

戦後日本の行政学研究の定量的検討：『年報行政研究』（1962～2018 年） 掲載論文題目のテキストマイニング手法に基づく統計的内容分析

茂 木 康 俊

はじめに

人文・社会科学系、理科系、医学系を問わず、テキストマイニング手法を用いたテキストデータの統計的研究が増えてきている。国内の論文データベースである国立情報学研究所（National Institute of Informatics: NII）の CiNii Articles で「テキストマイニング」をキーワードに検索すると公表時期が 2018 年の文献は 250 件であり、10 年前の 2008 年が公表時期であるものは 135 件であり、2 倍近くになっている⁽¹⁾。2019 年度日本政治学会総会・研究会（2019 年 10 月 5 日～6 日開催）で「計量テキスト分析」の分科会が設定されていることから窺えるように、後に先行研究の検討で具体的に説明するが、政治学や行政学においてもテキストマイニング手法を用いた研究が増えつつあり、官僚制の特徴として文書主義があるなど行政文書が行政運営にとって重要なことを考えると行政学においてもその利用が一層進む可能性がある。文章を扱うという意味では質的研究に位置づけられることもあるが、形態素解析の後、抽出語をカウントし多変量解析などの統計分析含め数的な処理を加えるという意味では統計的研究の一種であると言える。

テキストデータの分析手法は、以前は専ら内容分析と呼ばれ、現在よりも手動的に処理がなされていた⁽²⁾。各種のソフトウェアが有償・無償を問わず増えてきた現在と比べ、分析のための手間がかかり、かつテキストデータの分析方法や可視化（Data Visualization）の方法が限られていたため、それほど広く使われてはいなかった。最近のテキストマイニング手法の興隆は対応す

るソフトウェアの増加や可視化を含む分析機能の向上によるものが大きいと言える。

本稿は、我が国の戦後の行政学の研究動向を定量的に把握するために、日本行政学会の学会誌である『年報行政研究』について、1962年（昭和37年）発行の第1号から2018年（平成30年）発行の第54号までの戦後の昭和および平成年代の57年分の掲載論文の題目を集計し統計的に分析することで、我が国の行政学の60年近くの研究関心の変化を明らかにしようとするものである。後述するように、特定の行政学者によって、定性的に我が国の行政学会（行政学界）の研究動向を検討し論述している研究は少なくない。学識や経験に基づき研究動向の変化を明らかにし、他の研究者に示唆を与えていることは十分にあり、定期的に学会発表や依頼論文の形で戦後の日本行政学の研究動向については特定の学者による論考が発表されている。しかし、そのような論考は一定の示唆を与えうるものの、研究者個人の学識や経験に左右されるものであり、(1) 研究者個人の不得意な領域や研究に十分に取り組んでいない領域に関して言及が手薄になってしまう、(2) 研究者の行政学観や行政学理論に基づいて特定の立場からの研究に着目しすぎたり、逆に特定の立場からの研究を取り上げることに消極的になったりする可能性がある。このような着目対象や取り上げる文献の数に関する偏り（バイアス）は行政学の研究動向の客観的な把握という意味で信頼性や再現性に問題を生じさせるのである。定性的な経験に基づく学会動向の検討であれば、複数の研究を合わせて、いわばメタ分析のように統合的に検討することも不可能である。そのような意味で、検証可能性や複数の研究者間での対話可能性が保証されないのである。研究動向について定量的に把握しようという試みはアメリカ行政学においてはデルファイ法を用いたWald（1973）がある（今里，2000，p. 284）。我が国の行政学会に関連する主要な学術専門誌である『年報行政研究』および『季刊行政管理研究』を見る限り、そのような定量的な手法で行政学会の研究動向を把握しようとする試みは見当たらなかった。

本稿では近年自然科学のみならず社会科学において幅広く用いられるようになったテキストマイニングの手法を用いる。テキストマイニングとは、「文章の中から単語を切り出し、各単語の頻度、単語同士のつながりを数値化することにより統計的な分析を行う」(岩崎・中西・時岡, 2004, p. 184) 方法である。最近では国立の滋賀大学に2017年4月に日本で初めてデータサイエンス学部ができ、続いて広島大学で2018年4月に情報科学部が設置された。ビッグ・データとも関わるデータサイエンス、データマイニングはテキストマイニングとも関連がある。東京大学大学院情報理工学系研究科教授の山西健司はテキストマイニングについて、「自然言語処理とデータマイニング技法を結合した技術である。テキストマイニングの要素技術には、テキストクラスタリング、テキスト分類、相関性解析、情報抽出などがあ」(p. 213) と述べる。テキストマイニング手法はこのように、質的情報の定量的分析の方法であり、再検証の方法が明確になるよう本稿で行う学会誌の論文題目のデータ化については詳述している。後述するようにデータのテキスト化や処理の方法、分析方法を詳しく述べ、他の研究者に対し明らかにしているため、他の研究者が同様の手順を取って検証することが可能であるといえる。また、筆者あるいは別の研究者によって、後年同様の手順で例えば令和の年代のデータを付け加えて、三区分の時代区分で再検証することも可能である。本稿の学会誌の論文題目に関するテキストマイニングによる分析では、日本行政学会の研究動向に関して定量的に分析し、抽出語リストの確認やその後の統計分析で、特徴や規則性・法則性が定量的・視覚的に明らかにされることになる。行政学者の定性的な学会動向の検討結果と合わせ、我が国の行政学界の研究動向の特徴、その変化について多様な視点から把握することが可能になるものと考えられる。以上、テキストマイニングに基づく学会研究動向の利点をまとめると、(1) 再検証が可能であること、(2) 研究者の知識の傾向や関心に基づく選択バイアスがある程度押さえることができること、(3) 研究動向の定量的・視覚的な把握ができることを挙げられる。

1 先行研究の検討、我が国の行政学の研究動向に関する研究

先行研究の検討として本稿の用いる方法と内容との関連で、(1) 我が国の行政学における研究動向についての定性的な検討、(2) 隣接分野含めたテキストマイニング手法を用いた研究、(3) 特定学術分野の主要誌の掲載タイトル・内容を分析し、特定学術分野や学界の研究動向を検討する研究の順に取り扱う。分析方法の面では、(3) が先行研究として密接に関連があるものである。

(1) 我が国の行政学における研究動向については、定期的に学会誌である『年報行政研究』で回顧的観点からの依頼論文が掲載されている。また、我が国においては法学部を中心に行政学の科目が設置されているが、学部レベルで使用される教科書においても行政学の戦後の研究動向について言及される場合が多い。前者としては、代表的なものとして今里 (2011)、西尾 (2001b)、真淵 (2001)、村松 (2001) などがある⁽³⁾。今里 (2011) は著者の経験に基づく定性的な分析のみならず、日本行政学会の「行政学教育の現状と課題に関するアンケート調査委員会」実施のアンケート調査 (2009 年実施調査) に対応した研究動向の分析である⁽⁴⁾。後者として、行政学分野でよく使用される教科書である西尾 (2001) では、我が国の行政学の三つの視点として「制度学の視点」「管理学の視点」「政策学の視点」を挙げたのはあまりにも有名である (pp. 50-51)。上述したように、これらの研究は一定の意義があるものの、着目対象や取り上げる文献の数に関する偏りが避けられず、行政学の研究動向の客観的な把握という意味では不十分な面がある。

隣接分野含めたテキストマイニング手法を用いた研究については、齋藤 (2011) が社会科学のみならず自然科学も含め、発表時点での我が国においてテキストマイニングの手法を用いて行われた研究を幅広くレビューしている⁽⁵⁾。経営学、工学、経済学、心理学、教育学、文学、法・政治学、その他学術の各分野でテキストマイニング手法を用いている研究を挙げ、分析対象

のテキストや分析の概要を紹介している。冒頭で経営学あるいは経営の実場面での応用事例が最も多く見られたと述べている (p. 2)。経営学や経営実践の中で、企業のカスタマーセンター、コールセンターにおける顧客とのやり取りに関するテキストマイニング手法による分析事例が以前から存在していたことを指摘する。行政学に関連する「法・政治学」の分野として挙げているのは、判例の分析や特許文書の分析、政策研究として答申書の研究事例である。金(2009)は第1章の「統計的テキストマイニング」の章で、「小史」としてこれまでのテキストマイニング手法の応用を1887年のオハイオ州立大学のメンデンホールの光学におけるスペクトル分析方法を単語に応用した研究事例を端緒として紹介している (pp. 3-4)。また、同書は数理的な解説とともに統計言語であるRの文法を紹介しながら、テキストのクラスター分析、時系列分析、アソシエーション分析を詳述している。

本稿の筆者が把握した2012年以降の行政学・政治学分野でのテキストマイニングを用いた研究としては、まず地方議会の会議録に対する政治学の視点からのテキストマイニング手法を用いた研究がある(増田, 2012)。主な抽出語の検討から、市政における中心的な課題を把握し、「子ども・教育分野」に議員の関心が集まっていることを確認している。山下(2012)は北海道山越郡長万部町の広報誌および長万部町まちづくり総合計画を分析対象のテキストデータとし、Microsoft Excelのアドインソフトであるトレンドサーチ2008(株式会社社会情報サービス)を用いたテキストマイニング分析を行っている。爲我井(2015)は埼玉県越谷市を事例にし特例市移行をテーマに、計3年間の市議会の審議データを用いたテキストマイニング分析である。加えて、行政学・地方自治分野では茂木(2013)および茂木(2014)がある。国際政治学では、朝日新聞と読売新聞の社説のデータを基に日本社会の国際政治意識に関してテキストデータを用いた分析を行なう藤田(2012)がある。同研究では、大分類としてのグループおよび小分類としてのカテゴリーでキーワード(抽出語)のコーディング(コード化)を行っている。

研究事例の多い経営学の分野では、喜田（2007）、喜田（2008）が企業組織の経営研究にテキストマイニングの手法を取り入れたまとまった研究実績であるが、喜田（2007）は有価証券報告書のテキストデータを用いてアサヒビールの組織革新について分析している。その後、喜田（2008）は経営学分野での実際のテキストマイニングに基づく研究事例を複数紹介しテキストマイニングの手法と応用方法について広く論じている⁽⁶⁾。最近では、喜田（2018）は近年興隆しているデータサイエンス、ビッグ・データ、データヴィジュアライゼーション、そしてモデリングの観点から研究例を俯瞰的に説明している。また、前島・大江・柴（2019）は企業の経営理念に含まれる単語を対象にクラスター分析を行っている。

(3) 特定学術分野の主要誌の掲載タイトル・内容を分析し、当該学術分野や学界の研究動向を検討する研究としては、行政学・地方自治分野としては、必ずしもテキストマイニングの基づくわけではないが同様の発想によるものとして新藤・松本編（2010）、新藤・松本編（2012）がある。本稿と同様にテキストマイニングを用いた、学会での研究動向の分析については、日本神経学会の『臨床神経学』のタイトルを分析対象とした佐久嶋・佐々木・田代（2012）がある⁽⁷⁾。論文の書誌データ（とりわけ題目）を用いて研究動向を分析する先駆的な研究であり、他分野の研究ではあるが、本稿の分析において事前に詳細に参照し、テキストデータの準備・処理、分析方法の選択などで方法論的に参考にしている。その後、学会誌の掲載論文のテキストデータを用いた研究は各分野で行われるようになった。

表1で、各文献の対象のテキストデータ、形態素解析の使用ソフトウェア、各種処理・分析の実施有無についてまとめ、比較した。抽出語間の関係を探る共起ネットワーク分析や対応分析については各研究で行っているかは分かれるが、抽出語の上位リストはすべての研究で行っている。本稿でも行っている対応分析は4研究中3研究が行っている。ここで取り上げた文献の分野は神経学、保育学、保健学、宇宙科学とさまざまであるが、形態素解析の後に、

抽出語を検討し、共起ネットワーク・対応分析などのデータ分類や可視化の手法を用いて、それぞれ学会の研究動向を定量的に把握しているのである。

表1 テキストマイニング手法を用いた学会誌研究動向の分析（他分野）の構成要素

	対象のテキストデータ	形態素解析の使用ソフトウェア	強制抽出語の指定	抽出語の上位リストの検討	抽出語のコーディング	共起ネットワーク	クラスター分析	対応分析
佐久嶋・佐々木・田代 (2012)	論文題目	MeCab	○有り	○有り	○有り(疾患カテゴリー)	×無し	×無し	○有り
太田 (2015)	論文題目	不明 (KH Coder)	×無し	○有り	×無し	○有り	×無し	×無し
岩佐 (2019)	論文題目	Chasen	○(疾患名や専門用語)	△(論文内に表示せず)	○有り	×無し	×無し	○有り
水上・本田・中野 (2019)	論文要旨	不明 (KH Coder)	×無し	○有り	△グループ(コミュニティ)化は有り	×無し	×無し	○有り

出所：各文献の内容を筆者が確認の上、作表した。

本稿では、テキストマイニングの手法で品詞ごとの頻度に基づき語を抽出し、『年報行政研究』掲載論文の題目において、特徴的な語句を抽出する。次に、それらの語句同士の関係を把握するために共起ネットワーク分析、クラスター分析、対応分析などの手法を用いて検討を行う。それぞれの分析では、10年区切りの西暦の年代や昭和・平成という和暦を用いたほぼ均等の二区分の年代区分を外変数として使い、特徴的な抽出語がどのように変化しているか分析を行う。このように最近社会科学分野においてもしばしば用いられるようになったテキストマイニングの手法を日本行政学会の研究動向の変化

を定量的に把握するために活用する。研究動向の経年的な変化を抽出語に着目して分析する点やクラスター分析や対応分析を用いて抽出語の出現傾向に基づいてグループ分けを行い、学会としての研究全体の動向の変化を定量的に分析および分類する点が本稿の大きな特徴である。

2 学会誌論文題目のテキスト分析—複合語の処理、抽出語の分析

2.1 学会誌題目テキストデータの作成、データの概要

本章では上述の先行研究の検討を踏まえて、他分野の佐久嶋・佐々木・田代（2012）を方法論的に参考にし、日本行政学会誌『年報行政研究』の公表論文題目をデータとしてテキストマイニングの手法を用いて統計的な分析を行う。具体的には、国立情報学研究所の国内論文データベースである CiNii Articles でキーワードを「年報行政研究」とし、「200件ずつ表示」で「TSVで表示」を「実行」し、ウェブブラウザ上に表示された結果を TSV ファイルで保存した（2019年7月14日実施）⁽⁸⁾。この複数の TSV ファイルを Microsoft Excel に取り込み、単一のファイルとしたものをベースにしている。

この基本のファイルに対して、テキストマイニングが実行可能となるように以下のような修正を手動で行っている。まず、英語論文で、号数が発行年に入っているものについて修正を行った。次に、英語タイトルが日本語タイトルと重複して収録されているものについて、英語の分を削除した。さらに、日本語論文であるのに英語名で収録されているものについては、日本行政学会のウェブサイトですべて正確な書誌データを参考にし、日本語に変更した。

加えて、全体的に Excel のフィルタ機能を用いて、特定のデータ（ケース）を抜き出し、分析対象から除外した。①論文名に「書評」を含むものを除外、②著者名が空欄であるものを除外、③論文タイトルが「フルネーム著『…』」の形式のものは書評であるのでケースを削除、④「…」の形式であるものの、明らかに書評であるものを削除など論文でないものを分析対象から除外し、当初の 887 ケースから上述のフィルタリング処理の結果 458 論文がデー

タセットに残った。

上述のテキストデータを対象に、本稿ではテキストマイニングのための日本語処理・分析ソフトウェアとして立命館大学産業社会学部准教授の樋口耕一氏が文部科学省の科学研究費補助金を受けて作成した KH Coder を用いている⁽⁹⁾。KH Coder はインターフェイスを提供するいわばフロントエンドソフトウェアであり、文章を品詞の要素ごとに分解する形態素解析については Chasen (茶筌) と MeCab が選択可能であり、本稿ではデフォルトの Chasen を利用している。Chasen は国立奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座 (松本研究室) が著作権を有するソフトウェアである。一方、MeCab は京都大学情報学研究科 - 日本電信電話株式会社コミュニケーション科学基礎研究所 共同研究ユニットプロジェクトを通じて開発されているものである。KH Coder は形態素解析以外の処理や作図について統計言語の R を利用している⁽¹⁰⁾。

まずは Chasen を用いた形態素解析によって作成されたテキストデータ (ケース数 458) の概要について検討する。形態素解析の後に、Microsoft Excel の LEN 関数を用いて論文題目の文字数を数値化した⁽¹¹⁾。西暦の年代区分ごとの論文数と論文題目の文字数の推移は表 2 の通りである。

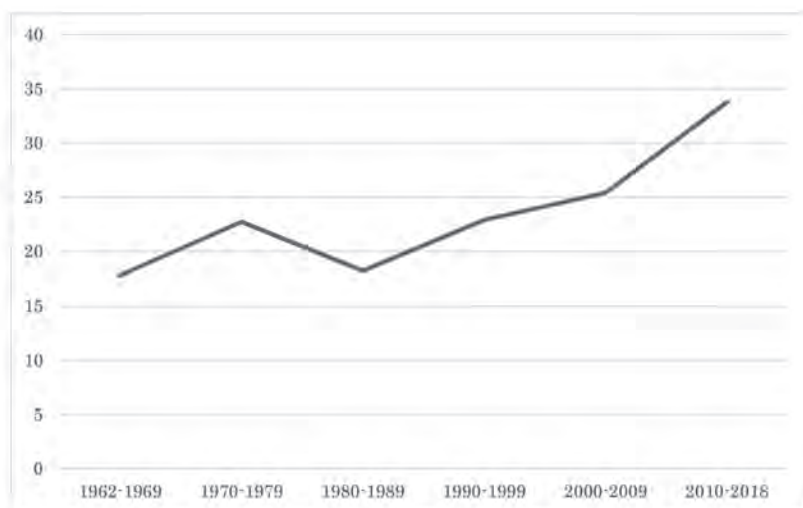
表 2 『年報行政研究』年代区分ごとの論文数、論文題目の文字数 (平均)

	論文数	論文題目の 文字数 (平均)
1962-1969	50	17.8
1970-1979	66	22.7
1980-1989	65	18.2
1990-1999	113	22.9
2000-2009	98	25.4
2010-2018	66	33.8
全体	458	23.7

出所：筆者作成。

1960年代と2010年代は対象年数が10年間でないため、論文数については比較できないが、1990年代で論文掲載数が1980年代の2倍弱の水準で増加している。2000年代は再び減少している。論文題目の文字数（平均）については、1960年代が最小で、多少の増減があるが、2010年代で最大となっている。図1で、この論文題目の文字数の推移をグラフ化した。

図1 『年報行政研究』論文題目の文字数（平均）



出所：筆者作成。

2.2 テキストデータの処理、複合語の処理

「強制抽出する語」（強制抽出語）の指定で複合語の処理を行う場合は一般的には、KH Coder に内蔵されているプログラム TermExtract を用いて、リストアップされた複合語候補を目視の上、リストから分析目的との関係で不適切な複合語候補を手作業で削除し、残ったものを強制抽出語に指定する⁽¹²⁾。この方法をとると、分析者の目が入るため複合語の取りこぼしを避けることができる反面、作業時間が長くなることと、分析者の知識の限界や恣意性が

入り込む可能性がある。また、ここでの作業内容が明確にならないために再現研究が行いにくくなってしまふ。本稿では再検証の可能性を最大限に高めるために、TermExtractと目視による強制抽出語指定に代え、我が国行政学でよく使用される教科書である西尾(2001)の索引のすべての語について手入力の上、次に示す微修正の上強制抽出語の対象リストを作成した。(1)人名のイニシャルは削除、(2)(・・・)が言い換えの場合は、言い換えて別語として入力、(3)―で省略されたものは、完全な表記に戻す、(4)(アメリカ)など国の表記は削除、(4)本文と一致していない索引語を修正(自己決定・自己責任の原則)、(5)「・」で区切っている複数語の列挙(例、勅任官・奏任官・判任官)については、それぞれ個別に入力、という微修正を手動で行った。その結果、西尾(2001)の索引に基づく強制抽出語は815語となった。

最後に、以上の強制抽出語の設定が適切かどうか確認的分析やデータセットのチェックを行った。その結果、最近の行政学において重要な概念である「ガバナンス」の語が適切に抽出されていないことが判明し、手動で「ガバナンス」の語を複合語(強制抽出語)のリストに追加した。

2.3 テキストマイニングに基づく抽出語の分析

形態素解析を基に表3で各品詞の抽出語リストを作成した⁽¹³⁾。文書数順というのは、ある語句が出現したかどうかを文書(論文)ごとにカウントしたものである。それぞれのリストは上位10位までを挙げている。名詞では「政治」や「制度」がそれぞれ文書数17と11で3位と4位で政治学としての日本での行政学の扱われ方が確認できる。サ変名詞では1位が「改革」で文書数25である。5位に「評価」があるが、これは通時的なものであるとは考えられない。この点については後の分析で確認する。複合語(強制抽出語)については、課題が3位で行政改革が4位である(それぞれの文書数は24と18である)。また、5位に「自治体」6位に「地方自治」がある。このことは伝統的に地方自治分野に注力してきた日本の行政学の特徴を表すものであると考えられる。

表3 抽出語リスト（名詞、サ変名詞、複合語）文書数順（『年報行政研究』論文題目）

	名詞	文書数	サ変名詞	文書数	複合語 (強制抽出)	文書数
1	中心	26	改革	25	行政	85
2	事例	18	研究	13	行政学	44
3	政治	17	変化	12	課題	24
4	制度	11	展開	8	行政改革	18
5	現状	10	形成	7	考察	14
	責任	10	組織	7	自治体	14
			評価	7		
6						
7	政策	9			地方自治	12
	役割	9				
8			展望	6	管理	11
					変容	11
					問題点	11
9	理論	8	構想	5		
			裁量	5		
			分析	5		
10	過程	7				
	概念	7				
	分権	7				

出所：筆者作成。

3 抽出語同士の関係分析——共起ネットワーク分析、クラスター分析、対応分析

抽出語について、各抽出語間の関係を分析するために本章では視覚的な分析手法である共起ネットワーク分析、クラスター分析、対応分析の各手法を用いる。特に西暦の10年ごとの区分で集計することで、各抽出語の出現頻度の推移や時代的特徴をここで把握する。それぞれの手法の着目する点や実際の分析で使用する外部変数が異なっても、同一のテキストデータを用いているためデータの顕著な特徴は共通して現れることになる。

注：中心性（媒介）を表示しており、昭和年代と平成年代の二区分を外部変数としている。出現数による語の取捨選択において、最小出現数を5に設定している。

出所：筆者作成。

この図を見ると、(1) 昭和年代と平成年代に共通する特徴語として「行政改革」「日本」「課題」といった語が見られる。(2) 昭和に特徴的な語として、「地方自治」「公務員制度」「住民参加」「管理」「公共性」「行政計画」「意思決定」「行政組織」「予算」「概念」「学会」などの語が見られる。これらの語は後の平成年代には、特徴的な語として見られなかったものである。(3) 平成年代に特徴的な語として、頻度の大きいものとしては「改革」「政治」「事例」「政策評価」「ガバナンス」「自治体」「評価」「変化」などが見られる。これらは図4の対応分析で右半分に位置づけられていた語であり、このように『年報行政研究』に掲載された論文題目の抽出語の変化からこれらの語に関連する研究テーマに研究動向がシフトしていることが窺われる。

3.2 クラスター分析に基づく論文の分類

クラスター分析の手法を用いて、抽出語の出現状況の特徴ごとに抽出語の分類を行う。ジャックカード係数を分類の基準とし、クラスター分析の指定のクラスター数を4として行った¹⁵⁾。クラスター数を4にしたのは、西尾(2001)の「制度学の視点」「管理学の視点」「政策学の視点」の3つの視点を念頭に、これ以外の視点を1つ想定してクラスター数4としたものである。クラスター数の設定次第で結果が変わる可能性があることには留意しなければならない。

表 4 クラスタ 1 の特徴語

	抽出語名	ジャックカード係数
1	行政学	0.7917
2	行政学教育	0.1190
3	日本	0.0943
4	行政法学	0.0930
5	責任	0.0833
6	課題	0.0820
7	制度化	0.0682
8	行政	0.0672
9	展望	0.0667
10	アイデンティティ	0.0476
11	時代状況	0.0476
12	対話	0.0476

出所：筆者作成。

表 5 クラスタ 2 の特徴語

	抽出語名	ジャックカード係数
1	行政	0.5567
2	政治	0.1857
3	役割	0.1364
4	員	0.0909
5	政策評価	0.0909
6	職	0.0870
7	責任	0.0857
8	制度	0.0845
9	裁量	0.0758
10	首長	0.0746

出所：筆者作成。

表6 クラスタ-3の特徴語

	抽出語名	ジャックカード係数
1	改革	0.6296
2	地方分権	0.3810
3	大都市制度	0.1500
4	構想	0.1429
5	公	0.1053
6	システム	0.0870

注：ジャックカード係数が0.0526で39個の抽出語が並ぶため、7位以下のこの表への掲載は見合わせた。

出所：筆者作成。

表7 クラスタ-4の特徴語

	抽出語名	ジャックカード係数
1	中心	0.0941
2	行政改革	0.0669
3	課題	0.0613
4	問題	0.0551
5	事例	0.0545
6	研究	0.0514
7	考察	0.0512
8	自治体	0.0512
9	管理	0.0435
10	問題点	0.0435
11	地方自治	0.0433

出所：筆者作成。

クラスタ分析やそもそものテキストマイニング分析は探索的な手法であるため、これ以降の分析は他の解釈の余地も残るものではあるが、これまでに述べた方法でデータや集計自体は再現や検証することが可能である。クラ

スター1は、今里(2011)も参考にして、「行政学のアイデンティティ」のクラスターと名付けたい。これは学としてのアイデンティティや教育に関わるクラスター(分類)である。クラスター2は、上位の「行政」「政治」「政策評価」「制度」の語から、「管理・制度・政策」のクラスター(分類)と捉えたい。クラスター3は、ジャカード係数が2位以下と比べ大きく離れている1位の「改革」および2位・3位を参考に「改革・地方自治」のクラスターとしたい。クラスター4はさまざまな抽出語が上位に位置付けられており、それぞれのジャカード係数は低い。3位までの抽出語に着目し「行政改革の課題」のクラスターと名付けることにする。ここまでをまとめると、クラスター1「行政学のアイデンティティ」、クラスター2「管理・制度・政策」、クラスター3「改革・地方自治」、クラスター4「行政改革の課題」と捉えることにした。これらの4クラスターの特徴語として表に示した語句を、それぞれコーディング(コード化)し、そのルールに基づいて、まず年代ごとに集計した(表8)。

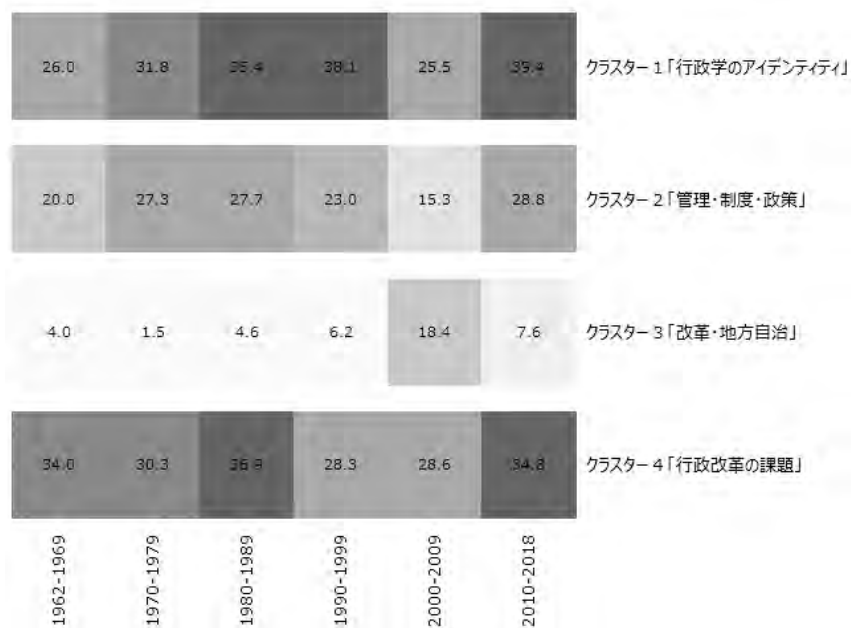
表8 年代ごとの研究分野の推移(クラスター分析)

	*クラスター1 「行政学のアイ デンティティ」	*クラスター2 「管理・制度・ 政策」	*クラスター3 「改革・地方自 治」	*クラスター4 「行政改革の 課題」	合計論文数
1962-1969	13 (26.00%)	10 (20.00%)	2 (4.00%)	17 (34.00%)	50 (100.0%)
1970-1979	21 (31.82%)	18 (27.27%)	1 (1.52%)	20 (30.30%)	66 (100.0%)
1980-1989	23 (35.38%)	18 (27.69%)	3 (4.62%)	24 (36.92%)	65 (100.0%)
1990-1999	43 (38.05%)	26 (23.01%)	7 (6.19%)	32 (28.32%)	113 (100.0%)
2000-2009	25 (25.51%)	15 (15.31%)	18 (18.37%)	28 (28.57%)	98 (100.0%)
2010-2018	26 (39.39%)	19 (28.79%)	5 (7.58%)	23 (34.85%)	66 (100.0%)
合計	151 (32.97%)	106 (23.14%)	36 (7.86%)	144 (31.44%)	458 (100.0%)

出所：筆者作成。

年代と研究テーマのクラスターの区分の二変数の関係を検討するために、カイ二乗検定を行ったところ、カイ二乗値が28.12、自由度が15で、5%水準で有意である（ただしクラスター3の列で期待度5未満のセルが2つある。）。つまり、年代とクラスターの区分との間で関連性があるということである。なお二変数間の関連の強さを示すクラメル係数は0.146であり、二変数間の関係はそれほど強くない。この西暦の年代ごとの区分での集計の数値をヒートマップとして作図した（図3）。ヒートマップとは集計表の可視化の手法であり、度数の多いセルが濃い色で描かれるものである。なお図示の向きについて、集計表とは異なり独立変数としての年代が列に移っている。

図3 年代ごとの研究分野の推移（ヒートマップ）



出所：筆者作成。

図を見ても各クラスターの年代ごとの分布の違いが分かる。クラスター1「行政学のアイデンティティ」は1980年代および1990年代に出現頻度が高く、再び2010年代に頻度が上がっている。クラスター3「改革・地方自治」は2000年代に頻度が上がっている。図4の対応分析で2000年代の改革や評価に関する抽出語が下部にまとまって描かれていることと同じ傾向が示されていると思われる。

3.3 年代に着目した対応分析

最後に対応分析 (correspondence analysis) を行う。対応分析は日本語でレスポネンデンス分析とも呼ばれ、「対応分析は多変量の (もっとも多くの事例では2変量の) カテゴリーデータをグラフィカルに表示するための典型的な技法として知られる。対応分析では、行と列それぞれの変数のカテゴリーを表す座標を導出し、これはさらに変数間の相関のパターンをグラフィカルに表示するためにプロットされることが多い。」と説明される (Everitt, 2006, p. 104. 邦訳、石田・石田・掛井訳, 2007, p. 113. 訳文は邦訳書のまま引用している。)。分析に用いる外部変数として、10年ごとの西暦の年代区分を用い、データの対象期間を5分割した。描画における軸となる成分は、外部変数の項目数から1を除いた第5成分まで抽出しており、それぞれの成分は相互に独立となっている。対応分析を行った結果が図4であり、布置図の中で近くに位置づけられている抽出語は似たような性格を持っている⁽¹⁶⁾。成分1と成分2の0が交差するところが原点であり、多くの抽出語が原点の近くに位置づけられている。この原点近くに位置づけられている語は大きな特徴を持たない語である (「行政」など)。

成分1は左右の軸であるが、西暦年代の変数が右に行くほど基本的に最近になるように位置づけられており (1970年代を除く)、時間に関する軸であると捉えられる。上下の軸の成分2は、抽出語の内容に関わるものと考えられる。2000年代の「改革」「評価」「政策評価」が下の方に位置づけられ、左

の上部に「住民参加」「公共性」「行政計画」が位置づけられているように、下部に行くほど評価や改革に関連し、上部に行くほど関連しなくなる、「評価・改革」に関する軸であると捉えたい。この2軸の第2成分までの累積寄与率は57.06%である。

図の周辺部分を見ると、右上で「東日本大震災」「政府中枢」、左上に「公共性」「住民参加」「行政計画」、右下に「ガバナンス」「評価」といった言葉が見られる。これらの語句は原点近くに位置づけられる語句とは異なる孤立した特徴的な性質を持つ語句であり、外部変数として設定した10年ごとの西暦の年代区分ごとの特徴をある程度表現している。すなわち(1)1970年代の論文の特徴として、「住民参加」「行政計画」「公共性」「政策決定」「意思決定」が挙げられるが、それ以降の年代にはこれらのテーマはあまり扱われていないのである。(2)2000年代の特徴的な語として「改革」「政策評価」「ガバナンス」が図の下部から見て取れる。(3)2010年以降としては、「東日本大震災」「首長」「沖縄」「国際比較」といった抽出語が右上に位置づけられている。東日本大震災の影響と考えられるが、2000年代から2010年代で特徴的な抽出語が示す研究の傾向が大きく変わったことが確認できる。対応分析だけで判断してはならないが、年代ごとの上下の推移を見ると、原点に近い1960年代から上部の「公共性」などに象徴される非「評価・改革」方向に1970年代に推移し、その後2000年代までは一貫して「評価・改革」方向の下部に移動している。一転して、2010年代は1970年代に匹敵するほど上部の非「評価・改革」の研究傾向を示していると言える。

注：描画設定の出現数による語の取捨選択において、最小出現数は 5 と設定している。描画の基準となるスコア（得点）の算出方法は「標準化」である。
出所：筆者作成。

4 結論——本稿の分析結果と今後の課題

本稿の調査・分析で明らかになった点をここで確認したい。抽出語に基づく統計分析の前に、テキストデータの概要を確認した。右肩上がりの論文題目の文字数（平均）を折れ線グラフでも確認した。抽出語についての分析としては、まず①抽出リストの作成によって、名詞では「政治」（文書数 17）や「制度」（文書数 11）でそれぞれ 3 位と 4 位で政治学と位置づけられる日本での行政学の特徴が確認できる。サ変名詞では 1 位が「改革」（文書数 25）であり、5 位に「評価」が見られる。複合語（強制抽出語）については、課題が 3 位（文書数 24）で行政改革が 4 位（文書数 18）である。行政改革や改革、評価が研究の中心的な課題であり、5 位に「自治体」6 位に「地方自治」があるように、我が国の行政学が地方自治分野に注力してきたことが確認される。

抽出語同士の関係を分析するため共起ネットワーク分析、クラスター分析、対応分析を行った。それぞれで確認された特徴、例えば 2000 年代を中心に改革や評価・政策評価に関する研究が多くなり、それが他の年代と異なった特徴を示していたこと等は共通して確認できた。

①抽出語の出現状況の特徴ごとに抽出語を分類するためクラスター分析を行った。ジャカード係数を基準に、クラスター数を 4 と設定し行ったところ、それぞれの特徴語がリスト化され、西尾（2001）や今里（2011）などの、学会の研究動向についての定性的な研究の提示する分類を参考に、クラスター 1 「行政学のアイデンティティ」、クラスター 2 「管理・制度・政策」、クラスター 3 「改革・地方自治」、クラスター 4 「行政改革の課題」と名付けた。この 4 クラスターの特徴語（表 4～7 に示したものを）をグループ化し、コー

ディング（コード化）し、その分類ルールに基づいて、西暦の年代ごとの区分との間で集計し、ヒートマップとして作図した（図3）。その結果、カイ二乗検定では年代とクラスターの区分との間で5%水準の有意な関連が示された。つまり、年代とクラスターの区分との間で関連性があるということである。集計表やヒートマップを見ると、クラスター1「行政学のアイデンティティ」は1980年代および1990年代に出現頻度が高く、再び2010年代に頻度が上がっている。クラスター3「改革・地方自治」は2000年代に頻度が上がっていたことなど、各抽出語の分類（コード）ごとの研究動向の重点の変化が明らかになり、2000年代の行政改革に関する研究の隆盛を視覚的に確認できた。

②10年ごとの西暦の年代区分を外部変数とした抽出語の対応分析では、各年代の特徴的な語が明らかになった。すなわち（1）1970年代の論文の特徴として、「住民参加」「行政計画」「公共性」「政策決定」「意思決定」が挙げられる。抽出語から見れば、それ以降の年代にはこれらのテーマは重視されていないといえる。（2）2000年代の特徴的な語として「改革」「政策評価」「ガバナンス」が見られる。このことは2000年代の橋本行革や政策評価法に基づく政策評価制度の導入や自治体における行政評価制度（三重県の事務事業評価制度を嚆矢とする政策評価制度）などの動きと関連があると思われる。（3）2010年以降としては、「東日本大震災」「首長」「沖縄」「国際比較」といった抽出語が位置づけられている。東日本大震災は2011年（平成23年）発生であり、時代区分で言えば2010年以降の行政学研究を大きく特徴付ける災害であったと言え、それらに関する研究は行政改革や評価・政策評価に関する2000年代の研究と異なった特徴を持っていることを確認し、対応分析での位置づけが大きく変わっていることから2000年代から2010年代で研究の傾向が大きく変化したことが明らかになった（2010年代の位置づけは1970年代と同様である）。

対応分析では第1成分と第2成分のみならず、第2成分と第3成分の描画

と検討も行った。その結果、各図から第1成分は時間、第2成分は「評価・改革」、第3成分は「政策プロセス」の軸であると解釈した。第1成分の時間を除く、第2成分および第3成分のテーマが日本の戦後行政学の研究の大きな柱であることが今回のデータからは推測される。クラスター分析の「行政学のアイデンティティ」「管理・制度・政策」および「改革・地方自治」と同様に、日本の戦後行政学の研究の中心が定量分析から抽出された。

本稿では、しばしば定性的に検討が行われる日本の行政学の研究動向の変化について、論文データベースである CiNii Articles のテキストデータを基にテキストマイニングの手法で検討を行い、戦後の我が国の行政学研究の特徴や変化について、一定の規則性や傾向を見いだすべく分析を行った。抽出語を基にした共起ネットワーク分析、クラスター分析、対応分析の各結果を基に上述の年代ごとの特徴と期間を通じた上述の研究の中心が明らかになった。

テキストマイニングは基本的に探索的な手法である。本稿では TermExtract 等のプログラムを用いた自動生成のリストを基に行う複合語の取捨選択を行っていない。つまり研究者（作業従事者）が分析視角との関係で意味がある語の組み合わせ（複合語）かどうかチェックする作業を避け、この段階での恣意性を排除しようとした。しかし、分析手法の選択や対応分析における各成分の解釈等に研究者自身の知識量による限界や判断上のバイアスが残る可能性がある⁽¹⁷⁾。また、本稿の分析の対象は論文（研究ノート含む）であり、特集論文（依頼論文）と投稿論文の両方を含んでいる。特集論文は学会全体の動向を反映している反面、会員歴の長い学者が執筆者の中心で、中堅・若手の研究動向を反映していない可能性ある。投稿論文のみに分析を絞れば、中堅・若手の問題意識と今後の学会の研究動向を推測することが可能になるかもしれない。しかし、投稿論文の著者は中堅・若手中心であり、研究テーマ・手法に関して学界全体の特徴と異なる可能性がある。その点で、本稿のようにまずは特集論文と投稿論文の両方を対象にしつつ、分析目的次第でどちら

かに分析対象を絞ることも想定される。

本稿の分析の後半で行った抽出語の年代ごとの出現の違いについては、傾向や規則性を把握し描写するに留まり、それをもたらす原因や要因までは十分には迫っていない。公務員数などで表される行政機能自体の変化や学会員数の推移や日本公共政策学会・地方自治関連学会などの近接学会の創設を含めた日本行政学会の学会自体の性格の変化など、学会における研究動向の変化に影響を与える各要因との関係については、今後の筆者の研究や定性的・定量的両面での他の研究者による検証に委ねられることになる⁽¹⁸⁾。筆者としては、今後これらの点について明らかにできるように、今回の分析結果を踏まえた定性・定量の両面での研究を進め、ある程度時間が経った後には昭和、平成、令和の3期間の比較での再分析・再検証も自ら行いたい。本稿が試みた行政学会の研究動向の定量的把握に関して、学界全体において検討や検証が進むことを期待している。

- (1) 2019年8月25日にCiNiiの検索を行った。出版年をそれぞれ2018年と2008年に設定した。このようにテキストマイニング手法を用いた研究が増えているのは後述するKH CoderやRなどのフリーソフトウェアの利用が広がっているためであろう。
- (2) 質問紙調査の自由記述を分析する目的の、テキストデータの分析としては最近では医学分野で2002年に行われた認知症の人と家族の会の会員2,700人を対象にした、郵送法による質問紙調査の、在宅環境配慮に関する具体的な経験について自由記述回答の分析を行った大島・見玉・後藤(2006)、炎症性腸疾患(IBD)患者に対し行った自記式質問紙調査の性的に幸福で満足できる状態に関する88名の自由回答についての分析を行った三木・前川・法橋(2018)など多数の研究例がある。自由回答について橋口(1999)は「社会調査における自由回答法は、探索的データの収集に適している。自由回答法は、回答選択肢法よりも回答者の真の意見を反映する可能性が高いと考えられるので、探索的研究では重要な研究方法の1つである。」と述べ、コーディング手続きのプログラム開発と定量化のための分析手法として数量化理論第Ⅲ類の解析例を示している。内容分析は新聞や宣伝(プロパガンダ)など多様な対象に対して行われており、Krippendorff(1980)はクロス表の作成、判別分析、クラスター分析の各手法を

既に提示しており、「コンピュータの利用」に基づく統計解析を単独の章で詳述している(邦訳書の 11 章 pp. 184-201. 原著の 11 章 pp. 119-128.)。

- (3) この他、佐藤 (2001) は日本行政学会の記念講演を基にした論考である。
- (4) 2009 年実施調査の詳細については原田 (2011) で論じられている。
- (5) 異なる分野におけるテキストマイニングの応用事例を紹介・検討する文献としては、村田・小木・高山・末吉・今村・測上 (2008) がある。「社説タイトルを利用した社会動向の把握」、公開特許文書を利用した技術傾向の分析、不動産アンケートを題材にした広告・営業の改善のための分析など 5 つの事例を挙げている。
- (6) テキストマイニングのためのソフトウェアである Clementine の操作法も紹介している。テキストマイニングの手法が広がる前には、内容分析として文章の言語的な分析は行われていた。喜田, 2008, p.30 は「内容分析は統計的手法のもとで仮説検証を行う方法であるといえる。」と述べている。最近は内容分析という言葉も用いられるが、コンピューターによる自然言語の解析を伴う分析を通常はテキストマイニングと呼ぶことが多い。
- (7) この他、宇宙科学分野の衛星プロジェクトの関連論文の要旨についてテキストマイニング手法に基づく分析を行って、学界の研究動向を分析しているものとして、水上・中村・家富・大島・中野 (2019) がある (これは公表論文ではなく、シンポジウムの報告概要資料である)。
- (8) TSV ファイルの TSV は Tab Separated Values の略であり、各データをタブ文字で区切って並べたファイル形式のことである。
- (9) KH Coder は 2014 年～2018 年に課題番号 26705006 で科学研究費補助金 (若手研究 (A)) を受けて開発されたものであり、科研費データベースによれば公費から合計 1,040 万円の助成を受けている (直接経費、間接経費の合計)。**KH Coder** の設計思想については、Higuchi (2016) の英語論文で説明されている。本稿で使用したバージョンは 3.Alpha.17c である。**KH Coder** は統計言語である R のいわばフロントエンドである。石田 (2008) は R を直接コマンドベースで操作しテキストマイニングを行う方法について詳述している。**KH Coder** の出力は直接 R を用いた場合と同様である。
- (10) R は統計学者であるオークランド大学統計学部の Ross Ihaka と Robert Gentleman によって開発されたものであり、「R は高水準言語であり、データ分析とグラフィックスの環境である。」(Crawley, 2012, p. vii) と位置づけられている。本稿の分析結果は、Chasen および R を直接用いて行った場合と同様である。
- (11) 単純な文字数ではなく、バイト数 (全角文字は 2 バイト) について LENB 関数で算出する方法もあるが、英文タイトルや算用数字の扱いが必ずしも統一されているとは限らないことを考慮し、文字数の LEN 関数を用いた。
- (12) TermExtract とは東京大学情報基盤センター中川研究室開発のソフトウェアである。

複合語とは例えば「独立行政法人」のように複数の語で構成される用語・概念であり、このような専門用語の複合語について「独立」「行政」「法人」のそれぞれでカウントし、集計すると概念の対応という意味で集計として不適切になってしまうため、複合語として取り扱う事前処理を行うのである。TermExtractを用いる場合は、通常はプログラムが算出したスコアを基準にした上位500位の複合語候補に関して、分析者が目を通し、不適切な複合語候補を手作業で取り除きリストを作成することになる。高知大学と北海道大学の研究者による宮崎・原田・草場(2019)も複合語の処理にTermExtractを用い、学習指導要領の解説理科編を対象にテキストマイニング手法で分析を行い、抽出語の上位リストの検討、共起ネットワーク分析を行っている。

- (13) 品詞の判定は機械的に行っており必ずしも文法的に正確でない場合があるが、その後の分析では品詞の違いを用いていないので、再現性を優先し人為的に修正することは行っていない。
- (14) 畑島・清野・井手(2019)は、教育学の分野でテキストマイニング手法を用いた研究であり、ICTを活用した「へき地教育」として遠隔交流を行い(長崎県対馬市)、長崎大学の学部生と研究対象の小学校の生徒との間のやりとりをビデオに撮影し、参加者の発語を文字に起こし、そのデータをテキストマイニング手法で分析し、抽出語について共起ネットワークで図示化している。
- (15) ジャッカード係数(Jaccard index)は0から1の間の値をとり、ジャッカード係数が大きいほど類似度(similarity)が高いことを示す。ジャッカード係数の数理的な定義については米ジョージア大学統計学科(Department of Statistics, University of Georgia)の教授らによるMcCormick, Lyons, & Hutcheson(1992)が数式を用いて詳述している。
- (16) 対応分析の布置図の解釈については、藤本(2017)は犯罪発生に関するモザイク図を作成した後、対称マップ(Symmetric)で対応分析を実際に行っている(p. 145)。対応分析の手法に関して世界的に著名なMichael Greenacre(M. グリーンエーカー)の2007年の著作を引用し、対応分析のグラフ表示に関する論争について説明している(p. 149)。
- (17) 本稿の複合語の処理では、よく使用される行政学の教科書の索引語を強制抽出語に指定する等で再現性の確保についてできる限りの対応を行っている。それに加え、テキストデータの作成、編集、分析手順を本稿で詳細かつ明確に述べているので、検証研究や追跡研究の可能性は十分に確保されている。
- (18) 本稿のテキストマイニングは論文題目を対象にしており、本文で取り扱われていても題目に表れていなければカウントされない。例えば、公共性や住民参加について2000年以降は論文題目の抽出語としてはほとんど出てこないが、本文の内容として言及されていることはあるだろう。そのような論文題目に表れない記述内容・テーマについては定性的な分析を併用せざるを得ないと思われる。

参考文献

- Crawley, M. J. (2012). *The R book*. John Wiley & Sons.
- Everitt, B. S. (2006). *An R and S-PLUS companion to multivariate analysis*. Springer Science & Business Media. (石田基広・石田和枝・掛井秀一訳、B. エヴェリット『R と S-PLUS による多変量解析』シュプリンガー・ジャパン、2007 年)
- Higuchi, K. (2016). A Two-step approach to quantitative content analysis: KH Coder tutorial using Anne of Green Gables (part I). *Ritsumeikan Social Science Review*, 52(3), 77-91.
- Krippendorff, K. (1980). *Content analysis; An introduction to its methodology*. Sage. (三上俊治・椎野英雄・橋元良明訳、K. クリッペンドルフ『メッセージ分析の技法：「内容分析」への招待』勁草書房、1989 年)
- McCormick, W. P., Lyons, N. I., & Hutcheson, K. (1992). Distributional properties of jaccard's index of similarity. *Communications in Statistics: Theory and Methods*, 21(1), 51-68.
- Wald, E. (1973). Toward a paradigm of future public administration. *Public Administration Review*, 33(4), 366-372.
- 石田基広 (2008) 『R によるテキストマイニング入門』森北出版。
- 李慧瑛・下高原理恵・緒方重光 (2018) 「テキストマイニングによる緩和ケア論文表題の可視化」『医療と社会』28 (2), 259-275。
- 今里滋 (1991) 「行政学教育に関するアンケート集計結果：行政学の「制度化」の現状」『年報行政研究』(26), 113-139。
- 今里滋 (2000) 『アメリカ行政の理論と実践』九州大学出版会。
- 今里滋 (2011) 「行政学と行政学教育：アイデンティティ、制度化、標準化」『年報行政研究』(46), 51-74。
- 岩崎学・中西寛子・時岡規夫 (2004) 『実用 統計用語事典』オーム社。
- 岩佐由美 (2019) 「女性の健康課題に関する研究パラダイムの変化：1980～2014 年 医学中央雑誌収録論文のタイトル分析から」『日本保健科学学会誌』21(4), 181-191。
- 大島千帆・児玉桂子・後藤隆 (2006) 「認知症高齢者の状態像に対応する在宅環境配慮評価軸の抽出：テキストマイニングを用いた家族介護者の自由記述回答の分析」『老年社会科学』28(3), 334-347。
- 太田雅代 (2015) 「わが国における障害児保育学研究の動向：テキストマイニングを用いた論文タイトルからの検討」『小田原短期大学研究紀要』(49), 173-178。
- 喜田昌樹 (2007) 『組織革新の認知的研究：認知変化・知識の可視化と組織科学へのテキストマイニングの導入』白桃書房。
- 喜田昌樹 (2008) 『テキストマイニング入門：経営研究での活用法』白桃書房。
- 喜田昌樹 (2018) 『新テキストマイニング入門：経営研究での「非構造化データ」の扱い方』白桃書房。

- 金明哲 (2009) 『テキストデータの統計科学入門』 岩波書店。
- 齋藤朗宏 (2011) 「日本におけるテキストマイニングの応用」『北九州市立大学経済学会 Working Paper Series』 2011-12 号。
- 佐久嶋研・佐々木秀直・田代邦雄 (2012) 「テキストマイニングを用いた学会誌論文タイトルの時系列分析：日本神経学会誌「臨床神経学」の分析」『医療情報学』 32(6), 315-321。
- 佐藤竺 (2001) 「戦後半世紀の行政学を回顧して」『年報行政研究』 (36), 9-18。
- 新藤宗幸・松本克夫編 (2010) 『雑誌『都市問題』にみる都市問題 1925-1945』 岩波書店。
- 新藤宗幸・松本克夫編 (2012) 『雑誌『都市問題』にみる都市問題 II 1950-1989』 岩波書店。
- 爲我井慎之介 (2015) 「計量テキスト分析による政策課題から見た特例市移行：埼玉県越谷市を事例として」『日本地域政策研究』 (14), 66-73。
- 西尾勝 (2001a) 『行政学 [新版]』 有斐閣。
- 西尾勝 (2001b) 「時代状況と日本の行政学の課題」『年報行政研究』 (36), 34-41。
- 橋口捷久 (1999) 「自由記述データの数量化理論第 III 類による解析例 (言語データの解析法の具体例について)」『日本教育心理学会総会発表論文集 [第 41 回総会]』 41, 18。
- 畑島英史・清野聡子・井手弘人 (2019) 「探究的な学習を充実させる遠隔交流の授業分析：テキストマイニングの手法を用いて」『長崎大学教育学部教育実践研究紀要』 18, 231-240。
- 原田久 (2011) 「「89 年調査」以降における行政学教育の変化」『年報行政研究』 (46), 93-106。
- 樋口耕一 (2014) 『社会調査のための計量テキスト分析：内容分析の継承と発展を目指して』 ナカニシヤ出版。
- 藤田泰昌 (2012) 「日本社会の国際政治意識の変容：テキストマイニングによる多面的検証」『政経研究』 (98), 107-121。
- 藤本一男 (2017) 「対応分析のグラフを適切に解釈する条件：Standard Coordinate, Principal Coordinate を理解する」『津田塾大学紀要』 49, 141-153。
- 前島誉・大江秋津・柴直樹 (2019) 「組織アイデンティティが早期離職の三大要因に与える影響：経営理念を用いたテキストマイニングによる実証研究」『経営情報学会誌』 27(4), 287-300。
- 増田正 (2012) 「地方議会の会議録に関するテキストマイニング分析：高崎市議会を事例として」『地域政策研究』 15 (1), 17-31。
- 真淵勝 (2001) 「行政研究：方法と課題」『年報行政研究』 (36), 145-154。
- 三木佳子・前川厚子・法橋尚宏 (2018) 「炎症性腸疾患患者の主観的セクシュアルウェルビーイングと属性別にみた特徴：テキストマイニングによる自由回答の分析」『日本看護科学会誌』 38, 46-55。
- 水上祐治・中村匡佑・家富紗妃・大島昭子・中野純司 (2019) 「テキストマイニング分析に

- よる研究動向把握の簡易的手法：衛星プロジェクトの関連論文の要旨分析』『平成 30 年度「宇宙科学情報解析シンポジウム」報告概要』会場：宇宙科学研究所。
- 水上祐治・本田啓介・中野純司 (2019) 「論文要旨のテキスト・マイニング分析によるホスピタリティ・マネジメント分野の研究傾向と論文誌別の採択傾向」『日本ホスピタリティ・マネジメント学会誌』(29), 59-66。
- 宮崎亮介・原田勇希・草場実 (2019) 「テキストマイニングによる現行及び次期学習指導要領の比較：共起ネットワークを用いた中学校学習指導要領解説理科編の可視化」『高知大学教育学部研究報告』79, 71-79。
- 村田真樹・小木しのぶ・高山泰博・末吉正成・今村誠・淵上美喜 (2008) 『事例で学ぶテキストマイニング』共立出版。
- 村松岐夫 (2001) 「日本行政学会の五〇年について」『年報行政研究』(36), 3-8。
- 茂木康俊 (2013) 「公共組織の目標の曖昧性に関する検討：テキストマイニングの手法を用いた独立行政法人の中期目標の分析を中心に」『季刊行政管理研究』(144), 31-48。
- 茂木康俊 (2014) 「地方自治体のミッションに関する定量的検討：自治基本条例前文のテキストマイニング分析を中心に」『地方自治研究：日本地方自治研究学会誌』29(2), 13-27。
- 山下良平 (2012) 「自治体が発信する情報の構造分析に対するテキストマイニングの可能性」『農村計画学会誌』31, 267-272。
- 山西健司 (2003) 「データとテキストのマイニング」甘利俊一・竹内啓・竹村彰通・伊庭幸人編『言語と心理の統計 [統計科学のフロンティア 10]』岩波書店。