

令和 6 年 6 月 11 日現在

機関番号：32689

研究種目：若手研究

研究期間：2019～2023

課題番号：19K13607

研究課題名（和文）全国紙の論調は極性化しているかー計量テキスト分析によるアプローチ

研究課題名（英文）Are National Newspapers Polarizing?-A Quantitative Text Analysis Approach

研究代表者

細貝 亮（Hosogai, Ryo）

早稲田大学・政治経済学術院・その他（招聘研究員）

研究者番号：30582259

交付決定額（研究期間全体）：（直接経費） 900,000円

研究成果の概要（和文）：本研究の目的は、近年進行していると指摘される全国新聞の極性化現象について、計量テキスト分析の手法を用いてアプローチするものであった。計量テキスト分析の一手法である教師あり機械学習を用いてテキストの極性化度を測定する「極性化指数」を提案し、全国紙4紙の記事に適用した。結果、4紙の論調は世間が共有している「リベラルな朝日・毎日新聞」対「保守的な読売・産経新聞」という対立図式沿ったものであることが判明したが、これらの差異は大きいものではなく、極性化が進展しているという積極的な証拠を見出すことはできなかった。

研究成果の学術的意義や社会的意義

政治的言論の多様性と相互の交流は民主主義において不可欠なものとされているが、近年、政治的な言論空間の「分極化」や「分断化」が指摘されることが多くなった。この文脈で、日本における言論の大きな担い手である全国新聞の論調もまた極性化していると指摘されることが多かったが、本研究では計量テキスト分析手法を用いて実証的にこの問題にアプローチした。本研究の知見は、全国新聞の極性化について実証的な基盤を与えると同時に、言論空間の「分断化」をめぐる近年の社会的な議論にも示唆をもたらすと考えられる。

研究成果の概要（英文）：This study is intended to approach polarization in national newspapers, which has been noted to be progressing in recent years, using a method of text analysis. Using supervised machine learning method, I proposed the Polarization Index, which measures the degree of text polarization, and applied it to articles in four national newspapers. The results showed that the tone of the four newspapers was in line with the conventional view of "liberal Asahi/Mainichi" versus "conservative Yomiuri/Sankei," but these differences were not large, and there was no evidence of polarization in the national newspaper text.

研究分野：政治学

キーワード：極性化 テキスト分析 内容分析 新聞

## 1. 研究開始当初の背景

近年、ジャーナリストや言論人を中心に、全国紙の論調の違いがこれまで以上に大きくなっていくとの指摘がなされるようになった。確かに「原発再稼働」「共謀罪(テロ等準備罪)法案」などの政策争点の報道にあたり、全国紙の論調の対立、具体的には「リベラルな朝日・毎日新聞」対「保守的な読売・産経新聞」という対立図式がクローズアップされる機会が多くなっている。背景には全国紙の部数減による論調の先鋭化や第二次安倍政権下における憲法改正論議のような国民意見を二分する争点の前面化などが指摘される。このような新聞論調の対立化ないし2極化を本研究では「極性化(polarization)」と呼ぶ。

全国新聞の論調の極性化は、それを読む読者の分断化にも影響を与える可能性があり、相互理解と合意が尊重される民主主義社会では必ずしも好ましい現象とは言えない。しかしながら、全国紙の論調の極性化が体系的に観察できるのか、またどのような争点で極性化が観察、進行してきたのかについて学術的コンセンサスは得られておらず、研究の余地があると考えられる。

## 2. 研究の目的

本研究は、一般に流布している「全国新聞の極性化現象」について新聞記事のテキスト分析から実証的にアプローチするものである。この研究には、(1)そもそも新聞記事の極性化の程度をどのように測定するか、(2)極性化はどの政治的争点に顕著にみられるか、(3)極性化がみられるとしてその大きさは近年変化しているのか、という3つのリサーチクエスチョンが含まれている。これらのクエスチョンに具体的な回答を提出することが本研究の目的である。

## 3. 研究の方法

新聞記事を計量テキスト分析の手法によって分析した。2000年以降の政治争点で、比較的継続性があり、また新聞社間で論調の差異が大きいと思われる10の争点について記事テキストを収集した。具体的には「原発」「憲法改正」「天皇」「アベノミクス」「消費税」「郵政民営化」「共謀罪法案」「安保法制」「北朝鮮問題」「自衛隊」である。各新聞社のデータベースで争点のキーワードを検索し、1面あるいは政治面の時期別の記事数を把握する。系統サンプリングによって実抽出記事を特定後、各新聞の縮刷版から記事を印刷し、データ化した。この時、社説記事と一般記事は区別せず、どちらも記事テキストとして扱う。ひとつの争点につき「朝日」「読売」「毎日」「産経」の4紙から各100件、合計400件を目標として記事を収集した。以上の手順で収集した記事は最終的に3,601件(平均662字)である。

記事の内容が極性化していることを示す「極性化指数」の算出には、教師あり機械学習手法のひとつであるランダムフォレストを利用した。まずは争点の記事の4分の1を層化サンプリングし、「朝日の記事か否か」を判定するモデルを作成する。次にこのモデルを利用して、残りの4分の3の記事が「朝日の記事である確率」を推定する(4分割交差検証)。この確率を4紙別に計算すると、各新聞の「朝日の類似度」を示す数値となる。4紙の数値の標準偏差が大きいことは、「朝日か否か」をはっきり識別できるという意味で、当該争点の記事は分極化していると判断される。逆に小さければ新聞間の違いはそれほどなく分極化していないと判断される。つまり、ある記事を「『朝日』と明瞭に識別できる程度」によって、極性化を測定しようという試みた。これが新聞記事テキストの「極性化指数」である。以上の手順を各争点において繰り返すことで、争点別の極性化の程度を測定した。

## 4. 研究成果

研究の成果は大きく3つに分けることができる。(1)テキスト分析から得られた極性化に関する知見、(2)テキスト分析の方法についての知見、(3)当初想定していなかったテキスト分析以外のデータ(世論調査データ)から得られた知見、である。

(1)新聞テキストを対象とした「極性化指数」による分析結果、得られた結論は以下である。

対象となった争点のうち、いくつかの争点では極性化指数が高く、新聞ごとの論調に差異を見出すことができた。極性化指数が高かった争点は、つまり新聞間の論調の違いが大きかった争点は、「共謀罪」(.312)「安保法制」(.289)「原発」(.276)であった。いずれの争点でも「朝日」を基準とした場合、最も近い距離にあるのは「毎日」であり、やや離れて「読売」「産経」が布置された(「ただし読売」「産経」の位置関係は争点によって入れ替わる)。世間的に共有されている「リベラルな朝日・毎日新聞」対「保守的な読売・産経新聞」という図式が、新聞記事テキストの分析でも確認できたと言える。一方、基本的にはこの図式を維持しつつも、ベースラインを超えて分極化していると判定できた争点は上記の3つのみであった。

全体として極性度の高い争点は当初想定したほど多くはなかった。いずれの争点も社説記事を読む限り差異があるようにみえるが、分析には社説以外の一般記事のテキストが多く含まれる。一般記事は事実関係の記述が多いため、新聞社間の差異が明確にならなかった可能性がある。

争点別の極性化指数を時系列でみた場合、いかなる特徴があるだろうか。結果から言えば、ほとんどの争点において時系列的な変化は確認できなかった。そもそも極性化が明瞭に観察できる争点は少なく、「共謀罪」「安保法制」など極性化指数が高い争点も、時間的な変化を追えるほど記事分量の広がりがなかった。唯一「原発」争点は興味深い変化があった。2012年以降一時的に「原発」争点の極性化指数が大きくなった。特に「朝日」と「読売」の違いは大きく、2011年の東日本大震災後、原発の将来性や再稼働をめぐる議論が2つの新聞記事上の言論の分断をもたらしたことが推測される。一方、原発再稼働の議論が落ち着きを見せると「原発」争点の極性化指数も低下した。この変化は興味深い事例であるものの、構造的なものであるもの、一時的なイベントの影響を受けたものと解釈できる。

今回収集した4つの新聞社の新聞記事の違いは、一部の争点を除けば、それほど大きくなかった。一時的なイベントによって露出が高まった争点や社説のみに注目すれば新聞社間の差異は見出せるが、一般記事まで含めたより広範な分析では、明確な差異はみられなかった。つまり、新聞記事の分極化を示す積極的な証拠は得られなかった。これが本研究の暫定的な結論である。

(2)以上のサブスタンスな知見とは別に、教師あり機会学習を用いたテキスト分析手法を適用するにあたって、本研究が直面した方法上の工夫や課題についても、この分野の今後の進展のために報告しておく。

テキスト分析では、分析のノイズを取り除くために、単語を取捨選択するのが一般的である。先行研究では「出現頻度が低い単語10~20%を削除する」などのアドホックな対応が採られているが、取捨選択の方法は分析結果に影響を与える可能性があるため、何らかの基準を設けるべきである。本研究では、単語マトリクス of 残差分析を利用して、統計的に出現頻度が少ない単語だけではなく、「多すぎる」単語も分析から除外し、精度の向上を図った。

新聞社の政治的立場に由来する差異以上に、争点の掲載の「時期」に由来する差異の方が大きく、記事の掲載時期を統計モデルでどのようにコントロールするかが、常に悩みの種であった。争点の中には「原発」や「天皇制」など比較的長期に渡って記事に現れるものが存在するが、そもそも記事の件数は特定の時期の政治イベントの発生に依存するため、「新聞社の政治的立場を反映した単語」ではなく「当該時期の政治イベントを反映した単語」が多くなる。つまり本来観察したい新聞社の論調の差異ではなく、アドホックな時期の差異の方が前面化してしまう。時期の影響をコントロールするため、単語行列とは別に「時期ダミー変数」を統計モデルに投入したが、その設定基準(「年」「月」「週」など)によって分極化指数の精度にも差が生まれる。本研究では時期ダミー変数を暫定的に「半年」単位で統一したが、全争点に共通する一般的な基準を定めることが難しかった。

教師あり学習によるテキストの分類は、申請者にとって初めての試みであったため、手法の確立に時間がかかると予想していたが、実際、想定以上の時間がかかった。単語の前処理や機械学習モデルの選定、チューニングの設定、それらの検証に多くの時間を要した。最終的にパフォーマンスが比較的良好だったランダムフォレストを分類手法として採用したが、なぜ当該手法が他の手法よりも優れているのか、「新聞記事」というテキストの特殊性がどの程度アウトプット精度に依存するのか、などの論点については今後の課題である。

機械学習によるテキスト分析手法は近年増加しているが、どの種類のテキストをどのように分析すべきかの細かいノウハウについては発展段階のため、本研究でも適用に際してもいくつも困難があった。分析の各過程で選択肢が多いことは教師あり機械学習の魅力である。しかし、得られた指数が実態を測定したものなのか、それとも分析手法に依存するアーティファクト(人工物)なのか、判断に迷うことが多かった。実際、上記の知見の一部を研究会で発表した際、最も質問が多かったのは機械学習のノウハウに関する部分であり、多くの研究者が試行錯誤していることが分かった。これら(ネガティブな内容も含む)ノウハウを開示し、他研究者と共有することも、研究の発展段階においては重要であると考えている。

(3)コロナ禍によって本研究の基礎的なデータである新聞記事の収集に制限がかかったことは、当初想定しえなかったことである。しかし、その制限期間において本研究に刺激を受けるかたちで極性化についての関連研究を進められたことは想定外の収穫であった。以下では、極性化に関連する副次的な成果について報告する。

2021年度に実施した世論調査データを利用して、有権者が接触する複数のメディアの「分極化度」を測定した。新聞やニュース番組などを含む約20のメディアについてそれぞれ分極化係数を推定した結果、インターネット(特にソーシャルメディア)や対人ネットワークの分極化係数が比較的大きいことが分かった。一方、新聞やニュース番組は分極化係数が低く、比較的穏健なメディアと言えるが、接触度が他のメディアよりも高いため、メディア環境全体として

みた場合、その偏りは無視できないことが判明した。この研究は 2022 年日本選挙学会研究会で口頭発表している（優秀ポスター賞受賞）。

2000 年以降に実施された複数の世論調査データを用いて、有権者の「感情的分極化」の時系列変化を測定した。「感情的分極化」とは、自分の支持政党（内政党）とは異なる政党（外政党）についての感情的な距離（嫌悪）である。この距離は有権者の政治的対立の大きさ（分極化）と解釈することが可能である。有権者は報道の影響を受け、近年、政治態度を分極化させている可能性があることを念頭においている。分析の結果、2012 年の第二次安倍政権以降、感情的分極化が若干拡大しているものの大きな変化とまでは言えず、むしろ最近 20 年の分極化の程度は安定的に推移してきたことが判明した。この研究成果は 2021 年の日本マス・コミュニケーション学会（現・日本メディア学会）秋季研究大会と 2022 年日本政治学会分野別研究会で報告している。

上記の で示したもとの同一のデータ、すなわち 2000 年以降に実施された複数の世論調査データを用いて、有権者が接触するメディアの保守係数（保守/革新度）を測定し、時系列変化の特徴を調査した。データには有権者の保守イデオロギー態度、購読新聞が記録されており、これらを利用して購読新聞の保守/革新度を推定する方法を提案した。具体的には「読売新聞」の保守度を測定するためには、「(読売新聞購読者のうち自身を保守とする人数 ÷ 読売新聞購読者のうち自身を革新とする人数) ÷ (読売新聞購読者のうち自身を保守あるいは革新とする人)」を係数として利用すればよい。購読新聞別にこの係数を算出することで、各調査時点における各新聞のイデオロギー位置を推定できる。分析の結果、世間で認知されているように「朝日新聞」「毎日新聞」の係数の絶対値は小さく（革新に偏っている）、「読売新聞」「産経新聞」の係数が大きい（保守に偏っている）ことが判明した。一方、2000 年以降、購読新聞の保守係数は特段大きな変動はなく、近年新聞が極性化している証拠はなかった。この知見は今後、学会で発表する予定である。

以上、本研究は新聞記事のテキスト分析を中心としながら、当初想定していなかった世論調査データも分析に加えることで、結果として日本の政治的極性化現象を多面的に検証するものとなった。いずれの研究においても、当初想定していた「極性化」を積極的に示す実証的証拠は見いだせなかった。この知見は政治的極性化や「社会的分断」をめぐる近年の議論に示唆を与えるものであろう。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計6件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 細貝亮
2. 発表標題 オンライン/オフラインメディアの選択性
3. 学会等名 日本政治学会分野別研究会「近代現代日本政治研究フォーラム」
4. 発表年 2022年

1. 発表者名 細貝亮
2. 発表標題 日本における感情的分極化 - その変化と規定要因 -
3. 学会等名 日本選挙学会
4. 発表年 2022年

1. 発表者名 細貝亮
2. 発表標題 インターネットは選択性の高いメディアか？
3. 学会等名 日本マス・コミュニケーション学会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------