

マイ広報紙を用いた首長の性別による 記事数・文字数・テキスト内容の差異分析

蛇谷祐人¹ 木村泰知¹ 尾野嘉邦²¹小樽商科大学 ²早稲田大学

g1202519607@edu.otaru-uc.ac.jp kimura@res.otaru-uc.ac.jp

概要

本研究は、自治体が発行する「広報紙」を対象として、自治体ごとの情報発信の特徴および差異を明らかにすることを目的とする。まず、マイ広報紙サイトにおいてテキスト化され公開されている全国の広報紙を収集し、テキスト分析が可能な JSON 形式へと整理・構造化した。次に、構築したデータを用い、首長の性別に着目して、記事数や文字数といった量的指標に基づく比較分析に加え、テキスト内容に基づく比較分析を行った。

1 はじめに

自治体の広報は、市民とのコミュニケーションを通じて政策情報を伝達し、住民参加や住民理解を促進する重要な役割を担っている。とりわけ自治体広報紙は、行政による公式な情報提供の中核を担う媒体として、政策・施策から生活関連情報に至るまでを継続的に住民へ提供してきた。従来研究では、自治体広報一般は扱われてきたものの、自治体広報紙(広報紙)を対象として本文テキストの計量分析に基づく自治体間の定量比較を行った研究は限定的である [1]。さらに、政治学における代表制の議論では、属性の構成(記述的代表)が政策(実質的代表)にいかに関与するかが問われており [2]、首長の性別が政治的判断に有意な変化をもたらすことも実証されている [3]。広報実務自体は職員が担うが、政治的リーダーの選好が組織のアウトプットに動的に反映されることから [4]、広報紙においても首長の属性が間接的に波及すると考えられる。

そこで、本研究の目的は、自治体が発行する広報紙を対象として、首長の属性による情報発信の特徴および差異を明らかにすることである。

本研究の貢献は以下の3点である。

1. 全国の自治体広報紙を収集し、本文テキストを

含む形で JSON フォーマットへと整理した。

2. 整理したデータセットを用いて、自治体ごとの情報発信パターンを定量的に分析した。
3. 首長の性別などの自治体属性を説明変数とし、広報紙の記事量との関係を統計的に分析した。

2 マイ広報紙の収集と整理

2.1 マイ広報紙とは

広報紙は、地方自治体が発行する広報紙である。マイ広報紙とは、全国の広報紙を収集し、テキストデータとして公開するサービスである¹⁾。従来、広報紙は紙媒体や PDF 形式での公開が中心であり、検索や利活用に制約があった。マイ広報紙では、図 1 に示したように、自治体が発行する広報紙の記事ごとにテキストデータ化し、インターネットで無料配信することで、全文検索や記事単位での閲覧を可能とし、住民が関心のある自治体やテーマに応じて効率的に情報を得られる仕組みを提供している。

2.2 マイ広報紙の収集

本研究では、全国 1,955 自治体(行政区を含む)を対象に、マイ広報紙からテキスト化された広報紙を収集した。取得可能であった広報紙については、自治体ごとに JSON 形式で整理し、各記事の本文をテキストデータとして抽出した。

抽出したテキストは、表 1 に示す項目に基づいて整理した。具体的には、自治体名や自治体コード²⁾といった基本情報に加え、ファイルサイズ、広報紙数、記事数、総文字数などの量的指標を記録した。さらに、「暮らし」「しごと」「子育て」などのカテゴリ別に記事数を集計し、記事あたりの平均文字数やテキスト率を算出した。

1) <https://mykoho.jp/>

2) <https://www.soumu.go.jp/denshijiti/code.html>

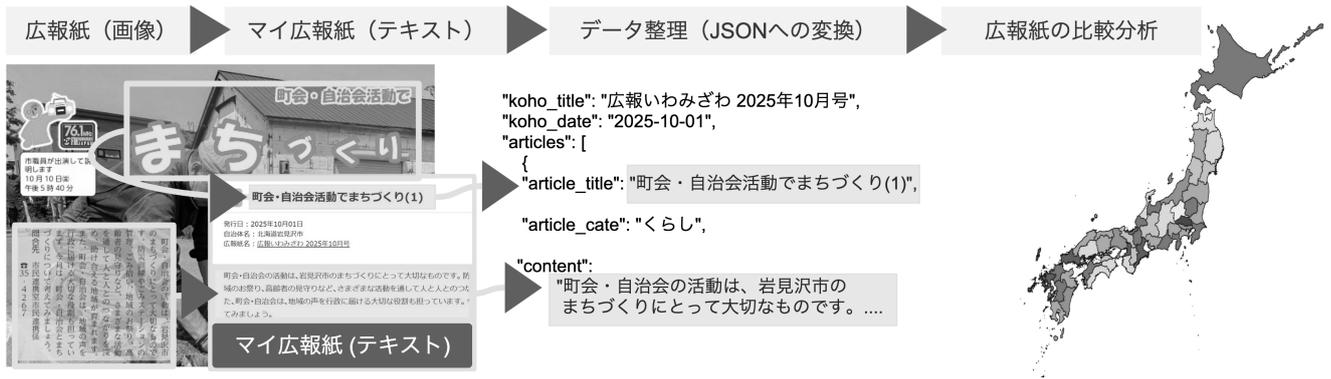


図1 研究概要：マイ広報紙報からデータ収集および広報内容の比較分析の流れ

表1 収集結果のデータ項目

項目名	説明
自治体名	地方自治体の名称
自治体コード	6桁の地方公共団体コード
ファイルサイズ	サイズの合計 (KB)
広報紙数	広報紙の総数
記事数	記事の総数
テキスト記事数	テキスト記事の総数
総文字数	広報紙に含まれる総文字数
掲載開始日	最も古い発行日
掲載終了日	最も新しい発行日
カテゴリ_くらし	「くらし」区分の記事数
カテゴリ_しごと	「しごと」区分の記事数
カテゴリ_その他	「その他」区分の記事数
カテゴリ_イベント	「イベント」区分の記事数
カテゴリ_スポーツ	「スポーツ」区分の記事数
カテゴリ_健康	「健康」区分の記事数
カテゴリ_子育て	「子育て」区分の記事数
カテゴリ_文化	「文化」区分の記事数
カテゴリ_講座	「講座」区分の記事数
平均文字数/記事	総文字数 ÷ 記事数
テキスト率 (%)	テキスト記事数 ÷ 記事数

2.3 マイ広報紙の整理

マイ広報紙サイトからは、1,955自治体のうち997自治体分の広報紙をダウンロード可能であり、取得率は50.9%であった。収集されたデータセットの総容量は約0.91GBで、総記事数は501,647件、そのうちテキスト化された記事は478,905件であった。総文字数は約3億1,699万文字に達する。

収集対象とした広報紙の発行期間は、2024年8月30日から2026年1月1日までである。記事内容の

カテゴリ別内訳を見ると、「くらし」に関する記事が最も多く261,781件を占め、次いで「イベント」53,111件、「子育て」45,663件の順となった。

3 首長の性別と広報内容

3.1 目的

本節では、全国の地方自治体が発行する広報紙における、首長の属性(性別)と広報内容の関連性について検証を行う。特に、自治体広報において住民の関心が高く、かつ政策的優先度が反映されやすい「子育て」カテゴリを対象として、量的指標に基づく比較分析およびテキスト内容に基づく比較分析を行い、さらに交絡要因を考慮した分析を通じて検討を行う。

3.2 分析対象のマイ広報紙

本研究では、季節変動による偏りを排除し、通年での広報戦略を捉えるため、分析期間を12カ月(1年間)と設定した。期間の選定にあたっては、予備調査の結果、2024年11月から2025年10月までの12カ月間において、完全なデータを保有する自治体数が最大(847件)となることが確認されたため、本期間を分析対象期間として採用した。広報紙データの掲載が確認された997自治体に対し、全国の現役首長一覧³⁾を用いて、首長属性との紐づけおよびデータ完備性によるフィルタリングを実施した。具体的な除外基準およびプロセスは以下の通りである。第一に、行政区および一部の特別地方公共団体を除外した(100件)。これらは選挙により選出された首長が存在しない、あるいは本研究の定義する首長属性の分析に適さないためである。第二に、対象

3) <https://uub.jp/lce/lce.html>

表2 首長の性別による広報紙(子育てカテゴリ)の量的指標に基づく比較

分析項目	男性首長 (n = 627)		女性首長 (n = 35)		検定結果	
	平均値	中央値	平均値	中央値	Welch	U 検定
記事数 (件)	51.0	44.0	60.5	44.0	.278	.524
総文字数 (文字)	29,191	24,887	34,686	23,820	.304	.707

注: Welch はウェルチの t 検定, U 検定はマン・ホイットニーの U 検定による両側確率を示す.

期間中に首長の交代があった,あるいは任期が期間全体をカバーしていない自治体を除外した(103件).これは,広報紙の内容に対する特定の首長の影響を正確に測定するためである.第三に,当該期間中にデータ欠損が存在し,12カ月分のテキストデータが揃わなかった自治体を除外した(131件).以上の手続きを経た結果,最終的に662自治体(男627・女35)を本研究の分析対象とした.

3.3 分析方法

量的指標に基づく比較分析:記事数,文字数

本節では,首長の性別によって「記事数」および「総文字数」に統計的な有意差が存在するかを検証する.分析に先立ち,各指標の分布形状を確認した上で,ウェルチの t 検定とマン・ホイットニーの U 検定客観的な数値に基づいて評価を行う.

テキスト内容に基づく比較分析:対数尤度比

本分析では,男性首長(627名)と女性首長(35名)というサンプルサイズの著しい不均衡が存在するため,対数尤度比[5]を用いて分析した.さらに,政策トピックの具体性を捉えるため複合名詞のみを分析対象とし[6],低頻度語の除外[7],バースティネスの考慮[8]によって性別グループとしての一般的な傾向を抽出することを目的に,特徴グループの文書頻度を3以上とした.また,意味的に除外するストップワード[6]を設定し,ノイズ除去を行った.これらの前処理を施した上で,男女それぞれに特徴的な語彙を抽出し比較する.

交絡要因を考慮した分析

交絡要因とは,分析対象の違いとは別に結果へ影響を与える可能性のある要因である.首長の性別に加え,年齢,在任期数,人口規模⁴⁾といった交絡要因を考慮するため,重回帰分析を用いて各要因を統制する.これにより,首長の性別が広報紙の文字数

などの量的指標に与える固有の影響を検証する.

3.4 結果

量的指標に基づく比較分析:記事数,文字数

表2に示した通り,子育てカテゴリに関する記事数および総文字数は,首長の性別によって有意な差を示さなかった.

テキスト内容に基づく比較分析:対数尤度比

表3に示した通り,対数尤度比(LLR)を用いた特徴語分析の結果,首長の性別によって子育て支援情報の提示における焦点に差異が認められた.男性首長の特徴語上位には,子育て支援センター,保育所など,組織・施設の名称が多く含まれていた.一方,女性首長の特徴語としては,組織・施設名は含まれるものの,児童育成手当,医療証などの制度の提示,思春期,乳幼児向けなど具体的な対象を指した語彙が抽出された.

交絡要因の統制:年齢,性別,人口,期数

表4に示した通り,広報量(総文字数)の決定要因を検証するために実施した重回帰分析の結果,首長の性別($p = .754$),年齢,在任期数といった個人的属性はいずれも広報量に対して統計的に有意な影響を示さなかった.このことは,首長個人の属性が広報の「量」を規定する主要因ではないことを示唆している.一方,自治体の人口規模(対数変換)は,0.01水準で有意な正の影響を示し($p < .001$),広報量に対して強い規定力を有することが明らかとなった.

3.5 考察

情報発信の量的規模に影響を与える要因の分析

記事数と文字数による量的指標に基づく比較分析において,首長の性別による有意差が認められなかった点については,重回帰分析の結果がその背景を説明している.分析の結果,記事数や文字数と

4) https://www.soumu.go.jp/main_sosiki/jichi_gyousei/daiyo/jinkou_jinkoudoutai-setaisuu.html

表3 首長の性別による特徴語上位12語とその分類(LLRスコア順)

順位	男性首長			女性首長		
	単語	LLR	カテゴリ	単語	LLR	カテゴリ
1	保健センター	53.5	組織・施設	児童センター	132.4	組織・施設
2	子育て支援センター	46.8	組織・施設	子ども家庭支援センター	65.3	組織・施設
3	保育所	44.6	組織・施設	児童育成手当	62.7	制度・事務
4	こども支援課	40.5	組織・施設	思春期	60.4	対象・属性
5	健康課	34.4	組織・施設	義務教育学校	57.5	組織・施設
6	離乳食教室	30.1	事業・状況	学務課	50.6	組織・施設
7	児童生徒	27.7	対象・属性	参加者全員	50.2	対象・属性
8	支援センター	25.9	組織・施設	実施園	48.2	事業・状況
9	ファミリーサポートセンター	25.5	組織・施設	幼児食	47.5	事業・状況
10	子どもたち	25.3	対象・属性	直接施設	35.4	制度・事務
11	番窓口	24.1	制度・事務	乳幼児向け	35.0	対象・属性
12	子育て支援課	23.8	組織・施設	医療証	34.9	制度・事務

表4 子育てカテゴリの1記事当たりの平均文字数に対する重回帰分析の結果

説明変数	係数(β)	標準誤差	p値	95%信頼区間
性別(女性ダミー)	-16.44	52.36	.754	[-119.34, 86.46]
年齢	-2.13	1.36	.119	[-4.80, 0.55]
期数	13.78	9.95	.167	[-5.77, 33.32]
人口(対数)	49.71	8.36	<.001	[33.29, 66.14]

いった広報量を規定していたのは、首長の性別や年齢といった個人的属性ではなく、自治体の人口規模という構造的要因であった。人口が多く、財政規模の大きな自治体ほど子育て支援事業の総量が増加し、それに伴って広報すべき情報量も増大するという構造が支配的であり、性別で差異を検出することは困難であったと考えられる。

広報内容における特徴的な単語(男女差)

量的側面において首長属性の影響が限定的であった一方、対数尤度比を用いた特徴語分析の結果からは、広報内容の質的構成に差異が確認された。

男性首長の特徴語として、子育て支援センターや保育所などの組織・施設名が上位を占めたことは、行政サービスの主体・場所を明確に示す語彙が見られた。一方、女性首長の特徴語には、施設名に加えて、児童育成手当や医療証といった具体的な制度名や、思春期、乳幼児向けといった特定の対象属性が抽出された点は注目に値する。この対比は、男性首長の「施設や組織といった支援環境の提示」対し、女性首長は「具体的な対象とサービスの提示」と解

釈できる。

これらの差異を解釈するにあたっては、交絡因子の可能性を考慮すべきである。例えば、女性首長が選出されている自治体が、都市部や特定の規模の自治体に偏在している場合、首長の性別そのものではなく、その自治体が抱える行政課題や人口構成を反映している可能性がある。今後、首長の性別が語彙選択に与える影響を検証することが課題となる。

4 おわりに

本研究では、「マイ広報紙」の全国の広報紙を対象として、情報発信の量的側面については文字数および記事数の観点から分析した。さらに、質的側面については対数尤度比を用いたテキスト分析を行った。その結果、広報量は主として人口規模といった構造的要因によって規定される一方で、広報内容の質的構成には首長の性別に応じた差異が認められた。今後は、自治体の人口、都市区分、財政力などの構造的要因を統制した上で、首長の属性が自治体コミュニケーションに与える因果構造を解明していく必要がある。

謝辞

本研究の実施にあたり「マイ広報紙」のデータを活用させていただきました。貴重なデータを提供してくださった関係者の皆様に深く感謝申し上げます。本研究は、JST, RISTEX, JPMJRS25L2 の支援を受けたものです。

参考文献

- [1] Masateru Noguchi. 地方自治体における広報紙の現代的な位置づけ. 公共コミュニケーション研究, Vol. 8, No. 1, pp. 31–44, 2023.
- [2] H.F. Pitkin. **The Concept of Representation**. Campus (Berkeley, Calif.). University of California Press, 1967.
- [3] Raghavendra Chattopadhyay and Esther Duflo. Women as policy makers: Evidence from a randomized policy experiment in india. **Econometrica**, Vol. 72, No. 5, pp. 1409–1443, 2004.
- [4] B. Dan Wood and Richard W. Waterman. The dynamics of political control of the bureaucracy. **The American Political Science Review**, Vol. 85, No. 3, pp. 801–828, 1991.
- [5] Ted Dunning. Accurate methods for the statistics of surprise and coincidence. **Computational Linguistics**, Vol. 19, pp. 61–74, 03 1993.
- [6] John Justeson and Slava Katz. Technical terminology: Some linguistic properties and an algorithm for identification in text. **Natural Language Engineering**, Vol. 1, pp. 9–, 03 1995.
- [7] Justin Grimmer and Brandon M. Stewart. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. **Political Analysis**, Vol. 21, No. 3, p. 267–297, 2013.
- [8] Kenneth Church and William Gale. Poisson mixtures. **Natural Language Engineering**, Vol. 1, , 09 1995.