

新聞社説論調の文章解析ソフトによる測定結果と短命内閣における支持率との連動性

吉田文彦[†]

森, 安倍, 福田, 麻生, 鳩山の5つの内閣を対象に, 4 全国紙の内閣関連社説内のポジティブな意味の語とネガティブな意味の語を数え時系列で論調を指数化したところ, どの内閣でも支持率と有意な相関があった.

Synchronized Movements of Short-Lived Cabinets' Approval Rates and Media Evaluations as Assessed by Text Analysis Software

Fumihiko Yoshida[†]

Analyzing editorials of four major newspapers on five cabinets in recent years, Mori, Abe, Fukuda, Aso, and Hatoyama cabinets, by morpheme analysis software, and identifying words with positive and negative denotation, five time-series indices were obtained regarding media evaluations of these five cabinets. This study demonstrates that “media evaluations of a given cabinet” assessed by this method is significantly correlated with “cabinet approval rates” obtained by opinion polls, suggesting a possibility that cabinet approval rates may be estimated from media evaluations of the cabinet.

1. はじめに

2000年4月から2010年6月に至るおよそ10年間に日本では7人の総理大臣が誕生している。在任期間が5年半近くに及んだ小泉内閣と本稿執筆時点での現職の菅内閣を除く5つの内閣はいずれも在任期間が1年前後の短命内閣であった。本稿では短命内閣が続く理由の一つは内閣に関するマスメディアの論調と内閣支持率という形で示される世論との連動性の強さにあるのではないかという仮説に立ち, まずは両者の相関関係を分析し連動性の強さを実証することをひとつの目的とする。この連動性は場合によってはマスメディアの論調が一般市民に影響を与えることから生じようが, マスメディア側も市民一般も考えることに大差はない場合が多いということからも説明できよう。もうひとつの目的は, 本稿で報告するような手法が社会状況の把握および分析のツールとなり得ることを実際の分析を通して示すとともに, 分析手法に関しさらなる改善を目指すには具体的にどのような必要があるかを詳しく論じることである。

2. データ生成と分析手法

当研究で用いた研究手法の概略は次のようなものである。研究対象として取り上げた5つの内閣に関して, それぞれ2つのタイプのデータを収集・作成した。ひとつは就任期間中に実施された世論調査から得られた内閣支持率のデータである。もうひとつは, それぞれの内閣に関するマスメディアの論調を示すデータである。どの内閣に関しても内閣支持率のデータは読売新聞社によるものを用いた。同社による調査結果を用いたのは, 比較的調査回数が多く, 近年のものはそのほとんどの結果がウェブ上でまとめて公開されていることによる。

もう一方のマスメディアの論調に関するデータはかなり複雑なプロセスを経て生成したものである。詳しくはこのあと順を追って説明するが, 大まかにいえば, 4つの全国紙の社説のうち, 時の内閣に触れたものをすべて選び出して, 首相に言及している文章だけを対象に文章解析ソフトウェアで解析し, その結果を用いて内閣に関するマスメディアの論調を示す「PN 指数」を在任期間中の個別の一日ごとに数値化した。マスメディアの論調と内閣支持率との関係は, 両者の時系列グラフの検討と回帰分析や相関分析などの一般的な統計手法を用いて分析した。以下では「マスメディアの論調」の数値化のプロセスを説明する。

マスメディアの論調の測定対象としては4つの新聞を選択した。また, 解析対象記

[†] 東海大学文学部広報メディア学科
Department of Media Studies, Faculty of Letters, Tokai University

事としては、各紙の社説を選んだ。実際に使用したのは、朝日新聞、日本経済新聞、毎日新聞、読売新聞の4紙である。これらは発行部数の多い主要全国紙であるとともに、データベース経由の記事内容の入手も容易であり、さらには一般的な日本国民の政治的な立場と極端に乖離していない立場に立つ新聞社であると考えられる。

分析対象として一般の記事ではなく社説を選んだのは、事実関係の報道ではなく、内閣の働きぶりに関するマスメディアの評価を本稿では分析の対象としているためである。いずれの内閣に関しても在任期間中に上記4紙に掲載された全社説のうち当該内閣に言及した全ての社説をまず選びだした。さらに、各社説の全ての文章から首相本人の名前あるいは肩書きで言及されている文章のみを選び出した。もちろんその際には、「歴代首相」や「首相公選制」などのように、現職の首相を指し示さない言葉を含んだ文章は除外した。なお本稿で分析対象とした社説および文章の数は内閣別に表1に示されている。

表1 分析した社説と文章の数

内閣	社説数	文章数
森喜朗内閣	562	2048
安倍晋三内閣	776	3110
福田康夫内閣	665	2265
麻生太郎内閣	598	2240
鳩山由紀夫内閣	630	2854
合計	3231	12517

選択した文章の解析には TeX-Ray という文章解析ソフトウェアを使用した。これは筆者が開発したソフトウェアであり、形態素解析、構文解析、語尾のモダリティ分析、価値評価語分析、語彙分析などが行える[a]。本研究では、TeX-Ray の各種機能のうち、主として形態素解析と価値評価語分析の機能を用いた。価値評価語分析は識別された形態素のうち、名詞、動詞、形容詞、副詞を対象に、「良い⇔悪い」および「友好的⇔敵対的」の二つの次元を用いてポジティブ語(P語)とネガティブ語(N語)の識別を行うものである。TeX-Ray にはおよそ2万語のP語とN語を収録した PND(Positive Negative Dictionary)と呼ばれるデータベースがあり、P語とN語の識別に使用される。

マスメディアの論調の数値化には、一定期間のP語とN語の出現数から次式によって得られる「PN指数」を用いた[b]：

$$PN \text{ 指数} = P / (P + N) - N / (P + N)$$

PN指数はすべての価値評価語がP語のみなら+1.0、逆にすべての価値評価語がN語のみなら-1.0となり、P語とN語の出現数が同数なら0.0の値をとる。なお本研究では、否定されているP語やN語はそれらの出現数のカウントからは除外されている。

不特定時期に実施される世論調査の結果とマスメディアの論調との関係を探るうえで、任期中のどの時点であれ、内閣に関するその日その日のマスメディアの論調がデータ化されていることが望ましい。そうであれば、さまざまな時点で実施された世論調査の結果(内閣支持率)とともに、世論調査が行われた時点での内閣に関するマスメディアの論調のデータも存在することになる。

そこで、本研究では、どの内閣の場合も首相の在任期間中の全ての個別の一日に関して、その日を含めて過去21日間(3週間)のマスメディアの論調のデータを集計し、その期間中のPN指数を算出し、それをその日のマスメディアの論調の数値とする方式をとった。

当然ながら、なぜ21日なのかという疑問が生じようが、これは一般に新聞社による内閣支持率を含む政治問題に関する世論調査はほぼ1ヶ月の間隔で実施されていることを勘案したものである。3週間という期間は前回の調査以降の時期のおよそ4分の3にあたるわけであり、その間の出来事がそれ以前のものよりもおそらく世論の変動により強く影響するのではないかという想定に基づくものである[c]。また、この方式はデータの平滑化に寄与する利点もある。

3. 分析の結果

3.1 森喜朗内閣

どの内閣の場合も分析結果の報告にあたっては、2つのグラフで結果を示すことにする。ひとつは当該内閣に関する新聞4紙の社説の分析から得たPN指数と内閣支持率を時系列の折れ線グラフの形で示したものである。もうひとつは上記PN指数と内閣支持率の関係を散布図の形で表示したものである。まずは図1に示されている結果を見ることにする。おおまかにみて、破線で示されている内閣支持率は、実線で示されている新聞4紙のPN指数の動きとかなり連動しているように見える。実際、両者

a 詳しくは文献[1]および[2]を参照されたい。

b 極めて単純な指数なので多くの先行使用例があるはずだが、現時点では見つからない。

c 森政権のデータを用いて調べてみたところ、調査実施日も含め過去7日間のマスメディアの論調データを用いた場合、内閣支持率との相関が最大になることがわかった。これは世論形成に関する記憶モデルの検証という側面にも関わってきて興味深いテーマであるが、本研究ではごく限られた範囲内での実験しか行っていない。

の相関は図2に示されているようにかかなり高い ($r=0.706, n=11, p<0.05$).

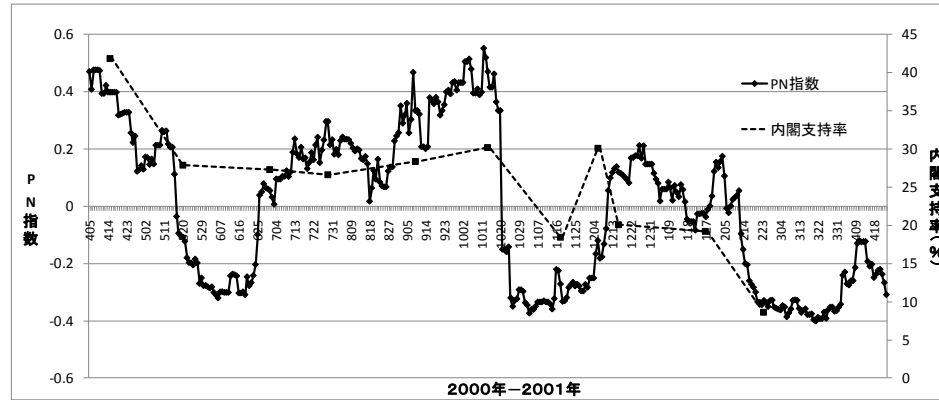


図1 森内閣のPN指数と内閣支持率の推移

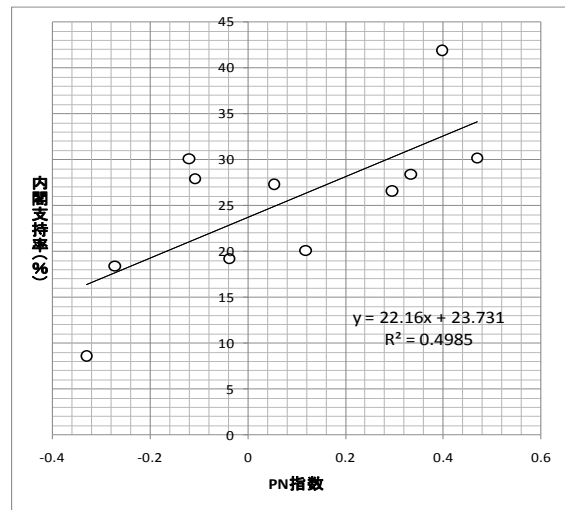


図2 森内閣のPN指数と内閣支持率の関係

3.2 安倍晋三内閣

図3の任期中の支持率とPN指数の時系列のグラフであるが、既に見た森内閣の場合と同様に、支持率とPN指数はほぼ同様のトレンドを示しており、両者間にかかなり密な関係性が存在することを伺わせる。そのことは図4の散布図を見れば明らかであり、両者の間には強い相関関係が存在する ($r=0.843, n=18, p<0.001$).

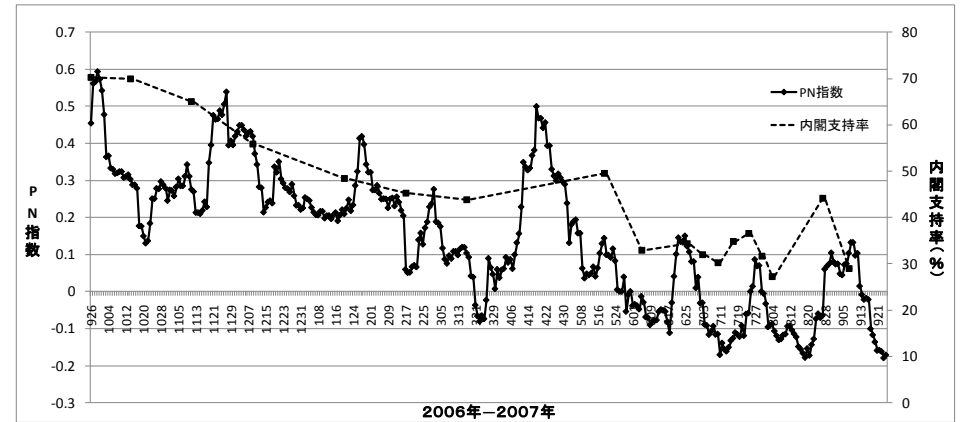


図3 安倍内閣のPN指数と内閣支持率の推移

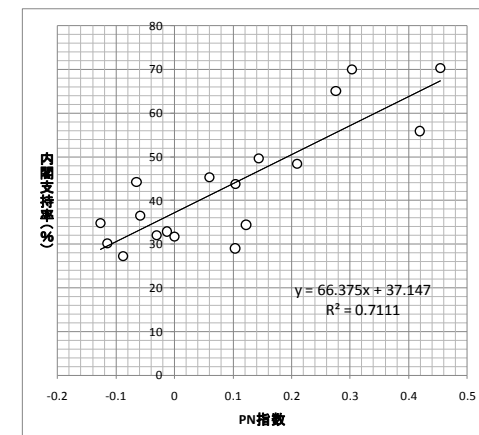


図4 安倍内閣のPN指数と内閣支持率の関係

3.3 福田康夫内閣

図6に示されているように、安倍政権の場合に比べると相関係数はやや低くなっており、有意レベルにも違いが見られる ($r=0.685$, $n=13$, $p<0.01$). とはいえ、政権末期に行われた内閣改造で一時的に内閣支持率が上昇したことや最終的な内閣支持率とともに30%を切るあたりで終わっていることも含め、全般的な傾向はほぼ類似しているといえよう.

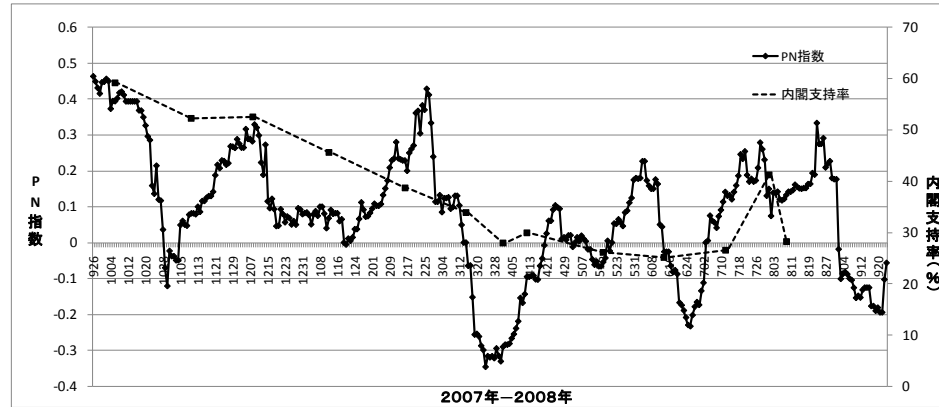


図5 福田内閣のPN指数と内閣支持率の推移

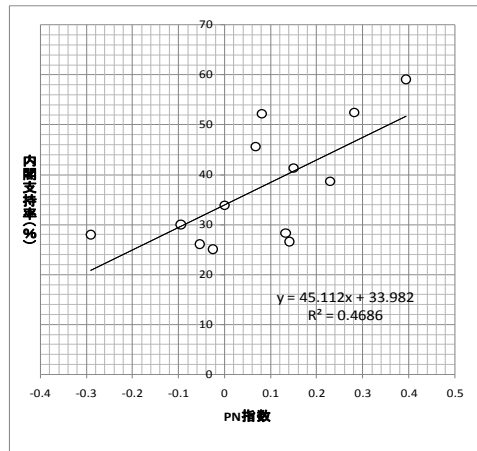


図6 福田内閣のPN指数と内閣支持率の関係

3.4 麻生太郎内閣

支持率の動きのパターンとしては内閣改造によって支持率の一時的な上昇を見た森内閣の場合に近いともいえるが、図7に見られるように、内閣支持率の動きと新聞4紙のPN指数の動きにかなりの連動性が見られるという点では、これまで見てきた3内閣の場合と共通性が見られることは確かである. 図8のPN指数と内閣支持率の散布図によれば、両者の間には比較的強い正の相関関係が存在することがわかる ($r=0.744$, $n=21$, $p<0.001$).

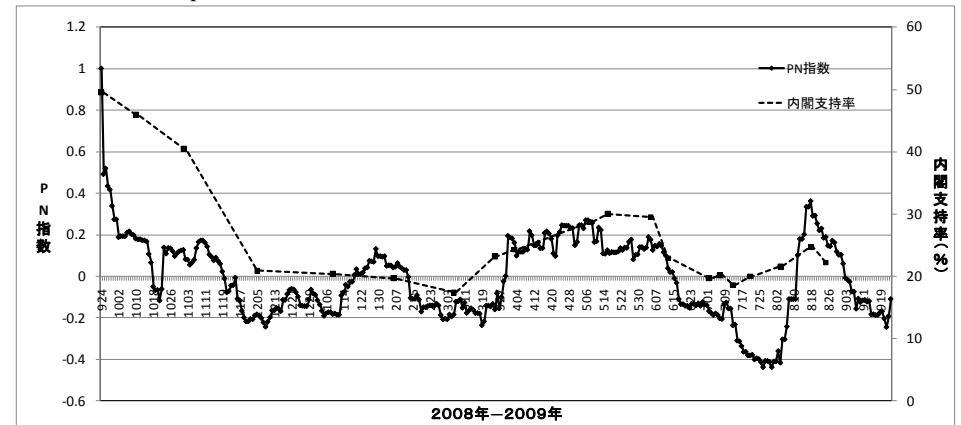


図7 麻生内閣のPN指数と内閣支持率の推移

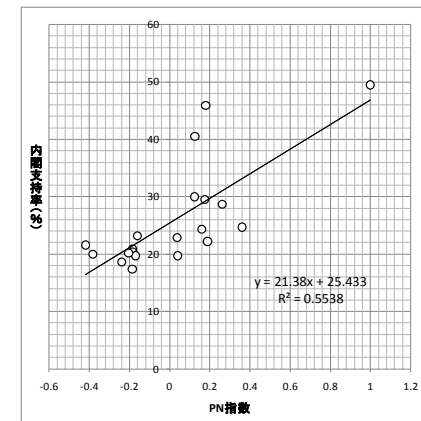


図8 麻生内閣のPN指数と内閣支持率の関係

3.5 鳩山由紀夫内閣

図9では内閣支持率を示す破線とPN指数をあらわす実線とがほぼ並行するような形で示されており、両者の間にはかなり強い正の相関が存在することが予想される。図10は鳩山内閣に関するPN指数と内閣支持率との関係を散布図の形で表しているが、両者の間にはたしかにかなり強い正の相関関係が存在する ($r=0.776, n=12, p<0.01$)。

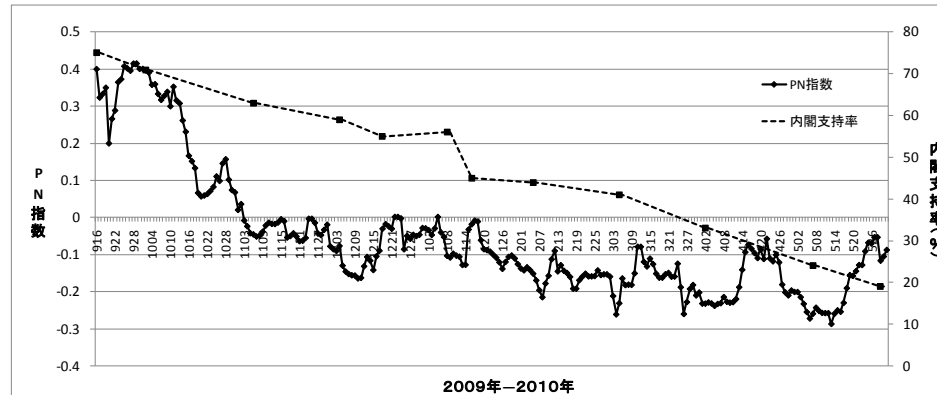


図9 鳩山内閣のPN指数と内閣支持率の推移

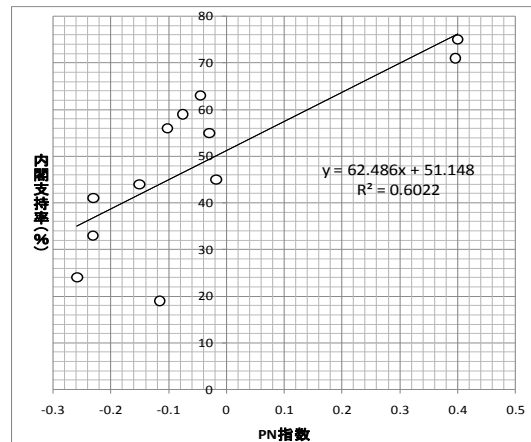


図10 鳩山内閣のPN指数と内閣支持率の関係

4. 考察とまとめ

分析の結果が示すように、5つの事例のいずれにおいても4つの新聞の社説から得られたPN指数の値と内閣支持率の相関係数は統計的に有意であった。しかも、新聞社説のPN指数は内閣支持率の調査日に先立つ21日間の期間中に掲載されたものから得られており、このことは内閣支持率の予測の可能性を示唆するものである。ただし、本稿で対象とした内閣は、過去10年間の5つの短命内閣のみであり、この結果は必ずしも一般化できるものではない。

ひとつには、上記の2変数間の相関が短命内閣特有の特徴であるならば、当然すべての内閣に関し一般化することはできない。さらには、他の内閣においてもほぼ同様の状況がある程度存在するとしても、両者間の相関が高くない場合もある。はじめの点に関しては、著者の事例研究では竹下内閣と橋本内閣に関しては、有意な相関が見つかっており、新聞社説のPN指数と内閣支持率の間での統計的に有意な相関の存在は必ずしも短命内閣に限られた現象であるとはいえない[d]。しかし、同じく著者の小泉内閣に関する事例研究では、両者間に有意な相関は見つかっておらず、新聞社説のPN指数のみから内閣支持率を予測することはかなり困難なことのようである[e]。

しかしながら、本稿で用いた新聞社説の論調の数値化の手法に関する限りはさらに改善の余地があり、長い目で見れば、広い意味でのマスメディアの論調からの世論の予測という研究分野の進展に寄与できるのではないかとと思われる。そこで、本稿の最後に、分析手法の更なる改善へ向けての指針を論じてみたいと思う。

本稿で用いた分析手法に関する反省点は大きく分けて2つある。そのひとつは、文章解析の基本手法に関するものである。本稿では、「首相」という語、あるいは首相の個人名が含まれる文章を解析対象文として選択した。そして、それらの文章に含まれるP語とN語を識別し、一定期間内の両者の出現数を基に、論調の指数化を行ったわけである。このような方式を採用したことによりいくつかの問題が起きてくる。

そのひとつは、「首相」という語が文中に含まれている限り、首相が行った行為も、首相に対して行われた行為も区別されずそれらはひとまとめに分析されてしまうということである。首相に対する不信任決議が提出されたという文章も、首相の発言や行動を示すその他の文章と同じように扱われるわけである。

もうひとつの問題は、首相その人を批判する文章で用いられる批判的な語と、政策課題などに関連して出てくるネガティブな意味合いを持つ語(たとえば、テロリズム、災害、拉致など)の両者が共にN語として扱われるため、ある時期に語としては特定のネガティブなニュアンスを持つ政策課題がしばしば文中で言及される場合には、必ずしも首相を批判する語が多用されていなくても、全体としてはネガティブな意味の

d 文献[3]参照。

e 文献[3]参照。

語の出現数が膨れ上がってしまう可能性がある。

これらの問題への対応策としては係り受け分析の結果を用いることが考えられる。首相が行動の主体なのか、あるいはその対象なのかの判定に役立つであろうし、ネガティブな意味を持つ語が首相の姿勢や言動の評価として用いられているかどうかの判定にも役立つものと思われる。こういった方向での係り受け分析の結果の利用法を今後考えていくことが必要であろう。

分析手法に関するもうひとつの反省点は、P語とN語のリスト(すなわち、PND)に関する事柄である。政治関係の社説に多く出てくる表現のひとつに「カネと政治」がある。「カネ」も「政治」も単独ではP語ともN語ともみなせないが、「カネと政治」となると事務所経費の不正経理や政治資金の虚偽申告などを総称するネガティブなニュアンスの表現となる。細かく見ていくとこういった政治の世界独自の価値判断を含む表現が新聞社説では多く用いられている可能性がある。したがって、本研究ではそれらをかかり見落としている可能性もあり、そういった表現を拾い上げデータベースに追加することで、解析精度が多少改善できる可能性もある。それとは別に、支持率の上昇や下落に先立つ期間に特に顕著に見られる表現を調べることも役立つはずである。

基本的な反省点は以上の2点であるが、さらに改善を要する点として、否定形についての対応が挙げられる。本稿で報告した分析に当たってはP語あるいはN語が否定形により否定されている場合は、P語およびN語のカウントから除外する処置をとった。これは「正しくない」が「正しい」としてカウントされたり、「誤っていない」が「誤っている」としてカウントされることを防ぐためであった。しかし、この処置は否定を示す部分がP語やN語の直後に現れるような判定しやすい場合にのみ限られており、「指導力が欠如している」とか「誠意のなさ」のように明確な否定のかたちをとらず、ある望ましいものが足りないとか欠けているといった表現には対応できていない。

また、この方式自体が論調を正確に捉える上で好ましいものであるかどうかとも問題である。たとえば、「首相の姿勢には賛成できない」という文章を例にとれば、たしかに「賛成できない」が「賛成」と記録されてしまうという問題は回避できるが、文章本来の批判的な論調が無視されてしまうわけである。この点に関してもさらなる工夫が必要である。

以上のような改善を重ねれば、マスメディアの論調と世論変動との関連性に関しては、より正確なデータ作成が可能となり、実態に即した議論に資するものと思われる。

参考文献

- 1) 上村龍太郎, 吉田文彦, 山下俊恵: 社会知能システム—内閣支持率は予測できるか?, 東海大学出版会(2007).
- 2) 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, 68号, 80-6(2006).
- 3) 吉田文彦: 新聞社説論調と内閣支持率の関係—竹下政権から安倍政権まで(未発表原稿).