

主辞に着目した政治問題の定義と注釈付け

木村 泰知
小樽商科大学

関根 聡
ニューヨーク大学

kimura@res.otaru-uc.ac.jp sekine@cs.nyu.edu

1 はじめに

近年、社会が抱える課題や問題などの情報を収集し、分析する手法の確立が望まれている [1]。社会が抱える課題や地域特有の問題などは、主に自治体が解決するものであり、地方議会において議論されている。

地方議会では、政治問題について、首長や議員が議論を行っており、7割の自治体がそれらの議論を記録した会議録をウェブ上に公開している [2]。地方議会会議録には、政治問題について、首長側からの提案や議員からの質問などが含まれており、異なる視点からの発言が記録されている。また、地方議会会議録は、通時的なデータであり、長い年月の議論が含まれているため、議論の流れを把握することができる。しかしながら、会議録は、首長や議員の発言をそのまま書き起こしたテキストデータであり、読みやすさを考慮していない。

会議録を対象とした研究には、話し言葉による読みづらさを解消するために、文末表現の冗長性を言い換える研究がある [3]。しかしながら、文末表現の変換だけでは、膨大な会議録を読むことが難しい。地方議会では複数の政治問題が複雑に絡み合っており、それを簡潔に表現し、一般の市民でも容易に理解できる形で示すことができれば、会議録を簡単に理解できるようになると考えられる。

そこで、本研究では、会議録の内容を簡単に理解できるように、政治問題の関係を可視化して表示することを最終目標とする。図 1 には、本研究の最終目標とその課題について示す。例えば、図の「新市立病院建設に係る起債償還について、20歳以上の市民1人当たり負担額は幾らになるのか」という文からは、「新市立病院の建設費の起債償還」が「20歳以上の市民1人当たり負担額」に「影響」するという関係が得られる。他にも、「対立」の関係からは、提案された政策の長所と短所を明らかにすることが可能であり、会議録の内容が理解しやすくなると考えられる。

政治問題は、固有表現に似ている。しかしながら、実際に会議録を見ると修飾語がついていたり、説明のための形容があり、節や句の単位になっている場合が

多い。このような政治問題は、一般的な固有表現に比べると表現が長くなるため、政治問題の関係抽出に向けて、説明的な表現を認識し分離することが必要となる。本稿では、政治問題の関係抽出を最終目標として、「政治問題」がどのように表現されるかという課題について考察し、固有表現の枠を超えた政治問題の定義について検討する。

2. では、会議録に含まれる政治問題と政治問題の定義について述べる。3. では、政治問題の主辞候補について述べる。4. では、政治問題の注釈付けについて述べる。5. では、本論文のまとめと今後の課題について述べる。

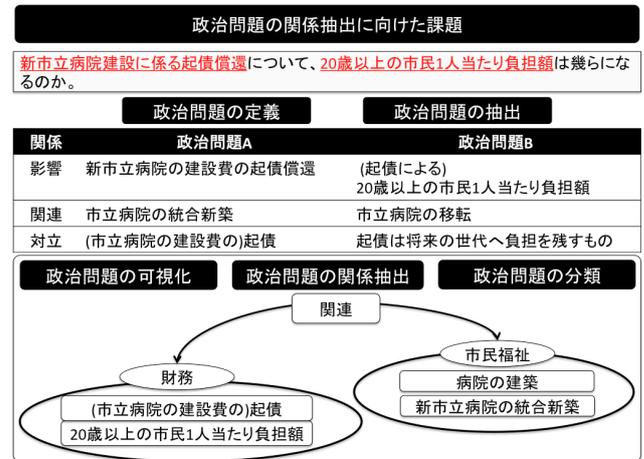


図 1: 最終目標のイメージと課題

2 会議録に含まれる政治問題

2.1 政治問題の定義

地方議会では、自治体で解決すべき諸問題に対して、首長と議員がそれらの問題を議題として取り上げ、議論を行っている。そこで、我々は、地方議会会議録に含まれる政治問題を下記の2つと定義する。

1. 自治体が動作主体となり解決する事柄

表 1: 政治問題の主辞

品詞	細分類	出現頻度	小計	割合	主辞の例
助動詞		28	28	0.040	ます (15) ない (6) た (3) ん (2)
動詞	接尾	16	104	0.145	れる (9) せる (3) れ (2)
	自立	69			する (9) し (7) 行う (4) 図る (3)
	非自立	19			いる (13) おり (3) しまい (1)
名詞	サ変接続	301	570	0.805	削減 (12) 減少 (11) 対策 (10) 整備 (10)
	ナイ形容詞語幹	19			問題 (19)
	一般	117			事業 (13) 効果 (5) 状況 (5)...
	接尾	82			化 (31) 者 (6) 金 (4) 性 (4) 法 (4) ...
	非自立	51			こと (41) もの (7) ところ (3)
記号	括弧閉	6	6	0.010	」 (5)) (1)
合計		708	708	1	

2. 議会において議題となっている事柄

2.2 コーパス中での政治問題の表現

我々は、上記の定義に従い、小樽市市議会会議録に含まれる政治問題の関係を含む文を調査することとした。従来の研究 [4] では、研究を開始する時点で、政治問題の表現をサ変名詞で終了する名詞句に限定していたが、政治問題の表現は、必ずしもサ変名詞で終了しているわけではない。そこで、本研究では、品詞を限定せずに、最終目標の関係を抽出できる政治問題の表現に対して、注釈付けを行うこととした。対象データは、小樽市市議会会議録の定例会の平成 11 年から平成 22 年までの会議録を対象として、ランダムに選択した 6,263 文である。

一般に、関係抽出の研究は 1 文内に 2 つの要素が現れ、その関係を見つけ出すという研究が多い。しかしながら、調査の結果、政治問題と注釈付けした文数は 316 文であり、政治問題の数は 708 表現であった。この調査結果から、「1 文中に 2 つ以上の政治問題が含まれ、かつ、政治問題の関係が記述されている」という文は、少ないことが明らかになった。一般の関係抽出は固有表現等の短い表現間関係を抽出するタスクが多いが、政治問題は固有表現に比べて長いという特徴があり、政治問題の関係が文を超えて現れることや、2 つの政治問題が含まれている場合でも照応表現になっていることが原因と考えられる。そこで、本研究では、「1 文中に 2 つ以上の政治問題が含まれ、かつ、政治問題の関係が記述されている」という条件で抽出するのではなく、1 文中に 1 つだけの政治問題が含まれる場合でも、政治問題を抽出し、その後で、政治問題の関係をみつけることとした。

また、この調査では、政治問題の主辞に特徴があることを確認した。政治問題の主辞を品詞で分類した結

果を表 1 に示す。政治問題の主辞は、名詞以外にも、動詞、助動詞、括弧が含まれることがわかったが、約 8 割が名詞であった。さらに、動詞の直前の品詞を調査した結果、「名詞のサ変接続」が 1 割程度存在していた。例えば、「... JR から 分離 される」「... 専門性を 向上 させる」などの表現がある。

さらに、これらの名詞の主辞は、政治特有の単語が繰り返し利用されている傾向があった。例えば、名詞-サ変接続では「削減」「減少」「対策」など、名詞-一般では、「事業」「効果」「状況」などがある。この特徴から、我々は政治問題の主辞を有限個に限定することが可能であると考えた。例えば、「促進」という単語が主辞となる表現は、図 2 のように数多く存在する。

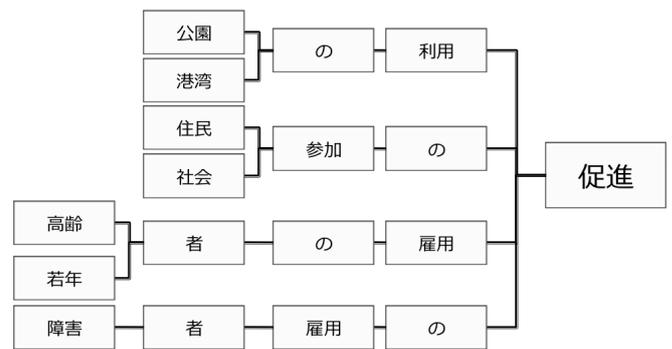


図 2: 「促進」が主辞となる表現の例

これらの結果から、我々は、政治問題の主辞を名詞に限定することとした。次節では、政治問題の主辞となる単語について調査を行う。

3 政治問題の主辞候補

3.1 主辞の必要性

前節で述べた理由に加え、政治問題の主辞をみつけることの必要性について説明する。地方議会会議録に含まれる政治問題の表現は、新たに発生した問題や地域特有の問題が含まれることから、予め政治問題として登録することが困難である。例えば、「鳥インフルエンザの感染」「小樽市立病院の移転」などは、新たに発生した問題や地域特有の問題といえる。また、地方議会に含まれる政治問題は、固有表現のように名詞だけの構成になることは少ないことから、政治問題の境界を判定することが困難である。例えば、政治問題を下記の3つの候補から、どの表現にするのかを決めることは難しい。

- (1)a. 労働者全体の賃下げ
- (1)b. 労働者全体の賃下げの悪循環
- (1)c. 労働者全体の賃下げの悪循環を招くこと

本研究では、会議録に含まれる政治問題の表現に対して、最終的な係り先となる「主辞」に着目することで、無数に存在する政治問題の表現を有限個の主辞に集約可能か検討する。例えば、「鳥インフルエンザの感染」「市立病院の移転」「高齢者の増加」のような場合、「感染」「移転」「増加」が主辞となる。

我々は、係り先となる主辞を有限個に限定することにより、「無数に存在する係り元の表現」と「政治問題の境界」の問題を解決することを試みる。

3.2 政治問題の主辞候補の収集

本節では、政治問題の主辞となる単語の候補を選択する方法について述べる。政治問題の主辞となる表現は、2.2の調査において、(直前の品詞も含めると)名詞である割合が約9割程度であったことから、名詞を対象に調査を行うこととした。主辞候補となる名詞は、政治問題の主辞となる可能性がある名詞を全て収集する。調査では、正規表現により議題となっている主辞を自動的に収集し、最終的には人手で確認する。正規表現は、地方議会会議録に含まれる「(接続詞)、○○について」や「、○○の問題」の○○を収集するために用いる。このように得られた「○○」部分の主辞を収集し、人手で確認する。調査の結果、表2のような主辞を収集した。

4 注釈付け

4.1 注釈付けの目的

注釈付けの目的は、政治問題の抽出の訓練データ及び評価データを作成することである。また、3. で取

表 2: 注釈付けの対象データに含まれる主辞

IPADIC の品詞分類	単語数	例
名詞:サ変接続	1139	向上, 削減...
名詞:一般	986	支払, 遅れ...
名詞:接尾	45	化, 性...
名詞:非自立	2	こと, もの
名詞:ナイ形容詞語幹	1	問題
合計	2,173	

集した政治問題の主辞候補は、政治問題になり得るのかを明らかにする。さらに、注釈者により、どの程度の違いが生じるのかを明らかにする。

4.2 注釈付けの概要

本研究では、下記の情報に注釈付けを行う。

1. 政治問題の有無
2. 政治問題の対象範囲
3. 政治問題を解決する動作主体
4. 具体的問題と抽象的問題

注釈付けを行う対象文は、表2の主辞を含む文とする。最初に、注釈者は政治問題の主辞を含む表現が政治問題であるかを判断する。政治問題として注釈付けできない場合、注釈者は注釈付けできない理由を選択する。一方、政治問題を含む場合、注釈者は政治問題の開始単語を判断する¹。また、注釈付けされた政治問題を解決する動作主体は「地方自治体」なのか「国」なのかを判断する。さらに、注釈付された問題は、具体的な政治問題なのか、抽象的な政治問題なのかを判断する。

4.3 注釈付けの準備

注釈付けの対象データは、2.2節において利用した小樽市市議会会議録からランダムに選択した6,263文である。この対象データの総単語数は336,285語、異なり単語数は11,419語、一文に含まれる平均の単語数は約54単語である。注釈付けの対象となる主辞候補の数は、32,525である。

会議録の注釈付けは、男性1名と女性3名の大学生4名が行う。本来であれば、同一データに対して複数人により注釈付けを行うべきであるが、対象データが多いことから、対象データを重複しないように割り当てている。

¹政治問題は、注釈付けした範囲だけを見て、政治問題として判断できる表現とする。また、政治問題は、指示代名詞などの省略表現が含まれていない表現とする。

4.4 注釈付けの結果

注釈付けの結果は、政治問題が含まれる主辞が11,410、政治問題が含まれない主辞が21,115となった。

表3は、主辞ごとの政治問題として注釈付けされた割合を示した例である。この例から、主辞により、政治問題になりやすい主辞と、なりづらい主辞とが存在することがわかる。今後の政治問題の抽出では、この点について考慮した手法を考える予定である。

表3: 政治問題として注釈付けされた割合を示した例

主辞	有/(有+無)	割合	政治問題有りの例
高騰	11/11	1.00	保育料の更なる <u>高騰</u>
衰退	3/4	0.75	小樽港の <u>衰退</u>
差別	2/4	0.50	障害に基づく <u>差別</u>
方策	5/17	0.29	小樽ブランドの <u>アピールの方策</u>
人員	0/7	0.00	-

4.5 注釈付けの評価

本節では、注釈者による注釈付けの違いを明らかにするために、同一データを対象として、4名の注釈者と著者の注釈付けの一致率を求める。注釈付けの評価データは、出現頻度の高い200単語の主辞を対象として、その主辞を含む表現の中から、ランダムに選択した。政治問題の有無について、注釈者1-4と著者のコーエンの一致係数は、注釈者順に、0.59, 0.71, 0.61, 0.44であり、全体的には高い一致率となった。表4は、第一著者と4人の注釈者の完全一致と不一致の例である。表4の「有」は政治問題として注釈付けした注釈者の数であり、「無」は注釈付けしなかった注釈者の数である。

不一致の理由としては、政治問題が含まれる文において、情報が不足していたり、照応表現が含まれることから、注釈者が判断に迷っていると考えられる。例えば、表4の例にある「国、道の支援策」の文脈は「今後、市といたしましても、商店街支援の立場から、国、道の支援策も活用...」となっているが、政治問題らしい表現であるものの、前方の文脈を読んでも、何に対する支援策なのか判断できないことから、注釈者が判断に迷っていると考えられる。また、表4の例にある「新しい定義に基づく調査」の文脈は「平成17年度の調査では、小学校3件、中学校6件の合計9件でしたが、平成18年度には新しい定義に基づく調査により、...」となっており、一文だけではどのような調査内容なのかかわからず、さらに、自治体が解決する問題とどのように関わっているのか判断できないことから、注釈者が迷っていると考えられる。

表4: 複数の注釈者による一致と不一致の例

完全一致の例	有	無
食品の安全性 <u>確保</u>	5	0
配偶者特別控除の <u>廃止</u>	5	0
…、その総額が大幅に <u>削減</u>	0	5
… 基本方針はまだ <u>策定</u>	0	5
不一致の例	有	無
環境 <u>整備</u>	4	1
国、道の支援 <u>策</u>	3	2
新しい定義に基づく <u>調査</u>	2	3
相談 <u>内容</u>	1	4

5 まとめ

本稿では、地方議会会議録を対象として地方政治問題の抽出を行うための定義と注釈付けの結果について述べた。政治問題は、固有表現に似ているが、節や句の単位になっていることが多く、固有表現に比べると長い説明的な表現であることから、我々は、固有表現の枠を超えた、表現単位についての考察を行った。その結果、政治問題の末尾である主辞に着目することで、政治問題を抽出するという考え方が良いことがわかった。

今後は、今回の調査結果を利用して、政治問題の自動抽出を行う予定である。

謝辞

本研究の一部は科研費 22300086 の助成を受けたものである。

参考文献

- [1] 橋本, 村上, 乾, 内海, 石川, 社会課題発見のための文書クラスタリングとクラスタ評価指標人工知能学会論文誌, Vol. 24, No. 4, pp.333-338, 2009.
- [2] 高丸, 木村, 渋谷, 全国の市町村議会会議録のウェブ公開とデータ提供の状況宇都宮共和大学 都市経済研究年報第11号, 2011.
- [3] 山本, 安達, 国会会議録を対象とする話し言葉要約自然言語処理, Vol.12, No.1, pp.51-78, 2005.
- [4] 木村, 渋谷, 高丸, 乙武, 小林, 森, 地方議員マッチングシステムにおける能動的質問のための質問生成手法人工知能学会論文誌, Vol. 26, No. 5, pp.580-593, 2011.