述べる。研究の限界として、計測可能な側面のみでの検討の結果である点、分析には候補となる著者の形態素解析済み電子テキストが必要である点が挙げられる。

現存している文化資源（テキスト）に、文章の数量分析という新たな切り口を活用することで、文章の意味内容や書誌情報を用いた研究では得られなかった知見が得られる可能性がある。しかし、村上（2016）は「客観的なデータを用いた分析というだけで、分析結果を過大に評価することは慎まなければならない。」と述べている。これは、文章の数量分析で得られた結果は、数量可能な側面を数値化し、分析を行っているため、物事の一面のみを切り取っている可能性があることが原因である。数量分析で得られた結果のみを根拠に判断するのではなく、これまで国文学研究の領域で積み重ねられてきた、意味内容、書誌情報等から得られた知見と数量的研究で新たに得られる知見の積み重ね、もしくは、数量的研究で得られた知見を国文学的観点から再検討されることによって、真実に近づくことができる。

文章の数量分析を行うためには、形態素解析済みの電子テキストが必要である。著者推定の問題では、候補として挙げた人物の中に真の著者が含まれていなければならない。報告者は、近世の浮世草子作品を中心とした文章の計量的研究を行っている。研究を行う際、問題となるのは、電子テキスト化された近世文学作品の資料が少ない点と浮世草子作品に対応した形態素解析辞書が存在しないという点である。近世文学作品を対象とした数量的な研究をはじめようと考えた場合、問題解明までかなりの時間を要することが予想される。今後、このような状況の改善が期待される。

参考文献

- [1]村上征勝.(1996).『計量統計学』9(1). pp.65～74.「計量文献学の歴史と課題」. 日本計量統計学会.
- [2]Guiraud,P(佐藤信夫 訳). (1959).『文体論—ことばのスタイル』. 白水社.
- [3]波多野完治.(1935).『文章心理学』. 三省堂.
- [4]De Morgan, S.E. (1882). Memoir of Augustus De Morgan by his wife Sophia Elizabeth De Morgan; with selections from his letters. London : Longmans, Green.
- [5]Mendehall, T.C.(1887). Science.IX(214). pp.237～246. The Characteristic Curves of Composition.
- [6]Mendehall, T.C. (1901). Popular Science Monthly. 60(7). pp.97～105. A mechanical Solution of a Literary problem. New York: The Science press.
- [7]Williams, C.B. (1975). Biometrika.62(1). pp.207～212. Mendenhall's studies of word-length distribution in the works of Shakespeare and Bacon. Oxford University Press
- [8]Cambell, L. (1867). The Sophistes and Politucus of Polato. Clarendon Press.
- [9]Mosteller,F. and Wallace,D.L. (1963). Journal of the American Statistical Association. 58(302). pp.275～309. Inference in an Authorship Problem. Journal of American Statistical Association.
- [10]Holmes,D.I. and Forsyth, R.S. (1995). Literary and Linguistic Computing.10(2). pp.111～127. The Federalist Revisited:New Directions in Authorship Attribution.Oxford University Press.
- [11]Brinegar,C.S. (1963). Journal of the American Statistical Association.58(301). pp.85～96. Mark Twain and the Quintus Curtius Snodgrass Letters: A Statistical Test of Authorship. Taylor&Francis,Ltd.
- [12]Binongo, J.N.G. (2003). CHANCE.16(2). pp.9～17. Who Wrote the 15th Book of Oz? An Application of Multivariate Analysis to Authorship Attribution.
- [13]Kestemont,M.・Moens,S.・Deploige,J.(2015). Digital Scholarship in the Humanities.30(2). pp.199～224. Collaborative authorship in the twelfth century: A stylometric study of Hildegard of Bingen and Guibert of Gembloux. Oxford University Press.
- [14]Grieve,J.(2007).Literary and Linguistic Computing.22(3). pp.251～270. Quantitative Authorship Attribution: An Evaluation of Techniques.
- [15]大野晋.(1956).『国文学』24.「基本語彙に関する二三の研究—日本の古典文学作品に於ける—」. pp.34～46.日本語学会.
- [16]水谷 静夫.(1965).『計量国語学』35.「大野の語彙法則について」. pp.1～13.
- [17]樺島忠夫,寿岳章子.(1965).『文体の科学』. 綜芸舎
- [18]安本美典.(1957).『文学・語学』.「宇治十帖の作者—文章心理学による作者推定—」. pp.27～33. 三省堂書店.
- [19]蕨沢 正.(1965).『計量国語学』33.「由良物語の作者の統計的判別」. pp.21～28. 計量国語学会.
- [20]伊藤瑞穂・村上征勝.(1992).『大崎学報』148. pp.1～52.「三大秘法稟承事の計量文献学的新研究」
- [21]金 明哲・樺島 忠夫・村上 征勝.(1993).『計量国語学』18(8). pp.382～391.「読点と書き手の個性」.
- [22]村上征勝・今西祐一郎.(1999).『情報処理学会論文誌』40(3). pp.774～782.「源氏物語の助動詞の計量分析」一般社団法人情報処理学会.
- [23]金明哲.(2009).行動計量学 36(2).p89～103.「文章の執筆時期の推定 - 芥川龍之介の作品を例として - 」.
- [24]村上征勝・岸野洋久・伊藤瑞穂.(1990).『文献情報のデータベースとその利用に関する研究会』. pp.311.「日蓮遺文の計量分析—思想変化と文体の変化—」.
- [25]村上征勝・岸野洋久.(1990).『文献情報のデータベースとその利用に関する研究会』.pp.309.「品詞の使用率からみた和文体・漢文体の特徴」.
- [26]村上征勝. (2004).『シェークスピアは誰ですか?—計量文献

- 学の世界ー』. 文春新書.
- [27]“青空文庫”. <http://www.aozora.gr.jp/> (参照 2016-10-2).
- [28]“Web 茶まめ”. <http://chamame.ninjal.ac.jp/> (参照 2016-10-2).
- [29] “NLPIR.”<http://ictclas.nlpir.org/> (参照 2016-10-4).
- [30] “Stanford Log-linear Part-Of-Speech Tagger”.
<http://nlp.stanford.edu/software/tagger.shtml> (参照
2016-10-4).
- [31]“MTmineR について”. <http://mj.in.doshisha.ac.jp/MTmineR/>
(参照 2016-10-2).
- [32]“多言語テキスト統計分析ツール MTmineR”.
<http://www.cis.doshisha.ac.jp/mjin/MTmineR/manual.pdf>, (参
照 2016-10-2).
- [33]金 明哲・張 信鵬.(2013). 『日本行動計量学会大会発表論文
抄録集』41. pp.360～363. 「テキストマイニングツール MTmineR
のコンセプトと機能」.
- [34] “KH Coder”. <http://khc.sourceforge.net/> (参照 2016-10-2).
- [35] “CasualConc”. <https://sites.google.com/site/casualconcej/> (参
照 2016-10-2).
- [36]今尾康裕.(2011). 『外国語教育メディア学会関西支部 メソド
ロジー研究部会 2011 年報告論集』. pp.121～178. 「Mac OS X
用コンコーダンサーCasualConc－基本的な使い方と用例検索
ツールとしての応用例ー」.
- [37]財津 亘・金明哲 (2015) . 『日本法科学技術学会誌』 20(1).
pp.1～14. 「テキストマイニングを用いた犯罪に関わる文書の
筆者識別法」.
- [38]神田宏.(2013). 『近畿大学法学』 61(2・3). pp.1～36. 「テキス
ト・マイニングの手法を用いた量刑理由の記述統計分析ー裁判
員裁判制度導入前後の死刑・無期懲役判例を素材にー」.
- [39]村上征勝・金明哲・土山玄・上阪彩香.(2016). 『計量文献学の
射程』. 勉誠出版株式会社.
- [40]上阪彩香.(2016). 「西鶴浮世草子の文章に関する数量的研究ー
遺稿集を中心とした著者の検討ー」. 博士論文. 同志社大学.
- [41]上阪彩香.(2016). 「日本語テキストの計量分析の現状と課題」.
『人文情報学月報』 No.62.巻頭言.