

都道府県議会会議録を対象とした議題・議案表現の 自動抽出に向けた検討

Investigation of Automatic Extraction of Agenda / Agenda Expression for Prefectural Assembly Meeting Proceedings

田中琢真*¹ 小林暁雄*¹ 坂地泰紀*² 内田ゆず*³ 乙武北斗*⁴ 高丸圭一*⁵
Takuma Tanaka Akio Kobayashi Hiroki Sakaji Yuzu Uchida Hokuto Ototake Keiichi Takamaru

木村泰知*⁶ 増山繁*¹
Taichi Kimura Shigeru Masuyama

*¹ 豊橋技術科学大学 Toyohashi University of Technology *² 成蹊大学 Seikei University *³ 北海学園大学 Hokkai-Gakuen University *⁴ 福岡大学 Fukuoka University
*⁵ 宇都宮共和大学 Utsunomiya Kyowa University *⁶ 小樽商科大学 Otaru University of Commerce

In recent years, it becomes important for political scientists to analyze local assembly minutes against heating discussions about revitalization of local communities. Many Japanese local autonomies provide access to various political documents on their websites. Such documents include basic urban development plans, local assembly minutes, and ordinances. Therefore, we began collecting these political documents and extracting important information. To analyze utterances particularly, agendas of discussions are needed. Hence, this paper proposes a method that extracts agendas of discussions from political documents.

1. はじめに

近年、地方創生などの議論が盛んになり、地方議会会議録の分析が重要性を増してきている。地方議会会議録では、地方議会におけるすべての発言は記録され、自治体ごとでウェブ公開されている。我々は、以前から、ウェブ上に公開されている地方議会会議録の収集および整理を行っており、「地方議会会議録コーパス」として研究者に限定公開をしている。「地方議会会議録コーパス」では、情報工学分野[木村ら 2013]や言語学分野[高丸ら 2015]などで成果が得られている。また、最近では、政治学や経済学分野の研究で必要となる統計情報を提供するために、対象範囲および対象年度を限定し、正確で網羅性の高いコーパスの構築を進めている。具体的には、全国 47 都道府県議会を対象として平成 23 年 4 月から平成 27 年 3 月までの本会議を対象とした「都道府県議会会議録コーパス」の構築を進めている[Kimura et al. 2016]。収録されているデータ例を図 1 に示す。

都道府県名	北海道
会議名	平成26年第4回定例会11月28日-03号
開催年	26
開催月	11
開催日	28
発言者	知事高橋はるみ君
発言	ごうした中、釧路市では、地元関係機関が参加する阿寒湖世界自然遺産登録地域連絡会議が設置され、マリモを育む阿寒湖とその周辺地域の登録を目指した、さまざまな取り組みが進められているところであります。

図1 「都道府県議会会議録コーパス」の収録データ例

連絡：田中琢真，豊橋技術科学大学，t-tanaka@la.cs.tut.ac.jp

「都道府県議会会議録コーパス」は正確で網羅性の高いコーパスとして整いつつあるものの、議会ごとの比較分析を行う際に、それぞれの議会で議論されている議題を抽出する必要がある。地方議会では予め提出された議案についての議論を行っている。しかしながら、それらの議案は、「議案番号」のみ発言されることも多く、議案番号と議題の対応関係を把握しなければ、議題を適切に抽出することが困難である。例えば、愛知県議会平成 26 年 6 月定例会(第5号)では、原よしのぶ議員が「... 第百二十一号議案及び第百二十二号議案は同意されたいという動議を提出いたします。」と発言している。この発言に含まれる第百二十一号議案は「公安委員会の委員の選任について」であり、第百二十二号議案は「人事委員会の委員の選任について」であるが、発言の前後を読んだだけでは議案を理解できない。会議録には、各自治体の会議ごとに議案番号と議案の対応が異なるため、「自治体名」「年度」「議会」「議案番号」「議案」の対応関係を抽出する必要がある。このような、議案の呼び方は、会議録に含まれる会議録特有の照応問題といえる。

そこで、本研究では、会議録特有の照応問題を解決するために、議案の記述についての傾向を明らかにする。また、議案の記述傾向を用いて自動で抽出する手法についても検討する。

2. 議題・議案表現抽出

議会では会議の初めに議長によって議題として議案が取り上げられる。次に議長の発言の例を以下に示す。

○副議長(伊藤勝人君) 第百八号議案平成二十六年度愛知県一般会計補正予算から第百二十二号議案人事委員会の委員の選任についてまでを一括議題といたします。これより一般質問並びに提出議案に対する質問を許します

この発言のように、一つの会議で複数の議案を議論する。議案とは議会の議決を得るために提出される条例や予算などの案である。議案の例を次に示す。

- ・平成 27 年度愛知県一般会計予算
- ・愛知県教育委員会教育長の勤務時間等に関する条例の制定について
- ・愛知県ふれあい広場条例の一部改正について

このように、議案は会計予算や、条例などについてもが多く、自治体の政策を表している。

2.1 議案表現抽出

議案について質疑や応答を行う発言者は議案番号と議案名を口述する傾向が高い。本稿では議案番号と議案名を口述する際の表現方法を議案表現とする。この議案表現には自治体ごとに特徴が存在する。次に議案表現を含んだ発言の例を示す。

私のほうからは、議第 48 号 岐阜県企画経済関係手数料徴収条例の一部を改正する条例について御説明をいたします。
(以下略)

議案名の語尾には「～条例」「～について」「～に関する件」などのパターンが存在する。このパターンは各自治体当たり 3~4 種類程度である。また、議案番号の表現方法は自治体ごとで違いはあるが、自治体内では統一されている。これらの特徴を用いて 47 都道府県すべての自治体の議案表現のパターンの調査を行う。議案番号の表現方法が自治体内で統一されているため、議案番号が記述されている発言だけを検索することが可能であり、これらの発言からパターンを調査する。さらに、それらの発言に対して、正規表現を用いたパターンマッチングを行うことで議案表現を抽出する。ただし、この手法で抽出できるのは議論された議案のみであり、議論されずに可決される議案は抽出できない。しかしながら、本研究の目的は、議論の議題を抽出することであるので、そのような議案は本研究の対象外とする。

3. 議案表現抽出実験

「都道府県議会会議録コーパス」を用いて、平成 23 年 4 月から平成 27 年 3 月の愛知県議会の本会議を対象として、議案表現の抽出実験を行った。語尾については「～条例」と「～について」の 2 パターンを用いた。「～について」で先にマッチングした後、マッチングがなければ「～条例」でマッチングするといった手法をとった。これは、「～条例改正について」などに対応するためである。議案番号と議案名のペアが正確に抽出できているものを正解とし、議案番号のみの場合を失敗とした。議案番号・議案名が重複して取得された場合(重なり有)とそうでない場合(重なり無)それぞれについて集計を行った。その結果を表 1 に示す。

表 1 議案表現抽出実験の結果

総抽出数	467
正解抽出数(重なり有)	352
正解抽出数(重なり無)	115
失敗抽出数	126

重なり有の場合での適合率は 0.75 となった。次に、正解と失敗の例を表 2 に示す。抽出失敗の例として、複数の議案を議案

番号で口述するような発言が多く見られた。一方で、「～について」が語尾についている場合についても、議案番号のみ記述されている発言も散見された。

表 2 抽出結果

議題と議案ペアが書かれた発言

- ・第八十七号議案職員の給与に関する条例の一部改正について
- ・第 114 号議案の高等学校等奨学金の長期滞納者に対する返還請求訴訟の提起について
- ・第一号議案、愛知県議会における政務調査費の交付に関する条例の一部改正について

議案番号のみ

- ・第 81 号議案まで及び専決第 13 号に対する討論の通告について
- ・第 99 号議案及び第 100 号議案について
- ・第 100 号議案の審議であるが、委員会審査並びに討論について

4. 考察

4.1 議案表現抽出

議案表現抽出実験において、複数の議案が含まれる場合の抽出ミスについては、ひとつの発言に複数の議案番号が存在するかどうかを確かめることで除くことができると考えられる。次に失敗の例として多かったのは、議案について話す際に議案番号しか口述していない場合である。コーパス中のほかの発言や目次等から取得する方法を検討する必要がある。

4.2 議題と議論の結び付け

議題から議論の検索や比較分析を行うためには、各発言を議論と結び付ける必要がある。つまり、すべての発言に対して、その発言に対する議題、もしくは、議題を含まない発言という情報を付与しなくてはならない。

5. おわりに

本稿では都道府県議会会議録を対象とした議題・議案表現の自動抽出に向けた検討を行った。この手法により、今後、市議会などの他の地方議会会議録などの、様々な会議録に応用できると考えられる。

謝辞

本研究はJSPS科研費 JP16H02912 の助成を受けたものです。

参考文献

- [木村ら 2013] 木村泰知, 関根聡: 主辞に基づく政治問題抽出手法, 人工知能学会論文誌, Vol.28, No.4, pp. 370-378, 2013.
- [高丸ら 2015] 高丸圭一, 内田ゆず, 乙武北斗, 木村泰知: 地方議会会議録コーパスにおけるオノマトペ -出現傾向と語義の分析-, 人工知能学会論文誌, Vol.30, No.1, pp. 306-318, 2015.
- [Kimura et al. 2016] Yasutomo Kimura, Keiichi Takamaru, Takuma Tanaka, Akio Kobayashi, Hiroki Sakaji, Yuzu Uchida, Hokuto Ototake, Shigeru Masuyama, Creating Japanese Political Corpus from local Assembly Minutes of 47 prefectures, Coling2016 ALR12, 2016.