原著論文

対数尤度比と政治語彙度を用いた 議員の議会活動の可視化†

内田 ゆず*1・高丸 圭一*2・乙武 北斗*3・木村 泰知*4

我々は、47 都道府県議会本会議を対象として地方議会会議録コーパスを構築した。このコーパスに収録されている発言文には、発言者の氏名・役職・選挙区・生年・性別等の情報が付与されている。したがって、このコーパスを利用すると全国の都道府県議会議員の議会活動の一端を可視化することができる。本論文では、議員の年齢・性別構成を明らかにしたうえで各議員の発言文字数を集計し、議員の属性による発言量の違いを考察する。また、対数尤度比を用いて議員の属性ごとの特徴語を抽出し、議員の属性による発言内容の違いを考察する。その際、政治に関連する用語を抽出する指標として「政治語彙度」を新たに提案する。

キーワード：地方議会会議録、コーパス、統計的指標

1. はじめに

インターネットの普及、および、情報公開の進展に伴い、地方議会の会議録は自治体のウェブサイト公開されるようになってきた。これまで、日本各地の議会事務局や図書館等へ出向かなければ、閲覧することができなかった議会会議録を居ながらにして、しかもコンピュータ上での分析が行いやすいテキストデータとして入手できるようになった。現在新たに、地方政治における議論の過程と結果、およびその背景に関連づけて考えるために、地方議会会議録（議論の過程）と条例（議論の結果）および新聞記事（議論の社会的背景）を結びつけた「地方政治コーパス」の構築が進められている [1]。このコーパスでは、比較分析を定量的に行うことができるように、収集条件（対象期間・対象自治体）を整えて、都道府県議会会議録の収集が行われている。

収集した会議録コーパスを用いることで、これまで焦点をあてられることが少なかった地方議会における議論を分析することが可能となる。[2] では、各都道府県議会で議論される政治課題が TF・IDF に基づいて抽出される特徴語にあらわれることを明らかにした。例えば、医療施設の名称（「ふくしま国際医療科学センター」（福島県）、「さいたま赤十字病院」（埼玉県）、「東千葉メディカルセンター」（千葉県））は、それぞれの施設の新設や移転にあたり繰り返し議論されたことにより、特

徴語として抽出された。このほか、茨城県では「森林湖沼環境税」（期間延長）、東京都では「二〇二〇年大会」（オリンピック）、長野県では「浅川ダム」「F-V 断層」（地震への対応）、奈良県では「奈良公園」「平城宮跡」「ホテル誘致」、鳥取県では「環境大学」（公立化）、滋賀県では「近江大橋」（無料化）などの語が特徴語として抽出された。これは会議録コーパスを用いて、議会の議論を可視化する研究事例の 1 つである。

地方議員の性別構成や年齢構成に目を向けると、分布に偏りがあり、住民の性別・年齢構成を十分に反映しているとは言えない（2 章、表 1 参照）。地方議員の性別・年齢構成の偏りは、取り組む政治課題のバランスや政策の判断に影響を与えている可能性があることが予想される。しかしながら、これまで地方議会における議員の性別・年齢別の発言内容を分析し、特徴を明らかにする研究は行われていない。[1] の地方議会会議録コーパスには、収録されているすべての発言文に対して、発言者名の情報が付与されている。さらに、議員については、選挙の際に公開された情報に基づき、属性情報（選挙区・生年・性別）が付与されている。これを利用することで、日本のすべての都道府県議会議員を対象に議会活動の可視化を行うことが可能である。なお、本研究では「議会活動」を地方自治法に規定されている会議（本会議、各種委員会）、その他の会議での議論と定義する。本論文では議会活動のうち本会議での発言を分析対象とする。

本論文の目的は、地方議会会議録コーパスから、議員の発言における性別・年齢別の特徴を可視化することである。具体的には以下の 3 点を明らかにする。

Q1 議会で活発に発言しているのはどの層の議員か？

Q2 議員の性別によって発言内容に違いはあるか？

Q3 議員の年齢層によって発言内容に違いはあるか？

先行研究 [3] において、政治課題に関連する幾つかの典型的なキーワードを対象として属性ごとの出現割合を比較し、Q2 および Q3 に対する予備的な調査を行った。性別の比較では、たとえば「増税」「貧困」「消費税」「年金」「保育」などのキーワードは、女性議員の方が多く使用している一方、男性議員は

† Visualization of Activity of Local Assembly Members Based on Log-Likelihood Ratio and Political Vocabulary Ratio
Yuzu UCHIDA, Keiichi TAKAMARU, Hokuto OTOTAKE,
and Yasutomo KIMURA

*1 北海学園大学
Hokkai-Gakuen University

*2 宇都宮共和大学
Utsunomiya Kyowa University

*3 福岡大学
Fukuoka University

*4 小樽商科大学
Otaru University of Commerce

「地方分権」「高速道路」「観光」などへの言及が多かった。また、年層による比較では、「インターネット」「待機児童」「保育」「英語」などのキーワードを若い議員ほど多く使用する傾向が見られた。本論文では、性別・年層別の発言内容を、網羅的・定量的に分析する。性別・年層別の特徴語の基準として、対数尤度比を導入する。抽出された語が政治課題に関連する語であるかどうかを判定する基準として新たな尺度「政治語彙度」を提案する。

本論文の構成は以下の通りである。2章では分析対象となるコーパスの概要、発言者名および属性情報の付与について述べる。3章では、発言文字数に基づく分析を行い、Q1についての分析結果を述べる。4章では、対数尤度比および政治語彙度を用いた発言内容可視化の方法を述べた上で、Q2、3についての分析結果を述べる。

2. 地方議会会議録コーパスの概要

2.1 収集方法

収集対象は全国47都道府県議会の本会議であり、収集期間は平成23年4月の統一地方選挙から平成27年4月の統一地方選挙の前（平成27年3月）までの4年分である。

都道府県議会はすべて、それぞれの自治体のWebサイトに会議録の全文検索システムを公開している。全国の都道府県のうち、42都道府県は全文検索システムを提供する主要4社のいずれかのシステムを採用している（シェア89%）。そこで、まずこの4システムに対応したクローラを作成した。ただし、自治体ごとに検索システムや表示部分をカスタマイズしているため個別対応が必要であり、場合によっては部分的に人手による作業も交えながら収集作業を行った。さらに、主要4社以外のシステムによって会議録を公開している5県については、1自治体ごとにクローラを作成して、収集を行った。

2.2 データ量

収集された全国47都道府県の本会議会議録は、約400万文、1.5GBであった。会議録を句点で1文単位に分割し、それぞれの発言の発言者名を抽出した。そして、会議の開催情報等と併せて1文を1レコードとして格納した。総収録データは4,395,876レコードであり、発言の総文字数は252,364,746文字である。

各都道府県の本会議における総発言数を図1に示す。発言数が最も多いのは鳥取県で、最も少ない山形県の約5倍に達する。本会議の開催回数が異なるため一律に比較はできないが、発言量は議会の活発さのおおよその目安になる。

2.3 発言者の同定

会議録から抽出した発言者名は、同一人物でも複数の表記をもつことがある。例えば、「梅澤佳一自然再生・循環社会対策特別委員長」「梅澤佳一議会運営委員長」「梅澤佳一委員」「五十番（梅澤佳一議員）」「五十六番（梅澤佳一議員）」はいずれもコーパスに出現し、一人の「梅澤佳一」氏を示している。しかし、これらの発言者名を機械的に処理すると同一人物だとみなされない。

表1 地方議会会議録コーパス収集期間の議員の構成

	25-44 歳	45-64 歳	65-84 歳	不明	計
男性	547	1,616	457	71	2,691
女性	51	174	25	5	255
計	598	1,790	482	76	2,946

表2 議員の属性別平均発言文字数

	25-44 歳	45-64 歳	65-84 歳	不明	計
男性	43,873	44,159	31,808	40,278	41,901
女性	43,901	61,465	51,081	45,450	56,620
計	43,875	45,841	32,808	40,618	43,175

議会での発言を議員単位で分析するためには、発言者を正確に同定する必要がある。したがって、発言者名が一つに定まらない発言者については、人手で名寄せの作業を行った。具体的には、発言者の氏名に付加された敬称や肩書、議席番号等を正規化した。また、旧字体や通称を用いる議員への対応や、会議録作成者による記載ミス等の修正も行った。名寄せの結果得られた発言者数は合計2,794名であった。

2.4 発言者の属性情報付与

会議録には議員の発言の他にも議長の発言（議事進行）や首長の発言、職員の発言、参考人の発言も記録されている。本コーパスには「発言者の役職」フィールドが設けてあり、各発言に役職の情報が付与されている。この情報を利用すると、議会での立場の違いによる発言の特徴を分析できる。

また、議員については生年や性別、対象行政区（選出選挙区）といった属性情報が分かると、発言内容の分析や言語的特徴の分析にとって有益であると考えられる。しかし、これらの情報は会議録中には記載されていないため、選挙の立候補者情報がまとめられたサイト（選挙ドットコム¹、政治山²）から収集し、発言を格納するテーブルとは別のテーブルに格納した（詳細は[2]参照）。

以上の処理により明らかになった都道府県議会議員の構成を表1に示す。表中の人数には会議録に発言が記録されていない議員も含まれている。なお、年齢は平成23年12月31日時点のものである。日本では国会の女性議員比率が世界的にみても低いことが知られている³。地方議会においても女性議員は男性議員の1/10に満たない人数であり、同様の傾向が見られた。

3. 議員属性による発言量の違い

各議員がどれほど活発に活動しているのかを知るための指標として、議会での発言量が考えられる。ここでは、発言文字数を属性別、個人別に分析する。

地方議会会議録における文字遣いは「発言記録作成標準[4]」などによって標準化されている。したがって、各都道府県議会では表記の方法が大きく異なることはないと判断し、分析の単位

1 <http://go2senkyo.com/articles/>

2 <https://seiji-yama.jp/>

3 国会の女性議員比率：衆議院 9.5%、参議院 15.7%（平成28年1月現在）

図1 各都道府県議会における発言数

を文字とした。

3.1 属性別

議員の属性別に発言の平均文字数を集計した結果を表2にまとめる。65～84歳の男性議員が最も発言文字数が少なく、45～64歳の女性議員が最も多い。また、どの年層でも男性議員よりも女性議員の方が多く発言している。ただし、女性議員の人数は非常に少ないため、本会議全体に占める発言の割合は少ない。

発言文字数による分析の範囲では、女性議員の方が活発に発言している。女性の方が丁寧な表現を使用して発言が長くなるなど、表現技法の違いが影響を与える可能性もある。今後、属性による文末表現の違いなども分析したい。

3.2 個人別

発言文字数ごとの議員数を図2のヒストグラムに示す。表2によると平均発言文字数は43,175字であり、多くの議員が10,000字以上発言している。一方、発言が少ない議員（1,000字未満）も183名存在する。そのうち、発言文字数が0文字、つまり会議録中に発言が全く記録されていない152名について詳しく調査したところ、83名はコーパス収集期間と任期が一致しておらず⁴、10名は議長として議事進行を行っている期間があった。したがって、それらの93名を除いた59名（全員男性）は4年間の任期中に本会議で一度も発言していないことになる。

上述の59名の議員について当選回数を調査した結果を図3に示す。比較のために平成23年の統一地方選挙で当選した議員の当選回数も併せて示している。当選回数のデータはWebサイト「読売オンライン⁵」の選挙情報ページから取得した。

議員全体の当選回数の傾向と比較して、発言のない議員は当選回数が多いことがわかる。発言のない議員のうち55名は5回以上当選しており、所謂「多選」の議員である。低い投票率や無投票選挙の増加⁶により議員の固定化が進む中、多選議員

図2 発言文字数別議員数

図3 議員の当選回数

は議会活動へのモチベーションが低下している可能性がある。

もちろん、本会議での発言が議員活動のすべてではない。ベテラン議員が若手議員に質問の機会を譲るケースもある。しかし、有権者から見ると議会で全く発言しない議員を評価することは難しいだろう。したがって、このような情報は有権者が投票先を選ぶ際に有益だと考えられる。

4. 議員属性による発言内容の違い

4.1 対数尤度比

先行研究[3]では、議員の属性によって発言内容に差があることが示唆された。したがって、より詳細な分析を行うために、統計的指標を用いた特徴語抽出を行う。ここでは、指標と

4 期間中に落選、当選、辞任、死亡、長期欠席した、あるいは逮捕、リコールされたため。

5 <http://www.yomiuri.co.jp/>

6 平成23年の統一地方選挙の道府県議選での平均投票率は48.15%、無投票当選率は17.6%。

して対数尤度比を採用する。

対数尤度比はコーパスから特徴語を抽出するタスクに有効だとされている。[5, 6] では、英語の教材における語彙選定を目的として、9種類の統計的指標（頻度、ダイス係数、対数尤度比、コサイン、イエーツの補正公式、カイ二乗値、補完類似度、自己相互情報量、マクニマーテスト）を用いて特徴語を抽出し、各指標の比較を行っている。その結果、対数尤度比は特徴度の高い語彙を精度良く抽出することが報告されている。この知見に基づき、対数尤度比を用いて特徴語を抽出する試みが複数分野で行われている。たとえば、[7] では医療コーパスから一般人向けの専門用語を抽出しており、[8] では新聞コーパスから特徴度の高い新聞用語を選定している。

表 3 に示すクロス集計表に関する対数尤度比 (Log-Likelihood Ratio) は以下のように定義される。

$$LLR = a \log \frac{aN}{(a+b)(a+c)} + b \log \frac{bN}{(a+b)(b+d)} + c \log \frac{cN}{(c+d)(a+c)} + d \log \frac{dN}{(c+d)(b+d)} \quad (1)$$

ただし、 $ad - bc < 0$ の場合は、 -1 を乗じて補正する。

男性、女性、25～44 歳、45～64 歳、65～84 歳の 5 つの属性毎に議員の発言をまとめ、コーパスとして利用する。以降では、それぞれ男性コーパス、女性コーパス、若年コーパス、中年コーパス、高年コーパスと呼ぶこととする。そして、男性-女性、若年-中年、中年-高年、高年-若年のコーパスの組を対象として対数尤度比を算出し、水準 0.1% で有意となる 10.83 より大きい単語を 1 つ目のコーパスの特徴語、-10.83 未満の単語を 2 つ目のコーパスの特徴語とする。

対数尤度比を算出する言語単位は名詞、あるいは複合名詞とする。発言の単語への分割には形態素解析器 MeCab [9]、およびその辞書として mecab-ipadic-NEologd⁷ [10] を用いた。また、その際に連続する名詞は接続し、複合名詞とした。

男性コーパス、女性コーパスの双方で 1 回以上出現した複合名詞 134,798 語に対して対数尤度比を算出した結果、特徴語の条件を満たした語は男性で 2,534 語、女性で 5,375 語となった。

同様に、若年コーパスと中年コーパスの双方で 1 回以上出現した複合名詞 140,115 語のうち、特徴語の条件を満たした語は若年で 1,465 語、中年で 0 語、中年コーパスと高年コーパスの双方で 1 回以上出現した複合名詞 127,277 語のうち、特徴語の条件を満たした語は中年で 250 語、高年で 2,135 語、高年コーパスと若年コーパスの双方で 1 回以上出現した複合名詞 97,984 語のうち、特徴語の条件を満たした語は高年で 2,567 語、若年で 1,848 語となった。

4.2 対数尤度比が高い語の特徴

各コーパスから特徴語として多数の語が抽出されたが、中には本研究の分析対象としては適切ではない語（例：「登壇」「県当局」「積極的」）も散見される。したがって、対数尤度比が高

表 3 クロス集計表

	コーパス A	コーパス B	計
単語 w の出現頻度	a	b	$a + b$
単語 w 以外の出現頻度	c	d	$c + d$
計	$a + c$	$b + d$	N

い語の概要を把握するために、男性コーパス、女性コーパスにおける対数尤度比上位 300 語を政治関連語、一般語、代名詞・固有名詞、場所表現、時間表現、その他の 6 カテゴリに分類した。分類結果を表 4 にまとめる。

ここから、対数尤度比が高い語であっても、政治課題を直接表現しない語が多く含まれることがわかる。議員の発言における分野の性差、あるいは年層差を明らかにするためには、政治関連語のみを分析対象とするべきである。したがって、これらの語から「政治関連語」に相当する語を抽出するために新たな指標を導入する。

4.3 政治語彙度

語が政治課題に関連するかどうかを客観的に図る尺度として「政治語彙度」を提案する。基本的な考え方は、政治に関連する文書においてよく使われる語は政治課題に関連する語である可能性が高いというものである。政治語彙度の計算には、国立国語研究所が開発した「現代日本語書き言葉均衡コーパス (BCCWJ)」[11] を用いる。BCCWJ には書籍全般、雑誌全般、新聞、白書、ブログ、ネット掲示板、教科書、法律などのジャンルから無作為に抽出された 1 億 430 万語のデータが格納されている。収録されたデータのうち、新聞、白書、広報誌、法律、雑誌（政治・経済・商業、政治／外交、国勢／民力、金融／財政）、教科書（社会）、Yahoo!知恵袋（ニュース、政治、国際情勢）、ブログ（政治）の各分野を政治関連コーパスと定め、次式によって政治語彙度を定義する。

$$\text{政治語彙度} = \frac{\text{政治関連コーパス中の出現数}}{\text{BCCWJ 全体での出現総数}}$$

4.1 節で述べた、対数尤度比によって抽出されたすべての特徴語を対象として政治語彙度を算出したところ、政治語彙度が 80% 以上の語（政治特徴語と呼ぶ）は男性議員には 194 語、女性議員には 386 語存在した。同様に若年の政治特徴語は、中年と比較すると 126 語、高年と比較すると 87 語、中年の政治特徴語は、高年と比較すると 21 語、高年の政治特徴語は、若年と比較すると 121 語、中年と比較すると 169 語存在する。

抽出された政治特徴語の語数と具体例を政治分野ごとにまとめた結果を表 5 から表 8 に示す。政治分野は、都道府県の公式 Web サイトにおける行政分類を参考にして作成したものである。表 6 は若年コーパスと中年コーパスを比較し、若年コーパスに特徴的な語を示している。中年には特徴語が存在しないため、表中に掲載されていない。同様にして、表 7 は中年コーパスと高年コーパスを、表 8 は高年コーパスと若年コーパスを比較し、それぞれのコーパスでの政治特徴語を示している。

「分類なし」には、「安部政権」「平成二十三年度」「聖域扱い」

⁷ <https://github.com/neologd/mecab-ipadic-neologd>

表4 対数尤度比が高い語のタイプ別語数と具体例

タイプ	男性	女性	例
政治関連語	93	158	中産間地域, 原発, 消費税増税, 介護
一般語	105	92	整備, 発生, 討論, 理由
代名詞・固有名詞	20	15	私たち, 日本共産党, 安倍首相
場所表現	15	13	首都圏, 成田空港, 愛宕山, 辺野古
時間表現	15	1	今後, 先般, 平成二十二年
その他	52	21	原案可決, いかが, こと, 非常

図4 性別による政治特徴語の出現割合の比較

図6 年層による政治特徴語の出現割合の比較（中年・高年）

図5 年層による政治特徴語の出現割合の比較（若年・中年）
※中年には政治特徴語が存在しない

図7 年層による政治特徴語の出現割合の比較（高年・若年）

「天草地域」「議会運営」などの語が該当する。これらの語は政治的な文脈で用いられることが多く、政治語彙度が高くなることは妥当である。しかし、単体で特定の政治課題を表現する語ではないため、分析対象からは除外する。

4.4 可視化

政治特徴語の性差・年層差を把握するために、政治特徴語を政治分野に分類した結果を可視化する（図4-7）。コーパスサイズが異なるため、政治特徴語の実数を直接比較することは適

切ではない。したがって、各コーパスにおける政治特徴語の分野別出現割合をグラフ化する。

図4は政治特徴語の分野別出現割合を性別ごとに示したものであり、男性議員における出現割合を降順にソートしている。つまり、グラフの左側には男性議員が多く発言する政治分野が配置され、右側には男性議員が発言しにくい政治分野（つまり、相対的に女性議員の発言が多い分野）が配置される。図5では、若年議員と中年議員の政治特徴語の出現割合を示している。ただし、中年議員には若年議員と比較した時の政治特徴語が存在しないため、若年議員のみのデータとなっている。図6では中年と高年の政治特徴語の出現割合を中年で降順にソート

表5 政治特徴語の政治分野別語数と具体例（性別）

政治分野	語数	男性	語数	女性
		例		例
くらし・すまい	6	定住促進, 地籍調査	13	市民活動, 食生活改善推進員
出産・子育て	1	民間移管	32	妊婦検診, 放課後児童クラブ, ファミリーサポートセンター, 乳幼児医療費助成
人権・男女共同参画	0	-	18	男女共同参画施策, デートDV, 配偶者暴力相談支援センター
教育・研究	4	全国学力・学習状況調査, 小中一貫教育	28	自然体験学習, スクールソーシャルワーカー, 複式学級, テクノカレッジ
健康・福祉	1	高齢者等	86	医療費無料制度, 新型インフルエンザ対策, オストメイト, 発達障がい, 補助犬
労働・商工業・エネルギー	12	次世代産業, 木質バイオマス発電	21	若者就労支援, イノベーション推進機構, 軽水炉, コージェネレーション設備
観光・特産品	10	観光戦略, 観光誘客, 県産品	0	-
歴史・文化・スポーツ	8	世界文化遺産登録, スポーツツーリズム	4	障害者スポーツ, プロスポーツチーム
危機管理・安全保障	1	国境離島	8	夜間訓練, 普天間問題
都市計画・交通	33	県土づくり, 合併処理浄化槽, バイパス整備, えちぜん鉄道, コミュニティバス	19	無電柱化, 海岸利用, ハッ場ダム
防犯・交通安全	12	暴力団排除, 交通事故死者数	1	青パト
防災・災害対策	22	消防本部, 海岸防災林, 津波対策, オフサイトセンター, DMA T	11	河川防災, モニタリングポスト, 復興公営住宅, 放射能対策
財政・税金	4	財政調整基金, 歳出総額	11	国保税, 消費税増税反対, 市場公募債
自然・環境	2	法定検査, 水源地域	7	産廃処分場建設, 騒音測定器
農林水産業	30	主食用米, 有害鳥獣対策, 新規就農者, 栽培漁業, 県産木材	4	青年農業士, 公共育成牧場
分類なし	48	企業局長, 平成二十三年度, 発生件数	123	巡回連絡, 1次試験, 安倍政権, 聖域扱い

して示している。同様にして、図7では、高年と若年の政治特徴語の出現割合を高年で降順にソートして示している。

4.5 政治特徴語の精度

政治語彙度の妥当性を検証するために政治特徴語の精度を求める。政治分野の用語は一般語との境界が曖昧な上に、固有名詞や新語も多く含むため、正解データを厳密に定めることは困難である。ここでは、「朝日新聞のキーワード」を疑似的な正解データとする。朝日新聞のキーワードは2006年1月以降の朝日新聞紙面に掲載されたもので、約22,000語がWeb上で公開されている⁸。

まず、①男性コーパスの対数尤度比上位100語、②男性コーパスの政治関連語93語、③男性コーパスの政治特徴語100語（194語からランダムサンプリング）が朝日新聞のキーワードに含まれているかを人手で確認する。このとき、字面が完全に一致する語だけでなく、同義表現や類義表現についても考慮する。例えば、同義表現には「川内原子力発電所／川内原発」や「観光戦略／ツーリズム戦略」などがあり、類義表現には「有害鳥獣対策／有害鳥獣の捕獲」や「学力向上対策／学力向上拠点形成事業」などがある。次に、完全一致、同義表現、類義表現の和が全体に占める割合を算出し、精度とする。

結果を表9に示す。政治語彙度によって抽出された政治特徴語が最も高い精度となった。なお、女性コーパスにおいても政治特徴語全体の精度は50.5%となり、類似した傾向をもつこ

とが確認された。さらに、「存在しない」に分類された政治特徴語の中にも「転出超過」「漁場環境」「人口減少対策」など、政治分野の用語として適切だと考えられるものが存在した。このことは、政治語彙度によって既存の辞書にはない表現も抽出できることを意味する。

以上のことから、政治語彙度の妥当性が一定程度示された。

4.6 政治特徴語の属性による差異

先行研究では、政治課題に関連するキーワードを人手で選択し、それらの性差・年層差を分析した[3]。分析結果を図8（各キーワードにおいて「男性議員を1とした時の女性議員の発言回数」と図9（45～64歳の年層を1とした時の他の年層の発言回数）に示す。図8から「増税」「貧困」「保育」などのキーワードが女性に多く使用され、「高速道路」「地方分権」などが男性に多く使用されていることが分かる。政治特徴語の分類（図4）では、「健康・福祉」や「出産・子育て」分野の割合が女性で高く、「都市計画・交通」分野の割合が男性で高いことから、結果の傾向は類似していると考えられる。

本研究で抽出した政治特徴語は複合名詞を対象としているため、図8、図9に示した人手で選択したキーワードはほとんど含まれていない。例えば「消費税」は政治特徴語として抽出されていないため、「消費税」の使用傾向を政治特徴語から明らかにすることができない。しかし、政治特徴語の中には「消費税増税関連法案」「消費税財源」「消費税非課税」など「消費税」を部分文字列として持つ語が87語存在する。政治語彙度を導入することで、人手によるキーワード設定では網羅できない政

8 <https://kotobank.jp/dictionary/news/>

表 6 政治特徴語の政治分野別語数と具体例（若年・中年）

政治分野	語数	若年 例
くらし・すまい	0	-
出産・子育て	8	MFICU, 子ども教室, 子育てサークル, 就園
人権・男女共同参画 教育・研究	0	-
健康・福祉	7	テクノカレッジ, 職業技術校, 不登校者数, 養護教員
労働・商工業・エネルギー	13	保険外併用療養, 休日夜間応急診療所, 救命講習, 臓器移植コーディネーター, 介護者支援
観光・特産品 歴史・文化・スポーツ	9	Uターン情報センター, 求人活動, 採石業, 商店街事業, 新エネルギー技術
観光・特産品	2	ふるさと大使, 観光バス駐車場
歴史・文化・スポーツ	5	ニュースポーツ, 国際芸術祭, 文化財建造物,
危機管理・安全保障	2	取締船, 予備自衛官
都市計画・交通	7	国内定期便, 電気自動車等, 東側ルート, 南北自由通路
防犯・交通安全	3	厚木警察署, 初動警察活動, 通信指令室
防災・災害対策	5	災害物資, 地域防災訓練, 連絡通報体制
財政・税金	0	-
自然・環境	3	侵略的外来種, 保健環境研究所, 緑の回廊
農林水産業	3	コスト林業, 乾燥材, 青年農業士
分類なし	59	各種手続, 世帯別, 滞納案件

※若年と比較すると中年には政治特徴語が存在しない。

図 8 性別によるキーワードの出現割合の比較（[3] より再掲）

治関連用語の使用傾向をつぶさに明らかにすることができると考えられる。

4.4 節で説明した通り、本研究では議員の属性による発言内容を比較するために、抽出された政治特徴語の政治分野の出現割合を可視化した（図 4～7）。以降では、抽出された政治特徴語の政治分野について考察し、具体的な例を挙げながら詳細を述べる。

まず、政治特徴語の性差を考察する（図 4、表 5 参照）。なお、具体例については表で示した語以外の例にも触れる。

図 9 年層によるキーワードの出現割合の比較（[3] より再掲）

男性議員は都市計画・交通、農林水産業、防災・災害対策に関する語を多く用いている。都市計画・交通分野では「合併処理浄化槽⁹」のようなインフラに関わる語や「コミュニティバス」のような交通機関に関わる語が抽出された。表には掲載されていない語としては「工業用水道事業」「北関東自動車道」「北陸新幹線開業」「海岸堤防」などがある。観光・特産品については、女性議員では 0 語であるのに対して、男性議員では 10 語が抽出された。具体的には「観光戦略」「観光誘客」「県産品」などの語が含まれる。

女性議員は健康・福祉、出産・子育て、教育・研究に関する語を多く用いている。健康・福祉分野では表で挙げた語のほかにも「がん検診」「生活保護費」「重粒子線治療」「高齢者虐待」などの語が抽出され、女性議員が予防医療から貧困問題、障害者支援、介護問題といった健康・福祉分野の幅広い内容を重視していることがわかる。出産・子育て分野では男性の「民間移管¹⁰」の 1 語に対して、「妊婦検診」「放課後児童クラブ」「新生児訪問指導」「待機児童解消」「子育て世帯」など 32 語が抽出された。人権・男女共同参画については、男性議員には特徴語がなかったが、女性では 18 語が抽出された。具体的には「男女共同参画施策」「デートDV」「配偶者暴力相談支援センター」「女性活用」「女性相談センター」などの語が含まれる。主として女性が抱える問題は、女性議員によって議会で取り上げられているといえる。ただし、2.4 節で述べた通り、女性議員の数は男性議員の 1/10 程度にとどまり、女性議員が重要視する政治課題に関する発言の絶対量は少ない。

次に、政治特徴語の年層差を考察する（図 5-7、表 6-8 参照）。

若年議員は、健康・福祉、労働・商工業・エネルギー、出産・子育て、教育・研究、防災・災害対策の割合が高い。特に、健康・福祉はどの年層に対しても割合が高く、「コンビニ受診¹¹」といった新しい概念や、「介護者支援」「地域包括ケア」といった介護を担う側の発言がみられる。労働・商工業・エネルギー分野には、「Uターン情報センター」「求人活動」「新エネルギー技術」などの語が含まれる。地方の働き手不足や、最新技術への関心の高さが示唆される。

9 し尿と生活雑排水を併せて処理する浄化槽。

10 公立保育園の民営化。

11 一般的に外来診療をしていない休日や夜間に緊急性のない軽症患者が病院の救急外来を自己都合で受診する行為。

表 7 政治特徴語の政治分野別語数と具体例（中年・高年）

政治分野	中年		高年	
	語数	例	語数	例
くらし・すまい	0	-	5	建築住宅センター，市民活動支援センター，分譲推進
出産・子育て	4	児童委員，保育ニーズ，民間移管	1	子育て体験
人権・男女共同参画	1	女性消防団員	3	児童虐待対応，男女共同参画センター
教育・研究	0	-	4	へき地教育，教科書展示会，長寿大学
健康・福祉	6	救急救命士，手話通訳者，重粒子線治療，障害者就業	2	字幕放送，城山病院
労働・商工業・エネルギー	2	若年無業者，任期付職員	11	核燃料サイクル政策，技術系人材，高速増殖炉，六ヶ所再処理工場
観光・特産品	0	-	4	コンベンション協会，観光地域づくり
歴史・文化・スポーツ	2	フィルムコミッション，教会群	4	ぎふ清流大会，スポーツ事業，神宮式年遷宮
危機管理・安全保障	0	-	2	危機管理センター，軍民共用化
都市計画・交通	1	工業用水道	20	モビリティ，橋りょう，経過地，交通結節機能，高度処理型浄化槽，能越自動車道
防犯・交通安全	0	-	2	移動交番車，保護観察対象者
防災・災害対策	0	-	9	チリ津波，異常気象時，住宅用火災報知器
財政・税金	0	-	5	国際経済交流，特別会計決算，特別徴収義務者
自然・環境	0	-	5	育樹祭，生活排水対策，地下水保全，底泥
農林水産業	1	地域材	12	婦農塾，魚道設置，収造林，葉草栽培，林内路網
分類なし	4	天草地域，二項目め，府内各地	80	統計調査，附属機関等

中年議員は特徴語の少なさが目立つ。高年議員と比較すると 21 語が抽出されたものの，若年議員との比較では 1 語もない。中年議員は他の年層と比較して人数が多いため，発言内容が多岐にわたるためだと考えられる。特徴語の語数が少ないため単純な比較はできないが，高年議員と比べると，健康・福祉，出産・子育て，歴史・文化・スポーツの分野の割合が高い。

高年議員は，どの年層に対しても農林水産業の割合が高い。農林水産業では就業者の高齢化が進んでいることもあり，「新規就農者」「畜産振興」「婦農塾」「6次産業」などの語が目立つ。同年代の議員が議会でこれらの問題を取り上げているものと考えられる。財政・税金，自然・環境への関心は高く，出産・子育てには極端に関心がないことも特徴的である。

このように，属性の異なる議員がお互いに別の分野に注力することで，政策の立案や決定に多様な国民の意見を的確に反映させる効果があると考えられる。これらの分析結果は，地方議会議員の多様性を確保することの重要性を示すために有効である。

5. おわりに

地方議会会議録コーパスを利用して，全都道府県議会の議員を対象に議員の発言における性別・年層別の特徴を可視化した。性別・年層別の特徴語の抽出には，統計的な尺度である対数尤度比を用いた。抽出した語が政治課題に関連する語であるかどうかを判定するために，新たな尺度「政治語彙度」を提案した。その結果，1章で提示した Q1～Q3 に対して以下の結論を得た。

Q1 議会で活発に発言しているのはどの層の議員か？

最も活発に発言している層は 45～64 歳の女性議員である。ただし，女性議員の人数が極端に少ないため，発言量の総数は少ない。また，どの年層でも男性議員は女性議員よりも発言量が少なく，特に 65～84 歳の男性議員は一人当たりの発言量が最も少ない。本会議で 4 年の任期中に一度も発言していない議員が 59 名存在し，その多くは多選の男性議員である。

Q2 議員の性別によって発言内容に違いはあるか？

議員の性別によって発言内容の分野の違いがあることが明らかになった。男性議員では都市計画・交通，農林水産業，防災・災害対策，観光・特産品，女性議員では健康・福祉，出産・子育て，教育・研究，人権・男女共同参画が特徴的であった。

Q3 議員の年齢層によって発言内容に違いはあるか？

議員の年層による発言内容の違いも明らかになった。若年議員は，健康・福祉，労働・商工業・エネルギー，出産・子育て，教育・研究，高年議員は，農林水産業，財政・税金分野に重点を置いている。中年議員は発言の内容が特定分野に特化していないことも明らかになった。これは，中年議員は人数が多く，発言内容の幅が広いためだと考えられる。

日本は議員の多様性が低く，特に女性議員は世界的にみても非常に少ない。平成 30 年 5 月 23 日には議会選挙で男女の候補者数をできる限り「均等」にするよう政党に求める「政治分野における男女共同参画推進法」（候補者男女均等法）が公布・施行されるなど，改善へ向けた取り組みが進みつつある。本研究で得られた各属性の議員がもつ議会活動の特徴は，議員の多様性確保に向けた議論の材料となるだろう。

従来，議員の言動は政治学や政策学の分野で分析されてきた。[12] では，国會議員の委員会での発言が，議員の再選，党

表 8 政治特徴語の政治分野別語数と具体例（高年・若年）

政治分野	高年		若年	
	語数	例	語数	例
くらし・すまい	3	食肉センター，人口減少対策，地籍調査	1	食育
出産・子育て	0	-	5	窓口無料化，多子世帯，病後児保育
人権・男女共同参画	1	男女共同参画計画	1	デートＤＶ
教育・研究	1	通学区域	5	コミュニティスクール，出前講座，中学校給食
健康・福祉	3	オストメイト，県民福祉，鳴門病院	16	コンビニ受診，自殺対策，手話通訳，地域包括ケア，認知症サポーター
労働・商工業・エネルギー	9	原子力施設，新規就業者，地方公務員給与	6	グローバルな人材，ジョブカフェ，テレワーク，通信事業者
観光・特産品	3	観光づくり，観光誘客，観光力	0	-
歴史・文化・スポーツ	3	ぎふ清流国体，屋内スケート場，国民文化祭	4	芝生化，障害者スポーツ，多文化共生，府立体育館
危機管理・安全保障	2	米軍再編，辺野古移設	0	-
都市計画・交通	9	中部横断自動車道，水道事業会計，送水管	5	市有地，自転車利用者，除排雪
防犯・交通安全	1	交通事故防止	1	刑法犯認知件数
防災・災害対策	18	S P E E D I，オフサイトセンター，広域防災拠点，山地災害，津波浸水	13	帰宅困難者，首都直下地震，消防団，防災教育
財政・税金	12	決算剰余金，県内景気，個人住民税	1	徴収率
自然・環境	5	高レベル放射性廃棄物，水源地域，法定検査	2	レジ袋，全国都市緑化フェア
農林水産業	21	集落営農組織，森林整備，肉用牛，分収林，薬用作物	0	-
分類なし	30	伊豆地域，議会運営，整備促進	27	パッケージ案，啓発活動，県民局長

表 9 政治特徴語の精度

	完全一致	同義	類義	存在しない	精度
①対数尤度比	5	2	3	90	10.0%
②政治関連語	14	3	17	59	36.6%
③政治特徴語	18	12	19	51	49.0%

内での出世，専門性の発揮に有用であることを論じている。[13]では，北関東主要7市議会の議員発言を対象として，政治と行政の関係を分析している。[14]では，3つの府県議会議員選挙および3つの市議会議員選挙の候補者が作成した選挙公報を対象として，女性政策を推進する政治家の特徴を分析している。このように，様々な観点から議員の活動を分析する試みがあるが，それらは国会議員を対象としたもの，一部の地方議員を対象としたもの，特定の政治課題のみを扱うものである。本研究は，全国の都道府県議会を対象とし，4年間の本会議における全発言を分析している点に優位性がある。

また，本研究で構築した地方議会会議録コーパスを利用した研究も進められている。計量経済学分野では，議会ごとの発言文字数を計量的に分析することにより，議会の活発さを明らかにしている[15]。[16]は，本コーパスを政治学的に利用する方法を検討しており，具体的な利用法として，被災地における「震災記憶等の風化」や「風評被害」の存在を確認すること，議員の公約の遵守状況を確認することを挙げている。さらに，本コーパスが“議員の資質を評価することにも応用可能と言えるかもしれない”と述べている。本研究で行った政治語彙度による分析は，これらの分野における研究への貢献も期待できる。

今後は，本論文で提案した政治語彙度を利用して，議会活動の地域差を可視化することを計画している。また，本研究で得

られる知見を有権者に提供することも視野に入れている。

謝辞

本研究は JSPS 科研費 JP16H02912, JP17K02739 の助成を受けたものである。本研究の一部は，国立国語研究所 共同研究プロジェクト 機関拠点型基幹研究プロジェクト「議会会議録を活用した日本語のスタイル変異研究」の研究助成を受けたものである。

参考文献

- [1] 木村泰知: “議論の背景・過程・結果を関連づける地方政治コーパスの構築に向けて,” 電子情報通信学会言語理解とコミュニケーション (NLC), 思考と言語 (TL) 合同研究会, 2016.
- [2] 高丸圭一, 内田ゆず, 木村泰知: “地方政治コーパスにおける都道府県議会会議録パネルデータの基礎分析,” 宇都宮共和大学シテライフ学論叢, Vol.18, pp. 136-155, 2017.
- [3] 内田ゆず, 高丸圭一, 乙武北斗, 木村泰知: “都道府県議会会議録コーパスを用いた議員の議会活動の可視化に向けて,” 人工知能学会全国大会 (第 32 回), 1E3-03, 2018.
- [4] 日本速記協会: 発言記録作成標準, 日本速記協会, 2007.
- [5] 内山将夫, 中條清美, 山本英子, 井佐原均: “英語教育のための分野特徴単語の選定尺度の比較,” 自然言語処理, Vol.11, No.3, pp. 165-197, 2004.
- [6] K. Chujo and M. Utiyama: “Selecting Level-Specific Specialized Vocabulary Using Statistical Measures”, *System*, Vol.34, No.2, pp. 255-269, 2006.
- [7] 金愛蘭, 桐生りか, 近藤明日子, 田中牧郎: “「一般向け専門用語」抽出の試み - 医療用語を例に -,” 日本語学会 2008 年度春季大会予稿集, pp. 199-206, 2008.
- [8] 中條清美, 田辺和子, 木下謙朗, 三橋麻子, 西垣知佳: “コーパスを活用した日本語教材作成の試み,” 日本大学生産工学部研究報告 B, Vol.42, pp. 43-52, 2009.
- [9] T. Kudo, K. Yamamoto, and Y. Matsumoto, “Applying Condi-

tional Random Fields to Japanese Morphological Analysis,” *Proc. of the 2004 Conf. on Empirical Methods in Natural Language Processing (EMNLP-2004)*, pp. 230-237, 2004.

- [10] T. Sato, T. Hashimoto, and Manabu Okumura: “Implementation of a word segmentation dictionary called mecab-ipadic-NEologd and study on how to use it effectively for information retrieval (in Japanese),” *Proc. of the Twenty-three Annual Meeting of the Association for Natural Language Processing, NLP2017-B6-1*, 2017.
- [11] 国立国語研究所コーパス開発センター: “『現代日本語書き言葉均衡コーパス利用の手引き』,” 第 1.0 版, 2011.
- [12] 松本俊太, 松尾見孝: “国会議員はなぜ委員会で発言するのか?,” 選挙研究, Vol.26, No.2, pp. 84-103, 2010.
- [13] 増田正: “我が国地方議会における政治・行政関係の計量テキスト分析,” 『地域政策研究』(高崎経済大学地域政策学会), Vol.20, No.3, pp. 1-19, 2018.
- [14] 坂本治也: “地方議会選挙と女性政策: 選挙公報を用いた試論的分析,” 関西大学法学研究所研究叢書第 58 冊『地方議会研究の新展開』, pp. 19-48, 2018.
- [15] A. Kawaura, Y. Kimura, K. Takamaru, and Y. Uchida: “Elected Officials in the Local Assembly: Analysis of Prefectural Plenary Session Transcripts,” 日本経済学会 2018 年度春季大会, 2018.
- [16] 河村和徳, 伊藤裕顕: “被災地選挙の諸相 (41) 都道府県議会会議録検索システム「ぎーみる」,” 月刊選挙, 2018 年 8 月号, pp. 6-12, 2018.

(2018 年 5 月 31 日 受付)
(2018 年 12 月 14 日 採録)

[問い合わせ先]

〒064-0926 札幌市中央区南 26 条西 11 丁目 1 番 1 号

北海学園大学工学部

内田 ゆず

TEL: 011-841-1161

E-mail: yuzu@hgu.jp

—— 著 者 紹 介 ——

うちだ ゆず [正会員]

2010 年北海道大学大学院情報科学研究科メディアネットワーク専攻博士後期課程修了。同年青山学院大学理工学部助手。2012 年同助教。2014 年北海学園大学工学部准教授。現在に至る。博士(情報科学)。自然言語処理, 特に日本語オノマトペ, 政治分野のテキストを対象とした研究に従事。日本知能情報ファジィ学会, 電子情報通信学会, 情報処理学会, 言語処理学会, 人工知能学会各会員。

たかまる けいち [非会員]

2004 年北海道大学大学院工学研究科博士後期課程単位修得退学。同年那須大学都市経済学部専任講師, 2013 年宇都宮共和大学シテイライフ学部准教授, 2018 年同教授(現在)。修士(工学), 博士(応用言語学)。自然言語処理, 音声学, 社会言語学, コーパス言語学などの研究に従事。電子情報通信学会, 人工知能学会, 社会言語科学会, 日本音声学会, 日本語学会各会員。

おとたけ ほとと [正会員]

2010 年北海道大学大学院情報科学研究科博士後期課程修了。同年, 福岡大学工学部助教着任。現在に至る。博士(情報科学)。自然言語処理, 教育工学などの研究に従事。言語処理学会, 情報処理学会, 人工知能学会, 日本知能情報ファジィ学会, 電子情報通信学会各会員。

きむら やすとも [正会員]

2004 年 09 月 北海道大学大学院工学研究科電子情報工学専攻 修了。2005 年 4 月-2007 年 3 月 小樽商科大学助教授。2007 年 4 月-2018 年 9 月現在 小樽商科大学 准教授。2018 年 10 月-現在 小樽商科大学教授。2010 年 10 月-2011 年 09 月 New York University 訪問研究員。2017 年 10 月-現在 理化学研究所 AIP センター客員研究員。自然言語処理の研究, 特に, 地方議会会議録を対象としたコーパス構築の整備に従事。

Visualization of Activity of Local Assembly Members Based on Log-Likelihood Ratio and Political Vocabulary Ratio

by

Yuzu UCHIDA, Keiichi TAKAMARU, Hokuto OTOTAKE, and Yasutomo KIMURA

Abstract:

We created a corpus for prefectural assembly minutes. Each record in the corpus has a field of "speaker's name." This field is manually checked to improve credibility of the corpus. The corpus also has attributes of speakers (namely, electoral district, birth year and gender) in the case of assembly members. We try to analyze the corpus and illustrate activities of all assembly members in Japan. This paper describes: 1) difference in the amount of speech depending on age and gender, 2) difference in the contents of speech depending on gender, 3) difference in the contents of speech depending on age.

Keywords: prefectural assembly minutes, corpus, statistical measures

Contact Address: Yuzu Uchida

Hokkai-Gakuen University

1-1 Nishi 11-chome, Minami 26-jo, Chuo-ku, Sapporo, Hokkaido, 064-0926, Japan

TEL: +81-11-841-1161

E-mail: yuzu@hgu.jp