

都道府県議会会議録コーパスの拡張 —2011期と2015期の比較—

内田ゆず
北海学園大学
yuzu@hgu.jp

高丸圭一
宇都宮共和大学
takamaru@kyowa-u.ac.jp

乙武北斗
福岡大学
ototake@fukuoka-u.ac.jp

木村泰知
小樽商科大学
kimura@res.otaru-uc.ac.jp

1 はじめに

近年、インターネットに公開された議会会議録を利用した学術研究がさまざまな分野で盛んに行われている [1][2]. 自然言語処理分野ではこれまでに要約 [3], トピック分析 [4], 意見抽出 [5] などの研究が行われている. 議会における発言の記録である会議録は, 地方自治におけるファクトチェックのための一次情報としての利用が期待される. これを指向して, NTCIR-15 QA Lab-PoliInfo-2 では, 東京都議会を対象として, 議案への賛否推定 (Stance classification), 質問と答弁の要約 (Dialog summarization) などのタスクが行われた [6].

筆者らは横断的に収集した会議録から「いつ」「どこで」「誰が」「何を」発言したのかを明らかにするために, すべての発言文にタグを付与した, 地方議会会議録コーパスの構築を進めてきた. 種々の分野で会議録を研究対象とするとき, 各種の統計量 (たとえば, 都道府県 P での 1 会議あたりの平均発言者数) を算出したり, ある一定の条件下での発言 (たとえば, 単語 w を含む議員 A の発言) を抽出したりするという利用方法が考えられる. これを実現するためには, 収集範囲および収集期間の統制がとれたコーパスサブセットが必要である. そこで, 収集範囲を「全国 47 都道府県議会の本会議」, 収集期間を「統一地方選挙を起点とした 4 年間」と定めた都道府県議会会議録コーパスサブセットを順次構築している. 2017 年に最初のサブセットである「47pref_2011-2015」(2011 期と呼ぶ) および横断検索システム「ぎ〜みる」を公開した. 新たに 2015 年 4 月から 2019 年 3 月までの 4 年間の会議録を収録した 2 期目 (2015 期と呼ぶ) のサブセット「47pref_2015-2019」を構築した. 本稿では, 2011 期と 2015 期のサブセットを比較し, 議員の構成, 発言量, キーワードの発言回数などについて述べる.

2 都道府県議会会議録コーパスの概要

2.1 データの収集

収集対象は全国 47 都道府県議会の本会議 (定例会および臨時会) であり, 収集期間は 2015 年 4 月の統一地方選挙から 2019 年 4 月の統一地方選挙の前 (2019 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. 各全文検索システムに対応するクローラを作成して, 会議録データの収集を行った. 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, 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.

収集された全国 47 都道府県の本会議会議録を句点で 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, 会議の名称や開催日等の情報と併せて 1 発言を 1 レコードとして格納した (詳細は [7] 参照). 総収録データは 3,873,026 レコードであり, 発言の総文字数は 248,281,520 文字である.

2.2 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, 衛藤明和氏は「衛藤明和議員」「衛藤明和文教警察委員長」「衛藤明和土木建築委員長」のように肩書が異なる表記や, 旧字体を使用した表記をもつ. さらに, 「岩切達哉/岩切達郎」「有岡浩一/有岡浩一」「太田清海/太田晴海¹⁾」のように, 会議録作成者の誤入力による表記揺れも存在する.

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

1) いずれも宮崎県議会会議録に実際に存在する例

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 7,606 名であった。

2.3 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を付与している。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 Web サイトである「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。さらに、議員には固有の ID を付与した。議員 ID によって、サブセット間で議員の発言を追跡することが可能となる。

以上の処理により、2015 期は、2,745 名の議員、55 名の知事、157 名の副知事、398 名の議長が発言していることが明らかになった。

3 2011 期と 2015 期の比較

2011 期、2015 期のデータについて、議員の構成、発言量、キーワードの発言回数の観点で比較する。なお、岩手県、宮城県、福島県、茨城県、東京都、沖縄県は統一地方選挙とは異なる日程で選挙が行われることに注意が必要である。

3.1 議員の構成

全国の都道府県議会議員の定数を合計すると 2,686 人だが、改選や補欠選挙による議員の入れ替わりが生じるため 2015 期の議員は 2,863 人存在する。改選による入れ替わりが 100 名、補欠選挙による入れ替わりが 77 名である。

議員の年齢³⁾・性別の構成を表 1 に示す。表中の括弧内の数字は、2011 期からの増減を表している。

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

3) 2015 年 12 月 31 日時点

高齢化が進んでいること、女性議員の割合はやや増加（8.7%→10.3%）していることが明らかになった。また、2011 期のデータを収集した時点では政治山から生年の情報を取得できない議員が 76 名存在したが、2015 期では 0 名であった。地方議員の Web 上での情報発信が促進されたことがうかがえる。

3.2 発言量

議会や議員の活動を可視化するために、発言量を都道府県別、属性別、個人別に分析する。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。都道府県議会会議録における文字遣いは「発言記録作成標準 [8]」などによって標準化されている。したがって、各都道府県で表記の方法が大きく異なることはないと判断し、分析の単位を文字とする。

全体 2011 期と比較して、総発言文字数は 252,364,746 から 248,281,520 へと微減した。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。発言数が最も多いのは鳥取県で、最も少ない山形県の約 4.5 倍に達する。これは、2011 期とほぼ同じ結果である。

2011 期と 2015 期を比較して発言文字数が大きく変化したのは、大阪府（25.1%減）、兵庫県（24.5%増）、長崎県（35.2%減）、沖縄県（34.9%増）であった。役職ごとに発言文字数を集計すると、大阪府は、知事の発言文字数が 611,715 から 247,961 と大きく減少している。沖縄県では、職員の発言文字数の増加が顕著である。沖縄県は、2011 期の職員の発言文字数が 3,088,869 で全都道府県議会中最多であったが、2015 期ではさらに 4,807,870 へと増加している。

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。2011 期と 2015 期を比較すると、高齢の男性議員の発言量が顕著に少ない点、女性議員の発言量が多い点は共通している。発言量に大きな変化があったのは 45-64 歳の女性議員と 65-87 歳の女性議員である。45-64 歳の女性議員の発言量は減り、65-87 歳の女性議員の発言量は増えている。2015 期に 100,000 字以上発言している女性議員 32 名のうち、23 名は 2011 期でも議員を務めている。この 23 名について、2011 期での発言文字数を遡って調査したところ、89,107~311,442 字であった。2011 期で活発に発言した議員は 2015 期でも発言量が多いといえる。2011 期の 25-44 歳と

	25-44歳	45-64歳	65-87歳	不明	計
男	361 (-186)	1,555 (-61)	653 (+196)	0 (-71)	2,569 (-122)
女	47 (-4)	202 (+28)	45 (+20)	0 (-5)	294 (+39)
計	408 (-190)	1,757 (-33)	698 (+216)	0 (-76)	2,863 (-83)

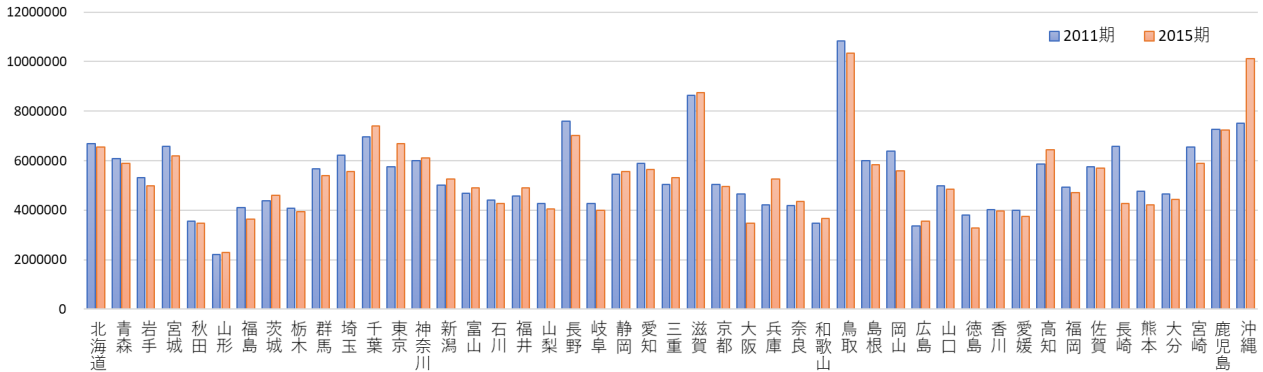


図1 各都道府県議会における発言文字数

45-64歳の女性議員の発言量の傾向と併せて考えると、女性議員の発言量の変化は議員が年齢を重ねたことが要因だと考えられる。

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によると全議員の平均発言文字数は42,661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119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118名について詳しく調査したところ、21名はコーパス収集期間と任期が一致しておらず⁴⁾、21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42名を除いた76名は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。このうち、男性は73名、女性は3名である。

無発言の76名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2015期に実施された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政治山、Webサイト「読売オンライン」の地方選挙情報ページ⁵⁾などから取得した。全議員では当選回数が増えるほど人数は減る一方、無発言の議員は当選回数が多い傾向があることがわかる。2011期においても、発言のなかった59名のうち58名は多選（4回以上）の男性議員であり、同様の傾向がみられた。

4) 期間中に落選、当選、辞職、失職、死亡、あるいは逮捕されたため。

5) <https://www.yomiuri.co.jp/election/local/>

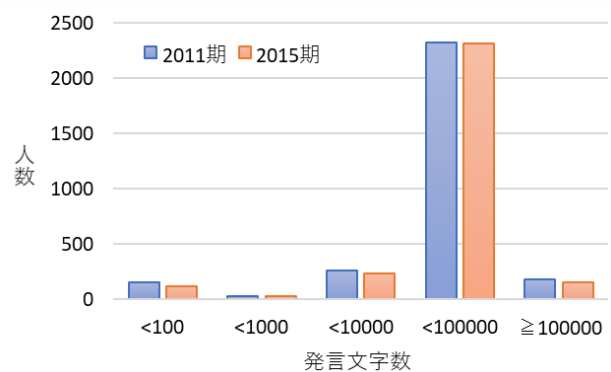


図2 発言文字数別議員数

3.3 キーワードの発言回数

我々はこれまでに、2011期のデータを対象として、議員の属性と政治課題に関するキーワードの出現回数の関係を調査し、特定のキーワードの出現傾向が議員の性別や年齢層によって大きく異なることを明らかにした[9]。2015期のデータについても[9]と同じキーワードを用いて同様の分析を行う。

図4は各キーワードの「男性議員を1とした時の女性議員の発言回数の割合」を示している。2011期で女性議員が「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倍であった。2015期ではさらにその差が開き、6.5倍となっている。「貧困」「就労支援」など、男性の発言割合が相対的に増えているキーワードも存在するが、生活に密着した政治課題への言及は依然として女性議員による言及が

表2 2011期と2015期の議員属性別平均発言文字数

	2011期					2015期				
	25-44歳	45-64歳	65-87歳	不明	計	25-44歳	45-64歳	65-87歳	不明	計
男	43,873	44,159	31,808	40,278	41,901	46,056	45,193	29,888	0	41,424
女	43,901	61,465	51,081	45,450	56,620	42,680	54,590	59,717	0	53,471
計	43,875	45,841	32,808	40,618	43,175	45,667	46,273	31,811	0	42,661

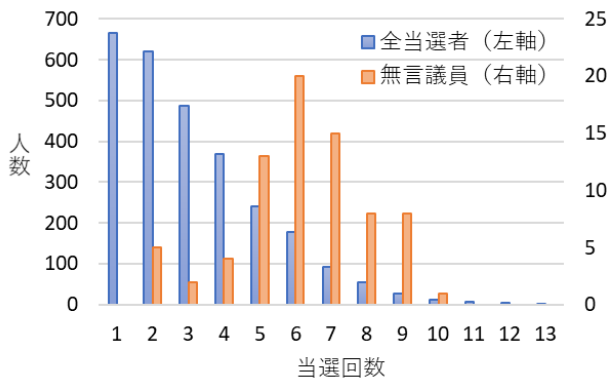


図3 議員の当選回数

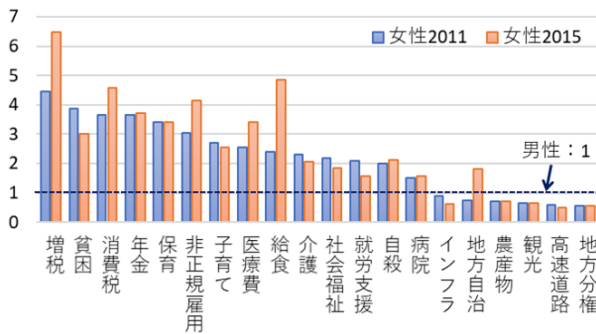


図4 性別によるキーワードの出現割合の比較

多い。男性議員は「地方分権」「高速道路」「インフラ」「観光」など、行政や産業分野への言及が多く、これも2011期と同様の特徴である。

以上より、議員の性別によって発言内容の特徴が異なることが示唆される。また、女性議員の数が圧倒的に少ないため、議会で取り上げられる政治課題に偏りが生じている可能性がある。今後、議員の年齢層による発言内容の特徴も分析したい。

4 おわりに

2015年4月から2019年3月に開催された全国47都道府県議会の会議録データを収集し、都道府県議会会議録コーパスを拡充した。本稿では、新たに追加したデータの概要と、2011期と2015期のデータの比較について報告した。今後は、発言文字数や単語の出現頻度などの表層的な分析だけでなく、より

詳細な分析を行いたい。また、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18K00632, 20K00576およびセコム科学技術振興財団の助成を受けたものである。

参考文献

- [1] 渡部春佳. 地方議会議事録分析による話題抽出についての一試論—地方議会・委員会での公の施設「劇場・音楽堂等」に関する議論を事例に. 社会情報学, Vol. 9, No. 1, pp. 1–15, 2020.
- [2] 桑田倫幸, 近田裕子, BATUER JULAITI, 森岡友紀, 山口明, 木佐谷康, 信田勝美, 板倉宏昭. 都市型観光における商店街の地域性—東京都品川区議会会議録のテキスト分析からの考察を中心に—. 第11回横幹連合コンファレンス, 2020.
- [3] 木村泰知, 関根聡, 乾健太郎. 地方議会会議録の要約に向けて. 言語処理学会第24回年次大会 (NLP2018), pp. 596–599, 2018.
- [4] 佐々木稔, 乙武北斗, 木村泰知. 大規模地方議会会議録の分散表現を用いた地方議会のトピック分析. 2020年度人工知能学会全国大会論文集 (第34回), 2020.
- [5] 葦原史敏, 木村泰知, 荒木健治. 地方議会会議録における節単位による議員の要望抽出. 電子情報通信学会論文誌, Vol. J98-D, No. 11, pp. 1390–1401, 2015.
- [6] Yasutomo Kimura, Hideyuki Shibuki, Hokuto Ototake, Yuzu Uchida, Keiichi Takamaru, Madoka Ishioroshi, Teruko Mitamura, Masaharu Yoshioka, Tomoyosi Akiba, Yasuhiro Ogawa, Minoru Sasaki, Kenichi Yokote, Tatsunori Mori, Kenji Araki, Satoshi Sekine, and Noriko Kando. Overview of the NTCIR-15 QA Lab-PoliInfo-2 Task. In *Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies*, 2020.
- [7] 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アライブ学論叢, Vol. 18, pp. 136–155.
- [8] 日本速記協会. 発言記録作成標準: 公正・正確な発言記録を目指して. 日本速記協会, 2007.
- [9] 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—What Can Be Seen from the Prefectural Assembly Minutes? 2018年度人工知能学会全国大会論文集 (第32回), 2018.

A 付録

都道府県議会会議録コーパス 2015 期のデータの概要を以下に示す。

都道府県	議員定数	会議録中の人数			発言レコード数				発言文字数			
		議員	知事	副知事	総数	議員	知事	副知事	総数	議員	知事	副知事
北海道	101	103	1	5	99,099	39,146	13,289	1,218	6,552,866	3,212,848	1,548,589	149,909
青森県	48	48	1	2	90,553	42,713	7,442	888	5,901,892	2,640,319	640,143	89,517
岩手県	48	53	1	2	76,887	31,869	6,530	197	4,979,445	2,016,231	683,462	22,280
宮城県	59	74	1	5	103,606	52,127	15,633	54	6,190,304	3,252,721	1,230,750	2,966
秋田県	43	46	1	1	55,796	25,087	11,947	14	3,468,511	1,627,936	1,089,338	876
山形県	44	45	1	2	33,290	14,037	3,550	14	2,279,419	959,611	292,322	674
福島県	58	72	1	2	62,668	34,014	4,511	95	3,647,078	2,108,170	331,601	7,006
茨城県	62	78	2	4	71,888	29,667	10,532	36	4,610,148	1,952,100	906,203	1,826
栃木県	50	52	1	5	49,190	12,867	3,086	96	3,936,740	2,166,974	433,317	14,860
群馬県	50	53	1	4	59,382	14,529	1,366	50	5,408,859	2,273,168	234,143	9,446
埼玉県	93	95	1	4	107,784	44,904	10,243	233	5,549,081	2,508,299	590,617	13,729
千葉県	95	97	1	0	70,816	28,412	3,615	0	7,408,536	4,284,295	534,949	0
東京都	127	175	2	9	135,680	43,909	14,786	666	6,695,650	2,527,824	925,912	33,989
神奈川県	105	105	1	3	108,550	56,278	24,201	79	6,114,639	3,649,304	1,498,938	4,536
新潟県	53	55	3	8	80,991	40,252	21,864	76	5,250,208	2,684,243	1,748,860	3,025
富山県	40	44	1	2	72,129	25,780	11,921	59	4,895,032	1,608,982	1,298,330	4,304
石川県	43	46	1	2	39,796	10,464	4,032	19	4,272,614	1,764,397	864,926	1,480
福井県	37	37	1	4	73,036	27,622	10,154	283	4,910,456	2,327,566	956,826	29,844
山梨県	38	40	2	3	72,315	26,134	9,770	9	4,061,280	1,766,912	820,859	295
長野県	58	60	1	2	84,561	26,540	10,346	417	7,013,601	3,057,202	1,122,850	41,093
岐阜県	46	46	1	5	62,492	35,686	6,705	109	3,982,943	2,420,557	520,042	7,504
静岡県	69	73	1	5	93,249	40,932	16,216	2,119	5,566,865	2,778,533	1,043,354	156,719
愛知県	102	102	1	7	83,592	49,538	8,696	147	5,651,399	3,443,272	680,092	9,320
三重県	51	52	1	3	84,093	43,368	7,228	136	5,305,625	2,843,589	590,841	7,085
滋賀県	44	46	1	3	93,524	30,831	17,001	189	8,747,162	4,226,431	2,057,885	21,839
京都府	60	63	2	3	71,463	40,833	12,357	20	4,949,577	2,851,039	1,212,267	995
大阪府	88	94	1	4	60,236	30,440	3,741	98	3,479,854	1,964,362	247,961	7,849
兵庫県	87	90	1	3	68,984	23,245	10,116	766	5,260,147	2,753,322	968,187	78,773
奈良県	44	44	1	5	73,989	36,414	20,562	62	4,353,413	2,305,394	1,324,109	2,989
和歌山県	42	43	1	1	65,608	34,906	5,118	27	3,672,947	2,165,253	438,490	1,852
鳥取県	35	35	1	2	156,995	63,057	60,336	219	10,328,438	3,822,897	4,359,090	18,431
島根県	37	37	1	2	106,565	45,263	10,478	298	5,837,247	2,645,101	717,656	18,326
岡山県	55	56	1	4	99,426	50,839	13,849	174	5,576,495	3,197,973	1,118,948	11,924
広島県	64	64	1	2	54,892	24,517	7,063	36	3,555,378	1,711,839	700,321	3,333
山口県	47	49	1	2	78,512	41,527	6,821	50	4,835,839	2,653,770	596,818	3,942
徳島県	39	39	1	3	53,264	24,755	6,970	983	3,291,457	1,544,583	741,796	104,433
香川県	41	43	1	2	64,981	31,053	9,517	23	3,978,312	2,276,056	948,711	1,157
愛媛県	47	49	1	5	53,418	17,596	4,375	131	3,756,833	1,987,629	532,359	11,941
高知県	37	38	1	1	88,149	38,964	15,366	216	6,430,533	2,688,822	1,333,994	18,939
福岡県	86	87	1	1	80,252	40,968	21,511	1,477	4,723,003	2,456,317	1,588,720	109,614
佐賀県	38	38	1	2	96,551	44,859	14,131	252	5,701,760	2,542,375	1,023,145	18,024
長崎県	46	47	1	4	69,218	35,933	6,620	141	4,266,846	2,185,641	577,954	12,842
熊本県	48	50	1	3	73,849	42,654	6,626	198	4,203,940	2,785,113	369,166	11,762
大分県	43	44	1	3	83,376	38,910	11,693	18	4,443,794	2,379,557	757,904	708
宮崎県	39	41	1	4	92,758	58,526	7,580	394	5,888,025	3,361,838	599,282	30,765
鹿児島県	51	51	2	5	118,743	61,632	10,316	137	7,225,976	3,739,742	734,469	9,293
沖縄県	48	64	2	4	196,830	87,211	10,328	1,957	10,121,353	4,018,681	752,538	130,551
合計	2,686	2,863	55	157	3,873,026	1,740,808	530,138	14,880	248,281,520	122,138,788	44,289,034	1,242,565