

メディアの計量的内容分析

稲増 一憲

メディア報道を実証的データに基づいて分析する内容分析は古い歴史を持つ手法であるが、コンピューター技術の発展とともに新たな広がりを見せている。本稿では、メディア研究を概観した後、実際の内容分析手順に沿って、手法の概要・留意すべき点などについて解説を行う。

キーワード：メディア，内容分析，社会調査

1. はじめに

近年、日本政治の研究分野において「メディアと政治」の関係性を問う書籍・論文が多く見られる（例えば、文献[1]～[3]）。このもっとも大きなきっかけは小泉純一郎という特異な首相の誕生であっただろう。「政治家らしくない」ライオンのたてがみのような独特の髪型やロックやオペラなどの趣味に加えて、「自民党をぶっ壊す」「聖域なき構造改革」「私に反対するものはすべて抵抗勢力」といったワンフレーズ・ポリティクスと評される短く・分かりやすい発言などは、メディアにとって格好の題材であり、メディアが政治を取り上げる機会は増加した。ことに、ワイドショー・スポーツ新聞などのこれまで政治を中心として扱ってこなかったメディアが政治を取り上げるようになったのは、大きな変化である。

一方で、小泉純一郎は風貌や発言が特異であっただけでなく、党内に有力な支持基盤を持たず世論の支持を背景に政権を運営していたという点も、それまでの多くの首相とは異なっていたが、これは小選挙区制という選挙制度によるところも大きい。中選挙区においては、同一選挙区において自民党の各派閥が候補者を当選させることが可能であり、仮に党公認が得られなくとも、派閥の応援があれば保守系無所属として当選することも比較的容易であった。しかし、党単位で戦う比例代表は言うに及ばず、自民党と民主党という二大政党への流れが進む中、小選挙区においても党執行部の公認を得なければ当選することは極めて困難である。結果として、派閥の領袖から党執行部へと権力が集中するとともに、「党のトップ」に対する国民的人

気が総選挙の勝敗を左右するようになった。つまり、メディアを通じて国民の支持を獲得することの意味が、それまで以上に大きくなったということである。小泉政権は任期満了に伴い2006年に退任したが、以降も「メディアと政治」についての関心は決して薄まってははいない。

2. メディアと政治をめぐる研究

近年大きな注目を集める「メディアと政治」の関係であるが、実証的な知見の蓄積を元に社会的な議論が行われているとは言い難い。例えば、講義で「マスメディアは中立だと思うか」と問えば、ほぼすべての学生が「中立ではない」と答え、また、学生に限らず、実際の政治や報道に携わる人々や研究者からも、マスメディアの中立性を批判する声は根強い。しかし、多くの視聴者を獲得するという経済的な理由からも権力ゲームに巻き込まれるのを避けるという理由からも、マスメディアは少なくとも客観・中立に見える必要があり、対立する複数の視点を取り上げるような報道を指向するのである[4][5]。実際、政治家やマスメディアの実務家たちを含む政治的エリート達を対象とした調査において、マスメディアは与野党、官僚、市民運動など多様な社会集団からアクセスされる存在であり、他の集団に比べて中立性が高いことが示されている[6]。

また、しばしば、孤立した有権者が受動的にメディアに接するといった想定に基づいて報道接触の効果が語られるが、実際の有権者は、自らが保有する先有態度に沿ったメッセージを選択し能動的に情報を処理する傾向を持つこと（選択的接触）、周囲に存在する対人的ネットワークとの相互作用の中でメッセージを処理することが知られている[7]。結果として、〇〇党寄りの報道への接触が有権者を〇〇党への投票に向かわせるといった説得効果が実際のデータにおいて示さ

いなます かずのり
武蔵大学 社会学部
〒176-8534 練馬区豊玉上1-26-1

れることは少なく、メディアへの接触効果は限定的なものと考えられる。一方で、メディアで多く取り上げられた政策争点を有権者が（賛否はどうか）重要と認知する議題設定効果[8]や、メディアが争点をどのような枠組みに基づいて報道するか（9.11.テロは「戦争」か「犯罪」か[9]、など）が有権者に影響を与えるフレーミング効果（文献[10]にレビュー）など、メディアの強力な効果を示した研究も70年代以降見られるようになった。ただし、これらの研究は、報道内容によって有権者の態度や行動が変化するという直接的説得効果ではなく、メディアは有権者が政治について考えるベースとなる情報環境を提供していることを示す研究といえる。メディアが連日起こるすべての政治的な出来事を伝えることは不可能であり、選択・編集に基づいて報道が製作される以上、製作者に特定の政党や候補者を応援する意図はなくとも、有権者に提供される情報環境の偏りを避けることは難しいのである（[メディアが有権者に与える影響については文献[11][12]を参照）。

さらには、特定の争点や政党に対する効果ではなく、「そもそもメディアは人々を政治に近づけるのか遠ざけるのか」という問いについても知見は定まっていない。メディアによる報道を通じて知識や関心を獲得することで有権者が政治に向かうという政治的動員効果（political mobilization）と、商業化・娯楽化された報道がむしろ人々に政治に対してネガティブな印象を与え人々に政治を忌避させる政治的倦怠効果（political malaise）という二つの対立する知見が存在しているのである（文献[13]を参照）。

3. 内容分析という手法

以上述べたように、「マスメディアは中立であるか」「有権者はマスメディアに影響されるのか」「メディアは人々を政治に近づけるのか遠ざけるのか」といった基本的な問いについても、知見が定まっているとはいえ、実証研究に基づく知見の蓄積が求められている。メディアを対象とした実証研究の手法として、欠かすことのできないものが内容分析（content analysis）と呼ばれる手法であり、上記の研究群の多くも内容分析を用いたものである。

メディアの「内容」を「分析」という観点から考えれば、対象となる文章をつぶさに読み込む、または映像・音声を見聞きすることで、傾向を把握し、推論を行うということも広義の内容分析に当たるといえ

る。そういう意味では、個人的な読書ですら内容分析に含めることは可能である。しかし、Krippendorffが、「データをもとに、それが組み込まれた文脈に関して再現可能でかつ妥当な推論を行うための調査技法である」と定義しているように[14]、社会科学の研究手法である以上、内容分析には客観性や再現可能性が求められるのである。結果として、Berelsonの「表明されたコミュニケーション内容の客観的・体系的・数量的記述のための調査技術」という初期の定義にも表れているように[15]、客観的に記述するためには何らかの数量的な記述を必要とする場合が多い。実際、18世紀にはすでに、『ジオンの歌』と呼ばれる讚美歌集の中の宗教的シンボルの数をスウェーデン国教会歌集と比較するという形で、数量的な内容分析が行われている[16]。

なお、質的内容分析と呼ばれる数量的な記述を伴わずに報道を比較・分析する手法も確立されているが、本稿では数量的な内容分析のみを取り上げる。さらに、数量的な内容分析においては、分析結果の再現可能性を求める以上、なるべく研究者の主観を排除した分析が求められ、また、系統的なサンプリングに基づく大量のデータを扱うことは、知見の一般化可能性を担保する一つの要因となる。これらの目的を達成する上で、内容分析におけるコンピューターの利用が急速に広まっており、本稿ではコンピューターを用いた手法を中心に扱う（内容分析におけるコンピューターの利用については文献[17][18]などを参照）。以降、実際の手順に沿って内容分析について解説していきたい。

3.1 データの取得

内容分析は歴史の古いメディア分析の手法であるが、近年の状況の変化の一つとして、電子化されたデータベースの整備が挙げられる。ことに、テキスト化されたデータについては、コンピューターを用いたコーディング・分析とともに、大量のデータを分析する手間を大きく削減するものである。

電子化された新聞記事の収集については、読売新聞の「ヨミダス文書館」、朝日新聞の「聞蔵Ⅱビジュアル」、日経4紙の「日経テレコン21」など、新聞社ごとに検索サービスが利用可能である。また、新聞社をまたぐ横断的な記事検索を行うことのできるデータベースとして、ニフティ株式会社が提供する新聞・雑誌記事データベース（<http://www.nifty.com/RXCN/>）や、ビジネス情報サイト「日経 goo」の新聞記事検索データベース（http://nikkei.goo.ne.jp/nkg/fnews_

detail_top.jsp) の二つを挙げる事ができる。また、映像情報であるテレビ報道を分析するには、現状では目視に頼らざるを得ず、膨大な手間を必要とするが、株式会社プロジェクト・株式会社エムデータが提供する TV メタデータは、報道内容がテキスト形式でまとめられており、コンピューターを用いた分析の適用が容易となっている。さらには、インターネット上に日々蓄積されるブログや twitter などのデータを分析対象としたいと考えるのは自然なことであろうが、これらウェブ上のデータを収集するためには、Tiny Web Crawler, Tiny Tweet Crawler などといったフリーソフトウェアが利用可能である (<http://mtmr.jp/index.html>)。これらのデータベース・ソフトウェア上では「選挙 and (年金 or 消費税)」といったクエリ(抽出条件文)を設定することで、条件に沿ったデータをダウンロードすることが可能である。

3.2 コーディング

テキスト形式のデータを一定のルールに基づくカテゴリー化を通じて数量的データに変換することをコーディング(coding)と呼ぶ。コーディングに際しては、まず、文章を単語単位もしくは形態素と呼ばれる日本語の最小単位に分割する。例えば、「単語の区切りが明確でない日本語は、コンピューターによる分析が難しい」という文章は、「単語/の/区切り/が/明確/でない/日本語/は/, /コンピューター/による/分析/が/難しい。」と分割される。ただし、言語学に基づく分析(例えば、文献[19]など)を除けば、「の」「が」といった助詞など、単独では意味を持たない語が分析に使用されることはまずなく、内容分析において用いられるのは名詞や動詞、形容詞などであろう。表1は2011年1月の新聞報道から抽出した文章であり、表2は政治の分野において重要と考えられる語について、対象となる語が含まれていれば1、含まれていなければ0という数字をコードした結果である。一度、表2のような形にデータが変換されてしまえば、通常の量的データと同様に統計的手法を用いた分析が適用可能

表1 変換前のテキストデータ

文章1	TPPへの参加と消費税引き上げに取り組む
文章2	消費税増税は解散後が筋である
文章3	解散はまったく考えていない

表2 変換後の数量データ

	TPP	消費税	増税	解散
文章1	1	1	0	0
文章2	0	1	1	1
文章3	0	0	0	1

となる。

ただし、コーディングは自動的に一意に決まるとは限らない点に注意が必要である。例えば、文章1に含まれる「引き上げ」と文章2に含まれる「増税」という語は、異なる語であるが、どちらも消費税という語にかかる語であり、この文脈における意味は、ほぼ同じである。従って、表2においては文章1の増税というセルには0という値が入力されているが、「引き上げ」という語を「増税」という語と同一とみなした場合、1という値が入力される。また、報道の分析において、「自由民主党」と「自民党」という語を別々の語として扱うことの意味は小さいだろう。このように、表記の異なる語を同一のものとして扱うかどうかという規則をコーディングルールと呼び、コンピューターによってコーディングを行うソフトウェアの多くは、同義語・類義語等のルールを設定する辞書機能を備えている。コーディングに際して注意すべき点としては、扱うテキストデータの量が多くなると、数千から数万種類の語が含まれている場合も珍しくないということがある。数値化されたデータの列の数はデータに登場する語の種類に等しいため、列の数が膨大になってしまうのである。数千、数万の変数を扱いながら分析を進めることが困難であることはいままでもないだろう。また、実際に分析に際しては、「ある」「する」「人」といった一般的な語であっても、出現頻度はせいぜい10%程度であって、分析において重要な意味を持つ語の多くは、登場する頻度が決して高くないということに気づくであろう。例えば、ある語が含まれる文章の割合が1%であれば、1という値が入力されるセルは全体の1%であり、残りの99%のセルには0が入力されるということである。結果として行列がスパース(ほとんどのセルの値が0)になってしまい、統計的な分析を用いることが難しくなってしまうという問題もある。これらの問題に対処するためには、コーディングルールを作成する際に、語の意味が変わらない範囲でなるべく多くの類義語をまとめるようにすることが必要になる。また、現在の自然言語処理の技術によれば、単語以上の単位、例えば、主語と述語といった係り受けを一つの単位とすることも十分可能であるが、その場合にはさらに表現のバリエーションが増し、行列がよりスパースになりやすいという点に注意する必要がある。

3.3 コーディング後の分析

内容分析の目的は大きく二つに分けることが可能で

ある。一つは研究者の仮説をデータに基づいて検証するという目的、もう一つは探索的にデータを解析することで新たな発見を得るという目的である。前者は Dictionary-based アプローチ、後者は Correlational アプローチとも呼ばれている[20][21]。また、仮説検証ではなく探索型のテキストデータ解析手法はテキストマイニングと呼ばれており、探索型のアプローチはテキストマイニングのメディア報道への適用例とみなすことも可能であろう。

前者の仮説検証型は、同一のコーディングルールを用いて、異なるデータの数量化を行い、文章間の単語の出現率を比較する、もしくは、分析対象となる語が、特定の単語と同一の文章内に共起する割合の違いを検討するといった手法によって行われる。語の出現頻度や共起頻度の偏りを調べ、それが統計的に有意なものであるかを検討するのである。また、後述の通り、報道内容のデータだけでなく、外部データを合わせて用いることで、仮説が検証されることも多い。

後者の探索的データ解析においては、新たな発見を得るのはあくまで分析者自身であるが、膨大なローデータから法則性を読み取ることは人間の能力では不可能であり、コーディングによって得られた行列を対象にデータを可視化する分析手法が用いられる。具体的には、主成分分析・数量化三類といった統計解析、および決定木・ナイーブベイズ分類器・自己組織化モデルといったデータマイニングで用いられる手法が一般的である[22]~[24]。

3.4 外部データの利用

政治とメディアを扱った研究、特に仮説検証型研究の多くは、メディアへの接触が有権者にもたらす影響に関心がある。しかし、内容分析単独で明らかになるのは、あくまでメディアの内容がどのようなものであったかという点だけであり、有権者にもたらす影響を扱うことはできない。従って、内容分析と社会調査などから得られた外部データを組み合わせることが必要となるのである。

外部データは大きく分けて、新聞社の世論調査などから得られる内閣や政党の支持率といったマクロデータ、および、社会調査によって得られる回答者ごとのマイクロデータの2種類が存在する。

マクロデータの利用例としては、内容分析の結果から支持率などを予測するシミュレーションモデルを構築し、モデルによる予測値と実際の時系列データの比較を行うといった手法を挙げることができる[25]。

次にマイクロデータ利用例を2例紹介したい。社会調査から得られた回答者ごとの番組視聴の有無を表す行列と、内容分析から得られた番組ごとの分析対象となる語の累計登場数を表す行列の積を取ることで、回答者ごとの分析対象となる語への接触量を表すデータを作成することが可能である。これにより、例えば、有権者の2005年衆院選時における「刺客候補」に対する報道への接触量の影響を検討するといった分析が可能になる。

$$\begin{bmatrix} \text{番組 A 視聴}_1 & \text{番組 B 視聴}_1 & \text{番組 C 視聴}_1 \\ \vdots & \vdots & \vdots \\ \text{番組 A 視聴}_i & \text{番組 B 視聴}_i & \text{番組 C 視聴}_i \\ \hline \text{単語登場数}_1(A) & \cdots & \text{単語登場数}_j(A) \\ \text{単語登場数}_1(B) & \cdots & \text{単語登場数}_j(B) \\ \text{単語登場数}_1(C) & \cdots & \text{単語登場数}_j(C) \end{bmatrix}$$

i (標本番号) = 1, ..., n ; j (単語) = 1, ..., m

2005年衆院選時の報道を分析した結果、「刺客候補」に対する報道への接触は、「刺客報道」が小泉自民党を勝たせたといった巷説に反して、自民党への投票に対する影響を持っておらず、政治全体に対する関心を高めるのみであったといった知見が明らかになっている[26]。

また、単独の番組や語についての接触効果の検証ではなく、主成分分析・クラスター分析などのデータ縮約手法を用いて、番組や単語の分類を行い、その結果を投票や個人ごとの政治関心といったマイクロデータの統計的分析と結びつける研究も実施可能である。例えば、クラスター分析を用いて報道関連番組を娯楽的な要素を含むソフトニュースとハードニュース[27]に分類し、それぞれへの接触がもたらす影響を検討するといった分析が可能となる。その結果、娯楽的なソフトニュースは、これまで政治知識を持ってこなかった人々の選挙への関心を高める補完的メディアとしての役割を担っていることが示唆された[13]。

このように外部データを組み合わせることで、報道内容の特徴の把握を超えて、メディア接触がもたらす影響を検討することが可能になるのである。

4. 展望

インターネットの普及は内容分析の対象となるデータを爆発的に増加させた。しかし、それに伴いインターネット上のデータを用いた内容分析研究から注目すべき知見が数多く得られているとは言いがたい。これには、マスメディアの報道と異なり、テキストデータ

と組み合わせることが可能な外部データの取得が困難なこと、単語や表現の多様性が非常に高くコーディングが難しいことが挙げられるだろう。しかし、インターネット上のブログ記事などから民意を予測するような分析サービスの提供は、可能性を秘めている (<http://shuugi.in/>)。世論調査を通じて有権者の民意が測定されることは、民意を政策に反映する上で重要であるが、世論調査の乱発は調査の質の低下を招く危険性を併せ持つため、将来的にこれを一部代替しうる可能性を秘めた分析に対する期待は大きい。

また、今回取り上げた研究は、テキスト化されたデータを対象としたものであったが、画像解析の技術が進歩し、映像を直接コンピューターを用いたコーディングの対象とすることが可能になれば、扱うことのできるデータの範囲は大きく拡大すると考えられる。

参考文献

- [1] 草野厚：テレビは政治を動かすか (NTT 出版, 2006).
- [2] 星浩・逢坂巖：テレビ政治—国会報道からテレビタックルまで (朝日新聞社, 2006).
- [3] ポプキン, サミュエル・蒲島郁夫・谷口将紀：メディアが変える政治 (東京大学出版会, 2008).
- [4] D. McQuail: Mass media in the public interest. In J. Curran, and M. Gurevitch (Eds.): *Mass media and society*. (Edward Arnold, New York, 68-81, 1991).
- [5] T. Cook: *Governing with the news: The news media as a political institution* (University of Chicago Press, Chicago, 1998).
- [6] 蒲島郁夫：マスメディアと政治, *レヴアイアサン*, 7 (1990), 7-29.
- [7] T. Klapper: *The effects of mass communication*. (Glencoe, Ill, Free Press, 1960) (NHK 放送学研究室 (訳): マス・コミュニケーションの効果 NHK 放送出版協会, 1966).
- [8] M. McCombs and L. Shaw: The agenda-setting function of mass media *Public Opinion Quarterly*, 36 (1972) 176-187.
- [9] A. Edy and C. Meirick: Wanted, dead or alive—Media frame adoption, and support for the war in Afghanistan, *Journal of Communication*, 57 (2007), 119-141.
- [10] D. Chong and N. Druckman: Framing theory, *Annual Review of Political science*, 10 (2007), 103-126.
- [11] 蒲島郁夫・竹下俊郎・芹川洋一：メディアと政治 (改訂版) (有斐閣, 2010).
- [12] 池田謙一・唐沢穰・工藤恵理子・村本由紀子：社会心理学, (有斐閣, 2010).
- [13] 稲増一憲・池田謙一：多様化するテレビ報道と、有権者の選挙への関心および政治への関与との関連：選挙報道の内容分析と大規模社会調査の融合を通して *社会心理学研究*, 25 (2009), 42-52.
- [14] K. Krippendorff: *Content analysis: an introduction to its methodology*. (Beverly Hills, Sage Publications, 1980). (三上俊治・椎野信雄・橋本良明 (訳): メッセージ分析の技法内容分析』への招待 勁草書房, 1989).
- [15] B. Berelson: *Content Analysis in Communication Research*. (New York: Free Press, 1952) (稲葉三千男・金圭煥 (訳): 内容分析, みすず書房, 1957).
- [16] K. Dovring: Quantitative Semantics in 18th Century Sweden, *The Public Opinion Quarterly*, 18 (1954-1955), 389-394.
- [17] R. Popping: *Computer-assisted text analysis* (London, Sage, 2000).
- [18] K. Krippendorff and M. Bock: *The content analysis reader* (London, Sage, 2008).
- [19] 東照二：歴代首相の言語力を診断する (研究社, 2006).
- [20] J. Stone: Thematic text analysis—New agendas for analyzing text content. In C. Robert (ed.) *Text analysis for the social sciences—Methods for drawing statistical inferences from texts and transcripts*. (Mahwah, Lawrence Erlbaum, 35-54, 1997).
- [21] 樋口耕一：テキスト型データの計量的分析—2つのアプローチの峻別と統合 理論と方法, 19 (2004), 101-115.
- [22] 那須川哲哉：テキストマイニングを使う技術/作る技術 (東京電機大学出版局, 2006).
- [23] R. Feldman and J. Sanger: *Text mining handbook advanced approaches in analyzing unstructured data*. (Cambridge, Cambridge University Press, 2007).
- [24] 松村真宏・三浦麻子：人文・社会科学のためのテキストマイニング (誠信書房, 2009).
- [25] 吉田文彦：文章解析ソフトウェア Tex-ray の概要と応用事例—小泉内閣関連新聞社説の内容分析結果による内閣支持率の予測 マスコミュニケーション研究, 64 (2006), 41-69.
- [26] 稲増一憲・池田謙一：マスメディアと小泉の選挙—メディアはコトバを与えたか、関心を高めたか 池田謙一 (編) 政治のリアリティと社会心理—平成小泉政治のダイナミクス (木鐸社, 2007).
- [27] M. Baum: *Soft news goes to war—Public opinion and American foreign policy in the new media age*. (Princeton, New Jersey: Princeton University Press, 2003).