

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の試み

A Preliminary Study of Construction of Japanese Local Political Corpus Connecting Backgrounds, Processes and Results

木村泰知 *¹
Yasutomo Kimura

小林暁雄 *²
Akio Kobayashi

坂地泰紀 *³
Hiroki Sakaji

内田ゆず *⁴
Yuzu Uchida

乙武北斗 *⁵
Hokuto Ototake

吉田光男 *²
Mitsuo Yoshida

高丸圭一 *⁶
Keiichi Takamaru

荒木健治 *⁷
Kenji Araki

*¹小樽商科大学
Otaru University of Commerce

*²豊橋技術科学大学
Toyohashi University of Technology

*³成蹊大学
Seikei University

*⁴北海学園大学
Hokkai-Gakuen University

*⁵福岡大学
Fukuoka University

*⁶宇都宮共和大学
Utsunomiya Kyowa University

*⁷北海道大学
Hokkaido University

This paper describes a preliminary study of construction of Japanese local political corpus for connecting among backgrounds, processes and results. We have proceeded with constructing the local political corpus which can be used by researchers. Furthermore, we have aimed at research activation in related to the division of local politics. We tried to expand our local political corpus for such as Politics, Economics, Social Language and Information Engineering. However, the corpus has not been used in the area of Politics and Economics because of insufficient information of backgrounds and results. Therefore, we will construct a new local political corpus which connects backgrounds, processes and results.

1. はじめに

近年、地方分権や地方創生などの議論が盛んになっており、地方政治が注目を集めている。このため、地方議会におけるすべての発言が書き記された地方議会会議録の分析も重要性を増している。幾つかの研究分野において、自治体がウェブ公開している地方議会会議録を対象とした調査・分析が行われている [増田 2012]。このような研究では、自治体によって個別に公開された議会会議録を逐一収集しており、調査・分析を始めるまでの準備に時間と労力がかかるという問題がある。

そこで、我々は、種々の分野における地方議会会議録を用いた全国規模の研究を推進することを目指して、平成 22 年から全国の自治体の地方議会会議録の収集を始め、整理する手法を確立し、「地方議会会議録コーパス」の構築を進めた *¹。その結果、情報工学の分野において、地方議会会議録コーパスを利用した応用研究が成果をあげている [木村ら 2013][草原ら 2015][木村ら 2011]。また、社会言語学の分野においても、我々の構築したコーパスを利用したさまざまな研究が進められている [井上ら 2013][二階堂ら 2015][高丸ら 2010]。

しかしながら、政治学および経済学等の分野における研究では、これまでに収集したコーパスをそのまま利用することが困難であることが明らかになった。地方議会会議録コーパスを利用した研究をさらに学際的に推進するために、政治学および経済学分野の研究者に意見を求め、学問的要求を調査したところ、下記の 2 点が明らかになった。

1. 議論の過程としての議会会議録に加えて、議論の結果（制定された条例や、採決における投票結果）の情報があることが望ましいこと [政治経済学分野]

2. 地方独自の課題や自治体の地域特性など、議論の背景となるものに関する情報が必要であること [政治学]

そこで、我々は、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（議論の背景）を関連づけるコーパスを構築するために、地方議会会議録、条例、新聞記事の収集・整理を行う。その上で、議論の過程、結果、背景を関連づけるために、会議録に含まれる議論と条例、および、新聞記事を地域・時間・課題の 3 つの観点から結びつける。

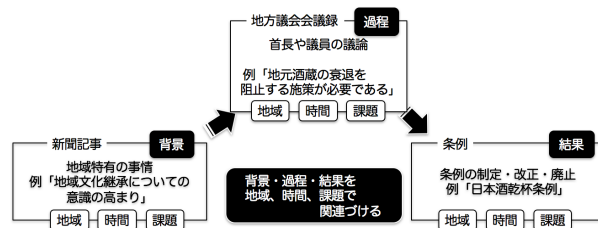


図 1: 議論の背景・過程・結果のイメージ

例えば、図 1 のように、議会における「地元酒蔵の衰退を阻止する施策が必要である」という議論は「日本酒乾杯条例」の制定という結果に結びついている。その議論には「地域文化継承についての意識の高まり」という背景がある。このような議論の結果や背景は、条例や新聞記事などの言語資源に書かれているが、それらの言語資源を個別に眺めているだけでは議論との関連性が見えにくい。そのため、「議論と結果」、および、「議論とその背景」を言語資源の有機的な連携によって示すことができれば、地方議会会議録コーパスに豊富に含まれる地域の課題に関する議論を政治学、経済学等の研究において有効に活用できると予想される。

本研究の目的は、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（議論の背景）

連絡先: 木村泰知, 小樽商科大学, kimura@res.otaru-uc.ac.jp

*¹ <http://local-politics.jp>

を結びつけることである。また、議論の過程、結果、背景を関連づけるために、会議録に含まれる議論と条例、および、新聞記事を地域・時間・課題の3つの観点によって結びつけたコーパスを構築することを目指している。

本稿では、地方政治コーパスの概要を説明するとともに、新聞記事（議論の背景）、地方議会会議録（議論の過程）、条例（議論の結果）の収集・整理の進捗状況を報告する。また、地方政治コーパスを用いて、議論の背景・過程・整理を結びける具体例を示す。

2. 地方政治コーパスについて

本章では、地方政治コーパスにおける「新聞記事」「地方議会会議録」「条例」の関係とデータの整理方法について説明する。

2.1 新聞記事

新聞記事は、地域特有の事情や議論の背景が記述されているコーパスである。新聞記事コーパスは、毎日新聞、朝日新聞、読売新聞などから購入可能であり、既に整備されていることから、容易に「時間」の抽出を行うことができる。また、地方紙版は、地域情報も掲載されていることから「地域」を特定することができる。なお、「課題」については、[木村ら 2013]の手法を参考に抽出する予定である。

2.2 地方議会会議録

地方議会会議録は、「いつ」「どこで」「だれが」「なに」を発言したのかを明らかにできる言語資源である。図2に会議録の例を示す。

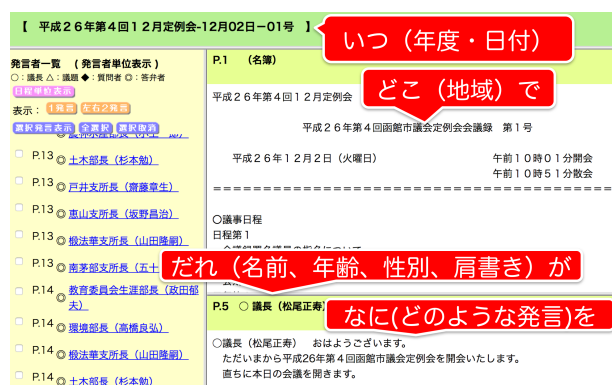


図 2: 地方議会会議録の公開例

地方議会会議録は、ウェブ上に公開する自治体が増えており、現在では約 86%の自治体がウェブ上に公開している。表1に会議録のウェブ公開数と自治体数を示す。

表 1: 会議録のウェブ公開数と自治体数

	ウェブ公開数	自治体数	公開率
都道府県	47	47	100%
市	790	790	100%
特別区	23	23	100%
町村	682	928	73%
合計	1,542	1,788	86%

地方議員の数は、都道府県議会、市町村議会、東京の特別区議会を合わせると、33,514 人である。表2に地方議員数を

示す*2。

表 2: 地方議員数

都道府県	市	特別区	町村	合計
2,687	18,654	902	11,271	33,514

我々はウェブ上に公開されている会議録を収集・整理することを進めている。地方議会会議録コーパスを構築する手順を下記に示す。

Step1 会議録 URL リストの作成

Step2 会議録の収集

Step3 会議録の整形

Step4 会議録の名寄せ

Step5 会議録の登録

Step6 会議録コーパスのウェブ公開

上記の処理項目ごとに処理内容を簡単に説明する。

[Step1] 会議録 URL リストの作成は、人出により、各自治体が会議録をウェブで公開しているかを確認し、URL を収集する作業である。会議録をウェブ公開しているサイトの URL は、既に収集しており、本プロジェクトのウェブサイト*3でも公開している。

地方議会会議録のウェブ公開は、主に4つのサービス提供会社がサポートしている。4つのサービス会社を下記に示す。

1. Discuss Net Premium 株式会社議事録発行センター
2. VOICES 株式会社フューチャーイン
3. DB-Search 株式会社大和速記情報センター
4. Sophia 神戸総合速記株式会社

都道府県議会の会議録は、89% (= 42/47) が、4つのサービス会社のいずれかによってウェブ公開されている。

[Step2] 会議録の収集では、上記の4つの提供会社に適したクローラーを作成し、収集を行っている。他のサービスを利用している会議録は、個別対応が必要であり、必要に応じて収集を進める。

[Step3] 会議録の整形は、異なる自治体の会議録を統一したフォーマットに変換することである。我々は、データベースのスキーマを定義し、17項目からなる「発言テーブル」、および、9項目からなる「議員テーブル」へ格納している。

発言テーブル

1. 識別子: 発言 ID
2. 都道府県名: 「北海道」「青森県」...
3. 回: 定例会の回数
4. 号: 各定例会の中の号数
5. 年: 開催年 「24」「25」...
6. 月: 開催月 「1」～「12」
7. 日: 開催日 「1」～「31」
8. 開催期間:
9. 表題: 「第 21 回岩手県議会臨時会」...

*2 http://www.soumu.go.jp/senkyo/senkyo_s/data/syozoku/h27.html

*3 <http://local-politics.jp/>

10. 発言者の役割: 「議長」「質問者」「答弁者」
11. 発言者 ID: 議員テーブルを参照する外部キー
12. 発言者名: 「工藤孝男」「佐々木博」...
13. 発言者の役職: 「知事」「教育長」...
14. 発言文: 「次に、冬の節電対策について伺います。」
15. 発言以外の記録文: 「(拍手)」 「[…君登壇]」...
16. 原本 URL:
17. HTML ファイルの保存場所:

議員テーブル

1. 議員 ID: 発言テーブルの 11 番目の項目と対応
2. 都道府県名: 「北海道」「青森県」...
3. 投票日: 「2011 年 4 月 10 日」
4. 対象行政区: 「札幌市中央区」
5. 氏名: 「千葉 英守」
6. フリガナ: 「チバ ヒデモリ」
7. 年齢: 「61」
8. 性別: 「男」「女」
9. 現/新: 新=新人, 現=現職

[Step4] 会議録の名寄せでは、異なる表記の同一発言者を同定する。会議録には、機械的に発言者を同定することが困難な場合がある。具体的には、自治体ごとに異なる書式、同一人物の異なる表記、政治家以外の発言者名、発言者名の記述ミスなどがある。例えば、同一人物の異なる表記名には「加藤鉦一議員」「加藤(鉦)委員」のような例がある。他にも、政治家以外の発言者名として、首長の代理として説明する自治体の職員名が存在する。さらに、発言者の記述ミスとして、同一人物に対して「遠藤達雄」「遠藤達雄」「遠藤達夫」の 3 つの表記が混在している例がある。機械的に処理できない場合には、最終的に、人手で発言者の同定を行う。

[Step5] 会議録の登録 では、整形されたデータのバージョン管理をするとともに、レコードの重複を防ぎながらデータベースへ登録を行う。

[Step6] 会議録コーパスのウェブ公開では、ウェブサイト上で「登録数の表示」「発言検索」「クロス表検索」「KWIC 検索」「地図による可視化」の実装を行っている。

2.2.1 パネルデータ作成の意義

本研究におけるコーパス構築の目的は、政治学および経済学分野の研究に対して、有用なコーパスを提供することである。政治学および経済学の統計情報を扱う分野では、欠損があるデータは研究対象に利用できず、ある条件を満たす全てのデータを正確に集める必要がある。これらの分野では、時系列データ、横断面データ、あるいは、それらを合わせたパネルデータが利用される [北村 2006]。

時系列データ (タイムシリーズデータ) は、ある現象の時間的な変化を、連続的に観測して得られた値であり、横断面データ (クロスセクションデータ) は、時間を一時点に固定して各地点で起こっているデータを記録した値である。パネルデータは、時系列データと横断面データを合わせたものであり、同一の対象を継続的に記録したデータであることから、より多くの変数を同時に織り込んだ分析が可能となる。

現在、本研究では、直近の 2 回の統一地方選挙に挟まれた期間、すなわち、平成 23 年度 (平成 23 年 4 月) から平成 26 年度 (平成 27 年 3 月) までの 4 年間の「時系列データ」と 47 都道府県議会会議録の「横断面データ」を合わせたパネルデー

タを構築している。図 3 に本研究におけるパネルデータのイメージを示す。

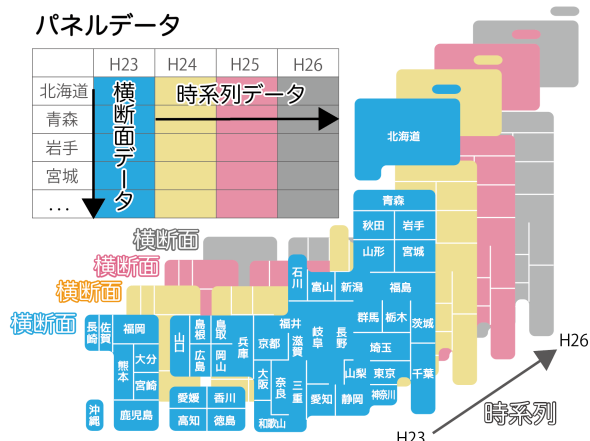


図 3: 本研究におけるパネルデータのイメージ

政治学および経済学分野における統計情報を扱う分野において利用するパネルデータは「数値」のデータセットであることから、収集した会議録テキストを「文字数」「単語の出現頻度」などへ変換する必要がある。また、収集した会議録テキストを正確に整理することが前提となる。

図 4 は、現時点の (名寄せ処理が未完成の) 会議録データを用いて、「原発」を検索した際の、平成 23 年度から平成 26 年度までのパネルデータの例である。

	都道府県\年度	23	24	25	26
1	北海道	875	401	351	199
2	青森	335	244	168	107
3	岩手	236	189	80	43
4	宮城	373	223	298	253
5	秋田	141	83	25	14
6	山形	132	67	6	8
7	福島	296	800	689	627
8	茨城	355	202	117	81
9	栃木	84	12	11	5
10	群馬	144	60	21	27

図 4: 「原発」のパネルデータの例

2.3 条例

条例は、約 94%の自治体がウェブ上に公開している。表 3 は、条例のウェブ公開率である。条例の収集・整理については、制定日、施行日、内容などに分けて整理することを検討している。

表 3: 条例のウェブ公開状況

	ウェブ公開数	自治体数	公開率
都道府県	34	47	72%
市	789	790	99%
特別区	23	23	100%
町村	827	928	73%
合計	1,673	1,788	94%

3. 議論の関連づけの例

本研究では、3つの言語資源を用いて、地域課題に関する議題の「背景」「過程」「結果」を「時間（発言日時、新聞の日付など）」「地域（自治体名・地域区分名など）」「課題（キーワード・固有名詞など）」の観点から結びつける予定である。

本章では議論を関連づける例として、平成24年12月に制定された「徳島県南海トラフ巨大地震等に係る震災に強い社会づくり条例」*4を「結果」として、「背景」や「過程」をどのように結びつけるのか説明する。

「徳島県南海トラフ巨大地震等に係る震災に強い社会づくり条例」は南海トラフや中央構造線活断層帯を震源とする直下型地震を想定して考えられた条例である。

まず、議論の背景としては、平成23年頃の記事「南海トラフ」で検索すると、平成23年12月27日の日本経済新聞に、内閣府の有識者検討会が報告した内容が記載されている。他にも、南海トラフの地域である瀬戸内海付近では、災害に対しての意識が高まり、新聞でも数多くの記事をみつけることができる。

次に、議論の過程としては、地方議会会議録の検索エンジンを用いて、議会の議論を確認することができる。図5は、発言検索を用いて「南海トラフ」を検索したKWICの結果である。検索条件の設定では、徳島県議会の地方議会会議録を対象として、平成23年度から平成26年度までに限定している。図5からわかるように、「南海トラフ」を含む発言数は531であり、前後の文脈をみると、「被害想定」「備蓄方針」などについて議論されている。

最後に、議論の結果としては、平成24年12月に「徳島県南海トラフ巨大地震等に係る震災に強い社会づくり条例」が制定されたことになる。下記に徳島県の条例を示す。

徳島県の条例

平成二十四年十二月二十一日 徳島県条例第六十四号
...
徳島県南海トラフ巨大地震等に係る震災に強い社会づくり条例

目次

第一章 総則

第一条 (目的)

この条例は、... することを目的とする。

第二条 (定義)

この条例において、次の各号に掲げる用語の意義は、当該各号に定めるところによる。...

第三条 (基本理念)...

新聞記事(背景)・議会会議録(過程)・条例(結果)のコーパスを整備し、横断的にテキストマイニング等を行い、上の例のような関係を自動的に見つけることにより、背景・過程・結果を関連づけた新たなコーパスを構築することができると考えられる。

4. おわりに

本稿では、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(議論の背景)を結

#	属性	前文脈	キーワード	後文脈
1.1	徳島県 24		南海トラフ	巨大地震対策についてであります。
2.1	徳島県 26		南海トラフ	地震防災対策特別措置法における強化指定地域に
3.1	徳島県 24	まず、	南海トラフ	巨大地震の被害想定についてであります。
4.1	徳島県 24	議第 四 号	南海トラフ	の巨大地震に関する津波浸水予測の早期公表を求める意見書 // //
5.1	徳島県 25	まず、	南海トラフ	地震などに対応した備蓄方針についてであります。
6.1	徳島県 26	まずは、想定	南海トラフ	三連動地震の経済的被害額についてであります。

図 5: KWIC による「南海トラフ」検索結果の例

びつける方法について述べた。また、現在進めている地方議会会議録のパネルデータの作成について説明するとともに、作成における課題について述べた。

本研究の学術的な特色は、学際的な応用研究を目指し、複数分野で利用できるコーパスを構築する点であり、本コーパスを政治学・経済学・言語学・情報工学の分野で利用してもらうために、異なる専門分野の研究者と連携して作成を進めることは、大きな意義があると考えている。

今後の課題は、地方政治コーパスの信頼性を高めるために名寄せ処理および発言と非発言の切り分けなどがある。また、パネルデータの情報として、性別・会派の情報を追加することを検討している。

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

- [増田 2012] 増田正: 地方議会の会議録に関するテキストマイニング分析 - 高崎市議会を事例として - 『地域政策研究』(高崎経済大学地域政策学会), Vol.15 No.1, pp.17-31, 2012.
- [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.
- [草原ら 2015] 草原史敏, 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.
- [木村ら 2011] 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.
- [井上ら 2013] 井上史雄: 「去った〇日」『ことばの散歩道』明治書院, pp.154-155, 2013.
- [二階堂ら 2015] 二階堂整, 川瀬卓, 高丸圭一, 田附敏尚, 松田謙次郎: 地方議会会議録による方言研究—セミフォーマルと気づかない方言—, 方言の研究, 1, pp.299-324, 2015.
- [高丸ら 2010] 高丸圭一, 木村泰知: 栃木県の地方議会会議録における整文についての基礎分析—本会議のウェブ配信と会議録との比較—, 都市経済研究年報, (10) pp.74-86, 2010.
- [北村 2006] 北村行伸: パネルデータの意義とその活用, 日本労働研究雑誌 551, pp.6-16, 2006.

*4 http://our.pref.tokushima.jp/reiki/reiki_honbun/o001RG00001609.html