

地方政治情報コーパスへの賛否の積極性に関する注釈作業

筒井 貴士¹ 我満 拓弥² 渋谷 英潔³ 木村 泰知⁴ 森 辰則⁵

^{*1*2} 横浜国立大学 工学部 電子情報工学科 ^{*3*5} 横浜国立大学 大学院 環境情報研究院

^{*4} 小樽商科大学 商学部 社会情報学科

E-mail: {^{*1}tkstti, ^{*2}gaman, ^{*3}shib, ^{*5}mori}@forest.eis.ynu.ac.jp ^{*4}kimura@res.otaru-uc.ac.jp

1 はじめに

地方政治における議員やその活動に関する住民の認知度は、国政と比較して一般に低い傾向がある。マスメディアで取り上げられる政治情報は国政に関するものが多く、地方政治に関する情報は少ない。地方政治に関する情報源としては地方議会会議録があり、検索システムとして公開している自治体も多い。しかしながら、公開されている会議録は1回の会議に限っても膨大な量のテキストデータであり、その全てを読むには多大な労力がかかるため、容易に目的の情報へのアクセスを可能とするためのシステムが望まれる。そうした現状を受け、我々は地方政治に関する情報、特に地方議会会議録における議員の発言を中心とした政治情報システムの開発に取り組んでいる[1]。また、その開発に向けて、分析・評価用データとなる地方議会会議録コーパスの収集[2]および注釈付け[3]を行っている。

我々は、大城ら[3]と同じく、会議録における発言を基に利用者と政治的に近い考えをもつ議員を判断して提示するシステムの開発を最終的な目的としている。政治的な考えの近さは、一般に、施策や事業などへの賛否の一致度合いにより推測できると考えられ、上神ら[4]などのポートマッチシステムでも、この考え方に基づいている。それゆえ、大城ら[3]では、施策や事業に対する賛否の意見を、構造的に捉えるための注釈付けタグセットを提案し、その有効性を確認した。しかしながら、同じ賛成の立場をとる議員の間でも、その賛成度合いには差が存在しており、例えば、「昨年度は などの事業に取り組んできた」と発言した議員と、「 などの事業を行うのもやむを得ない」と発言した議員では、前者の方が積極的に賛成であると考えられる。積極的に賛成である議員の方が、消極的に賛成である議員よりも、彼らが賛成する施策や事業の実現に向けて尽力すると考えられるため、当該の施策や事業を実現してほしい利用者には、積極的に賛成である議員の方を提示することが望ましい。否定の立場をとる議員においても同様の考えが成り立つ。したがっ

て、このような賛否の積極性を考慮して、利用者と近い考えをもつ議員を判断する必要がある。しかしながら、大城らのタグセットでは賛否の積極性に関する情報を注釈するには不十分であり、また、コーパスへの注釈作業も1名の作業員による1都市の会議録への付与という極めて小規模なものであった。以上の背景から、本稿では、複数名による複数の都市の会議録への、賛否の積極性に関連する情報の注釈作業を行い、注釈された結果の基礎的な分析を行う。

本稿の構成は以下の通りである。2章では、賛否の積極性を判断する上で必要となる情報に関する考察を行い、注釈すべき情報を定義する。3章では、注釈対象となるデータと注釈作業の内容について説明する。4章では、注釈情報の基礎的な分析を行い、その結果を考察する。5章は、まとめである。

2 賛否の積極性に関する情報

議会会議録における賛否を表明している議員の発言を読み、その賛否への積極性を判断する場合を内省すると、以下の3点が積極性を判断するための手掛かりであることが多いと思われる。

1点目は、発言中の「やむを得ない」などの表現である。この表現がある場合、消極的な賛否であることが読み取れる。しかしながら、会議録の場合、一般的には積極性を示すはずの表現の存在が必ずしもその積極性を示すとは限らない。例えば、「取り組まねばならないと考えている」という表現の場合、一般的には積極的な賛否を示していると考えられるが、会議録における議員の発言においては、この表現だけをもって、積極的な賛否であると判断することは不適切である。会議録には特有の表現や言い回しがあり、議員は施策や事業に積極的に取り組むことが当然として捉えられているため、積極であることを示す表現が常時の表現となっていることが多い。一方、消極であることを示す表現に関しては、実際に消極であることが多いと思われ、積極的か消極的かを示す表現で非対称性があるように思われる。

2点目は、言及された施策や事業の具体性である。積極的であるように読み取れる場合、その施策や事業への言及が具体的であることが多い。例えば、単に「取り組まねばならない」という発言よりも、「
という理由により 月までに取り組まねばならない」という発言の方が、積極的に取り組むという意味を読み取ることができる。具体性の有無を判断する手掛かりとして幾つか考えられるが、本稿では理由に着目し、理由が述べられている文は具体性があると考えて注釈作業に反映させた。

3点目は、言及された施策や事業の実現度合いである。すなわち、これから取り組みたいという意味を表明しているだけなのか、それとも、既に施策や事業の一部に着手しているのか、といった違いにより、賛否の積極性を読み取ることができると考えられる。

以上の3点に関する情報を、施策や事業に対する賛否の積極性と共に注釈付けすることとした。また、付与された注釈情報を用いることで、賛否の積極性を判断する上で有効な手掛かりとなりうるか分析可能となる。

3 賛否の積極性に関する注釈の付与

本稿では、地方議会会議録コーパス [2] に収録されている、札幌市、横浜市、京都市、北九州市の4市の2010年の第2回定例会を対象に注釈付けを行った。この4市を対象とした理由は、政令指定都市であること、全国に散らばっていること、同一の記述形式の会議録を採用していることの3点による。本稿では、注釈を付与する単位として「発話」「段落」「文」「文字列」の4つの単位を用い、「文」「段落」「発話」を以下のように定義した。まず、4市の会議録の記述形式では、全ての文の最後が句点で終わっていることから、句点を「文」の境界とした。次に、同一の話題に関する文は1つの段落にまとめて記述されており、全ての段落の最初は字下げされていることから、字下げを「段落」の境界とした。最後に、文の発言者に関する情報がコーパス中に収録されているため、発言者が同一人物である文の連続を1つの「発話」とした。

2章での議論を基に、表1に示す11種類の注釈を付与することとした。各注釈の説明を以下に述べる。

注釈(1)は、発言がどのようなシーンでなされたかの注釈であり、発話単位で付与する。議員が意見を述べることが多いシーンとして、「質問」「回答」「討論」「説明」の4シーンを想定し、各シーンを以下のように定義した。

- 「質問」: 発言中に他者に対して回答を求める文が存在している。
- 「回答」: 発言中に他者からの質問に対する回答

表 1: 注釈の一覧

番号	内容	付与単位
(1)	発言されたシーン	発話
(2)	施策・事業	文字列
(3)	発言内容のカテゴリー	段落
(4)	質問と回答の対応付け	段落 段落
(5)	疑問文	文
(6)	意見性の有無	文
(7)	意見の本人性	文
(8)	発言内容の中核となる文	文
(9)	理由の対応付け	文 文(文字列)
(10)	意見の実現度	文
(11)	意見の説得性	文

となる文が存在している。

- 「討論」: 自分の意見を一方的に表明している文が存在している。
- 「説明」: 議案等の内容を説明する文が存在している。

上記の4シーンに当てはまらないシーンは「その他」として注釈付けした。作業効率の観点から、「その他」のシーンにおける発言には(2)以降の注釈を付与しなかった。

注釈(2)は、発言者の関心がある施策や事業に関する注釈であり、文字列単位で付与する。発言中に含まれる施策や事業を示す文字列を抽出し、「賛成(推進)」、「反対(廃止)」、「その他」の何れかの極性を付与する。極性の判断は前後の文脈を考慮した作業者の主観に基づいて行われた。

注釈(3)は、発言内容のカテゴリーに関する注釈であり、段落単位で付与する。カテゴリーは、木村ら [1] の政治的カテゴリーを参考に、比較的議題に挙げられることが多い「医療」「教育」「環境」「観光」「防災」「公共」の6カテゴリーを対象とした。1つの段落に複数のカテゴリーを付与することを許可している。また、発言内容がどのカテゴリーにも属さない場合には「その他」として注釈付けした。

注釈(4)は、質問と回答の対応付けに関する注釈であり、段落単位で付与する。質問の段落から回答の段落へと1対1で対応付けており、もしも、回答が複数の段落にまたがっている場合は最初の段落に対応付けを行った。

注釈(5)は、疑問文かどうかの注釈であり、文単位で付与する。本稿での疑問文とは、他者の回答を要求する文と定義しており、「
についてお聞かせ願いたい」といった表現であっても疑問文とした。

発言者	発言内容	タグ	注釈	注釈内容	注釈理由	注釈結果
発言者	図1: タグ付けツール					
発言者	図2: XML データの例					

図 1: タグ付けツール

図 2: XML データの例

注釈 (6) は、意見性がある文かどうかの注釈であり、文単位で付与する。本稿での意見性がある文とは、「すべきだ」、「の方が良いと考えられる」といった意見であることが明確に示されている文と定義している。

注釈 (7) は、発言者本人の意見である文かどうかの注釈であり、文単位で付与する。

注釈 (8) は、発言内容の中核となる文に関する注釈である。本稿での中核となる文とは、発言内容を端的に述べている文と定義している。我々は、システムが利用者に発言内容を提示する際には、発言内容の整理・要約を行う必要があると考えており、整理・要約を行うための情報として利用することを想定している。(2) の施策・事業を注釈した文、または、(6) の意見性がある文を含む段落には、最低でも 1 文は中核となる文を付与することとした。

注釈 (9) は、(2) の施策・事業の極性または (6) の意見性がある文と、その理由となる文との対応付けに関する注釈である。理由となる文から、(2) で選択された施策・事業または (6) の意見性がある文へと多対多で対応付けを行っている。

注釈 (10) は、発言時点で文中の意見がどの程度実現されているかの注釈であり、文単位で付与する。実現の程度として、「表明」、「着手」、「完了」、「拡大」の 4 つの状態を以下のように定義した。

- 「表明」: 何も実現できていない状態。やるべきという意味を表明しただけの状態。
- 「着手」: 実現のために行動を開始した状態。現在進行中であり、目標は達成されていない。
- 「完了」: すでに目標を達成した状態。現在は行動していない。

<Paragraph Id="P178" CorrespondingAnswerParagraphId="None" Category="医療">
<Sentence Id="S432" Member=" (中山大輔君)" IsQuestion="False"
IsOpinion="True" IsPrincipal="True" Actualization="不明"
CorrespondingConclusive="P:434_25_17:感染症病床を方面別に新たに整備する"
IsPersuasive="False" IsCoreSentence="True"> 他の医療機関に協力を依頼することにより一時的に入院患者への対応はできるとしていますが、現行では厳重な感染対策を講じ、さらに、人工呼吸器の装着が必要な重症患者が増加をすれば通常の診療に影響がでることなど、強毒性のインフルエンザの入院治療必要患者を受け入れられる余力が一般の病院にあるのか、懸念されることです。</Sentence>
<Sentence Id="S433" Member=" (中山大輔君)" IsQuestion="False"
IsOpinion="True" IsPrincipal="True" Actualization="不明"
CorrespondingConclusive="P:434_25_17:感染症病床を方面別に新たに整備する"
IsPersuasive="False" IsCoreSentence="False">現段階のように、新型インフルエンザは毎年流行するインフルエンザと同じ弱毒性との見解を示されていますが、しかしながら、世界じゅうの大半の人が経験をしたことがないウイルスのため、免疫がなく、さらに、インフルエンザウイルスは変異をしやすい、流行中に強毒性に変わることも考えられるため、警戒は依然怠れないと考えます。</Sentence>
<Sentence Id="S434" Member=" (中山大輔君)" IsQuestion="False"
IsOpinion="True" IsPrincipal="True" Actualization="表明"
CorrespondingConclusive="P:434_25_17:感染症病床を方面別に新たに整備する"
IsPersuasive="True" IsCoreSentence="True">そうした中、パンデミックの状態になった場合に備え、<Policy Polarity="賛成">感染症病床を方面別に新たに整備する</Policy>などの対策も必要ではないかと考えます。</Sentence>
</Paragraph>

- 「拡大」: すでに目標を達成しており、さらなる成果を求めて行動したい (している) 状態。

(2) の施策・事業を注釈した文には必ず付与することとした。

注釈 (11) は、総合的に見て、文中の意見がどの程度説得性がありそうか (目標を実現できそうか) に関する注釈であり、文単位で付与する。説得性の判断は前後の文脈を考慮した作業者の主観に基づいて行われ、(2) の施策・事業に関する注釈付けを行った文には必ず付与することとした。

注釈者の注釈付け誤りを軽減するために、図 1 に示す、専用のタグ付けツールを開発し、ツールを通して上記の注釈付けを行った。注釈情報は、図 2 に示すような XML 形式で付与される。

4 注釈情報の基礎分析

注釈作業は現在継続中であるため、本稿では、先行して作業が完了している横浜市と札幌市の 1 名による注釈情報の基礎的な分析を行う。表 2 は、会議録中の発言の数と (1) で付与されたシーンの内訳を示す。表 3 は、会議録中の段落の数、(3) のカテゴリーが付与された段落の数および内訳を示す。「その他」のシーンの段落にはカテゴリーが付与されておらず、複数のカテゴリーの付与を許可していることに注意されたい。表 4 は、会議録中の文の数、(1) で付与されたシーンの内訳、(5) の疑問文の数、(6) の意見性のある文の数、(7) の本人の意見である文の数、(8) の中核となる文の数、(9) の理由となる文の数、(10) の実現度を有する文の数および内訳、(11) の説得性がある文の数を示す。表 5 は、(2) の抽出された施策・事業の数および極性の内訳を示す。各表中の括弧内の値は割合を示している。まず、両市を比較すると、注釈数の分布に大きな差

表 2: 発話単位の注釈結果

	横浜市	札幌市
発話数	155	151
質問	22 (.142)	11 (.073)
回答	32 (.206)	32 (.212)
討論	6 (.039)	2 (.013)
説明	6 (.039)	3 (.020)
その他	89 (.574)	103 (.682)

表 3: 段落単位の注釈結果

	横浜市	札幌市
段落数	1,618	1,797
カテゴリーを有する段落数	1,222	1,090
医療	243 (.223)	242 (.198)
教育	195 (.179)	151 (.124)
環境	75 (.069)	80 (.065)
観光	66 (.061)	82 (.067)
防災	39 (.036)	19 (.016)
公共	192 (.176)	186 (.152)
その他	553 (.507)	397 (.325)

は見られない。次に、表 5 の施策・事業の極性の内訳を見ると、賛成の割合が 9 割を占めていることが分かる。したがって、極性の判定において、否定判定の精度を高めることが重要になると考えられる。最後に表外の結果による考察として、現時点での狭い範囲ではあるが、理由が述べられている文や、実現度が「着手」や「完了」や「拡大」である文が説得性がある文となっている可能性が高く、2 章で述べた手掛かりが有効そうであることが確認できた。ただし、表 4 の実現度の内訳を見ると、「着手」「完了」「拡大」の占める割合は合計で 3 割程度であり、説得性がある文の数より大幅に少ない。そのため、実現度が「表明」である文の説得性を判断するための方法が必要になると考えられる。

5 まとめ

本稿では、札幌市、横浜市、京都市、北九州市の 2010 年の第 2 回定例会を対象に賛否の積極性に関連する情報の注釈作業を行い、札幌市と横浜市の注釈結果の基礎的な分析を行った。今後、全ての注釈結果を用いて分析を行い、その知見を基に賛否の積極性を判定する手法を開発したいと考えている。

謝辞

本研究の一部は、科学研究費 22300086 および放送文化基金の助成を受けたものである。

表 4: 文単位の注釈結果

	横浜市	札幌市
文数	3,766	2,480
質問	1,677 (.445)	1,152 (.465)
回答	1,075 (.285)	549 (.221)
討論	328 (.087)	97 (.039)
説明	40 (.011)	65 (.026)
その他	646 (.172)	617 (.249)
疑問文	277	160
意見性のある文	951	524
本人の意見である文	916	496
中核となる文	885	605
理由となる文	719	534
実現度を有する文	688	467
表明	466 (.677)	318 (.681)
着手	165 (.240)	117 (.250)
完了	13 (.019)	5 (.011)
拡大	44 (.064)	27 (.058)
説得性がある文	413	225

表 5: 施策・事業の注釈結果

	横浜市	札幌市
施策・事業数	424	336
賛成	387 (.913)	301 (.896)
反対	35 (.082)	31 (.092)
その他	2 (.005)	4 (.012)

参考文献

- [1] 木村泰知, 渋谷英潔, 高丸圭一, 乙武北斗, 小林哲郎, 森辰則: 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. (特集「実践 AI システム」).
- [2] 木村泰知, 渋谷英潔, 高丸圭一, 乙武北斗, 森辰則: 地方議会会議録コーパスの構築とその利用, 第 26 回人工知能学会全国大会, 3B3-NFC-4-3, 2012
- [3] 大城卓, 渡邊裕斗, 渋谷英潔, 木村泰知, 森辰則: 地方政治情報システムのための地方議会会議録への注釈付けタグセットの提案. 言語処理学会第 18 回年次大会発表論文集, pp.P3-9, 2012.
- [4] 上神貴佳, 佐藤哲也: 政党や政治家の政策的な立場を推定する—コンピュータによる自動コーディングの試み—, 選挙研究, Vol. 25, No. 1, pp. 61–73, 2009.