

計量テキスト分析による地方紙社説の独自性の検証

Originality of Editorials of Local Newspapers in Japan: A Quantitative Text Analysis

◎金子 智樹¹
Tomoki KANEKO

¹ 東京大学大学院法学政治学研究科博士課程・日本学術振興会特別研究員 (DC2)
Graduate Schools for Law and Politics at the University of Tokyo; Research Fellow of JSPS

要旨・・・日本の地方紙の社説に関しては、「共同通信の『資料版論説』を引き写したのが多い」という言説が存在しており、地方紙社説の論調と関連付けて批判の対象とされることもある。しかし、主要地方紙の社説が実際にどの程度、共同通信の「資料版論説」をベースに書かれていると考えられるのかを、網羅的に分析した研究は乏しい。本研究では、2017年～2018年にかけての10ヶ月間の主要地方紙の社説テキストをデータ化した上で、計量テキスト分析の手法を用いて、各地方紙の「共同通信資料利用頻度」を数値化することを試みた。結果として、「共同通信資料利用頻度」が極端に高い地方紙は確かに存在するものの、それは地方紙全般の特徴とは言えないことや、社説のトピックによって「共同通信資料利用頻度」には多様性が見られることなどが分かった。

キーワード 地方紙、社説、共同通信、テキスト分析、政治コミュニケーション

1. はじめに

日本における政治コミュニケーションを考える上で地方紙は重要な存在である(金子2018)。特に各都道府県で高い普及率を占める県紙やブロック紙の報道内容は、地域の有権者の政治意識に影響を与えている可能性も指摘されている(金子2017)。

各新聞の政治的立ち位置が最も象徴的に反映されているのは、社説の文章であろう。全国紙の社説の特徴や論調に関しては、主に定性的な内容分析に基づく研究蓄積がなされている。しかし地方紙の社説に関しては、その重要性に比して十分に研究が行われているとは言いがたい。その理由の1つとして、後述するように「地方紙の社説は、共同通信の資料を引き写している」というイメージや言説が存在していることがあると思われる。もし日本の地方紙の社説の一般的な特徴が、通信社の配信する資料のコピーであるとすれば、各地方紙の社説の内容を精査する意義も薄れてしまうかもしれない。

しかし、こういった地方紙の社説観が実際にどの程度妥当なのか、客観的・網羅的に分析した研究は乏しい。そこで本研究では、各地方紙の社説テキストを可能な限り網羅的に収集した上で、計量テキスト分析を行い、「各地方紙の社説文章がどの程度オリジナルなものだと推測されるか」を数値化することを試みる。

2. 先行研究の検討

地方紙の社説や論調に関しては、憲法改正やTPP、原発問題などの個別の争点やテーマを対象として内容分析した先行研究や知見がいくつか存在する(丸山2006; 北原2012; 藤田2013; 住吉2014; 梶居2017; 農山漁村文化協会(編)2017)。個々の内容に言及する紙幅の余裕は無いが、多くの地方紙の政治的立ち位置は、「全国紙に比べると中央の政治権力に距離を置いており、比較的リベラルである」と評価されることが多いと考えられる¹。

しかし、地方紙の社説に関しては、「共同通信の配信している『資料版論説』の引き写しをしているのではないか」という重要な指摘が存在する。共同通信の論説資料は、「共同通信論説委員会が、契約地方新聞用に配信する社説用『雛形』」であり、地方紙がどのように使っても構わないとされている(朝倉2010: 41)。この点は、上述の「地方紙がリベラル寄りである」というイメージと合わせて、保守派の論客から批判材料として用いられることもある(日下(編)2011)が、地方紙の社説が共同通信の資料を流用することで紋切り型になっているという指摘は護憲派の立場からもなされているところである(丸山2006)。他紙の社説の盗用を行ったとされる山梨日日新聞の元論説委員長の手記でも、その遠因として共同通信の「資料版論説」を自社の社説として掲載することへの抵抗があったことが述べられている(小林2007)。学術論文としては、2001年8月15日の地方

紙社説のうち通信社社説に基づくものが4分の1程度存在していること（石川2007: 28-30）や、2017年5月の憲法関連社説において、共同通信の配信資料を利用したと推測される地方紙が一定数存在していたこと²（梶居2017: 53-54）が報告されている。

これらの知見や先行研究の指摘は重要であるが、本研究の問題関心においては以下のような限界があると思われる。まず、評論家による地方紙社説の「引き写し」論は、地方紙の論調への批判を前提としている場合もあるため、客観的な立場からの分析が必要である。また学術研究においても、特定のトピックやテーマに絞って社説の内容分析をしているケースが専らであり、多くの場合は分析対象日・期間もかなり限定されているため、その知見を地方紙社説の特徴として一般化できる保証は無い。さらに、「共同通信の参考資料を流用しているかどうか」の判定を専らヒューマン・コーディングに頼る場合、判定基準の客観性を欠く可能性がある。地方紙は共同通信の「資料版論説」の全体ではなく一部を切り貼りして社説を書く場合もあるため、特に大量のテキストの内容分析を行う場合、人間の判断によって「共同通信のコピー」かどうかを評価することは現実的に難しい。すなわち、日本の主要地方紙の社説テキストを可能な限り網羅的に収集した上で、何らかの機械的な方法によって各社説文章の「オリジナル度」を分析する必要がある。

3. データと分析方法

以上の検討を踏まえた上で、本研究では、各新聞社のウェブサイトにて毎日掲載される社説テキストをウェブ・スクレイピングによって収集・データ化し、計量テキスト分析の手法によって各社説の「オリジナル度」を数値化するという着想に至った。

具体的には、2017年10月1日～2018年7月31日の10ヶ月間を分析対象期間として、有力地方紙（少なくともいずれかの都道府県で世帯普及率上位3紙に入っている地方紙+沖縄の2紙）と全国紙5紙、政党機関紙2紙の社説テキスト（計14368個）を収集した。地方紙以外の新聞も収集対象としているのは、後述する指標の妥当性をチェックするためである。収集対象の新聞は表1に示した通りである（「社説」以外のタイトルを冠している新聞はカッコ内に記載）。なお社説の収集においては、政治に係る社説を選び出すのではなく、掲載された全ての社説を対象としている。

表1 分析対象の新聞一覧

北海道新聞	北日本新聞	中国新聞	琉球新報
東奥日報（時論）	北國・富山新聞	徳島新聞	
デーリー東北（時評）	福井新聞（論説）	愛媛新聞	<全国紙>
岩手日報（論説）	山梨日日新聞（論説）	高知新聞	読売新聞
河北新報	信濃毎日新聞	西日本新聞	朝日新聞
秋田魁新報	中日新聞	佐賀新聞（論説）	毎日新聞
山形新聞	京都新聞	大分合同新聞（論説）	日本経済新聞
福島民報（論説）	神戸新聞	熊本日日新聞	産経新聞（主張）
福島民友	日本海新聞	宮崎日日新聞	<政党機関紙>
茨城新聞（論説）	山陰中央新報（論説）	南日本新聞	しんぶん赤旗（主張）
新潟日報	山陽新聞	沖縄タイムス	公明新聞（主張）

次に、各地方紙の社説が、共同通信の「資料版論説」をどの程度参考に行っていると考えられるのかを数値化することが目標になるが、「資料版論説」自体は残念ながら一般公開されていないため、「文章が似通った他紙社説が存在するかどうか」を調べることによって代替する。ただし、2つの地方紙の別の日の社説が同一の「資料版論説」を用いているケースがあるため、「それぞれの社説の掲載日と前後3日間ずつの計7日間」について、他紙社説を参照することにした。

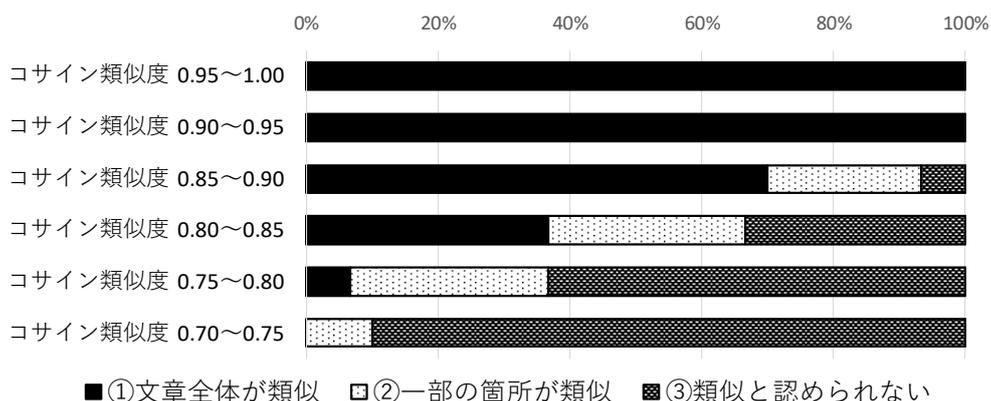
2つの社説テキストの類似性を測る際には、上述したように機械的な手法で行うことが望ましい。そこで本研究では、文章の類似度の代表的な指標である「コサイン類似度」を採用する。具体的には、各テキストの前処理を行った上で、統計分析ソフトRの`quanteda`パッケージを利用してトークン化し、各社説における各単語の出現頻度を算出する⁴。

コサイン類似度がどの程度高いとほぼ同じ文章だと見なせるのか、さらには共同通信の「資料版論説」という同一の「元ネタ」を用いていると考えられるのかについては、別途人間の判断を要する。ここでは、佐賀新聞の各日の社説に関して、コサイン類似度が一定の範囲内（6階級）であるような他紙社説をランダムに1つ選び（無い場合はスキップ）、選ばれた佐賀新聞社説と他紙社説のペアからさらにランダムに30個抽出した上で、計180ペアの社説テキストをヒューマン・コーディングすることにした。すなわち、①「文章全体の構成から考えて、明らかに同一の『資料版論説』の引き写しだと判断できるペア」、②「一部の箇所では同じ表現が見られるなど、同一の『資料版論説』を部分的に用いているが、全体としては『コピー』とは言

えないペア」、③「同一の表現が少なく、それぞれ完全に別個のテキストだと考えられるペア」、の3グループに分類する。

図1は、コサイン類似度ごとにヒューマン・コーディングの結果をまとめたものである。まず、コサイン類似度0.9以上の社説については、文章全体が類似しており、「資料版論説」の「引き写し」と判断して良いと思われる。コサイン類似度0.8～0.9では、「引き写し」だけでなく、「一部利用している」パターンも見受けられる。コサイン類似度0.8以下では「引き写し」パターンはほとんど見受けられない。以上の検討より、コサイン類似度0.9以上の社説は両者が「資料版論説」を流用している可能性が極めて高く、0.8以上の社説は少なくとも部分的に参考にしている可能性がある程度高い、と推測できることになる。

図1 2つの社説テキスト間のコサイン類似度と、ヒューマン・コーディングの結果の比較



そして、各地方紙が全体としてどの程度共同通信の「資料版論説」を使用していると考えられるかについては、「分析期間中に各新聞で社説が掲載された日のうち、コサイン類似度が一定値以上の他紙社説が7日間以内に存在していた日の割合」を算出することによって求め、これを便宜上「共同通信資料利用頻度」と呼ぶことにする。

4. 地方紙社説の「共同通信資料利用頻度」

各地方紙の「共同通信資料利用頻度」をまとめたのが図2である。判断基準のコサイン類似度を0.9以上とした場合の値が黒のグラフ、0.8以上とした場合の値が斜線のグラフで示されている。まず、コサイン類似度0.9以上の「共同通信の引き写し」という意味での「共同通信資料利用頻度」に注目すると、山陰中央新報、茨城新聞、大分合同新聞、佐賀新聞、東奥日報、宮崎日日新聞、山形新聞、山梨日日新聞など、頻繁に共同通信の「資料版論説」を流用していると考えられる地方紙が一定数存在することが分かった。特に山陰中央新報、茨城新聞、大分合同新聞は「共同通信資料利用頻度」が極めて高く、オリジナルな社説はあまり掲載されていない。また、基準をコサイン類似度0.8以上にした場合も全体的な傾向は大きく変わらない。しかし、逆にその他の多くの地方紙は、「共同通信資料利用頻度」はいずれも低く、多くの場合は10%以下であった。

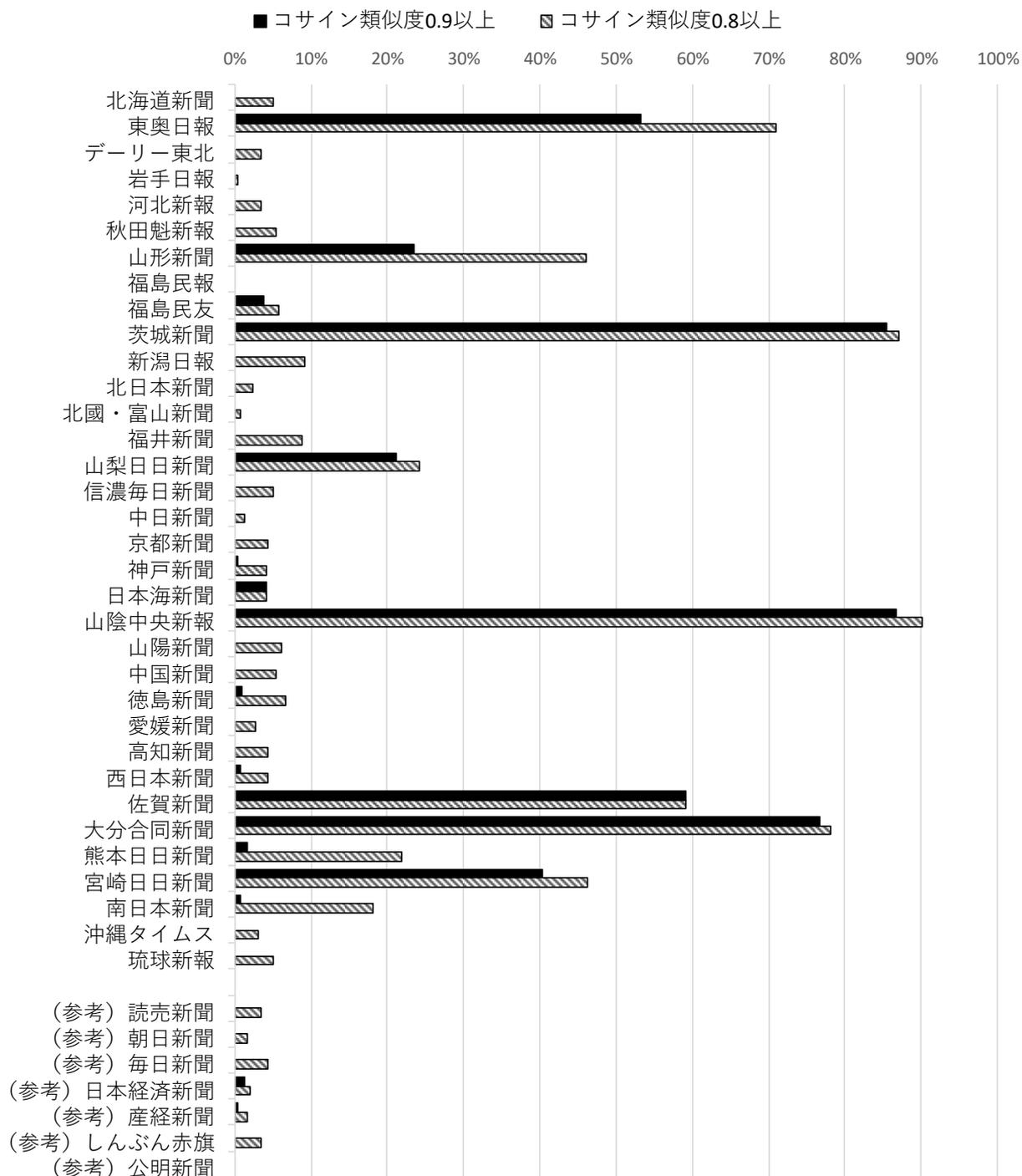
すなわち、共同通信の「資料版論説」を引き写した社説を主として掲載する地方紙は限定的であり、主要地方紙一般の傾向とは考えにくいと言えるだろう。またコサイン類似度0.8以上の結果を見る限り、「資料版論説」を部分的に使用するケースもそれほど多くはないと推測される。

地方紙を個別に見ていくと、憲法社説に関する梶居(2017)で挙げられていた新聞は「共同通信資料利用頻度」も概して高いが、日本海新聞のように必ずしも当てはまらない新聞もある。さらに、表1で記載したように「社説」以外のタイトルを付けている地方紙は「共同通信資料利用頻度」が高い傾向にある⁹⁾。「論説」のような名称で「社論」としての意味合いを薄めることで、共同通信の「資料版論説」を活用することを暗に認めているとも考えられる。この点に関連して、東奥日報は2017年3月まで「社説」として掲載していたが、同月末日の告知で、同欄を「時論」へと名称変更し「共同通信の論説資料も活用しながら」執筆される旨が発表されたことも象徴的だと言えるだろう。

最後に、どのようなトピックの社説が「共同通信資料利用頻度」が高い傾向にあるのかも報告する。紙幅が限られているため分析手順の詳細については省略するが、潜在トピックモデルと呼ばれる手法を用い、各社説テキストの単語分布からトピックの分布を推定し、事後分布の確率が最大となるような1つのトピックをその社説テキストのトピックであると判断している¹⁰⁾。

図3は、全社説テキスト内の出現頻度が上位30位以内の各トピックの社説に関して、コサイン類似度が0.9以上の他紙社説が存在している割合(共同通信資料利用頻度)をまとめたものである(頻度順に記載)¹¹⁾。それぞれのトピックのラベル名につい

図2 各地方紙の社説において、文章の類似度が高い他紙社説が存在する日の割合（共同通信資料利用頻度）

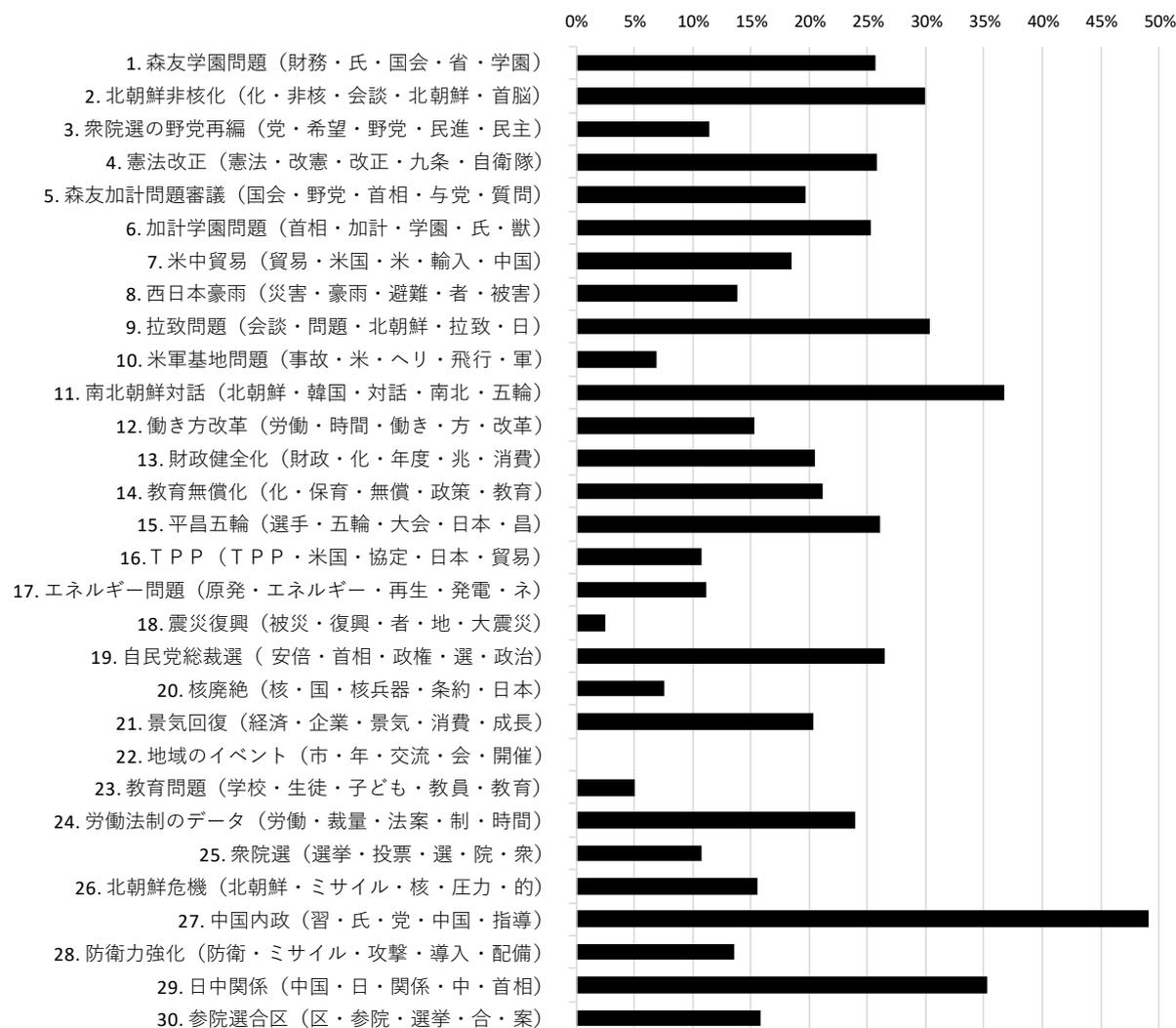


ては、特徴的な出現語句のリストを確認・解釈した上で決定した。カッコ内には、各トピックに特徴的な上位5個の語句を参考として記載している¹²。

図3からまず読み取れるのは、<27：中国内政>や<11：南北朝鮮対話>などのように、外交や国際政治に関わるトピックについては「共同通信資料利用頻度」が高いことである。このようなトピックは各地方紙の論説担当者だけでは対応が難しく、共同通信の「資料版論説」に頼らざるを得ない場合があることを示唆している。逆に、当然ではあるが、<22：地域のイベント>のような地域ごとのトピックについては各地方紙が自前で社説を執筆していることも分かる。それに加えて、<10：米軍基地問題>や<18：震災復興>のように各都道府県に固有かつ政治的に重要な問題については「共同通信資料利用頻度」が低

く、当該地域の地方紙だけでなく他地域の地方紙も、「地方」という独自の視点で社説を執筆していると考えられる。2017年の衆院選関連のトピックについても、各都道府県の政治動向を含めたオリジナルな社説が書かれる傾向が見られる。また、政権のスキャンダルに関する各トピックについても「共同通信資料利用頻度」は30%以下とそれほど高いわけではなく、独自の社説を書いている地方紙が多数であることは、やや意外な結果と言えるかもしれない。

図3 出現頻度が上位30位のトピックごとの「共同通信資料利用頻度」(コサイン類似度 ≥ 0.9)



5. おわりに

以上の知見を簡潔にまとめよう。主要地方紙の中には社説執筆において共同通信の「資料版論説」に大部分を依存していると考えられる新聞も確かに存在するが、それは地方紙全般に当てはまる傾向とは言えず、多くの地方紙は独自の視点で社説を書くことが多いと推量される。またトピックごとに見ると、共同通信の資料を引き写して社説が書かれる(と思われる)割合はトピックごとに大きく異なる。換言すれば、地方紙ごとやトピックごとに、各社独自の社説が掲載される割合に関しては多様性が存在するのであり、ステレオタイプ的に論評することは妥当性を欠く可能性があると言える。

本研究にも課題が無いわけではない。まず、公式サイトから社説を入手できない地方紙(上毛新聞、下野新聞、岐阜新聞など)については、共同通信の「資料版論説」を多用していることを背景に社説を掲載していないかもしれない。また、「共同通信資料利用頻度」が高い地方紙は他にも存在する可能性がある。また、2つの社説の間に共通の「元ネタ」が存在するかどうかをコサイン類似度という指標だけで正確に判定できるわけではない。少なくとも社説のテーマを決める際には、各地方紙は共同通信

の「資料版論説」を何らかの形で参考していると考えられるのも自然であり、本研究は大多数の地方紙が完全にオリジナルな社説を専ら書いていることを主張するものではない。

しかし、地方紙の社説が共同通信の資料の「切り貼り」に過ぎないというイメージはやはり一面的であり、多くの地方紙の社説には各社の独自の視点が盛り込まれていると考えられる。地方紙の社説の論調の分析は別稿に譲らざるを得ないが、仮に多くの地方紙の論調がリベラル寄りだったとしても、その原因を共同通信の強大な影響力に求めるのはやや早計と思われる。換言すれば、それぞれの地方紙の論調や報道内容を個別に分析・検討していく意義も小さくないのではないだろうか。

補注

※本研究は、JSPS 科研費 18112661 の助成を受けたものです。また本研究で用いた社説テキストは、三輪洋文・学習院大学法学部政治学科准教授との共同プロジェクトにおいて収集されたデータであることを付記します。

- 1)ただし、北國・富山新聞の社説は比較的保守寄りだと見なされることが多いなど、実際には地方紙内でも多様性が存在している。
- 2)具体的には、東奥日報、茨城新聞、上毛新聞、岐阜新聞、日本海新聞、山陰中央新報、佐賀新聞、宮崎日日新聞、山形新聞も一部利用していると述べられている。
- 3)日本 ABC 協会 (2018) における各紙朝刊部数を、総務省「住民基本台帳に基づく人口、人口動態及び世帯数調査」における各都道府県世帯数で割ることによって算出。四国新聞のように社説自体が無い地方紙や、公式サイトで社説を閲覧することができない地方紙は除外される。
- 4)ひらがなや算用数字のみの単語を除外するなどの処理を別途行っている。
- 5)佐賀新聞を選択したのは、図2に示したように、同紙社説の「共同通信資料利用頻度」が比較的高いことによる。
- 6)厳密には複数のコーダーを用意することが望ましいが、ここでは筆者自身がコーディングを行った。なお、コーディングの際には各々のコサイン類似度を伏せた形式で行っている。
- 7)単に文の順番を入れ替えただけのペアや、部分的にオリジナルの文が挿入されている程度の社説も含まれる。
- 8)ただし、コサイン類似度 0.8 以上を判断基準にすると、共同通信の資料を参考にしていないとは考えにくい全国紙や政党機関紙も「共同通信資料利用頻度」が 0 以上となるケースが生じていることから分かるように、「共同通信資料利用頻度」はもちろん絶対的な指標ではない。
- 9)なお、「社説」以外の名称を用いても「共同通信資料利用頻度」が低い地方紙もある点には留意されたい。
- 10)具体的には、R の `stm` パッケージを用いて構造的トピックモデルにより推定した。トピック数の指定は、今回は 160 としている。
- 11)図3の分析においては、1日に1つの社説が掲載される地方紙にサンプルを限定している。
- 12)ただし、ラベル名を付ける際にはより下位の語句も確認している。

参考文献

- 1)朝倉敏夫 (2010) 『論説入門』中央公論新社。
- 2)石川徳幸 (2007) 「新聞社説に関する一考察：社説におけるテーマ設定の独自性」『選挙学会紀要』825-42。
- 3)梶居佳広 (2017) 「日本国憲法をめぐる新聞論説：施行70年の憲法記念日を中心に」『社会システム研究』35:47-74。
- 4)金子智樹 (2017) 「地方紙読者の安倍政権評価」『生活経済政策』247:26-30。
- 5)金子智樹 (2018) 「地方紙の普及率低下は投票率を下落させるのか?：鹿児島新聞の廃刊という自然実験的事例と参院選パネルデータの分析」『年報政治学』69(1):202-224。
- 6)北原斗紀彦 (2012) 「原発立地県の地方紙の論調分析：<3.11後>をどう論じてきたか」『尚美学園大学総合政策論集』14:43-65。
- 7)日下公人 (編) (2011) 『誰も書かなかった「反日」地方紙の正体』産経新聞出版。
- 8)小林広 (2007) 「手記私になぜ社説を盗用したか：不覚の結末」『論壇』146:184-195。
- 9)住吉主税 (2014) 「テキストマイニングを用いた都道府県の新聞論調の比較内容分析：自民党政権に対する分析」『慶応義塾大学大学院法学研究科論文集』54:93-137。
- 10)日本 ABC 協会 (2018) 『新聞発行社レポート・月別府県 (2018年4月)』日本 ABC 協会。
- 11)農山漁村文化協会 (編) (2017) 『地方紙の眼力：改憲・安全保障・震災復興・原発・TPP・地方創生』農山漁村文化協会。
- 12)藤田博司 (2013) 「新聞の立ち位置が問われている：TPP報道で露わになった全国紙と地方紙の乖離」『世界』844:154-162。
- 13)丸山重威 (2006) 『新聞は憲法を捨てていいのか』新日本出版社。