

## 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

高丸 圭一 (宇都宮共和大学)

内田 ゆず (北海学園大学)

木村 泰知 (小樽商科大学)

### Fundamental Analysis of Panel Data of Regional Assembly Minutes for the Local Politics Corpus

Keiichi Takamaru (Utsunomiya Kyowa University)

Yuzu Uchida (Hokkai-gakuen University)

Yasutomo Kimura (Otaru University of Commerce)

**要旨** 本論文では、都道府県議会本会議を対象としたパネルデータの収集・構築について述べる。まず、地方議会の議員数や地方議会会議録の公開状況等の基本データを示す。次に、会議録の収集やコーパス構築に向けた整理の方法と問題点について発言の抽出、発言者名の抽出、発言者の名寄せの観点から述べる。最後に、地方議会会議録から抽出した特徴語に基づく、発言内容の地域差、時間差、個人差の基礎的な分析結果について述べる。

キーワード 地方議会会議録 コーパス パネルデータ 特徴語

## 1 はじめに

インターネットの普及、および、情報公開の進展に伴い、地方議会の会議録は自治体のウェブサイト公開されるようになってきた。これまで、日本各地の議会事務局や図書館等へ出向かなければ、閲覧することができなかった議会会議録を居ながらにして、しかもコンピュータ上での分析が行いやすいテキストデータとして入手出来るようになった。このことにより、地方議会会議録を活用した研究が行われるようになってきた。例えば、1つの地方議会の会議録を対象とした特定分野の政治課題に対する取組状況の分析(松本 2008, 上田他 2012)や、学校教育への応用(國原 2017)などの研究が試みられている。また、増田(2016)は北関東の7つの市議会会議録をテキストマイニングし、政策課題の比較分析を試みている。種々の研究分野において地方議会会議録を用いた研究の要求があるものの、自治体によって個別に公開されているものであるため、1つまたは一部の地方議会を対象とした事例研究にとどまっているのが現状である。日本全国を対象として研究を行うためには、調査・分析を始めるまでの準備に多大な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・整理を進め、423自治体の議会会議録を収録した「地方議会会議録コーパス」(約

1億3400万文、約80GB)を構築した。このコーパスを活用し、情報工学の分野における情報抽出手法の開発(木村他(2013), 葦原他(2015), 木村他(2011))や、言語学分野における言語使用や言語変異の実証的研究が進められてきた(高丸他(2015), 井上(2013), 二階堂他(2015))。情報工学分野、言語学分野において、本コーパスは有用な大規模言語資源としての機能することが明らかとなった。しかし、人的・時間的コストの許す範囲で可能な限り収集するという方針で、データ収集を進めた結果、収集対象自治体、および、収集対象期間にばらつきがある(図1, 2)。このことは、定量的な比較研究を行う際に問題となる場合がある。

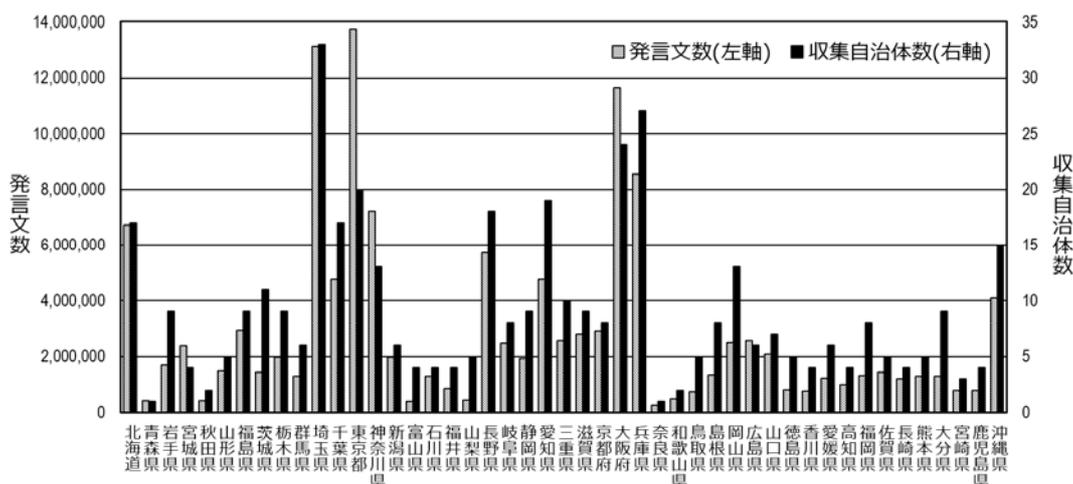


図1 都道府県ごとの収集自治体数(右軸), および, 収集文数(左軸)

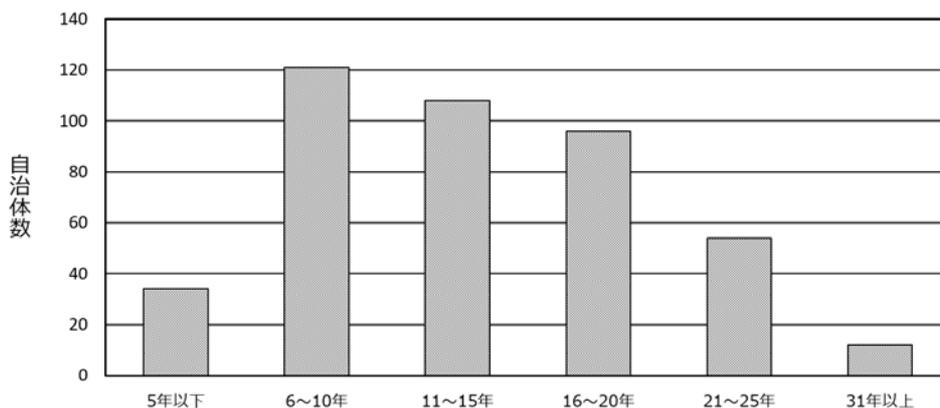


図2 収集対象期間の分布

筆者らは現在新たに、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(議論の社会的背景)を結びつけた「地方政治コーパス」の構築を進めている。このコーパスでは、比較分析を定量的に行うことができるように、地方議会会議録の収集条件(対象期間・対象自治体)を整えた収集を行う。新たに構築する会議録コーパスは、データ

規模は小さくなるものの、時系列（時間に伴う変化）、および、横断面（ある時点の地域による変異）の分析が可能なパネルデータとして使用することができる。

本稿では、平成23年4月から平成27年3月までの4年分の都道府県議会本会議を対象としたパネルデータの収集・構築、および、その基礎分析の結果について述べる。まず第2章で、コーパスの前提となる地方議会の議員数や地方議会会議録の公開状況等の基本データを示す。次に第3章で、会議録の収集やコーパス構築に向けた整理の方法と問題点について述べる。あらかじめ電子化され、テキストデータとしてウェブ上に存在する会議録を機械的に収集し、データベースに登録するコーパス構築の過程は一見すると「コピペ」作業の繰り返しであり、技術的な難易度が高くないように思われる。しかし、後から機械的な処理を行うことを想定して作成されていないテキストデータの整理には、様々な困難がともなう。特に発言の抽出、発言者名の抽出、発言者の名寄せの3点に焦点を当て、問題点を述べる。最後に第4章において、地方議会会議録から抽出した特徴語に基づく、発言内容の地域差、時間差、個人差についての基礎分析の結果を述べる。

## 2 地方議会会議録と地方議会議員

表1に自治体数、地方議会会議録のウェブ公開率、議員定数、議員実数（男女別）をそれぞれ「都道府県」「市区」「町村」別に集計した値を示す。このうち、会議録のウェブ公開率は小樽商科大学木村研究室による調査の結果であり、2017年3月現在のものである。議員定数等は総務省「地方公共団体の議会の議員及び長の所属党派別人員調等（平成27年12月31日現在）」<sup>1)</sup>に基づくものである。

表1 全国の自治体数、会議録のウェブ公開率、および、議員数

	都道府県	市区	町村	合計
自治体数	47	813	928	1,788
会議録のウェブ公開率	100%	100%	73%	86%
議員定数	2,684	19,556	11,271	33,514
男性議員数（実数）	2,414	16,541	10,083	29,038
女性議員数（実数）	261	2,802	1,064	4,127

### 2.1 会議録ウェブ公開の状況

都道府県議会については、2007年2月の時点ですべての自治体が会議録をウェブ公開していることが指摘されている（大山 2007）。それから10年を経た2017年3月現在では、都道府県議会に加えてすべての市議会と区議会が会議録をウェブ公開している。一方、町村レベルにおいては、全928町村のうち246町村でウェブサイトにおける会議録の公開が確認されず、公開率は73%にとどまる。

大多数の自治体では、会議録全文検索システムを議会のウェブページからリンクする方法で、会議録を公開している。市町村の中には、会議録のPDFやテキストファイルを

そのまま公開している自治体も存在する（例えば、小樽市<sup>2</sup>（北海道））。また、委員会の会議録の公開については、自治体により公開／非公開の対応が分かれている。

## 2.2 地方議会議員の概要

全国1,788地方自治体における、地方議会議員の総数は33,165人（欠員349）である。議員数がもっとも多い自治体は東京都議会の124人（欠員3）であり、もっとも少ない自治体は音威子府村議会（北海道）、与那国町議会（沖縄県）などの6人である。会議録を公開していない町村の割合で町村議員数を按分して差し引いても、約3万人の地方議会議員の発言がウェブに公開されている計算になる。議会では、議員に加え、首長や理事者（役所の職員）、参考人等も発言を行うため、公開された会議録に収録される発言の発言者総数はさらに多い。

次に、議員の属性（性別・年齢・出身地）について考える。表1に示した通り、議員に占める女性の割合は低く、全体で12%である。自治体区分ごとに見ると、市区は14%であり、都道府県、町村（いずれも10%）と比べやや高い。地方議会議員の年齢構成および出身地については、竹安(2004)の2002年時点での地方議会議員に対するアンケート調査がある。竹安(2004)の表Ⅲ-1に基づいて計算すると、25歳以上39歳未満が2%（有効回答16,842のうち339人）、40歳以上59歳未満が47%（同7,787人）、60以上が52%（同8,716人）である。また、竹安(2004)の表Ⅲ-2に基づいて計算すると、出生地と現在居住する都道府県が等しい議員は88%（有効回答16,844のうち14,740人）であった。このうちの86%の議員（同12,709人、全体の75%）は出生地と現在居住する市区町村が等しかった。

地方議会議員は高年層の男性が多く、自分の生まれ育ったまちにおいて議員を務めているということは直感的に明白であるが、このことがこれらのデータにより裏付けられる。

## 3 会議録の収集と整理

### 3.1 都道府県議会会議録パネルデータの収集

本節では、都道府県議会会議録パネルデータの収集について述べる。収集対象は全国47都道府県議会の本会議、収集期間は平成23年4月の統一地方選挙から平成27年4月の統一地方選挙の前（平成27年3月）までの4年間である。統一地方選挙の時期以外に議員選挙を行う議会も存在するが、収集範囲をこの4年間に設定することにより、同じ議員の4年間の発言を追跡できるという利点がある。

都道府県議会はすべて、それぞれの自治体のウェブサイトにて会議録の全文検索システムを公開している。会議録の収集はクローラと呼ばれるウェブページの収集プログラムを作成、実行して行われる。表2に示すように、全文検索システムを提供する主要4社が89%のシェアを持つ。そこでまず、この4システムに対応したクローラを作成する。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、個別対応するようにプログラムを調整したり、部分的に人手による作業を挟みながら収集作業を進める。さらに、主要4社以外のシステムによって会議録を公開している5県については、1自治体ごとにクローラを作成して、収集を行う。この作業によって収集された全国47の都道府県の本会議会議録は、約400万文、1.5GBであった。

表2 都道府県議会における会議録検索システムの採用状況

会議録検索システム	採用都道府県数	販売元
discuss	18	NTT-AT
DB-Search	16	大和速記情報センター
VOICES	6	フューチャーイン
Sophia	2	神戸総合速記
その他	5	(不明)

### 3.2 会議録の整理の考え方

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（例えば、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（例えば、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。しかし、会議録の本文には、発言以外にも目次、名簿、資料、状況記載（登壇や拍手）などが区別なく記述されている。さらに、地方議会会議録の公開方法は自治体によって差異があり、会議録のフォーマットも統一されていない。

したがって、「地方政治コーパス」を構築するにあたり、会議録を統一された形式に整理しなおす必要がある。そこで本プロジェクトでは、以下の方針で会議録の整理を進める。

- ・発言は1文単位に分割し、非発言と明確に区別すること
- ・会議録に含まれる情報を過不足なく抽出すること
- ・議員に対して、選挙時に公表されている属性情報を付与すること

以上の方針を踏まえて、都道府県議会会議録を「発言テーブル」と「発言者テーブル」で構成される関係データベースの形で整理する。発言テーブルは発言文1文を1レコードとし、以下の17項目のフィールドをもつ。

1. 識別子（主キー、発言文に固有に振られたID）
2. 自治体名（「北海道」「青森県」「宮城県」…）
3. 回（定例会の回数）
4. 号（各定例会の中の号数）
5. 年（開催年（和暦））
6. 月（開催月）
7. 日（開催日）
8. 開催期間
9. 表題（例：「平成24年 2月 定例会（第335回）-02月27日-04号」）
10. 発言者の役割（「議長」「質問者」「答弁者」）
11. 発言者ID（発言者テーブルを参照する外部キー）
12. 発言者名（例：「畠山和純君」「畑正芳君」）
13. 発言者の役職（例：「議長」「議会事務局議事課長」「知事」）
14. 発言文（例：「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（例：「(拍手)」「…君登壇」「-----（区切り線）」）
16. 原本URL
17. HTMLファイルのパス

発言者テーブルは発言者1名を1レコードとし、以下の9項目のフィールドをもつ。

1. **発言者 ID** (主キー、発言テーブルの 11 番目の項目と対応)
2. **都道府県名** (例:「北海道」「青森県」「宮城県」)
3. **自治体名** (例:「宮城県」)
4. **氏名** (例:「**島山 和純**」)
5. **フリガナ** (例:「**ハタケヤマ カズヨシ**」)
6. **対象行政区** (例:「**気仙沼市**」)
7. **生年** (投票日の西暦から投票日時点の年齢を減じて算出)
8. **性別** (「男」or「女」)
9. **職名** (例:「**議会事務局議事課長」「知事」「会計管理者」「教育長**」)

発言テーブル

都道府県別 議別子	都道府県 府県	開催 回	年月日 日	開催 表題	発言者の 役職	発言者 ID	発言者名	発言者の役職	発言文	発言以外の記録文	原本URL	ファイル 保存場所
						04000200006	島山和純君	議長	第三百三十一回宮城県議会を開 会いたします。			
						04000200006	島山和純君	議長	これより本日の会議を開きます。			
						04000200006	島山和純君	議長	本日の議事日程は、お手元に配布 のとおりであります。			
						04000210004	畑正秀君	議会事務局議事課長	御起立願います。			
						04000210004	畑正秀君	議会事務局議事課長	黙祷。			
									(起立・黙祷)			
						04000210009	村井嘉浩君	知事	本日ここに第三百三十一回宮城 県議会が開会され…			
						04000210009	村井嘉浩君	知事	初めに、我が国の歴史上いまだかつ てない大震災により犠牲と…			
						04000200003	安藤俊成君	四十九番	請願番号三三一の一、水産軍復 興特区創設の撤回に…			
						04000200003	安藤俊成君	四十九番	先ほど委員長報告にありましたと おり、産業経済委員会は…			
						04000200007	ゆきみゆき君	四十一番	二十四年度は復興元年として、宮 城県にとって新たな歴史を刻むス タートの年です。			
						04000200007	ゆきみゆき君	四十一番	今議会には、県の長期総合計画で あります宮城の将来ビジョンと…			

発言者テーブル

選挙ドットコムから得た情報

発言者 ID	都道府県名	自治体名	氏名	フリガナ	対象行政区	生年	性別	職名
04000200001	宮城県	宮城県	渡美 雄	アツミ ヨシ	東松島市	1948	男	
04000200002	宮城県	宮城県	安部 孝	アベ タカシ	宮城県 松島町 利府町	1956	男	
04000200003	宮城県	宮城県	安藤 俊成	アンドウ トシタケ	白石市、刈田郡 荒玉町、七ヶ宿町	1958	男	
04000200004	宮城県	宮城県	伊藤 和博	イトウ カズヒロ	仙台市、仙台市泉区	1960	男	
04000200005	宮城県	宮城県	須藤 哲	スドウ シロ	栗田郡、大河原町、村田町、川崎町	1944	男	
04000200006	宮城県	宮城県	島山 和純	ハタケヤマ カズヨシ	気仙沼市	1947	男	
04000200007	宮城県	宮城県	遊佐 美由紀	ユサ ミユキ	仙台市、仙台市青葉区	1964	女	
04000210001	宮城県	宮城県						会計管理者
04000210002	宮城県	宮城県						会計管理者兼出納局長
04000210003	宮城県	宮城県						監査委員
04000210004	宮城県	宮城県						議会事務局議事課長
04000210005	宮城県	宮城県						議会事務局長
04000210006	宮城県	宮城県						教育委員会委員長
04000210007	宮城県	宮城県						教育長
04000210008	宮城県	宮城県						総務部長
04000210009	宮城県	宮城県						知事

図3 発言テーブルと発言者テーブルの関係

議員の年齢や性別、対象行政区（選出選挙区）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、「4. 氏名」～「8. 性別」は選挙の立候補者情報がまとめられたサイト（例えば、選挙ドットコム<sup>3</sup>）から情報を収集し、格納する。なお、発言者が議員である場合、「9. 職名」は空欄となる。一方、発言者が議員ではない場合、「4. 氏名」～「8. 性別」は空欄とし、職名に対して発言者IDを付与する。図3に発言テーブルと発言者テーブルの関係を示す。

### 3.3 会議録の整理の流れ

会議録は従来印刷物として作成していたものであるため、ウェブ上で公開されている

会議録も、紙に出力した時の体裁に準じたフォーマットになっていることが多い。また、公開されたデータをさらに別の用途で用いることは全く考慮していないため、データが構造化されていない。会議録を収集し、発言テーブルの各フィールドに値を格納する作業を行うには、それぞれの自治体における会議録のフォーマットを逐一確認する必要がある。完全に自動化することはできない。特に、発言テーブルの「11. 発言者ID」～「15. 発言以外の記録文」を完成させるためには、「発言の抽出」「発言者の抽出」「発言者の名寄せ」を適切に行う必要がある。

以下では、それぞれの作業の具体的な手順を説明すると同時に、作業の自動化を妨げている会議録の特徴を具体的な事例を挙げて説明する。

### 3.3.1 発言の抽出

会議録に記載されている全テキストを発言と非発言を分離し、発言テーブルの「14. 発言文」、「15. 発言以外の記録文」にデータを格納する。具体的な作業手順として、まず、キーワードや区切り線、正規表現を利用して発言/非発言の判断を行う。次に、発言は句点で1文単位に分割して1レコードとし、非発言は1行を1レコードとして発言テーブルに格納する。会議録には発言以外の記録文が多く含まれる。実際の会議録の例を図4～6に示す。図中の網掛け部分が発言であり、その他は全て発言以外の記述である。発言者の入退場や登壇、不規則発言などの状況説明を〔〕内に記載している自治体が多い。また、図5のように会議中に配布された資料がそのまま会議録中に転載されることもある。これらの特徴から、発言の抽出を難しくするいくつかの要因が生じる。

1つ目は、発言/非発言の判断が難しい記述である。通常、配布資料が会議録に転載されている場合、非発言とみなすことができる。しかし、図4に含まれている資料「意見書案提出について」は、〔議事課長朗読〕となっている。この資料は会議中に読み上げられているため、発言であるとも考えることもできる。このような資料に加えて、〔「いいぞ、頑張り」と呼ぶ者あり〕などの不規則発言等の取り扱いを決める必要がある。

2つ目は、発言と非発言の境界が明らかでないフォーマットである。多くの自治体では「-」（全角ハイフンマイナス、例えば図4の1行目）や「—」（水平線、例えば、図5の2行目）の連続で区切り線を表現し、発言と非発言の境界を示している。しかし、会議録作成者が誤って異なる記号を使用したり、突然記号を変更することもある。さらに、全く区切り線がない会議録もある。この場合、手作業で発言と非発言を切り分ける必要がある。

3つ目は、プレーンテキスト以外のデータである。会議録の中に、図5、6のように、罫線素片やHTMLのTABLEタグを用いた表が組まれることがある。また、資料や表が画像ファイルとして挿入されることもある。これらのデータについて、レコードの境界や格納順を個別に決める必要がある。

△議員提出議案の報告

○小島信昭議長 議員から議案の提出がありましたので、報告いたします。

議事課長に朗読させます。

(議事課長朗読)

平成二十五年三月二十七日

埼玉県議会議長 小島信昭様

埼玉県議会議員 宮崎栄治郎  
ほか十六名

意見書案提出について

本議会に付議する意見書案を次のとおり提出いたします。

議第四号議案 医師不足の解消に不可欠な医学部新設の方針決定等を求める意見書

議第五号議案 中国からの大気汚染物質飛来への対処を求める意見書

議第六号議案 在外邦人等の更なる安全確保に向けた危機管理体制の充実・強化を求める意見書

○小島信昭議長 ただ今報告いたしました議案は、お手元に配布しておきましたから、御了承願います。

(参照 - (三四八) ページ)

図4 会議録の例 (添付資料の挿入)

○神谷幸伸 議長 日程第四 石坂真一議員(ほか十四名の議員)から議案が提出されております。

議第一号

栃木県議会議事規則の一部改正について

右の議案を別紙のとおり、栃木県議会議事規則第十五条第一項の規定により提出します。

(中略)

栃木県議会議事規則(昭和三十七年栃木県議会議事規則第一号)の一部を次のように改正する。

別表栃木県議会議事情報公開審査会の項の次に次のように加える。

議会あり方検討会	議会のあり方に関する調査及び協議	会派から選出された議員	会長
----------	------------------	-------------	----

附 則

この規則は、公布の日から施行する。

○神谷幸伸 議長 議第一号を議題とし、審議に入ります。

この際、お諮りいたします。議第一号は、提出者の説明及び委員会の付託を省略し、直ちに採決したいと思います  
が、ご異議ありませんか。

(「異議なし」と呼ぶ者あり)

図5 会議録の例 (野線素片による表組み)

○議長（佐々木一榮君） 次に、発議案1件が提出になっております。お手元に配付いたしてありますから、御了承願います。

発議案第1号

平成23年6月30日

岩手県議会議長 佐々木 一 榮 様

**(中略)**

岩手県議会委員会条例（昭和31年岩手県条例第43号）の一部を次のように改正する。

改正前	改正後
<p>（<u>常任委員会の名称，委員定数及び所管</u>）</p> <p>第2条 常任委員会の名称，委員定数及び所管は、次のとおりとする。</p> <p>（1） 総務委員会 委員10人以内</p> <p>秘書広報室の分掌に属する事項，総務部の分掌に属する事項のうち教育に関する事項を除く事項，政策地域部の分掌に属する事項，出納局の分掌に属する事項，公安委員会，選挙管理委員会，監査委員及び人事委員会の所管に属する事項並びに他の常任委員会の所管に属しない事項</p> <p>（2）～（5） [略]</p>	<p>（<u>常任委員会の名称，委員定数及び所管</u>）</p> <p>第2条 常任委員会の名称，委員定数及び所管は、次のとおりとする。</p> <p>（1） 総務委員会 委員10人以内</p> <p>秘書広報室の分掌に属する事項，総務部の分掌に属する事項のうち教育に関する事項を除く事項，政策地域部の分掌に属する事項，復興局の分掌に属する事項，出納局の分掌に属する事項，公安委員会，選挙管理委員会，監査委員及び人事委員会の所管に属する事項並びに他の常任委員会の所管に属しない事項</p> <p>（2）～（5） [略]</p>

備考 改正部分は、下線の部分である。

**(中略)**

○議長（佐々木一榮君） 次に、知事から、法人の経営状況説明書を受理いたしました。お手元に配付いたしてありますから、御了承願います。

[法人の経営状況説明書の登載省略]

図6 会議録の例（TABLEタグによる表組み）

### 3.3.2 発言者の抽出

会議録から発言者の氏名と職名を抽出し、全発言文の発言者を同定し、発言テーブルの「12. 発言者名」および「13. 発言者の役職」に値を格納する。発言者の氏名と職名の抽出処理は、発言者を示す記号（多くの場合「○」「◎」「◆」）を手がかりに、正規表現を利用して自動的に行う。表3に正規表現の具体例を挙げる。なお、網掛け部分が抽出対象である。

表3から明らかなように、自治体によって発言者を記述するパターンが異なるため、それぞれの自治体のフォーマットに合わせた正規表現を用意する必要がある。ただし、会議録作成者のミスや、書式の変更により、同一自治体の会議録であっても発言者パターンと一致しない記述がある（例えば、半角スペースと全角スペースの混在）。その場合は、個別に手作業で対応する。

会議録検索システム「Discuss Net Premium」では、左フレームに発言者（議長、質問者、答弁者）が明記されるため、発言者の抽出が容易である（図7）。しかし、自治体によってはこの欄を独自の目的で使用し、発言者以外の情報が記載している場合がある（図8）。この場合も、会議録本文中から正規表現を利用して発言者を抽出する。

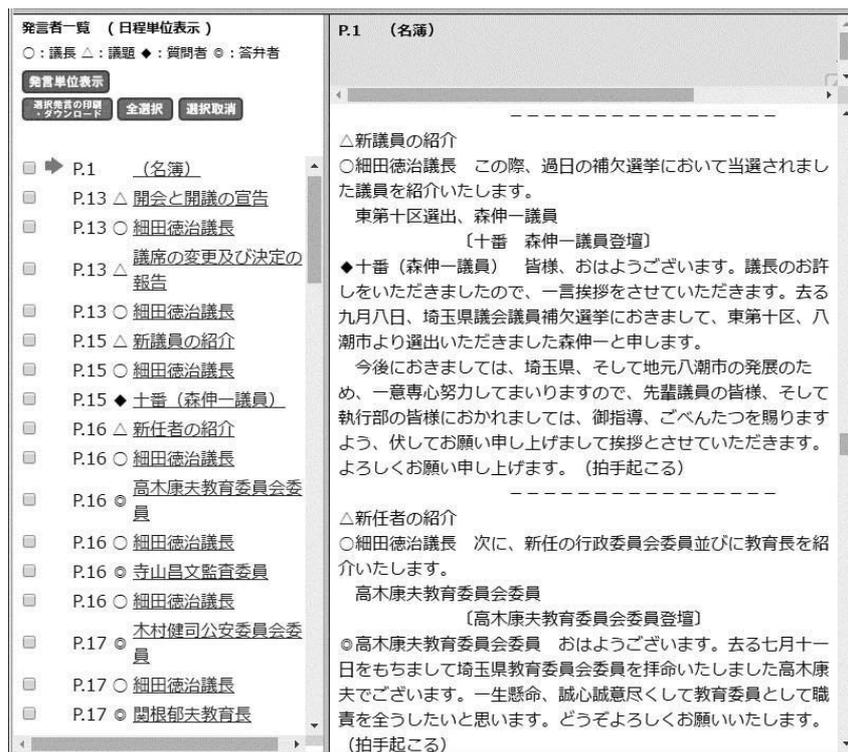


図7 左フレームの項目を本来の用途どおりに使用している例

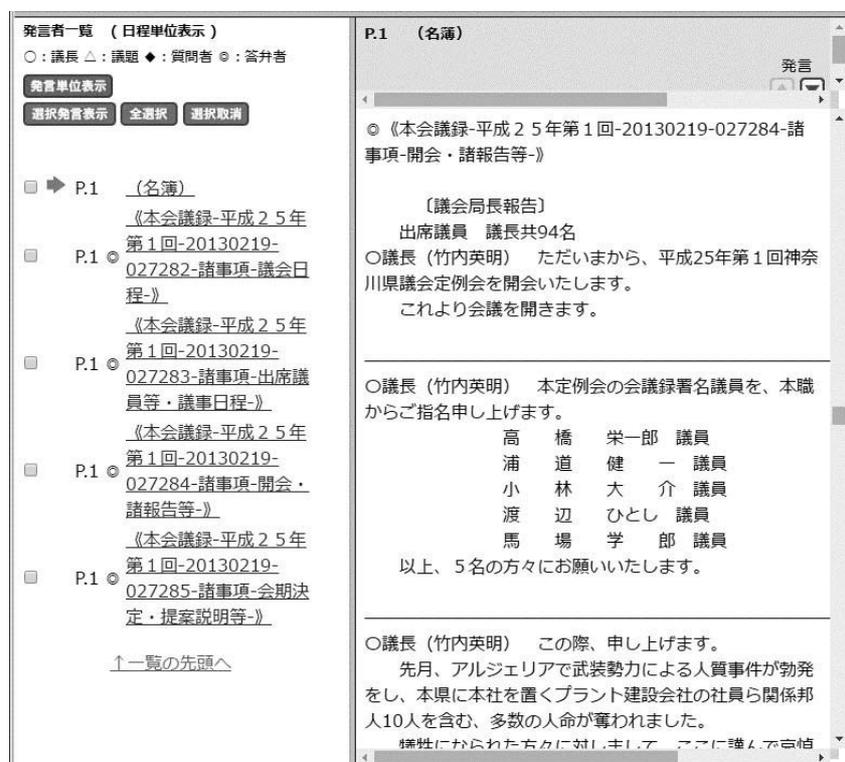


図8 左フレームの項目を本来の用途どおりに使用していない例

表3 発言者を抽出する正規表現

	発言者パターン	例
北海道	○（〔役職〕〔姓名〕君）〔全角空白〕〔本文〕	○（議長石井孝一君） 御異議なしと…
	◆（〔議席番号〕番〔姓名〕君）〔全角空白〕〔本文〕	◆（14番堀井学君）（登壇・拍手）…
	◎（〔役職〕〔姓名〕君）〔全角空白〕〔本文〕	◎（知事高橋はるみ君）（登壇）ただいま…
岩手県	○〔役職〕（〔姓名〕君）〔全角空白〕〔本文〕	○議長（佐々木博君） 次に、石川…
	○〔議席番号〕番（〔姓名〕君）〔全角空白〕〔本文〕	○41番（伊藤勢至君） それぞれ御答弁を…
	○〔姓〕〔名〕〔役職〕〔全角空白〕〔本文〕	○佐々木大和年長委員 御異議なしと…
	○〔姓名〕〔役職〕（続）〔全角空白〕〔本文〕	○千葉伝委員（続） 今、お話を伺い…
長野県	○〔役職〕（〔姓名〕〔全角空白〕君）〔本文〕	○副議長（竹内久幸 君） 休憩前に…
	◆〔議席番号〕番（〔姓名〕〔全角空白〕君）〔本文〕	◆6番（清水純子 君） 県民クラブ…
	◎〔役職〕（〔姓名〕〔全角空白〕君）〔本文〕	◎環境部長（荒井英彦 君） 地下水の…

### 3.3.3 発言者の名寄せ

各発言者に固有の「発言者ID」を与えるため、会議録から抽出した発言者を対象として、同一人物を特定する。ここでの「名寄せ」作業には2つの側面がある。

1つ目は、会議録内での同一人物の特定である。会議録から抽出した発言者は、同一人物でも複数の表記をもつことがある。例えば、「梅澤佳一自然再生・循環社会対策特別委員長」「梅澤佳一議会運営委員長」「梅澤佳一委員」「五十番（梅澤佳一議員）」「五十六番（梅澤佳一議員）」は、いずれも同一人物の「梅澤佳一」氏を示している。このような場合、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する処理を行う。

2つ目は、会議録と発言者テーブルの情報源である選挙ドットコムとの間での人物の同定である。発言者テーブルの「4. 氏名」は、選挙ドットコムの表記を採用しているが、この表記が会議録中の表記と異なることがある。この場合、人手を介して発言者名の同定を行う。

会議録と選挙ドットコムとの間で人物の同定が自動化できない原因の一つとして「通称」の問題がある。特に、氏名の一部（あるいは全部）をひらがなやカタカナ表記にする議員が多く見られる。これは「通称」として公職選挙法で認められており、会議録中にも通称が記載されることがある。一方、選挙ドットコムに登録されている氏名は、戸籍名と通称が混在している。したがって、選挙ドットコムの氏名と会議録中の氏名の表記が異なるという問題が生じる場合がある。

図9のように、選挙ドットコムでは戸籍名で表記されているが、会議録では通称が使用されている例は多い。字面が一致しない文字列を機械で自動的に同定することはできないため、選挙ドットコム掲載の氏名のフリガナから通称を導出してマッチングを行う。これとは逆に、会議録中では戸籍名が使用され、選挙ドットコムでは通称で表記されている例もある。例えば、「瑞慶覧功」議員は会議録には左記のと通りの戸籍名で表記されているが、選挙ドットコムには「ズケラン 功」と表記されている。この場合、選挙ドットコムの「戸籍名」欄を手がかりにして名寄せを行う方法が考えられる。しかし、「戸籍名」

が記載されていないことも多く、人手での処理が必要となる。

「通称」の問題以外にも、会議録や選挙ドットコム の作成者によるエラーや旧字体が名寄せの自動化を妨げる原因になっている。例えば、選挙ドットコムにおいて「萩原」を「萩原」と誤って入力しているページが存在する。この場合は、手作業で正しい表記を発言者リストに登録する。また、会議録作成の時点で発言者名の漢字変換が誤っている例(正:「達雄」、誤:「達夫」「達雄」など)も多い。この場合は、発言テーブルの「12. 発言者名」の誤表記は修正せず、発言者テーブルには正しい表記のみを登録する。そして、同一人物には同一IDを与えることで名寄せを実現する。

現在一部表示を調整中のためご迷惑をお掛けしています。

**柳ヶ瀬 裕文** ヤナガセ ヒロフミ 42歳(男)

選挙	東京都議会議員 選挙	選挙区	大田区選挙区
党派	日本維新の会	肩書き	東京都議会議員 2017年都議選 立候補 予定者
学歴	早稲田大学卒業	職歴	株式会社ジェイアール東日本企画(広告 会社) 大田区議会議員
戸籍名	柳ヶ瀬 裕文	戸籍名(カナ)	ヤナガセ ヒロフミ

(<http://go2senkyo.com/seijika/20428>)

○議長 (吉野利明君) 二十七番やながせ裕文君。  
(二十七番やながせ裕文君登壇)

○二十七番 (やながせ裕文君) 私は、日本維新の会東京都議団を代表して、今定例会に提出された全  
ます、第百九十三号議案から百九十七号議案まで、交通局における来年四月からの消費税率の引き上  
国の消費税率引き上げに伴う公共料金の改定に当たっては、国民全てに平等に課税され、税負担が適  
の改定に当たっては、都営交通が置かれている経営の現状、これを考慮し、消費税にかかわる料金転嫁  
仮に、転嫁を行わなかった場合には、健全な経営を損ない、その結果、料金の一層の値上げという事  
ことのないよう、転嫁することはやむを得ないものと考えますけれども、転嫁に当たっては、公営企業

図9 氏名の表記が異なる例 (戸籍名と通称)

## 4 TF・IDFによる特徴語比較

### 4.1 方法

本章では、TF・IDF法を用いて、都道府県別、発言時期別、議員別の特徴語を抽出し、それぞれの特徴比較を試みる。基本的には、ある文書において出現頻度 (TF) が高い語がその文書における特徴語であると考えられる。ただし、その語が複数の文書で共通に用いられる語であるとき、語はその文書における特徴語といえないため、他の文書で共通に使われる語の順位が下がるような係数 (IDF) をTFに乘じ、単語を順位付けする手法である。都道府県別の特徴語抽出では、1つの議会における発言全体を1文書として、47文書 (都道府県議会) の特徴語をそれぞれ求める。発言時期別の特徴語では、全議会における四半期 (3カ月) の間の発言を1文書にまとめ、全16文書 (4年×四半期) の特徴語を

求める。議員別の特徴語は1議員の全発言を1文書とし、議会ごとに発言議員の人数分の文書の特徴語を求める。

文書*j*における単語*i*の $TF_{ij}$ および、単語*i*の $IDF_i$ はに次式で求める。ただし、 $N_j$ は文書*j*の総単語数、 $n_{ij}$ は文書*j*における単語*i*の出現頻度、 $D$ は総文書数、 $df_i$ は単語*i*を含む文書数である。

$$TF_{ij} = \frac{n_{ij}}{N_j} \quad IDF_i = \frac{D}{df_i}$$

特徴語の抽出では、まず形態素解析を行い、各単語の頻度を求める。一般的な形態素解析を行うと、例えば、「離島航空路線」は「離島」「航空」「路線」の3語に分割され、「地域医療」は「地域」「医療」の2語に分割される。このまま「路線」や「地域」といった語の頻度を数えても、特徴を捉えることは難しいため、名詞連続（複合名詞）については、ひとまとまりの語として取り扱うことが適当である。そこで、本稿では中・長単位解析ツール「Comainu<sup>4</sup>」を用い、長単位<sup>5</sup>を基準として、品詞情報が「名詞-普通名詞-一般」であるものを対象に、特徴語の抽出を行う。形態素（短単位）解析と長単位解析の例を図10に示す。ゴシック体で示した語が抽出対象となる「名詞-普通名詞-一般」の長単位である。

<b>【入力文】</b>															
人口減少、少子高齢化の進行は、労働力人口の減少等による経済規模の縮小、社会保障給付費の増大、さらには、地域コミュニティの弱体化など、国民生活全般に大きな影響を及ぼす重要な課題であります。															
↓															
<b>短単位(上段)</b>	人口	減少	、	少子	高齢化	の	進行	は	、	労働	力人口	の	減少	等	
<b>長単位(下段)</b>	人口減少		、	少子高齢化			の	進行	は	、	労働力人口		の	減少	等
による															
経済規模															
の															
縮小															
、															
社会															
保障															
給付費															
の															
増大															
、															
さら															
には															
、															
地域															
コミュニティ															
の															
弱体化															
など															
、															
国民															
生活															
全般															
に															
大きな															
影響															
を															
及ぼす															
地域															
コミュニティ															
の															
弱体化															
など															
、															
国民															
生活															
全般															
に															
大きな															
影響															
を															
及ぼす															
重要															
な															
課題															
で															
あり															
ます															
。															
重要															
な															
課題															
で															
あり															
ます															
。															

図10 長単位解析の例(平成26年度栃木県議会第324回臨時会議06月03日-03号/福田富一 知事)

3.3.1節で述べたように、収集した会議録には、発言と発言以外の記述が混在する。現段階では手作業による区別が完了していないため、発言とそれ以外の記述を簡易的に区別する手がかりとして句点「。」を用いる。句点で終わる文を発言と見做し、分析対象とする。また、議事進行等の定型発言を除外するために、発言者名の文字列に「議長」「事務局」を含むものを対象外とする。

## 4.2 都道府県別の特徴語の比較

表4に、TF・IDFによって求めた都道府県別の特徴語上位5語を示す。特徴語から以下のような傾向および問題点を読み取ることができる。

### ① 各地域の政治課題

分析対象期間における各地の政治課題が多く見られる。特に、医療施設（例えば「ふくしま国際医療科学センター」「さいたま赤十字病院」「東千葉メディカルセンター」）は新設や移転にあたりそれぞれの地域で活発な議論があり、上位に挙げられた。茨城県の「森林湖沼環境税」（期間延長）、東京都の「二〇二〇年大会」（オリンピック）、長野県の「浅川ダム」「F-V断層」（地震への対応）、奈良県の「奈良公園」「平城宮跡」「ホテル誘致」、鳥取県の「環境大学」（公立化）、滋賀県の「近江大橋」（無料化）なども、分析対象期間に議論になった政治課題であるといえる。また、沖縄県の「在沖海兵隊」は、恒常的な地域特有の政治課題である。

### ② 地域の特産物・ブランドの名称

各都道府県で売り出している特産物の名称が見られる（例えば「鳴門わかめ」「すだち」）。特にブランド牛に力を入れている自治体が多い様子が見られる（例えば「仙台牛」「あか牛」「豊後牛」）。ゆるキャラを含む施設名（例えば栃木県の「とちまるショップ」（東京スカイツリーへの出店））も見られる。

### ③ 都・道・府の呼称

都道府県のうち、「都」「道」はそれぞれ1つ、「府」は2つしかないことから、「県」に置き換えると全国的に一般的な語が都道府において特徴語の上位に挙げられた。特に北海道の特徴語は「道教委」「本道経済」「道民生活」と続くが、これらが「県教委」「本県経済」「県民生活」であれば、上位には挙げられないと考えられる。東京都の第1位「都有地」なども同様であると考えられる。

### ④ 「(都道府県名)づくり」「オール(都道府県名)」

複数の都道府県で「(都道府県名) + づくり」あるいは「オール + (都道府県名)」という特徴語が複数見られた。（例えば「いばらきづくり」「京都づくり」「大分県づくり」「オール秋田」「オールとちぎ体制」）

### ⑤ 長単位の解析誤り等

栃木県の第2位「ちぎづくり」（→とちぎづくり）、和歌山県の第2位「やま大会」（→わかやま大会）は長単位解析に失敗している例である。また、富山県の第1位「湾クラブ」（→世界で最も美しい湾クラブ）、京都府の第2位「海病院」（→京都府立与謝の海病院）は助詞等を含むより長い固有表現の一部が切り取られたものであるため、長単位だけでは意味を捉えることができない。

#### 4.3 発言期間別の特徴語の比較

表5, 6に鳥取県, および, 宮城県における四半期単位の発言期間別の特徴語を示す。鳥取県では, 平成23年度第2四半期に「生レバー」が第1位に挙がっている。これは平成23年4月に発生した食中毒事件に関連していると考えられる。また, 宮城県の平成23年度第1～2四半期には, 東日本大震災関連の語が上位に多く見られる(例えば「東日本大震災対策」「震災復興計画案」「長期避難世帯」)。このほか, それぞれの次期に議論になった特徴語が見られると考えられるものの, 1つの四半期には概ね1回の定例会しか行われないため, 各期に発言される議員数が限定される。したがって, その期間に質問を行った特定の議員の発言内容(関心分野)が発言期間別の特徴語に反映されてしまっている可能性が考えられる。この点については, 議員別の特徴語と照らし合わせて, 今後確認を進めたい。

#### 4.4 議員別の特徴語の比較

表7に, TF・IDFによって求めた栃木県議会議員の議員別の特徴語上位5語を示す。対象の期間に議長等ではなく, 一議員の立場で発言した議員は51名である。このうち, 発言数が少なく, 特徴語を抽出できなかった2名を除いた49名を示した。発言特徴語は多岐にわたっており, 自治体が抱える政治課題が多岐にわたっている様子がうかがえる。表4の栃木県全体の特徴語上位5語は, 議員別の特徴語には挙がっていない。これらの語は複数の議員が使用したため, 議員別の計算ではTF・IDF値が低くなったためであると考えられる。全国における都道府県別の特徴語, 県内の議員別の特徴語がそれぞれ適切に抽出されていると考えられる。

議員ごとの特徴語をみると, 各議員が関心を持って取り組んでいる課題が見えてくる。例えば, 「モニタリングポスト」「干しシイタケ」「二重被害」「キノコ生産者」「クマザサ」(渡辺渡議員)という特徴語から, 放射能汚染の被害についての関心をもって発言していることがわかる。また, 都道府県議会議員は県全体の政治課題に取り組む必要がある一方で, 選出選挙区の代表という側面を持つため, 特定地域の地名, 施設名が特徴語として多く現れる傾向も見られる。例えば, 斎藤淳一郎議員の特徴語には「矢板南産業団地」「シャープ」といった矢板地域に関わりの深い語が見られる。女性議員に子ども関連の特徴語が多いことも全国的に見られる傾向である。例えば, 阿部博美議員の特徴語には「不育症」「障害児」「里親制度」などが見られる。

有権者(一般市民)が議会会議録の全て読むことは労力がかかるため, 現実的ではないが, このように特徴語を抽出することで, それぞれ議員の議会における発言傾向を簡単につかむことが可能となる。

## 5 まとめ

本稿では、まず、地方議会会議録のウェブ公開状況を示した。公開率は2010年（高丸他2011）の調査から13ポイント上昇した。情報公開が進んだことと市町村合併により小規模な自治体が減少したことの2つの要因が考えられる。次に、地方議会における主要な発言者である地方議会議員の属性について、データに基づく集計結果を述べた。高年層の男性が多く、生まれ育ったまちにおいて議員を務めていることがデータにより裏付けられた。発言者の属性（バックグラウンド）は発言内容の分析においても、言語的特徴の分析においても重要な手がかりとなるものである。現在構築中のコーパスでは、会議録の発言文（発言テーブル）に加えて、発言者の情報（発言者テーブル）を収集し、この要求への対応を試みている。現在のところ、議員の所属する党派（会派）に関する情報は付与していないが、政治学的な分析を進めるために必要な情報であると考えられるため、今後、会派情報の付与についても検討を進めたい。

次に、収集した会議録をデータベースへの登録するために必要となるデータ整理における問題点について、発言の抽出、発言者の抽出、発言者の名寄せの3つの観点から述べた。今後、会議録の収集対象範囲を広げつつ、将来にわたってデータの収集と整理を続けるためには、収集からデータベースへの登録に至るまでの処理をプログラムによって自動化することが望ましい。しかしながら、フォーマットが不統一で、構造化されていないテキストデータを対象にした自動処理には解決困難な問題が多数存在することを明らかにした。解決方法の検討を進めるためには、問題点の所在を明らかにし、共有することが肝要である。今後、解決に向けた検討を進めたい。

最後に、地方議会会議録からTF・IDF法を用いて特徴語を抽出し、発言内容の地域差、時間差、個人差について基礎的な分析を試みた。このような統一的な比較分析を行えることが、期間と範囲をコントロールしたパネルデータの利点である。本稿の結果を出発点として、今後パネルデータについて、さまざまな観点から学際的に分析を進めていきたい。

なお、紙幅の都合で、発言期間別の特徴語は2つの県の結果のみ、議員別の特徴語は1つ県の結果のみを示した。また、いずれの表においても、特徴語は上位5語までしか示すことができなかった。本稿で示すことが出来なかった都道府県の特徴語、および6位以下の特徴語については、「地方議会会議録コーパスプロジェクト」のウェブサイト<sup>6</sup>において、順次公開する予定である。

## 謝辞

本研究は科学研究費基盤(C) No.26370498「学際的応用を考慮した地方議会会議録コーパスの言語学的研究」の助成を受けたものである。

## 参考文献

- [1] 松本直樹「地方議員の図書館への関心に関する予備的考察：埼玉県市議会の議会会議録分析を元に」『日本図書館情報学会誌』54(1), 39-56, 2008
- [2] 上田翔, 八木田浩史「地方議会議事録における環境用語の出現頻度に基づく自治体の環境問題対応の解析」『環境情報科学学術研究論文集』26, 283-288, 2012
- [3] 國原幸一郎「地方議会における争点をふまえた公民の授業：東海豪雨と東日本大震災を事例として」『名古屋学院大学論集 人文・自然科学篇』53(2), 93-106, 2017
- [4] 増田正「北関東地方における政策課題と地方議会改革—主要7市議会会議録のテキストマイニング分析—」『地域政策研究』18(2・3), 33-49, 2016
- [5] 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』28(4), 370-378, 2013
- [6] 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』J98-D(11), 1390-1401, 2015
- [7] 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』26(5), 580-593, 2011
- [8] 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』30(1), SP2-K, 306-318, 2015
- [9] 井上史雄「去った〇日」『ことばの散歩道』明治書院, 154-155, 2013
- [10] 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, 299-324, 2015
- [11] 大山英久「地方議会の公開と会議録をめぐって」『レファレンス』57(6), 31-46, 2007
- [12] 小椋秀樹, 富士池優美「第4章 形態論情報」国立国語研究所コーパス開発センター編『現代日本語書き言葉均衡コーパス利用の手引』, 2011
- [13] 高丸圭一, 渋谷英潔, 木村泰知「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」『都市経済研究年報』, 11, 47-72, 2011

## 注

- 1 [http://www.soumu.go.jp/senkyo/senkyo\\_s/data/syozoku/h27.html](http://www.soumu.go.jp/senkyo/senkyo_s/data/syozoku/h27.html) (アクセス日:2017/03/25)
- 2 [https://www.city.otaru.lg.jp/sisei\\_tokei/gikai/honkaigi.html](https://www.city.otaru.lg.jp/sisei_tokei/gikai/honkaigi.html) (アクセス日:2017/03/25)
- 3 <http://go2senkyo.com/> (アクセス日:2017/03/28)
- 4 <http://comainu.org/> (アクセス日:2017/03/25)
- 5 小椋他(2001)「長単位は構文的な機能に着目した, 各媒体・各ジャンルの資料的な性

格を反映する単位であり、言語的特徴を解明するという目的にかなうものと言える。」

6 <http://local-politics.jp/>

表4 TF・IDFによって求めた都道府県別特徴語（上位5語）

都道府県	特徴語				
	1	2	3	4	5
北海道	道教委	本道経済	道民生活	道警察	道総研
青森県	青森県基本計画未来	防災公共	あおり産業総合支援センター	下北半島縦貫道路	地域県民局
岩手県	大震災津波	医療局	岩手県一般会計補正予算	第2期アクションプラン	医療法人白光
宮城県	仙台牛	貞山運河	水産業復興特区	みやぎ発展税	富県戦略
秋田県	第2期ふるさと秋田元気創造プラン	阿仁熊牧場	内陸線	オール秋田	総合戦略産業
山形県	総合支庁	最上地域	山形デザインেশンキャンペーン	短期アクションプラン	文教公安常任委員会
福島県	新生ふくしま	ふくしま国際医療科学センター	議案審査報告書	イノベーション・コース構想	廃炉安全監視協議会
茨城県	森林湖沼環境税	いばらきづくり	つくば国際戦略総合特区	県北山間地域	中本委員会所管事項
栃木県	新とちぎ元気プラン	ちぎづくり	とちまるショップ	ちぎ元気プラン	オール栃木体制
群馬県	東国文化	東毛広域幹線道路	群馬DC	事務・事業見直し委員会	高崎競馬場跡地
埼玉県	さいたま赤十字病院	危機管理防災部長	継続審査事項	通商産業政策	民主党・無所属
千葉県	東千葉メディカルセンター	北千葉道路	千葉ニュータウン	千葉県一般会計補正予算	袖ヶ浦福祉センター
東京都	都有地	都内中小企業	二〇二〇年大会	島しょ地域	都市東京
神奈川県	緊急財政対策	マグネット神奈川	保健福祉局長	ヘルスケア・ニューフロンティア	さがみ縦貫道路
新潟県	県央基幹病院	趣旨弁明	議案採決	夢おこし政策プラン	少子化対策モデル事業
富山県	湾クラブ	新湊大橋	第1条歳入歳出予算	経営計画概要	付け提出し
石川県	小松基地	のと里山海道	IRいしかわ鉄道	奥能登地域	里山海道
福井県	足羽川ダム	嶺南地域	県都デザイン戦略	エネルギー研究開発拠点化計画	民主・みらい
山梨県	自民党・県民クラブ	富士・東部地域	防災新館	富士北麓地域	峡東地域
長野県	林務部長	浅川ダム	30人規模学級	改革・新風	F-V断層
岐阜県	国ぎふJづくり	ぎふ清流国体・ぎふ清流大会	森林・環境税	行財政改革アクションプラン	未来会館
静岡県	内陸フロンティア	地震・津波対策アクションプログラム	グランシップ	富国有徳	静岡式
愛知県	あいちトリエンナーレ	重点改革プログラム	第五次行革大綱	暫時休憩さ	あいち森
三重県	みえ県民力ビジョン	東紀州	みえ産業振興戦略	北勢地域	自民みらい
滋賀県	近江大橋	本委員会所管部分	経過ならび	環境こだわり農業	新生美術館
京都府	京都経済	海病院	府立医科大学	京都づくり	府北部地域
大阪府	咲洲庁舎	府民文化部長	グランドデザイン・大阪	環境農林水産部長	OTK
兵庫県	第2次行革プラン	兵庫づくり	民主党・県民連合議員団	舞台・兵庫	21世紀兵庫長期ビジョン
奈良県	奈良公園	若草山	平城宮跡	議員お述べ	ホテル誘致
和歌山県	議案付託表	やま大会	コスモパーク加太	紀伊半島一周高速道路	経済警察委員会
鳥取県	子育て王国	環境大学	まんが博	環境管理事業センター	砂場議員
島根県	民主県民クラブ	一般事件案	自民党議員連盟	まめネット	農水商工委員会
岡山県	第3次おかやま夢づくりプラン	夢づくりプラン	おかやまマラソン	事業再点検	事件案件
広島県	国際平和拠点ひろしま構想	広島市東部地区連続立体交差事業	ひろしまブランド	チャレンジビジョン	未来創造支援事業
山口県	愛宕山開発用地	産業戦略本部	産業力・観光力	岩国基地問題	未来開拓チャレンジプラン
徳島県	農林水産総合技術支援センター	鳴門わかめ	すだち	次世代林業プロジェクト	vs東京
香川県	希少糖	香川づくり	せとち田園都市香川創造プラン	香川県産業成長戦略	さぬき讚フルーツ
愛媛県	えひめ国体	チーム愛媛	委員会審査報告書	いやし博	愛媛づくり
高知県	地域福祉部長	林業振興・環境部長	農業振興部長	永国寺キャンパス	地産外商公社
福岡県	民主党・県政クラブ県議団	沖端川	矢部川水系	まごころ製品	アンビシャス広場
佐賀県	所管事項一般	好生館	総額調整ルール	国際線専用施設	城原川
長崎県	県民所得向上対策	石木ダム	連立会派	アジア・国際戦略	長崎県づくり
熊本県	新4カ年戦略	4カ年戦略	条例等関係	立野ダム	あか牛
大分県	安心・活力・発展プラン	大分県づくり	豊後牛	大分県一般会計補正予算	対面席
宮崎県	県民政策部長	県立宮崎病院	防災拠点庁舎	議案・請願委員会審査結果表	県立延岡病院
鹿児島県	行財政運営戦略	県政刷新大綱	マリンポートかごしま	共生・協働	県開発促進協議会
沖縄県	21世紀ビジョン	下地島空港	軍転協	在沖海兵隊	入域観光客数

表5 TF・IDFによって求めた鳥取県における発言期間(四半期単位)別特徴語(上位5語)

年度	四半期	特徴語				
		1	2	3	4	5
23	1	防災監	福祉避難所	郵政改革法案	地上系	サマータイム
	2	生レバー	カジノ	シイタケ生産	職員住宅	青ナシ
	3	脳脊髄液減少症	ノロウイルス	認定NPO法人	科研費	地どり
	4	差額ベッド料	シーリング材	芝生化	地域点	砂防林
24	1	漁船法	電子自治体	ネーミングライツ	漁船検認業務	漁船登録
	2	入園券	シンナー	施設型農業	平均正答数	意思決定ルール
	3	両三柳工区	プレハブ	ボートパーク	土木遺産	仲卸市場
	4	空中散布	文学館	文学者	地域づくりセンター	ボランティア・市民活動支援センター
25	1	ペインクリニック	大山賛歌	GNI	性被害	最大多数
	2	義務負担	同種工事実績	不払い残業	品確法	判断水位
	3	軽自動車税	以西小学校	日本語対应手話	紙ごみ	自然遺産
	4	ペアレンタルコントロール	無料低額診療事業	魔法遣い	地方共同税	設計額
26	1	タンデム自転車	シーニックバイウェイ	ピロリ菌	海上警察権	防霜対策
	2	みどり	行政刑法	社会常識	義務教育費国庫負担制度	行政法
	3	高架化	寡婦控除	オオサンショウウオ	マタハラ	社会貢献活動
	4	市町村コンシェルジュ	地方創生特区	旅行券	ロケットスタート	地方創生事業

表6 TF・IDFによって求めた宮城県における発言期間(四半期単位)別特徴語(上位5語)

年度	四半期	特徴語				
		1	2	3	4	5
23	1	東日本大震災対策	中央受け付け分	早期支給	建ぺい率	長期避難世帯
	2	震災復興計画案	津波防御施設	援護寮	収納スペース	教育レベル
	3	ハートバッジ	PICU	園芸農業	JFEエンジニアリング	相互支援体制
	4	民生委員・児童委員	エネルギー供給体制	塩釜支所	海洋エネルギー	被災民生委員・児童委員
24	1	サイレン音	不妊治療	森林・林業	入札調書	慶長遣欧使節四百年記念事業
	2	地域主導	築館工業団地	復興関係予算	精神保健医療福祉	精神科救急部会
	3	非常勤講師	地域医療支援病院	尖閣問題	大震災復旧・復興対策	第六次地域医療計画中間案
	4	二項道路	狭隘道路	港湾運営会社	ワンストップ支援センター	女性相談交番
25	1	田川ダム	風しん	GPS装置	足こぎ車いす	ダム事業
	2	宗教法人	司法面接	不活動宗教法人	交通指導取り締まり	歴史的題材
	3	コンバインドサイクル	開成包括ケアセンター	議会選出監査委員	宮城復興局側	農業予算
	4	地域公共交通	人事評価	伊達文化	スケート場	みやぎ財政運営戦略
26	1	医学部附属病院	かけそば	防災キャビネット	集中架電作戦	サメ漁
	2	水土里ネット	全員協議会	中高一貫校	受動喫煙防止	手話言語法
	3	土砂災害対策	助産師外来	都市計画道路	がん患者	ワッチ体制
	4	教育運営協議会	医療法人K	日本地図	武器等製造法	県有化

表7 TF・IDFによって求めた議員別特徴語（栃木県議会の例）（上位5語）

議員名	特徴語				
	1	2	3	4	5
阿久津憲二	危険負担	復興院	一カ月半	満タン	
阿部寿一	E S C O事業	日光国立公園	木質バイオマス発電	サテライトオフィス	家庭教育オピニオンリーダー
阿部博美	不育症	障害児	性感染症	里親制度	ご夫婦
一木弘司	芳賀郡市	県版学力テスト	二宮遊水地	全国学力テスト	穴川用水
横松盛人	物損事故情報管理システム	同地域	大谷橋	一時保護所	計画的・効率的活用
加藤正一	防災士	国宝指定	地域づくり団体	鏝阿寺本堂	広域物流拠点
花塚隆志	中心的機能	大谷石塀	東日本全体	中枢機能	さくら市
角田まさのぶ	大学進学	不合格	消防防災ヘリ	生活保護世帯	授業中
梶克之	再建計画	物産センター	オリンピック競技大会	臨時雇用	パラリンピック競技大会
関谷暢之	子ども被災者支援法	発注数	低線量メニュー	しごき	住宅除染
岩崎信	伝統工芸品	茅葺き屋根	益子焼	旬菜店	窯業技術支援センター
亀田清	ジェネリック医薬品	訪問指導	県産木材	県道佐野田沼線	林道等
琴寄昌男	統一マスケットキャラクター	歴史文化	足尾鉍毒事件	先輩方	トレー
金子裕	農業高校	歯科健診	二次交通	成人期	渡良瀬架橋
五十畑一幸	非農用地	乙女大橋	キャンプ地誘致	八年後	農家負担
五十嵐清	事業協同組合	農業水利施設	拉致被害者	母国語	企業局等
高橋修司	サイバー攻撃	当該補助金	産業復興	高額療養費	創造元年
佐原吉大	緩和ケア	抽せん	抗インフルエンザウイルス薬	復職プログラム	精神疾患患者
佐藤栄	県有林	社会的企業	県単事業	釣り人	高齢者施設等
佐藤良	通年会期	定期予防接種	定例日	八ッ場ダム建設事業	サポートステーション
三森文徳	片づけ	コントラクター	飼料自給率	道徳教育	国際的人材
山形修治	乾しシイタケ	土砂災害警戒区域	定期点検	県有建築物	生産工程管理
山口恒夫	好き嫌い	脳脊髄液減少症	復興推進基金	栄養・食生活	栄養バランス
若林和雄	新法人	所得水準	校舎・校庭等	野焼き	十五項目
小林幹夫	ロコモティブシンドローム	アルツハイマー病	出馬宣言	遊休農地	健康教育
松井正一	病気等	運転免許	複合的利用	意見書案	被疑者
神谷幸伸	B型・C型肝炎	当該プラン	救済対象	地域材	県内経済活性化
神林秀治	家庭用蓄電池	プレイベント	出現率		県立中高一貫校
西村しんじ	国際公会	計基準	産業振興協議会	加入率	コンクリート診断士
斉藤孝明	通常分	情報戦略	外向け	大規模陥没	国民保護計画
石坂真一	県内放送メディア	機動センター	特別選考	リニューアル	全面オープン
早川けいこ	生涯スポーツ	骨髄移植	クリニカルラダー	県立美術館	文化・芸術
早川尚秀	厚生年金基金	障害者スポーツ	自然減	国家的緊急事態	六観地
相馬憲一	口蹄疫	日本史	病児・病後児保育施設	中学校長	医薬品
相馬政二	要保護児童	ちぎ観光リーダー育成塾	誘客回復	各部署	児童虐待等
増淵三津男	難視世帯	横川東小学校	徴収体制	地すべり	一期計画
池田忠	地熱発電	栃木元気	獣医師	シビックプライド	既存ダム
中川幹雄	空手道	養護老人ホーム	信用保証協会	同原子力発電所	医学部設置
中島宏	少子高齢・人口減少社会	新潟県立国際情報高等学校	ニート等	高等課程	地域学習
鶴貝大祐	主幹教諭	足利市教育委員会	リバースオークション	新生児マス・スクリーニング検査	タンDEMマス・スクリーニング検査
渡辺さちこ	売電先	公衆無線LAN	新電力	フィルムコミッション	一般競争入札
渡辺渡	モニタリングポスト	干しシイタケ	二重被害	キノコ生産者	クマザサ
白石資隆	筆記試験	面接官	公営住宅	一次面接	生徒指導能力
板橋一好	トレーニングセンター	県道昇格	鳥山最終処分場	環境保全公社	平池議員
平池秀光	臨時採用	危機管理・国土保全	逃げ場	議員さん	七十人
保母欽一郎	老健施設	竜巻対策	出産適齢期	他府県	少子化問題
野澤和一	G S世代	ピロリ菌	胃がん	中小・小規模企業	保育サービス
螺良昭人	政務活動費	屋外プール	慢性疲労症候群	積算線量計	東北地方太平洋沖地震
齋藤淳一郎	矢板南産業団地	災害時要援護者避難支援プラン	シャープ	矢板市内	再生可能エネルギー産業