

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 What Can Be Seen from the Prefectural Assembly Minutes?

内田 ゆず^{*1}

Yuzu Uchida

^{*1} 北海学園大学

Hokkai-Gakuen University

乙武 北斗^{*3}

Hokuto Ototake

^{*3} 福岡大学

Fukuoka University

高丸 圭一^{*2}

Keiichi Takamaru

^{*2} 宇都宮共和大学

Utsunomiya Kyowa University

木村 泰知^{*4}

Yasutomo Kimura

^{*4} 小樽商科大学／理研 AIP

Otaru University of Commerce / RIKEN AIP

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) the number of assembly members in Japan, 2) age and gender composition of assembly members, 3) difference in the amount of speech depending on age and gender, 4) difference in the contents of speech depending on gender, 5) difference in the contents of speech depending on age.

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出来るようになった。著者らは現在新たに、地方政治における議論の過程と結果、およびその背景を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」の構築を進めている[木村 2016]。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を収集している。

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。[高丸 2017]では、各都道府県議会での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。このほか、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する研究を進めている[Kawaura 2018]。

著者らが構築した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)を付与している。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。

本稿で可視化を試みる都道府県議会議員に対する具体的な問いは以下のとおりである。

- Q1: 全国の都道府県議会には何人の議員がいるか？
- Q2: 議員の年齢・性別構成はどうなっているか？
- Q3: 議会で活発に発言しているのはどの層の議員か？
- Q4: 議員の性別によって発言内容に違いはあるか？
- Q5: 議員の年齢層によって発言内容に違いはあるか？

これらの問いについて、機械学習などの手法を適用する以前の基礎分析として、発言者情報や単語頻度を単純な方法で集計し、構築したコーパスからの議員活動の可視化について、見通しを示すことを本稿の目的とする。2 章ではコーパスの概要、発言者名および属性情報の付与について述べる。その上で、属性情報を集計し Q1, 2 についての分析結果を述べる。3 章では、発言文字数に基づく分析を行い、Q3 についての分析結果を述べる。4 章では、Q4, 5 に答えるための基礎分析の結果を示す。

2. 地方議会会議録コーパスの概要

2.1 収集方法

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。それらをクロールして、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 年分の全国 47 都道府県議会の本会議会議録を収集した。

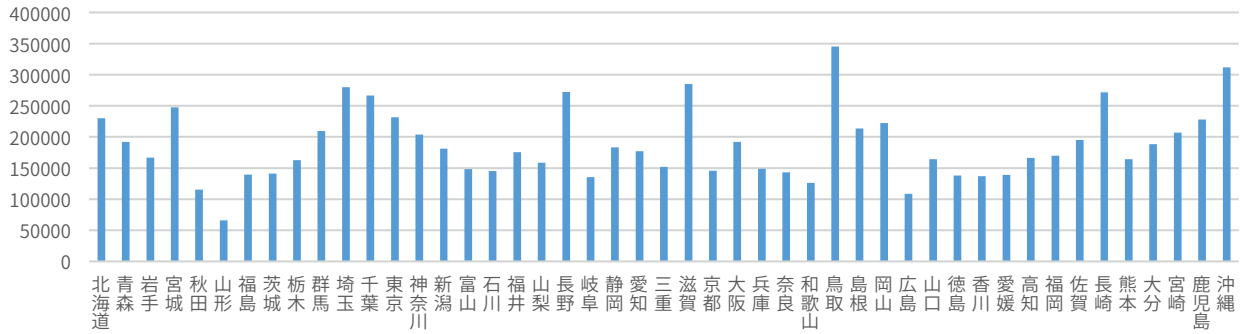


図 1 各都道府県議会における発言文字数

2.2 データ量

会議録を句点で1発言ごとに分割し、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発言を1レコードとして格納した。総収録データは1,651,676レコードであり、発言の総文字数は92,290,795文字である。

各都道府県の本会議における総発言文字数を図1に示す。発言量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はみなされない。

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2,793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(例えば、選挙ドットコム¹⁾)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[高丸 2017])。

¹ <http://go2senkyo.com/articles/>

² 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)

表 1 地方議会会議録コーパス収集期間の議員の構成

発言者	25-44 歳	45-64 歳	65-84 歳	不明	計
男	547	1,616	457	71	2,691
女	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

発言者	25-44 歳	45-64 歳	65-84 歳	不明	全体
男	43,794	44,168	31,888	41,308	41,931
女	43,901	61,465	51,081	39,124	56,496
全体	43,803	45,849	32,884	41,164	43,192

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低い²ことが知られている。地方議会においても女性議員は男性議員の1/10程度にとどまり、同様の傾向が見られた。

3. 議員活動の分析(1) 発言文字数

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。

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、どの年齢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。

ここでは単純に発言文字数のみを比較しているため、この結果から「女性議員の方が活発に発言している」と結論付けることはできない。女性の方が丁寧な表現を使用するため、発言が長くなる傾向があるとも考えられる。今後、属性による文末表現の違いなどを分析したい。

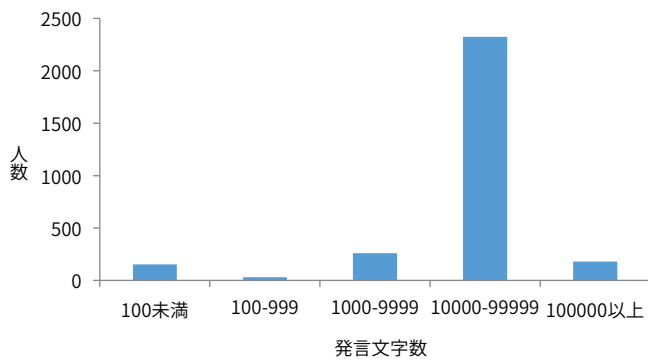


図2 発言文字数別議員数

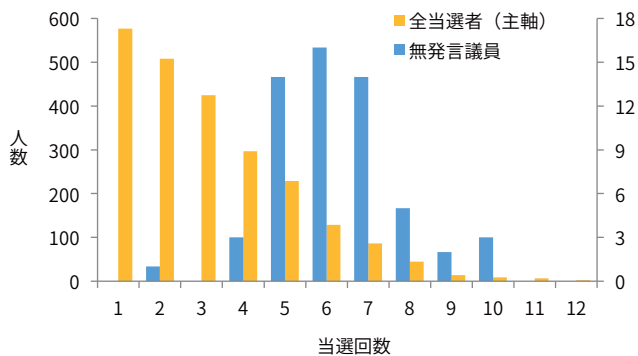


図3 議員の当選回数

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43,192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182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151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³、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58名は4年間の任期中に議員として一度も発言していないことになる。

上述の58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」⁴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54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⁵により議員の固定化が進む中、多選議員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ときに有益だと考えられる。

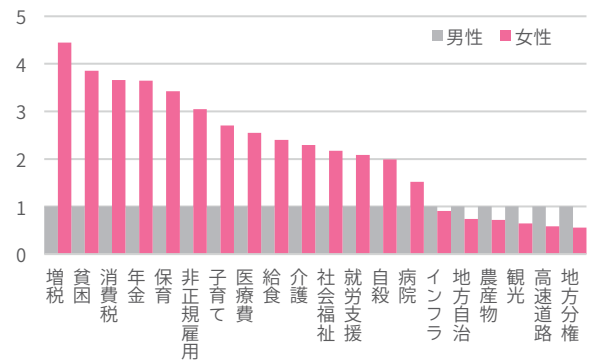


図4 性別によるキーワードの出現割合の比較

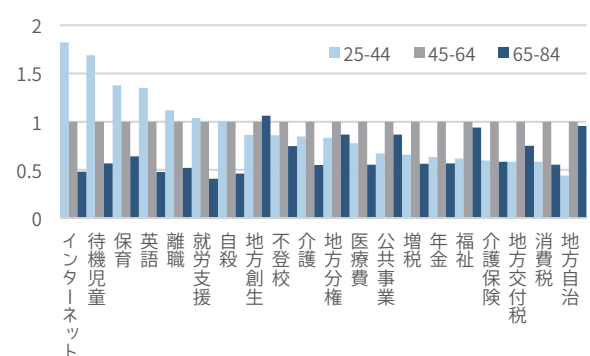


図5 年層によるキーワードの出現割合の比較

4. 議員活動の分析(2) キーワードの発言回数

議員によって関心のある政治課題は異なる。ここでは、政治課題に関連するキーワードの出現割合を属性ごとに比較する。キーワードは第1著者が選択した。

4.1 性別で比較

図4は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」を示している。女性議員が「増税」というキーワードを発言する回数は、男性議員の4.4倍にのぼる。その他、「貧困」「消費税」「年金」「保育」など、生活に密着した問題を取り上げるのは女性議員の特徴である。一方、男性議員は「地方分権」「高速道路」「観光」など、行政や産業分野への言及が多いことがわかる。

経済や産業も重要な政治課題であるものの、市民にとっては健康や子育て、家計のような身近な課題も大きな関心事である。現状では都道府県議会議員の9割以上が男性議員であり、政策の決定過程において女性の割合が極端に少ない。この分析結果は、「女性議員を増やすと身近な政治課題が議会で取り上げられ、民意が政策に反映されやすくなる」といった議論の材料になると考えられる。

4.2 年齢層で比較

図5は、各キーワードにおいて「45～64歳の年齢層を1とした時の他の年齢層の発言回数」を示している。たとえば、45～64

³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。

⁴ <http://www.yomiuri.co.jp/>

⁵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%。

歳の議員と比べて 25～44 歳の議員は「インターネット」というキーワードを 1.8 倍発言しており, 65～84 歳の議員は 0.5 倍発言していることがわかる。「待機児童」「保育」「英語」も若い議員ほど多く発言しているという同じ傾向が見られる。頻出キーワードをから, 25～44 歳の議員は子育て世代を代表した発言が多いと推測される。45～64 歳の議員は「年金」「介護保険」「消費税」などのキーワードを発言する回数が多く, 働き盛り世代が関心をもつ話題に言及していることがわかる。65～84 歳の議員には, 特徴的なキーワードが見つからなかった。表 1, 2 で示した通り, この年層の議員は人数が少ない上に発言量も少ない。そのため, 発言回数に差が生じるようなキーワードが出現しにくいと考えられる。

5. おわりに

著者らが構築した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会活動の可視化を目指した分析を行った。その結果, 最も発言量が多いのは 45～64 歳の女性議員であるが女性議員の人数が極端に少ないこと, 本会議で任期中に一度も発言していない議員がいること, 性別や年層によって発言内容に違いがあることなどが明らかになった。

本稿では発言文字数や単語の出現頻度などのシンプルな指標を用いた基礎分析を行ったが, 今後はより詳細な分析を行いたい。特に, キーワード分析については, 対数尤度比などの指標を用いて, より客観的な分析を行う予定である。さらに, 国会会議録との比較も検討している。

謝辞

本研究は JSPS 科研費 JP16H02912, JP17K02739 の助成を受けたものである。

参考文献

- [木村 2016] 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.
- [高丸 2017] 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ィライフ学論叢, 18, 136-155, 2017.
- [Kawaura 2018] Akihiko Kawaura, Yasutomo Kimura, Keiichi Takamaru, Yuzu 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(発表予定)