

2000年代における教育をめぐるイシューとアクター

—国会会議録検索システム検索用APIを用いた計量分析—

比較教育社会学コース	西村幸浩
比較教育社会学コース	寺田悠希
比較教育社会学コース	鎌田健太郎
放送大学	橋本鉦市

Issues and Actors Related to Education in the 2000s in Japan: A Quantitative Analysis of the Web API
for Full-text Database System for the Minutes of the Diet

Yukihiro NISHIMURA, Yuuki TERADA, Kentaro KAMADA, Koichi HASHIMOTO

In the 2000s, the decision-making process of Japan's education policy is said to have changed significantly. This study aims to overview the issues and actors related to education in Japan in the 2000s by means of quantitative text analysis, using the Web API (Application Programming Interface) for Full-text Database System for the Minutes of the Diet. This study also aims to provide new methods and perspectives for educational research by analyzing big data using API that have been developed and opened to the public in recent years.

目次

- はじめに
- 先行研究
- データと方法
 - データ
 - 方法
- 分析結果
 - 省庁再編後の教育をめぐるイシューとその変遷
 - トピック数の確定とトピックの内容
 - トピックの変遷
 - 省庁再編後の教育をめぐるアクターとその変遷
 - どのイシューがどのアクターによって語られたか
 - 政党間比較
 - 政党内比較
- 考察
 - 分析結果に関する考察
 - 今後の課題
 - 分析手法に関する課題
 - データに関する課題
 - 今後の展望

1. はじめに

1980年代後半からのわが国の行政改革は、2000年代に入って具現的な政策課題として結実し実施・施行されることとなった。2001年には中央省庁の再編ともなつて内閣機能の強化が目指され、内閣府が企画立案、総合調整を担うこととなった。また経済財政諮問会議や総合科学技術会議などの様々な合議制機関が内閣府におかれ、重要政策事項について調査・審議を行うこととなった。各省庁には政務次官に代わつて副大臣と大臣政務官が新たに置かれ（これにより官僚が大臣に代わつて国会の各委員会で説明する政府委員制度は廃止）、従来の倍近い数の政治任用ポストが各省庁に設けられることとなった。さらに、国家公務員改革のもとで内閣人事局による幹部人事の一元管理も進められ、次官・局長レベルの人事についても内閣の「了解」から「承認」を必要とされるようになった（荻原 2002; 羽田 2019, 2021）。

こうした官邸・内閣府を中心とする政治主導が進んだことにより、これまでの政策形成・決定のあり方も大きく変容した。すなわち、各省庁が政策の創発・共鳴などを行い、自由民主党との事前審査制によってボトムアップ的に形成してきたこれまでのプロセスから、政府・自由民主党ともにトップダウン型に移行し

たと言われる。文教政策について言えば、2001年に文部省と科学技術庁が統合されて文部科学省に再編され、教育・学術政策と科学技術政策が所掌されることとなり、多様なアクターが参入してこれまで支配的であった「教育政策共同体」(教育下位政府)は大きく変容を遂げ、また自由民主党文教族も世代交代が起こり、より官邸の影響力が強まったと言われる(城山・細野編 2002; 中北 2017; 河合 2019; 青木 2021; 日本高等教育学会高等教育政策プロジェクト研究グループ編 2021など)。

では、2000年代に入ってからのごとき教育政策を取り巻く環境変化の中で、国会での教育政策の議論では、実際にどのような議員(アクター)が、どのような政策課題(イシュー)を取り上げてきたのだろうか。そこで本研究では、国会会議録検索システム検索用APIを利用して、2000年代におけるわが国の教育をめぐるイシューとアクターについて、機械学習・自然言語処理の手法によって概観し、2000年代以降の政治主導への移行期に台頭してきた議員と彼らによって取り上げられた政策課題を抽出することを目的とする。本研究は、2000年代に台頭してきた新しい世代の自由民主党のみならず他政党における文教関連議員も特定し、彼らがどのような議論を切り広げてきたかを確認するが、この作業は、今後、政治主導の下で教育政策の形成・決定プロセスを分析していく際に不可欠であり、その基礎的作業として位置づけられる。また同時に、近年開発・公開されるようになったAPIを活用してビッグデータを分析することにより、教育政策研究に新しい手法と視座を提供することも目的としている。

2. 先行研究

2000年代の教育政策を対象とした研究は枚挙にいとまがないが、国会会議録を利用してイシュー・アクターの分析に取り組んだ計量テキスト分析としては、以下のものがある¹⁾。

まず、本研究と関心と対象を共有する研究を確認する。橋本(2007)は、国会会議録を利用して文教関連委員会の政策論議を計量テキスト分析のフリーソフトであるKH Coderによって考察している。60年にわたる高等教育政策のイシューとアクターを分析して、自由民主党文教族など主要アクターを抽出し、またイシューの内容とその推移などを解明している。橋本(2014)でも、同様に国会会議録を利用して、戦後の

文教関連議員とそのネットワーク、取り上げられた議題などについて計量分析を行い、長期スパンでの趨勢分析を行っている。

次に、特定のイシューに着目した研究をみていくと、歌川(2014, 2015)は、道徳教育について、それらに言及した議員や論じられたテーマなどをトレースしている。また山田・青田(2015)は1947年から2013年末までの「英語教育」の議論を対象として、画期となる時期、特徴的な言葉、関連するアクターなどを考察している。

また計量テキスト分析を行っているわけではないが国会会議録を利用した教育政策研究として、e-Learningに関連する論点とアクターを考察した張(2018)、教育無償化のイシューを扱った渡部(2020, 2021, 2022)などがある。

本研究の課題と目的も、国会会議録を題材にイシュー・アクターの解明を行う点で、これらの研究の延長線上にあるものだが、新たにAPIを利用して大量のテキストデータを計量的に分析する点で、より包括的な分析を目指す点に特徴がある。

3. データと方法

A. データ

本研究で用いるデータは、2001年1月31日から2021年12月31日までに行われた、衆議院「文部科学委員会」と参議院「文教科学委員会」ならびにそれらの連合審査会・公聴会における、各議員の発言である。参考人として招致された者や官僚など、議員以外の発言は除いている。各議員の発言のテキストデータは「国会会議録検索システム」の検索用APIを用いて収集を行い、各議員の発言単位でデータを入手した²⁾。本データには各議員の発言のほか、発言者の氏名や所属会派、肩書き(政務三役)なども含まれている。上記の参考人等の発言および会議録情報を除いた上での合計の発言数は84,730件であった。

後述するLDAでの分析のために、各議員の発言は以下の手順によって前処理を行った。まず、発言冒頭の発言者名及び数値や記号を取り除いた上で形態素解析によって分かち書き³⁾、各単語を原形に変換し、名詞のみを抽出した⁴⁾。またその際、「こと」、「もの」、「それ」などの一般的な用語や数詞などの他、議事にかかわる形式的な用語(条、号、本案)など独自にリストアップした単語をストップワードとして取り除いている。処理の結果、実際の分析に用いるデータは、

語彙（異なり語）が5万3,865語、総単語数は377万9,361語となった。

B. 方法

委員会における各議員の発言から潜在的なトピックを抽出するために、本稿ではLatent Dirichlet Allocation (LDA: 潜在ディリクレ配分法) を用いる (Blei et al. 2003; Evans & Aceves 2016; 小田中・中井 2019; 瀧川 2019)。LDAはトピックモデルの一種であり、トピックモデルは、テキストデータを構成する各文書の背後に潜在的なトピックの存在を仮定し、そのトピックの析出を通じてテキストデータ全体の傾向を要約するモデルである。

モデルの詳細について以下で述べる。LDAでは、各文書がいくつかの潜在的なトピックから構成されていると考える。文書内の各単語は文書を構成する特定のトピックの確率分布に従って生成されると仮定する。具体的な手順は以下の通りである。

まず、文書*i*のトピック比率 θ_i はディリクレ分布に従う。

$$\theta_i \sim \text{Dirichlet}(\alpha)$$

また、トピック*k*における単語の確率分布 β_k はディリクレ分布に従う。

$$\beta_k \sim \text{Dirichlet}(\eta)$$

文書ごとのトピック比率 θ_i に基づいて、文書*i*の*j*番目の単語のトピック z_{ij} を決定する。

$$z_{ij} | \theta_i \sim \text{Multinomial}(\theta_i)$$

割り当てられたトピック z_{ij} の単語分布 $\beta_{k=z_{ij}}$ をもとに、文書*i*の*j*番目の単語 w_{ij} を決定する。

$$w_{ij} | z_{ij}, \beta_{k=z_{ij}} \sim \text{Multinomial}(\beta_{k=z_{ij}})$$

本稿の分析に即していえば、議員の各発言の背後には議題となるテーマのようなもの（トピック）が存在し、各発言はそのテーマの1つあるいは複数を参照しつつ行われると想定する。こうして抽出されたトピックには委員会において問題とされている政策課題が反映されていることが期待されるため、トピックの時系列的な推移を観察することにより、取り上げられる政策課題の内容の変遷を跡付けることが可能であると考えられる。

実際の分析は以下の手順により行なった。はじめに、前節で述べた前処理の結果 (Bag of Words) に対してLDAを実行し、トピックの抽出及び解釈を行なった。LDAの実行にあたってはPythonパッケージのGensimを用いた。なお、LDAを含むトピックモデル

ではトピック数をモデルの外部で決定する必要があるが、本稿ではPerplexityおよびCoherenceといった量的指標を用いて8トピックから13トピックが妥当であると判断し、解釈可能性を考慮した上で11トピックを最終モデルとした。次に、推定したLDAの学習済みモデルから各発言のトピックの構成比率を推定し、最も比率の大きいトピックをその発言のトピックとして割り当てた。最後に、委員会における議論の内容がどのように変化しているかを明らかにするために、年度ごとにトピックの割合を算出し、時系列比較を行った。

4. 分析結果

A. 省庁再編後の教育をめぐるイシューとその変遷

1 トピック数の確定とトピックの内容

前章で述べた通り、トピック数として妥当なのは8トピックから13トピックであった。

各トピックの解釈を行うにあたり、各トピックでの出現確率の高い単語を確認するとともに、瀧川 (2021) に挙げられたNelson (2020) を参考に、各トピックに含まれる確率の高い上位10発言を定性的に深く読み込んだ。その結果、各トピックを命名し、それを「トピック名」と呼ぶこととする。また、各発言が含まれる確率が最も高いトピックをその発言のトピックと考える。各トピック番号のトピック名、発言数とその構成比、出現確率の高い単語、各トピックに含まれる確率が最も高い発言⁵⁾ は以下の図表1のようになった。

トピック5および7は委員会進行上の定型化された発言が多く含まれており、具体的な内容ではなく委員会進行上の定型句であるとした。トピック8は、上位10発言を確認すると、そのうち5つの発言がオリンピックに関連した発言だったため「オリンピック」とし、上位10単語ではなく実際の発言を優先して解釈した。その他については出現確率の高い単語と各トピックに含まれる確率の高い上位10発言の内容がおおむね一致していた。

2 トピックの変遷

2001年以降の各年における発言のトピック構成比⁶⁾ を以下の図表2に示す。

図表2から明らかなように、トピック構成比は変化している。上述のトピック5及び7【委員会定型句】を除いて考える（以下、特に言及しない限りこれらのトピックは除いて結果を報告する）と、2003年・2019年にはトピック6【大学・高等教育】が突出して多く、2014年にはトピック0【制度・規則】が突出して多く

図表 1 トピッカー一覧

トピック番号	トピック名	発言数	構成比	上位10単語	最もそのトピックに入る確率の高い発言
0	制度・規則	12,097	14.3%	制度 法律 基準 改正 趣旨 措置 実施 検討 必要 指摘	制度、仕組みとして反映することは可能であると、可能性としてあるということでありまして、どんな結果が出るかこれはまだ分かりませんので、それはその結果を見て検討することになるかと存じます。
1	科学技術・独立行政法人	2,541	3.0%	研究 推進 教育 科学技術 整備 独立行政法人 行政 科学 機関 統合	統合後の法人に期待をしたいところでありませけれども、その長であられませ機構長等にはどのような人物がふさわしいというふうにお考えでしょうか。
2	初中等教育	9,441	11.1%	教育 子供 文部省 子供連 義務 教育 問題 制度 高校 勉強 中学校	具体的な数字がないのは仕方ないと思うんですけども、大体どれぐらいかかるかぐらいは全体像を把握されているのかなと思いますし、今後、一旦デジタル関係の投資をすつと更新費用がかかると思いますが、それから、何年後かに、恐らく、デジタルのタブレットとかは三年か四年したら買い換えなきゃいけないわけですね。そういった意味では、今後相当自治体の負担が増えると思うんですね。それをどのように文部省が考えていらっしゃるのか。特に、更新費用ですね。恐らく、子供がタブレットを使うので結構壊れやすいと思いますし、うちの小学生の子供も宿題のために重たいのを持ち歩いているんですけども、行き帰り持ち歩いて一年間使うと相当傷むと思うんですよ。大人以上に傷みが早いと思いますので、そういった更新費用を今後どうしていくのか、政府としてのお考えを聞きたいと思います。
3	諸問題・不祥事	2,932	3.5%	日本 アメリカ 問題 水準 初め 公務員 非常 女性 一番 海外	その出た時間とかどこで食べたかというのは記憶になくて、今回秘書さん、ドライバーをしていた秘書さんと話していて、確かに、向かっている途中でどこかで食べようよと言って向かっていて、どこかで入って長くおしゃべりしながら昼食を食べたと、そこまでの記憶しかございませんので、何時間分に出たとかレシートが残っているとかがそういうものはなくて、その場所含めて時間等が正確に分らないのが現状です。
4	文化・愛国心	6,067	7.2%	活動 社会 地域 体験 分野 我が国 重要 日本 文化 非常	それでは、国土に対する理解と愛情とはいかなることを意味するのか、歴史に対する理解と愛情とはいかなることを意味するのかということをお答えいただきたいと思います。
5	委員会定型句(開会)	15,831	18.7%	大臣 問題 質問 議論 答弁 先生 大変 教育 僕 お願い	全会一致と認めます。よって、斎藤さん提出の附帯決議案は全会一致をもって本委員会の決議とすることに決定いたしました。ただいまの決議に対し、萩生田文部科学大臣から発言を求められておりますので、この際、これを許します。萩生田大臣。
6	大学・高等教育	8,708	10.3%	大学 評価 法案 国立大学 学生 入学 制度 公立 振興 年度	給付対象の規模につきましては、対象者の現状等を踏まえ、制度設計の中で検討しているところであり、全学生に占める割合については、現時点で明確にお答えすることはできません。
7	委員会定型句(委員の異動・質疑)	8,789	10.4%	委員 異議 副大臣 質疑 報告 文部科学省 参考 本日 委員長 調査	委員の異動について御報告いたします。昨日までに、鰐淵洋子君、山本香苗君、市川一朗君、荻原健司君、中川義雄君、神取忍君、吉村剛太郎君、蓮舫君、広中和歌子君、鈴木寛君、佐藤公治君、武内則男君及び森田高君が委員を辞任され、その補欠として浮島とも子君、浜四津敏子君、西田昌司君、山谷えり子君、坂本由紀子君、義家弘介君、藤谷光信君、谷岡郁子君、大島九州男君、植松恵美子君、友近聡朗君、大久保潔重君及び私、関口昌一が選任されました。
8	オリンピック	5,034	5.9%	教職員 計画 配置 定数 決定 標準 生活 文部科学省 施設 連携	政府の分科会のメンバー、それから感染研の危機管理センターからも専門家が入って、組織委員会において、まず、東京二〇二〇大会における新型コロナウイルス対策のための専門家ラウンドテーブルから助言を得ておまして、ごこと、それから私どもと、それからコロナ対策室とという連携をしながら評価をしております。
9	教員	10,772	12.7%	学校 先生 教員 指導 学級 研修 現場 授業 小学校 学習	私も現場の声に常に耳を傾け、子供たちの考え、先生たちの考え、校長さんたちの考え、あるいは教育委員会、すべてのそういう関係する方々、特に現場のお声に私どもは常に耳を傾けなければこうしているんな政策は出てこないわけでありませ。
10	予算・財政	2,518	3.0%	予算 部分 地方 費 事業 財政 中身 計上 補助 基金	全額補助でしょうか。全額補助される仕組みでしょうか。
計		84,730	100%		

なっている。トピック6【大学・高等教育】に含まれる確率が高い発言を確認したところ、2003年は国立大学法人制度の導入に関する議論、2019年は大学入試センター試験への民間試験導入の議論が行われていた。同様に、2014年に突出して多くなったトピック0【制度・規則】を確認すると、これは2014年における地教行法改正に伴う教育委員会制度に関する議論で

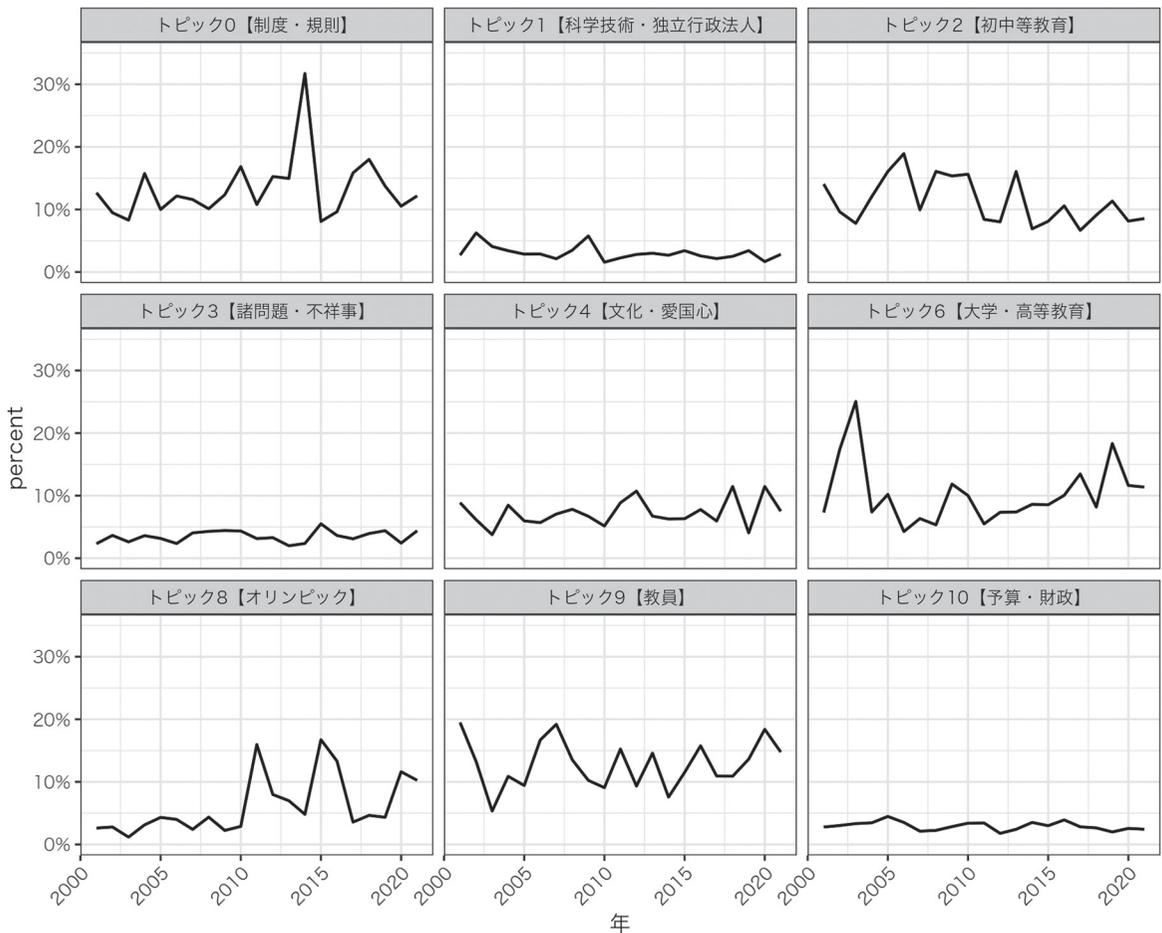
あった。また、トピック8【オリンピック】は2011年、15・16年、20・21年に構成比が高くなるといった周期的な変動が見て取れる。各年の発言内容を確認すると、2015・16・20・21年においては東京オリンピックに関連した発言が多くを占め、2011年はオリンピックに関連した発言に加え、東日本大震災に関連した発言を多く含むことが確認された。その他、トピッ

ク3【諸問題・不祥事】やトピック10【予算・財政】のように発言数が相対的に少なく、変動も大きくないトピックもある。

次に、衆議院と参議院でトピック構成比の変遷が異なるのかについて整理する(図3)。その結果、概ね両院とも年ごとのトピック構成比が似ていることが確認できる。しかし、両院で構成比に差があるトピックもある。5ポイント以上の差があった項目のうち、参

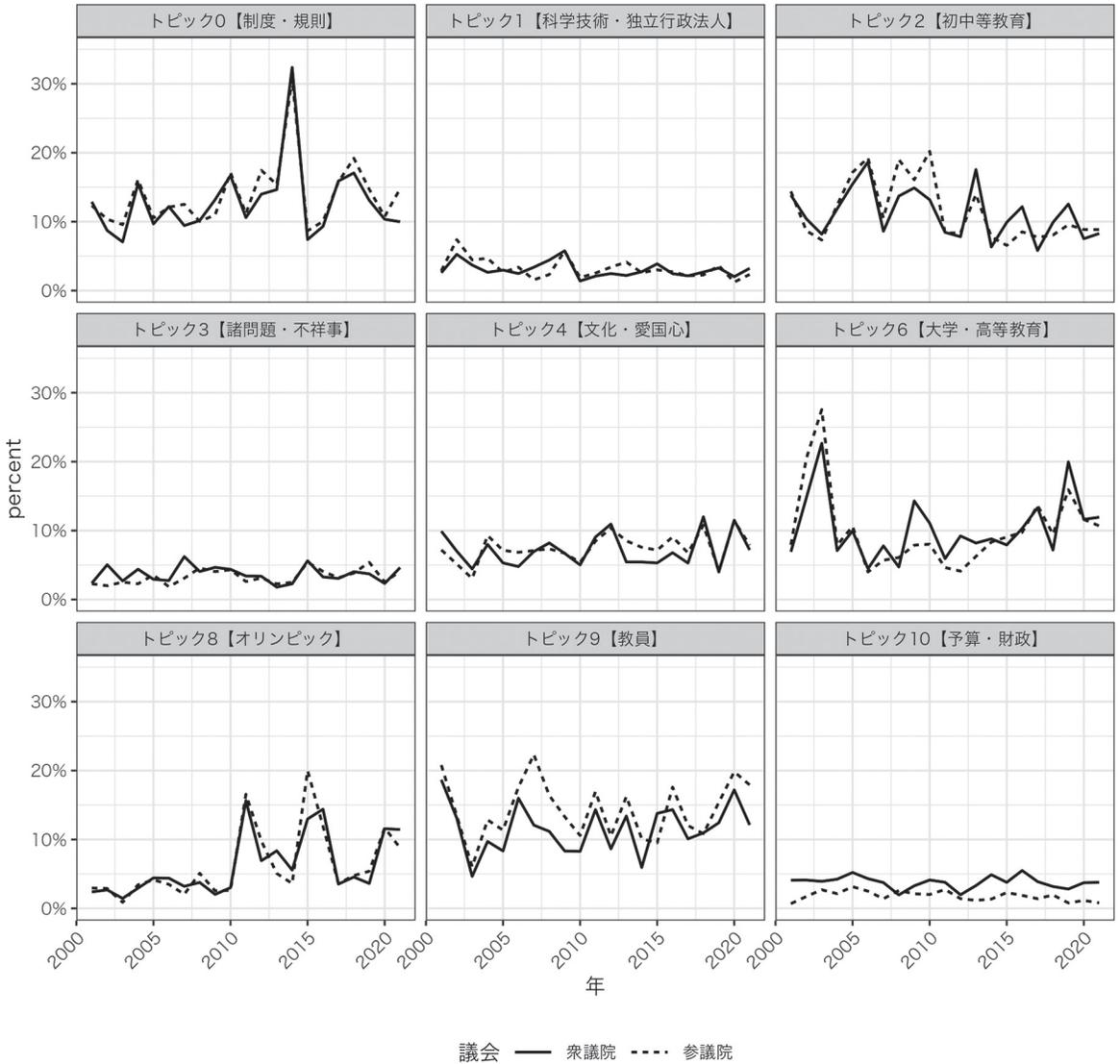
議院が衆議院よりも構成比が高いのが2008年と2010年のトピック2【初中等教育】、2002年のトピック6【大学・高等教育】、2015年のトピック8【オリンピック】、2007年と2008年のトピック9【教員】である。一方、衆議院が参議院よりも構成比が高いのは2009年と2012年のトピック6【大学・高等教育】のみである。

図表2 各発言のトピック構成比の変遷



【委員会定型句】のトピック5・7は省略

図表 3 衆議院と参議院のトピック構成比変遷の比較



【委員会定型句】のトピック5・7は省略

B. 省庁再編後の教育をめぐるアクターとその変遷

文教政策に対する議員の影響力に関する分析視角については、いくつかの方法が考えられるが、本研究では橋本 (2013) に倣い、国会の文教関連委員会での「発言量」に着目する。

前章で述べた通り、2001年から2021年までの衆議

院文部科学委員会、参議院文教科学委員会では、政治家による発言が84,730回あった。そのうち、発言回数が1,000回以上の議員を主要なアクターであると考え、主要なアクターは以下の16名となった (図表 4)。

図表 4 発言回数が1,000回以上の議員一覧

議員名	所属政党	発言回数	立場別の内訳				文部科学大臣在任期間
			文部科学大臣	文部科学副大臣・政務官	他省庁の政務三役	政務三役以外	
下村博文	自民	3,599	2,924	22	7	646	2012年12月26日～2015年10月7日
馳浩	自民※	1,951	453	242	0	1,256	2015年10月7日～2016年8月3日
萩生田光一	自民	1,887	1,835	8	19	25	2019年9月11日～2021年10月4日
河村建夫	自民※	1,810	875	702	0	233	2003年9月22日～2004年9月27日
畑野君枝	共産※	1,531	0	0	0	1,531	
石井郁子	共産※	1,447	0	0	0	1,447	
吉川元	立民	1,442	0	0	0	1,442	
笠浩史	立民	1,246	0	109	0	1,137	
松野博一	自民	1,207	904	8	0	295	2016年8月3日～2017年8月3日
宮本岳志	共産	1,135	0	0	0	1,135	
柴山昌彦	自民	1,135	1,107	0	0	28	2018年10月2日～2019年9月11日
池坊保子	公明※	1,124	0	175	0	949	
中山成彬	希望※	1,119	1,109	0	0	10	2004年9月27日～2005年10月31日
高木義明	民進※	1,107	1,084	0	0	23	2010年9月17日～2011年9月2日
鈴木寛	民主※	1,014	0	612	0	402	
平野博文	立民※	1,006	343	0	0	663	2012年1月13日～2012年10月1日

凡例：自民＝自由民主党、公明＝公明党、立民＝立憲民主党、共産＝日本共産党、民主＝民主党（当時）

民進＝民進党（当時）、希望＝希望の党（当時）

※は2022年の参議院通常選挙後に現役ではない議員が最後に所属した政党名

対象期間で最も発言回数が多かったのは下村博文（以下、敬称・身分など省略）であった。ただし、その発言回数の多くは文部科学大臣在任期間⁷⁾中の大臣としての発言である。下村博文は2001年以降で最も長く文部科学大臣を務めており、その影響で発言数が多くなっているものと思われる。その他、1,000回以上の発言回数があった議員16名のうち過半数である9名は文部科学大臣経験者であり、政務三役までを含めると12名が役職経験者であった。残りの4名は日本共産党の議員が3名、立憲民主党の議員が1名であ

り、対象期間中に与党経験のない日本共産党の議員が政務三役経験がないにも関わらず、発言回数が多いことが明らかとなった。

もちろん、大臣などの政務三役は文教政策における主要なアクターと言えるが、各委員会の質問に回答するために必然的に発言回数が多くなる。その点を考慮し、政務三役以外の立場での発言回数の上位20名の議員について、各年度における発言回数とともに以下の図表5のとおり整理した。

図表 5 政務三役以外として発言回数の多い上位20議員と発言のあった年度及びその回数

議員名	政党	総計	年度別発言数																				
			01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21
畑野君枝	共産※	1,531	143	117	154	24										86	55	118	230	263	138	203	
石井郁子	共産※	1,447	175	158	134	293	173	172	92	128	122												
吉川元	立民	1,442												101	244	188	139	170	236	265	29	70	
馳浩	自民※	1,256	21	11	16	10	24	9		14	90	265	381	150	41	37	7	10		24	77	25	44
笠浩史	立民	1,137				160	55	99	45	33	53	19			95	168	57	18	37	43	141	63	51
宮本岳志	共産	1,135									14	177	258	101	163	364	42		16				
池坊保子	公明※	949			12	418	50	85		9	54	92	153	76									
城井崇	立民	910				169	120				7	16								210	272	96	20
松沢成文	維新	817	23											11	116	119	85	103	125	94	68	73	
川内博史	立民※	807				145	170	27			16			52						113	214	46	24
斎藤嘉隆	立民	802											13	49	19	58	129	167	49	133	10	36	113
牧義夫	立民	722	32	29		110	55	63	38	46	25						57	24	38	20	88	48	49
義家弘介	自民	712							14	71	48	225	139	53	28	44	11		11	17	35	16	
田中真紀子	民主※	677						15			37	336	289										
平野博文	立民※	663	141	71	80	117	15			26						38	32	78	65				
吉良よし子	共産	659															30	97	183	173	69	107	
下村博文	自民	646									17	280	217	118								14	
小淵優子	自民	645	5		4	8		5		8					153	462							
田村智子	共産	643													67	248	224	63	23	10	8		
蓮舫	立民	639							76	98		15				101	104	28	187	30			

凡例：自民=自由民主党、公明=公明党、立民=立憲民主党、維新=日本維新の会、共産=日本共産党、民主=民主党（当時）

上位20議員のうち、日本共産党の議員が5名おり、自由民主党の4名を上回っている。このことから、与党であることが多い自由民主党の議員は、基本的には政務三役としての発言回数が多く、一般の委員としての発言回数は多くないことが見て取れる。馳浩・義家弘介の年度別発言回数を見ると、2009年から2012年にかけて自由民主党が野党だった時期には発言回数が急増していることから、自由民主党の議員は与党である時期においては一般の委員としての発言が少ないことが明白である。

C. どの 이슈がどのアクターによって語られたか

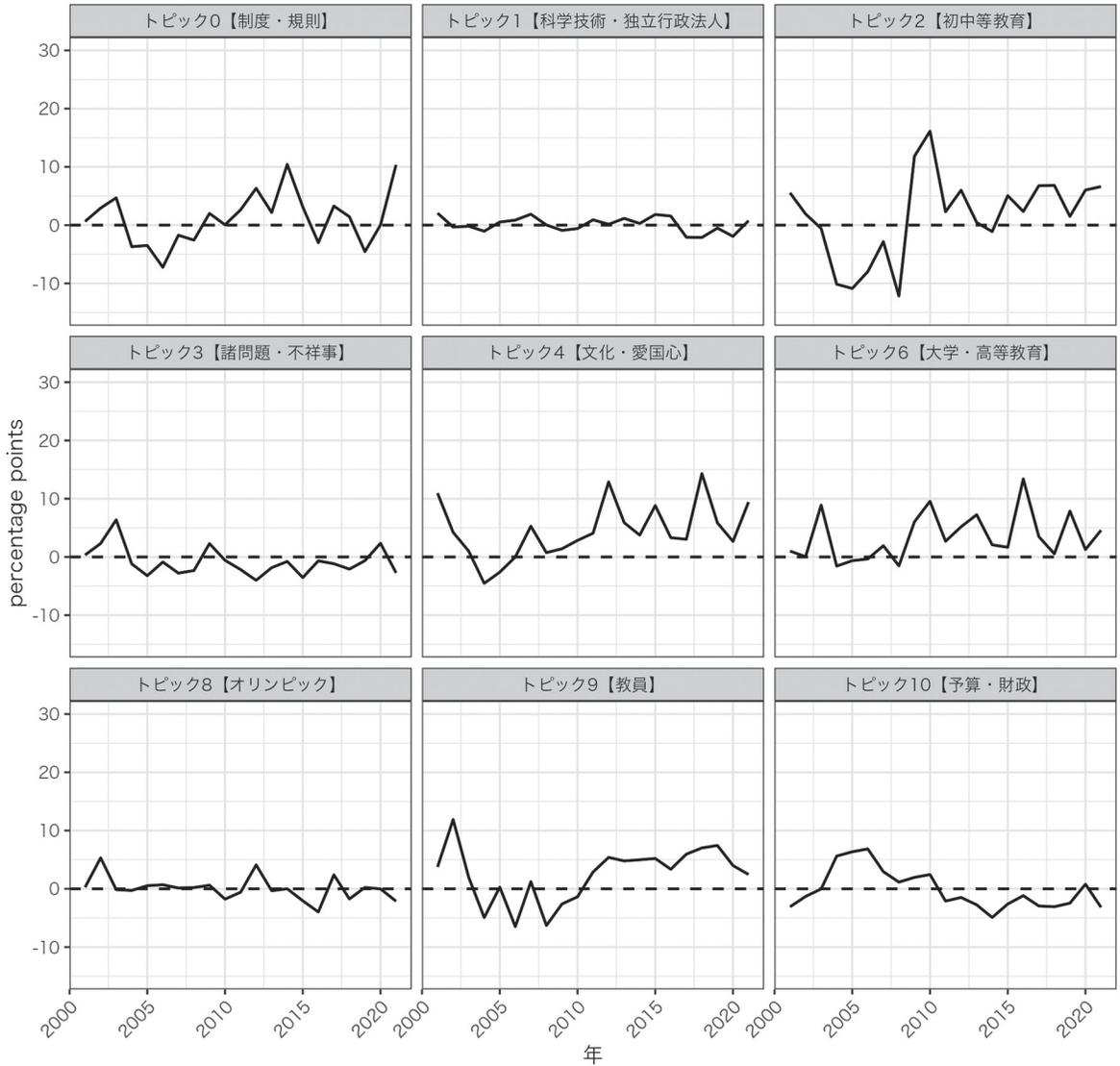
ここまで、省庁再編後の教育をめぐる 이슈およびアクターとその変遷について分析結果をまとめてきた。次に、アクターによって 이슈への関心が異なるかについて2つの観点から分析する。まず、政党間の比較を行い、政党による教育をめぐる 이슈の中での関心の濃淡を探る。次に、政党内の比較を行い、教育をめぐる 이슈の中で議員による役割分担が行われているのかについて探る。

1 政党間比較

省庁再編後において、どの政党がどういった教育をめぐる 이슈を語ってきたのか。本項では、自由民主党を含む会派と、自由民主党と連立政権を組む公明党、および対象期間中野党であり会派に変動のなかった日本共産党それぞれとの比較を行う。これら3会派を対象とするのは、2001年以降、政党名が変更されることなく2021年時点でも存在した政党のためである。ただし、自由民主党については様々な会派が結成されることがあり⁸⁾、自由民主党が含まれる会派をまとめて自由民主党として分析することとした。

自由民主党と公明党との比較結果は以下の通りである（図表6）。各年度において公明党の構成比から自由民主党の構成比を引き、差分のポイントをプロットした。トピック2【初中等教育】について2003年から2008年にかけては自由民主党のほうが構成比が高いが、2009年に逆転しその翌年まで相対的に大きな差がある。またトピック4【文化・愛国心】については公明党のほうが構成比が高い年が多いことが見て取れる。

図表6 自由民主党と公明党の年別トピック構成比の比較

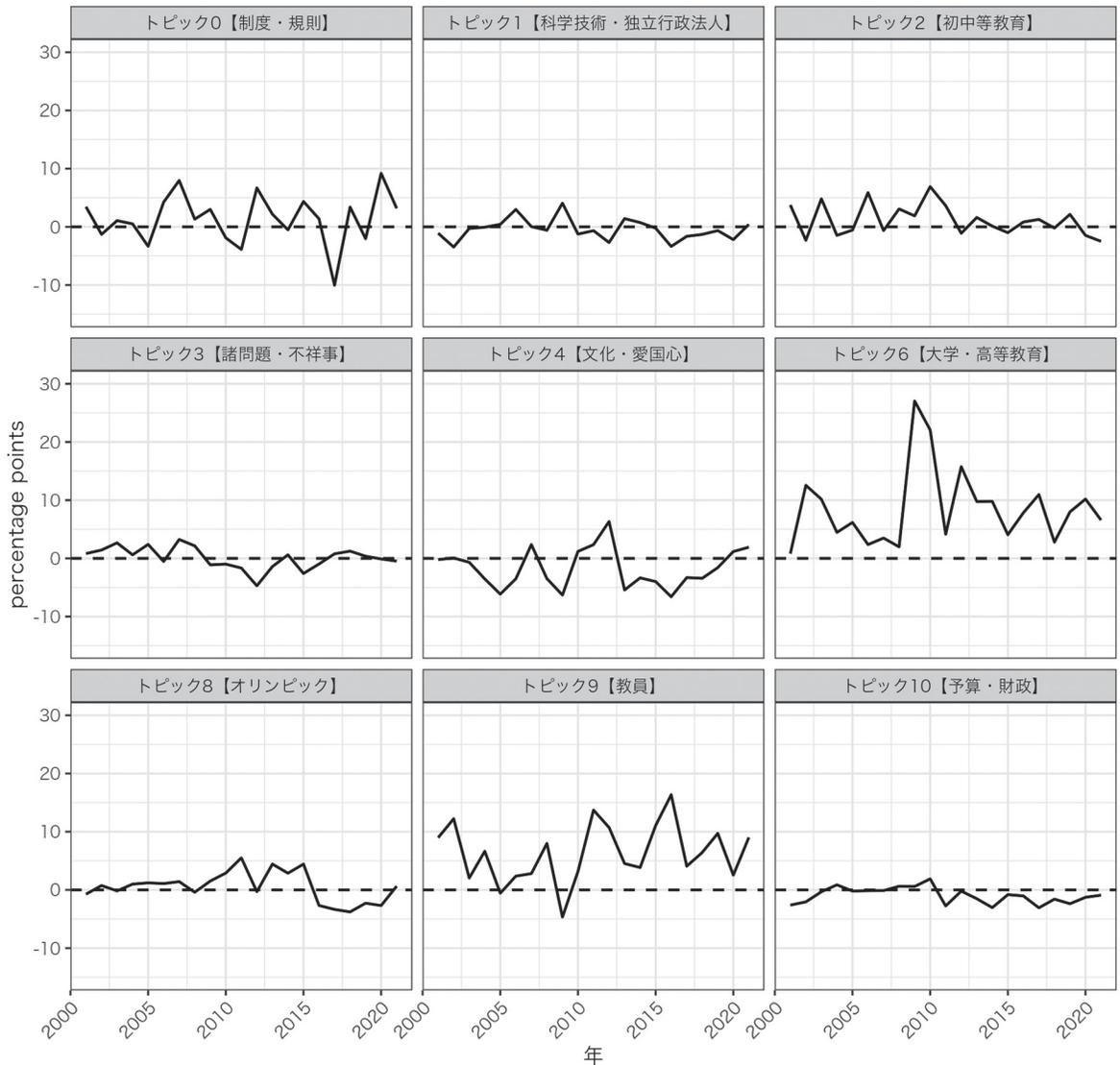


【委員会定型句】のトピック5・7は省略
縦軸が正であれば公明党の構成比が高い

次に、自由民主党と日本共産党の比較を行った結果が以下の図表7である。こちらも同様に、各年度において日本共産党の構成比から自由民主党の構成比を引き、差分のポイントをプロットした。図表7よりトピック6【大学・高等教育】およびトピック9【教員】

において両党の差が大きいことがわかる。特にトピック6【大学・高等教育】の日本共産党の構成比は常に自由民主党より高く、特に民主党政権下において相対的に多くなっている。

図表 7 自由民主党と日本共産党の年別トピック構成比の比較



【委員会定型句】のトピック5・7は省略
縦軸が正であれば共産党の構成比が高い

2 政党内比較

前項では、政党間で取り扱う 이슈の構成比が異なるのかについて分析を行った。本項では、同一政党内で議員によって取り扱う 이슈が異なるのかについて分析を行う。対象とするのは、B節の図表5に挙

げた政務三役以外として発言回数の多い上位20議員のうち、自由民主党と日本共産党の議員とする。なお、本項においては政務三役以外での発言のみを対象とする。

図表 8 政務三役以外の発言回数が多い自由民主党議員のトピック構成比

議員名	T0	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	総計
馳浩	15.3%	1.7%	7.7%	5.7%	5.7%	26.2%	5.6%	2.7%	10.5%	15.0%	4.0%	1256
義家弘介	14.3%	0.4%	15.3%	3.8%	2.4%	32.0%	3.2%	3.2%	2.2%	20.4%	2.7%	712
下村博文	15.2%	0.6%	8.7%	6.8%	3.3%	39.5%	5.4%	1.1%	7.3%	9.3%	2.9%	646
小淵優子	6.0%	2.5%	1.2%	0.9%	1.2%	8.8%	0.9%	51.6%	3.1%	1.7%	21.9%	645
自由民主党	8.4%	3.2%	6.6%	2.2%	4.8%	17.7%	5.4%	35.0%	4.1%	7.9%	4.8%	18193

図表 8 は自由民主党の議員のトピック構成比の比較である。トピック 5 と 7 【委員会定型句】以外で、自由民主党全体の発言の構成比よりも 5 ポイント以上差がある項目を黒く塗った。トピック 2 (図表では T2

と表記。以下、同様。)【初中等教育】やトピック 9 【教員】は義家弘介が多く、トピック 8 【オリンピック】は馳浩が多いことが見て取れる。

図表 9 政務三役以外の発言回数が多い日本共産党議員のトピック構成比

議員名	T0	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	総計
畑野君枝	13.6%	2.5%	8.0%	3.3%	7.4%	17.3%	15.5%	2.7%	5.6%	21.9%	2.1%	1531
石井郁子	12.0%	3.7%	13.7%	3.7%	5.5%	21.0%	18.6%	1.0%	3.7%	14.8%	2.4%	1447
宮本岳志	15.4%	1.9%	8.9%	2.5%	5.6%	17.6%	16.7%	1.3%	13.3%	14.2%	2.6%	1135
吉良よし子	15.0%	1.5%	7.9%	1.8%	6.8%	19.9%	18.1%	2.0%	2.3%	24.0%	0.8%	659
田村智子	21.6%	3.3%	8.9%	3.1%	2.2%	14.2%	14.2%	1.1%	11.8%	18.2%	1.6%	643
日本共産党	14.5%	3.1%	10.8%	3.0%	5.5%	18.6%	16.6%	1.7%	6.2%	18.1%	2.1%	7313

図表 9 は日本共産党の議員のトピック構成比の比較である。図表 8 と同様に日本共産党全体の発言の構成比よりも 5 ポイント以上差がある項目を黒く塗った。前項で日本共産党の議員はトピック 6 【大学・高等教育】の発言が自由民主党の議員に比べて多いことに触れたが、当該トピックについてはどの議員も構成比が高い。議員間で差が大きいのはトピック 8 【オリンピック】である。

5. 考察

A. 分析結果に関する考察

前章では、省庁再編後の教育をめぐるイシューとアクターおよびその変遷について分析結果を示した。まず、イシューについてはトピック構成比に変化があり、時点により取り扱うイシューが変化していることが示された。前章で確認したとおり、2003年・2019年にはトピック 6 【大学・高等教育】が突出して多く、2003年は国立大学法人制度の導入に関する議論、2019年は大学入試センター試験への民間試験導入の議論であった。2014年にはトピック 0 【制度・規則】

が突出して多かったが、これは地教行法改正に伴う教育委員会制度に関する議論であった。また、トピック 8 【オリンピック】は2011年、15・16年、20・21年に構成比が高くなるといった周期的な変動を指摘したが、2015・16・20・21年においては東京オリンピックに関連した発言がされていた。また、2011年はオリンピックに関連した発言に加え、東日本大震災に関連した発言を多く含んでいた。以上、イシューの変遷については、トピック構成比と各トピックに含まれる上位の発言内容を組み合わせることで、ある年にどういったイシューが議論されているかについて、一定程度概観できることが分かった。

次に、衆議院と参議院でトピックの構成比の変遷が異なるのかについて分析した結果からは、第一に両院の構成比は概ね同じになることが確認された。第二にいくつかのトピックについては両院で差が見られ、トピック構成比の増減のタイミングが異なることも確認された。この点について、片方の議院で話題となったことがもう一方の議院で遅れて議論される可能性もあるが、これについてはより詳細にテキストを確認する必要がある。

アクターに着目した分析では、発言回数が多い議員を文教政策に関心の高い議員と見ることができた。これらの議員は自由民主党のみならず他政党も含めた文教関連議員と見ることのできるであろう。なお、発言回数の多い議員は政務三役としての発言が多いという点に留意する必要があることが確認できた。そして、政務三役ではない発言回数の分析からは、自由民主党の議員が委員数を勘案すると相対的に少なく、与党の場合には発言回数が少ないことが明らかとなった。ただし、自由民主党においても野党時代には発言回数が多くなっていることから、与党の場合は委員会の場ではなく、事前に政府と与党間でイシューが調整されているため委員会の場での発言数が少ないことと言えるかもしれない。また、日本共産党の議員のように政務三役経験のない委員の発言回数が多く、委員会の場での野党の役割を果たしていると考えられることもできる。しかしながら、それらの発言が実際の政策にどのような影響を与えたのかについては、本研究では対象としていない文部科学省などの政策文書について分析する必要がある。

政党間の比較においても、いくつかのトピックで差異が見られ、政党によって取り上げるイシューが異なることが示された。特に自由民主党と日本共産党の間ではトピック 6【大学・高等教育】について常に日本共産党のほうが自由民主党よりも発言構成比が高かった⁹⁾。さらに、同じ政党内での比較では、自由民主党、日本共産党ともに議員によって発言しているトピックに偏りがあることが分かった。特に自由民主党では義家弘介がトピック 2【初中等教育】とトピック 9【教員】といういずれも初等中等教育に関連するトピックに、馳浩がトピック 8【オリンピック】のオリンピックに関連するトピックについて多くの発言をしていた。これは議員になる前に義家弘介が高校教員であったこと、馳浩がスポーツ選手であったことと関連していると考えられる。

以上、トピック分析によりイシューの変遷のみならず、政党間での関心の差や政党内での役割分担についても確認することができた。

B. 今後の課題

1 分析手法に関する課題

本研究ではトピックモデルの中でも最も基本的な LDA を用いたが、本稿のデータの持つ特徴を踏まえると、時系列構造への対処とネスト構造への対処といった点で改善の余地がある。前者に関しては、ト

ピック分布や単語分布にマルコフ性を導入することで時系列構造を表現する Dynamic Topic Model (Blei & Lafferty 2006) や、年度をトピック比率の共変量として考慮する構造トピックモデル (Roberts et al. 2016; 瀧川 2019) などが考えられる。後者に関しては、本稿のデータでは議員に発言がネストされており、同じ議員によってなされた発言のトピックには相関があることが考えられる。こうしたデータ構造のモデリングは Author-Topic Model (Rosen-Zvi et al. 2004) によって行うことが可能である。

また、本研究では恣意的な扱いを避けるため、発言一つ一つを単位とした分析を行った。そのため、人間の目で確認するとある一連の議論であっても、発言ごとに最も入る確率が高いトピックが変わってしまうことが確認された。データの単位、区切り方の扱い方をどのようにすべきかも今後の課題である。

2 データに関する課題

省庁再編後のイシューやアクターの変動を見ようという目的のため、本研究で扱ったデータは 2001 年以降のみとした。そのため、省庁再編前後でのトピックの変化は見ることができていない。また、文教政策に関心の高い議員について、発言回数を政務三役か否かで峻別することを試みたが、政務三役となっている議員は今回扱った期間以前に一般の委員として委員会で発言している可能性がある。実際に、文教政策に関心の高い議員がその後に政務三役になったのか、一般の委員としての発言と政務三役となった際の発言は異なるのかといった分析のためにはより長期間にわたるデータを扱う必要がある。さらに、「発言者肩書き」では委員長がわからない。委員長は会議の進行上、定型的な発言が多くなると考えられ、発言数の多寡で文教政策に関心の高い議員を判断する場合において、過剰に評価してしまう可能性がある。

また、本研究で用いた国会会議録検索システムから得たデータでは、各議員の属性情報は所属党派と肩書きしかわからない。そのため、当選回数や選出選挙区などの情報は別途用意する必要がある。国会会議録とは別に過去の全国会議員のデータベースを構築し、データ結合することで議員の属性に応じた分析を行える可能性がある。

本研究の分析では、議員によって言及するトピックの偏りについて、議員になる前の職業に関係する可能性があるという指摘したが、党議拘束による可能性もある。この点についてテキストのみから判断することは難しい。

さらに、国会会議録による分析という観点では、トピックの移り変わりや、議員によって言及するトピックに偏りがあるという点が、文教政策以外でも言えるのかについて検証が必要である。すなわち、厚生労働委員会や国土交通委員会など別の政策分野を扱う委員会との比較が必要ではある。

C. 今後の展望

このように、分析手法やデータの扱い方について課題は残るものの、本研究ではAPIによって国会会議録のデータを機械的に収集し、LDAによって分析を行うことで、イシューやアクターの変遷を解明する一連の方法について示すことができた。

次に、データ取得の観点から教育政策研究に対する示唆を述べる。国会会議録を計量的に分析するには、国会会議録検索システム検索用API公開以前は手動でデータを収集する必要があったのに対し、本研究ではAPIを使用して自動で取得した。イシューやアクターの変遷を分析するには、発言内容だけでなく、氏名、肩書き、政党、所属会派、日付、発言者役割、院名などのメタデータを用意する必要がある。本研究で分析対象となった発言の総数は84,730件あり、各発言に付随するメタデータを含めると収集が必要なデータは100万近くに及び、手動での収集では現実的な時間内に作業を終えることができないうえに、入力ミスといったエラーが発生する可能性が非常に大きい。こうした難点を克服するために、APIを使用したプログラミングによってデータ収集を実行し、手作業では数ヶ月～数年が必要なデータ収集の瞬時かつ正確な遂行が可能となった。

このように、従来の方法では困難だった大規模データセットの構築・分析が可能になったことは大きな転換点であり、ビッグデータを活用した教育政策研究の新たな可能性を本研究は示したといえる。

本研究を通して、2000年代以降の教育領域の政策形成・決定のプロセスの解明に必須であるイシューとアクターが抽出できた。今後は、これらの知見を踏まえ、個々のイシューとそれに関わるアクターを取り上げて、より詳細に政策プロセスを解明するとともに、分析手法もより深化・精緻化させていきたいが、これらの課題については他日を期したい。

【謝辞】

本研究は、JSPS 科研費21K18489の助成を受けたものである。

【注】

- 1) なお、国会会議録検索システム検索用APIの様々な領域での利用意義については川瀬・清水 (2015)、岡田 (2018) などに詳しい。また国会会議録を他の領域で利用・分析した近年の成果として、山本 (2011)、掛谷・大南 (2019) などがある。
- 2) 検索の際には2001年1月31日から2021年12月31日までの発言を収集するように設定を行い、会議名に「文部科学委員会」または「文教科学委員会」が含まれるものを検索した。当該期間の最初の発言は2001年2月15日、最後の発言は2021年12月21日である。データの収集は2022年2月25日に実施した。
- 3) 形態素解析にはMeCabを用い、分かち書きの際に使用する辞書にはmecab-ipadic-NEologd (佐藤ほか 2016, 2017) を使用した。mecab-ipadic-NEologdは、通常のシステム辞書では誤った単語の分割が行われる新語や固有表現を正しく分割できるように最適化された辞書である。
- 4) 表記揺れを防ぐため、Unicode正規化を行い、アルファベットは小文字に統一し、旧字体は新字体に変換した。
- 5) Nelson (2020) ではオンラインのAppendixで各トピック上位2文書を示しているが、本研究では紙幅の都合上、上位1発言のみを示した。
- 6) あるトピックのある年の構成比とは、(ある年における当該トピック発言数) / (ある年の発言数全て) である。
- 7) 文部科学省ウェブサイト「歴代文部科学大臣」ページ (https://www.mext.go.jp/b_menu/soshiki/rekiday/daijin.htm) にて確認した。
- 8) 例えば次のような会派が組まれていた。「自由民主党・無所属の会」、「自由民主党・国民の声」、「自由民主党・こころ」、「自由民主党・たちあがれ日本・無所属の会」、「自由民主党・保守党」。
- 9) この点は、小川 (2017) における義務教育の充実と大学進学機会の確保を比較した際に、右派政党支持に比べて「政党支持が中道・左派になるにつれ、大学教育を優先すべきと感じている」という分析結果を踏まえると、政党支持者の政策選好が現れた結果であると考えられることもできる。

【参考・引用文献】

- 青木栄一, 2021, 『文部科学省』中央公論新社。
- Blei, David M. & John D. Lafferty, 2006, "Dynamic Topic Models," *Proceedings of the 23rd International Conference on Machine Learning*, 113-120.
- Blei, David M., Andrew Y. Ng & Michael I. Jordan, 2003, "Latent Dirichlet Allocation," *The Journal of Machine Learning Research*, 3: 993-1022.
- Evans, James A. & Pedro Aceves, 2016, "Machine Translation: Mining Text for Social Theory," *Annual Review of Sociology*, 42(1): 21-50.
- 羽田貴史, 2019, 「高等教育政策の現在」『日本高等教育学会第22回大会』自由発表レジュメ (2019年6月9日, 金沢大学)。

- , 2021, 「2010年代高等教育政策決定過程の変容」日本高等教育学会高等教育政策プロジェクト研究グループ編『高等教育政策決定過程の変容と高等教育政策』広島大学高等教育開発センター『高等教育研究叢書』159, 9-31頁.
- 橋本鉾市, 2007, 「戦後高等教育政策におけるイシューとアクター」『東北大学大学院教育学研究科年報』56(1), 71-87頁.
- , 2013, 「戦後日本における高等教育関連議員の構造分析」『大学論集』44, 163-178頁.
- , 2014, 『高等教育の政策過程』玉川大学出版部.
- 掛谷英紀・大南勝, 2019, 「国会会議録に基づく短命議員・短命大臣の特徴分析」『知能と情報』31(2), 617-625頁.
- 河合晃一, 2019, 「文部科学省と官邸権力」青木栄一編『文部科学省の解剖』東信堂, 97-134頁.
- 川瀬直人・清水茉莉子, 2015, 「国会会議録フルテキスト・データベースWeb API開発の背景とその利用状況分析」『情報の科学と技術』65(2), 531-536頁.
- Laura K. Nelson, 2020, "Computational Grounded Theory: A Methodological Framework," *Sociological Methods & Research*, 49(1): 3-42.
- 中北浩爾, 2017, 『自民党——「一強」の実像』中央公論新社.
- 日本高等教育学会高等教育政策プロジェクト研究グループ編, 2021, 『高等教育政策決定過程の変容と高等教育政策』広島大学高等教育開発センター『高等教育研究叢書』159.
- 小田中悠・中井豊, 2019, 「意味世界の計算社会科学的分析に向けて——社会学におけるトピックモデルの意義の検討」『理論と方法』34(2), 280-295頁.
- 小川和孝, 2017, 「日本社会における教育政策への人々の選好に関する研究——公的支出の水準／配分の区別に焦点を当てて」『教育社会学研究』100, 225-244頁.
- 萩原克男, 2002, 「中央省庁改革と教育政策形成形式の変容」『日本教育政策学会年報』9, 47-63頁.
- 岡田祥平, 2018, 「日本語コーパスとしての『国会会議録検索システム検索用API』——計量的研究の精緻化・深化の可能性」『新潟大学教育学部研究紀要 人文・社会科学編』11(1), 31-51頁.
- Roberts, Margaret E., Brandon M. Stewart & Edoardo M. Airoidi, 2016, "A Model of Text for Experimentation in the Social Sciences," *Journal of the American Statistical Association*, 111(515): 988-1003.
- Rosen-Zvi, Michal, Thomas Griffiths, Mark Steyvers & Padhraic Smyth, 2004, "The Author-Topic Model for Authors and Documents," *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, 487-494.
- 佐藤敏紀・橋本泰一・奥村学, 2016, 「単語分かち書き用辞書生成システムNEologdの運用——文書分類を例にして」『第229回自然言語処理研究会研究報告』1-14頁.
- , 2017, 「単語分かち書き辞書mecab-ipadic-NEologdの実装と情報検索における効果的な使用方法の検討」『言語処理学会第23回年次大会発表論文集』875-878頁.
- 城山英明・細野助博編, 2002, 『統・中央省庁の政策形成過程』中央大学出版部.
- 瀧川裕貴, 2019, 「戦後日本社会学のトピックダイナミクス——『社会学評論』の構造トピックモデル分析」『理論と方法』34(2), 238-261頁.
- , 2021, 「第6章 テキスト分析——データとしてのテキスト」鳥海不二夫編, 『計算社会科学入門』丸善出版, 133-160頁.
- 歌川光一, 2014, 「国会会議録にみる道德教育関係者の諸相——『心のノート』『私たちの道德』への言及者」『学習院大学教育学・教育実践論叢』1, 119-127頁.
- , 2015, 「戦後の国会会議録における「道德教育」をめぐる計量テキスト分析」『学習院大学計算機センター年報』36, 59-65頁.
- 山田雄司・青田庄真, 2015, 「日本の国会における英語教育政策過程——時代区分・アクター・特徴語」『関東甲信越英語教育学会誌』29, 71-84頁.
- 山本冨里, 2011, 「国会における日本語教育関係議論のアクターと論点——国会会議録の計量テキスト分析からの概観」『日本語教育』149, 1-15頁.
- 渡部昭男, 2020, 「『教育無償化』論議の経緯と特徴(3)——2017年第194回～2019年第200回の国会審議から」『教育科学論集』23, 21-30頁.
- , 2021, 「『教育無償化』論議の経緯と特徴(4)——2020年第201回の国会審議から」『大阪成蹊大学紀要』7, 239-251頁.
- , 2022, 「コロナ禍の高等教育における学びの継続のための学生支援の在り方に係る論議——主に2020年第203回・2021年第204回の国会審議分析から」『大阪成蹊大学紀要』8, 297-306頁.
- 張文超, 2018, 「日本の国会の会議におけるe-Learningの言及について」『大阪大学言語文化学』27, 97-109頁.

(指導教員 本田由紀教授)