

地方議会では何が話題になっているのか — 宇都宮市議会会議録のテキストマイニング —

What topics are discussed in a local assembly?

— A Text-mining Analysis of the Minutes of Utsunomiya City Assembly —

高 丸 圭 一 (宇都宮共和大学 准教授)

目 次

- 1 はじめに
- 2 対象データと分析の方法
- 3 対応分析によるテキストマイニング
- 4 出現頻度の比較
- 5 まとめ

1 はじめに

地方議会会議録には、地方議会（都道府県議会、または市区町村議会）の本会議および委員会における議員や首長などの発言が書き起こされ記録されている。ウェブによる情報公開が進み、従来は文書（冊子）として作成され、役所等で閲覧することしかできなかった会議録はインターネットを介して誰でも閲覧することができるようになった。都道府県のレベルでは2007年の時点で全国47都道府県すべて（大山2007）が、市区町村のレベルでは2010年の時点で73.4%（高丸・渋谷・木村2011）が会議録をウェブサイトにて公開している。

筆者らは地方議会会議録の学際的な利用を促進するために、自治体のウェブページ上に公開された地方議会会議録を収集し、コーパスを構

築（齋藤・大城・菅原・永井・渋谷・木村・森2011）した上で、全文検索システムおよび機能的な横断検索を実行するウェブユーザインターフェイスの作成（乙武・高丸・渋谷・木村・森2013）を進めている¹。

近年、種々の研究分野で地方議会会議録が利用されはじめている。自然言語処理の分野では、地方議会会議録の要約支援（藤岡・渡邊・植崎2010）や地方議会会議録からの情報抽出（葦原・木村・荒木2012）などの研究が行われている。図書館情報学分野（松本2008）、環境学分野（上田・八木田2012）では、各分野における地方自治体の動向を探るために地方議会会議録を利用している。高丸（2013）では日本語学分野において地方議会会議録における文末表現の地域差を分析した。いずれの研究も、地方議会での発言（話しことば）を書き起こした膨大な文字列に含まれる情報に様々な観点から着目し、議論の内容や発言の傾向の抽出などを試みるものである。簡便に利用できるコーパスが公開されれば、地方議会の題材とした研究はさらに広がり

1 科研費基盤(B)No.22300086「地方議会会議録コーパスの構築とその学際的应用研究」
<http://www.local-politics.jp/>

を見せると予想される。

本研究では、宇都宮市議会を対象として、テキストマイニングの手法を利用することで、地方議会で何が話題になっているかの分析を行う。分析対象は、平成19年4月に行われた市議会議員選挙から平成23年4月に行われた直近の市議会議員選挙までの4年間に開催された本会議である。この期間の宇都宮市議会会議録については、星野ら(2011)において議員の発言と選挙公報等との比較が行われ、名詞出現頻度について部分的に分析されている。

地方議会会議録に対してテキストマイニング分析を行っている先行研究に増田(2012)がある。増田(2012)では、KHコーダー²を用いて高崎市議会会議録のテキストマイニングを行い、年度別に分類した名詞の出現傾向を単語頻度リスト、共起ネットワーク、階層的クラスター、多次元尺度構成法の各手法によって可視化している。本研究では増田(2012)と同様に、会議録における単語の出現頻度を用いた分析を行うが、異なる頻度算出方法、可視化手法を適用する。ただし、名詞の出現頻度については比較可能であるので、4章においてこれを述べる。

本論文の構成は以下のとおりである。2章では、対象とする宇都宮市議会会議録のデータの概要と分析の方法について述べる。3章では、対応分析を用いたテキストマイニングの結果について述べる。4章では、宇都宮市議会会議録と議会のウェブページに挙げられたトピック語(4.1節)、および、高崎市議会会議録(4.2節)との比較について述べる。

2 対象データと分析の方法

2.1. データの概要

宇都宮市議会のウェブサイト³で公開されている平成19年度から平成22年度までの本会議(臨時会、定例会)会議録のテキストデータをすべてダウンロードした上で利用する。会議録の発言内容の部分には、出席者の発言と議事進行の説明(例えば、「午前〇時開会」「〇〇君登壇」など)が併せて記されている。このうち議事進行の説明を削除し、出席者の発言のみをデータベース化した。データテーブルには「開催年(平成〇年)」、「開催回(第〇回)」、「開催日(〇日目)」、「発言者名」、「発言者肩書き」、「発言内容」の6種類のフィールドを設定し、会議録中の行頭から改行まで(通常1文に相当する長さ)を1レコードとしてすべての発言を登録した。

議員であっても、議長や委員長として発言する場合と、議員として質問を行う場合を区別して考えるべきであるので、「発言者肩書き」をあらかじめ、①「議事進行」(議長、副議長、臨時議長)、②「議員」(質問者として登壇した議員)、③「委員長」(委員長として登壇した議員)、④「市長」、⑤「行政」(答弁者として登壇した部局長等)、⑥「その他」(挨拶など)の6種類に分類した。

2.2. 整文

会議録は発言を一字一句忠実に記録するのではなく、読みやすさ等の観点から様々な修正が行われていることが知られている。これを整文という。宇都宮市議会会議録における整文の例を図1に示す。

整文によって略称が正式名称へ書き直されたり(例①)、フィラーや意味の重複する言い回しが削除される(例②)。例えば会議録から、「公職選挙法」は「公選法」と呼ばれているかや、

2 <http://khc.sourceforge.net/>

3 <http://www2.city.utsunomiya.tochigi.jp/utsunomiya/>

例① 略称の修正
整文前 …特定の候補者への応援は、<u>公選法</u>に抵触する可能性もあると思われ… 整文後 …特定の候補者への応援は、<u>公職選挙法</u>に抵触する可能性もあると思われ…
例② フィラー、意味が重複する言い回しの削除
整文前 …本<u>当</u>に<u>あ</u>の<u>一</u>、いろいろ<u>熟慮</u>しながら<u>考</u>えて<u>い</u>た<u>だ</u>い<u>て</u>… 整文後 …本<u>当</u>に、いろいろ<u>熟慮</u>して<u>い</u>た<u>だ</u>い<u>て</u>…

図1 宇都宮市議会における整文の例

フィラーがどの程度挿入されているかといった調査を行うことは難しい。しかし、整文による表記の統一は単語頻度を正確に計測することに寄与すると考えられる。

高丸(2011)では、平成22年6月15日の宇都宮市議会会議録の約2時間分(約30,000文字分)について、議会中継と会議録の比較を行った。宇都宮市議会における実際の発言と会議録とのレーベンシュタイン距離⁴は100文字あたり平均2文字であった。栃木県議会と那須塩原市議会における議会中継と会議録のレーベンシュタイン距離は100文字あたりそれぞれ約11文字、約18文字であったことから、宇都宮市議会は会議録作成における発言の修正量が比較的小さいといえる。会議録のレーベンシュタイン距離は、発言者が原稿にしたがって発言(朗読)する場面では非常に小さくなる傾向がある。宇都宮市議会では質疑の際に一括質問方式を採用しており、質問者、答弁者ともに、事前に用意した原稿の朗読が長く続くことが、レーベンシュタイン距離が小さい理由であると考えられる。

4 発言の書き起こしから会議録を作成する過程で「追加」「削除」または「置換」された文字数に相当する。編集距離ともいう。

5 発言者数は一人の議員が異なる肩書きで一例えば「委員長」と「議員」一発言しているものを分けて数えている。また、文字数には、句読点などの記号を含んでいる。

表1：会議録中の発言文字数

役職	発言者数	発言文字数
議長	6	138,204 (3.4%)
議員	50	2,085,348 (50.6%)
委員長	26	240,888 (5.8%)
市長	1	633,519 (15.4%)
行政	29	1,025,166 (24.9%)
その他	2	812 (0.0%)
計		4,123,937

本研究でテキストマイニングを行う会議録は、議会での発言—さらには事前に用意された原稿—に忠実なものである。

2.3. 発言文字数

データの規模を概観するために、発言者分類別の発言文字数を表1に示す。平成19～22年度の宇都宮市議会本会議には、114名による4,123,937文字の発言が記載されている⁵。個人としては「市長」の発言が633,519文字(全体の15.4%)と最も多い。発言者分類別では、「議員」による発言が約5割、「市長」「行政」が併せて約4割、「議長」の発言と「委員長」の報告が併せて約1割であり、本会議が主として「議員」の質問に対する「市長」「行政」の答弁によって構成されていることが発言文字数からもわかる。

2.4. 分析の方法

まず、「発言内容」フィールドに含まれるすべての発言を形態素解析ツール MeCab⁶によって、単語に分割する。解析結果から得られる品詞情報を利用し、名詞(一般)と名詞(サ変接続)のみを取り出す。会議録中の文に対する処理の一例を図2に示す。入力文(I)をMeCabに入力すると、形態素解析結果(II)が得られる。

6 <https://code.google.com/p/mecab/>

本研究では解析辞書としてUnidicを使用した。Unidicは国立国語研究所が規定した「短単位」を利用している。揺れがない分割単位であり、頻度情報を利用する本研究には適した解析辞書であると考えられる。<http://sourceforge.jp/projects/unidic/>

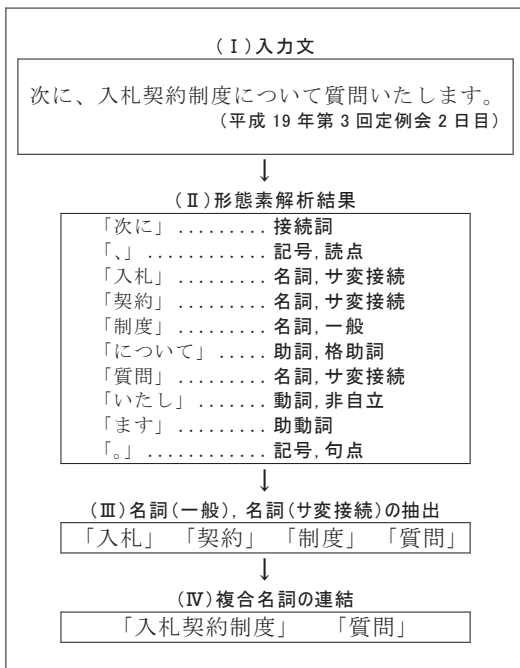


図2 形態素解析の例

ここから抜き出される名詞は(Ⅲ)に示す「入札」「契約」「制度」「質問」の4語である。

このうち例えば「制度」は、複合名詞の一部であり、会議録中には、図2の例の「入札契約制度」だけでなく、「指定管理者制度」「就学援助制度」「情報公開制度」等、様々な名詞と結びつき複合名詞として出現する。日本語ではこのように複数の名詞を連続して一単語として扱う複合名詞が多く見られることが知られている⁷。会議録のテキストマイニングを行う場合には、複合名詞の一部である名詞の出現頻度をそれぞれ数えるよりも、複合名詞としての出現頻度を数えた方がより話題の特徴をつかみやすいと考える。

そこで、形態素解析結果において、名詞(一般)、または、名詞(サ変接続)が連続して出

⁷ 会議録には「ICT(情報通信技術)環境整備状況」「バイオディーゼル燃料農業機械利用産地モデル事業」「新交通システム導入基本計画策定調査」「立体交差擁壁築造工事西側工区」などの長い複合名詞が存在する。

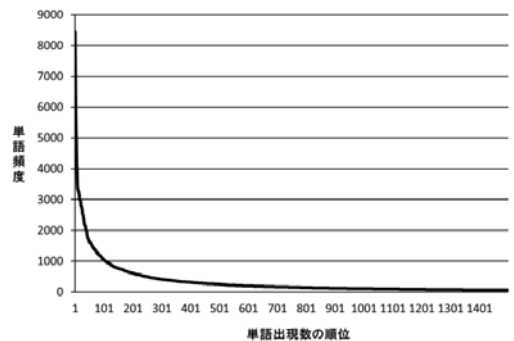


図3 名詞出現頻度の推移(上位 1,500 語の頻度数)

現している場合、これらを連結し一つの複合名詞として扱う。また、名詞(一般、サ変接続)の直前の接頭語、および、名詞(一般、サ変接続)の直後にある名詞(接尾)も連結し、複合名詞の一部として扱う。図2の例では、(Ⅳ)に示す「入札契約制度」と「質問」の2語が得られる。

このようにして会議録から抜き出した複合名詞について、年度別、発言者肩書き別、あるいは、発言者個人別に出現頻度を求める。複合名詞の出現頻度に基づいて、会議録中の話題の傾向を分析する。ただし、構文解析等を行わないため、話題となっている単語への立場や態度(賛否等)は分析の対象としない。

2.5. 名詞出現頻度の概要

複合名詞を作る前の名詞(図2の(Ⅲ)に相当)の出現頻度について概要を述べる。4年間の宇都宮市議会会議録の発言内容から抽出された名詞の総数は604,796語であった。また、重複を除いた異なり語数は12,203語であった。

図3に出現頻度上位1500語における出現頻度の推移を示す。一般に文書内の単語を出現頻度順に並べると、一部の語の出現頻度が非常に高い一方、出現頻度の低い語が大量に存在するロングテールの形状をなすといわれている。宇都宮市議会会議録においても最頻単語の「事業」の出現頻度が8,451回である一方、出現順

位 1500 番目の語（「市庁舎」「校庭」「電磁波」など 17 語）の出現頻度は 57 回である。また、出現頻度の下位約 3 分の 2 にあたる 8,112 語は出現頻度が 10 回未満である。

表 2（173 ページ）に、単語出現頻度の上位 100 語を示す。上位には、議会において多様な文脈で出現し得る「(1) 事業」「(2) 地域」「(3) 市民」「(4) 市」「(5) 質問」等の名詞が挙がっている。特定の分野に関わる名詞としては「(9) 教育」「(17) 学校」「(18) 環境」「(28) 交通」「(42) 児童」「(44) 高齢」「(45) 医療」「(55) 子供」などが上位に見られる。これを概観するだけでも、宇都宮市議会では、教育、環境、交通、高齢者、医療等の問題が頻繁に取り上げられていることがうかがえる。

2.6. 複合名詞の概要

会議録における複合名詞（図 2 の（Ⅳ）に相当）の総数は 465,012 語、重複を削除した異なり語数は 44,866 語である。表 3（173 ページ）に、複合名詞の出現頻度（上位 100 語）を示す。名詞単独で 8,451 回出現し、出現数がもっとも多かった「事業」は、複合名詞の頻度では 2,118 回（11 位）となっており、会議録中の「事業」の 75% が他の名詞と連結し複合名詞を形成していた（例えば、「事業者」408 回、「土地区画整理事業」326 回、「事業費」253 回、「公共事業」94 回）。このように複合名詞の一部になりやすい名詞（「事業」のほか、「計画」「制度」「委員」など）は名詞単独と比べ、相対的に出現順位が下がり、議事進行において複合名詞を形成せずに繰り返し出現する「議案」「市長」等は相対的に出現順位が上がっている。また、この 100 語の中に「まちづくり」「高齢者」「中心市街地」など具体的な意味をつかむことに役立つ複合名詞も見られる。

3 対応分析による テキストマイニング

3.1. 対応分析

年度別（→ 3.2 節）、発言者肩書き別（→ 3.3 節）、発言者別（→ 3.4 節）の発言の傾向をつかむために、複合名詞の出現頻度に対する対応分析（コレスポンデンス分析）の適用を試みる。対応分析には、統計解析ソフト R⁸（ver3.0.1）の MASS パッケージに収録されている `corresp` 関数を用いる。

3.2. 年度別の分布

総出現頻度が 50 以上の複合名詞 1,393 語を対象に、発言年度と出現頻度の対応分析を行った。寄与率は第 1～3 軸に対して、それぞれ、48.42%、27.29%、24.29% である。第 1、2 軸の散布図を図 4 に示す。発言年度（「H19」「H20」「H21」「H22」）は第 2 象限から反時計回りにプロットされている。複合名詞の多くは図の中央に集中している。これらの語は年度にかかわらず発言された語と捉えることができる。複合名詞のうち、図から読み取れるもの—すなわち、グラフ中央から外れた位置にプロットされたもの—は出現に偏り（年度差）が見られるものである。

例えば、平成 19 年度の付近にある「学校給食費」「連帯保証人」などの語から給食費の未納問題が取り上げられたことが分かる。また、「新市」「旧町」「市町合併」などの語から、河内町、上河内町との合併問題が取り上げられたことが分かる。平成 20 年度付近は、「定額給付金」がとりわけ特徴的であるが、この他にも「学校施設」「社会保険病院」「街路樹」などの語を読み取ることができる。また、平成 20～21 年にかけて

8 <http://www.r-project.org/>

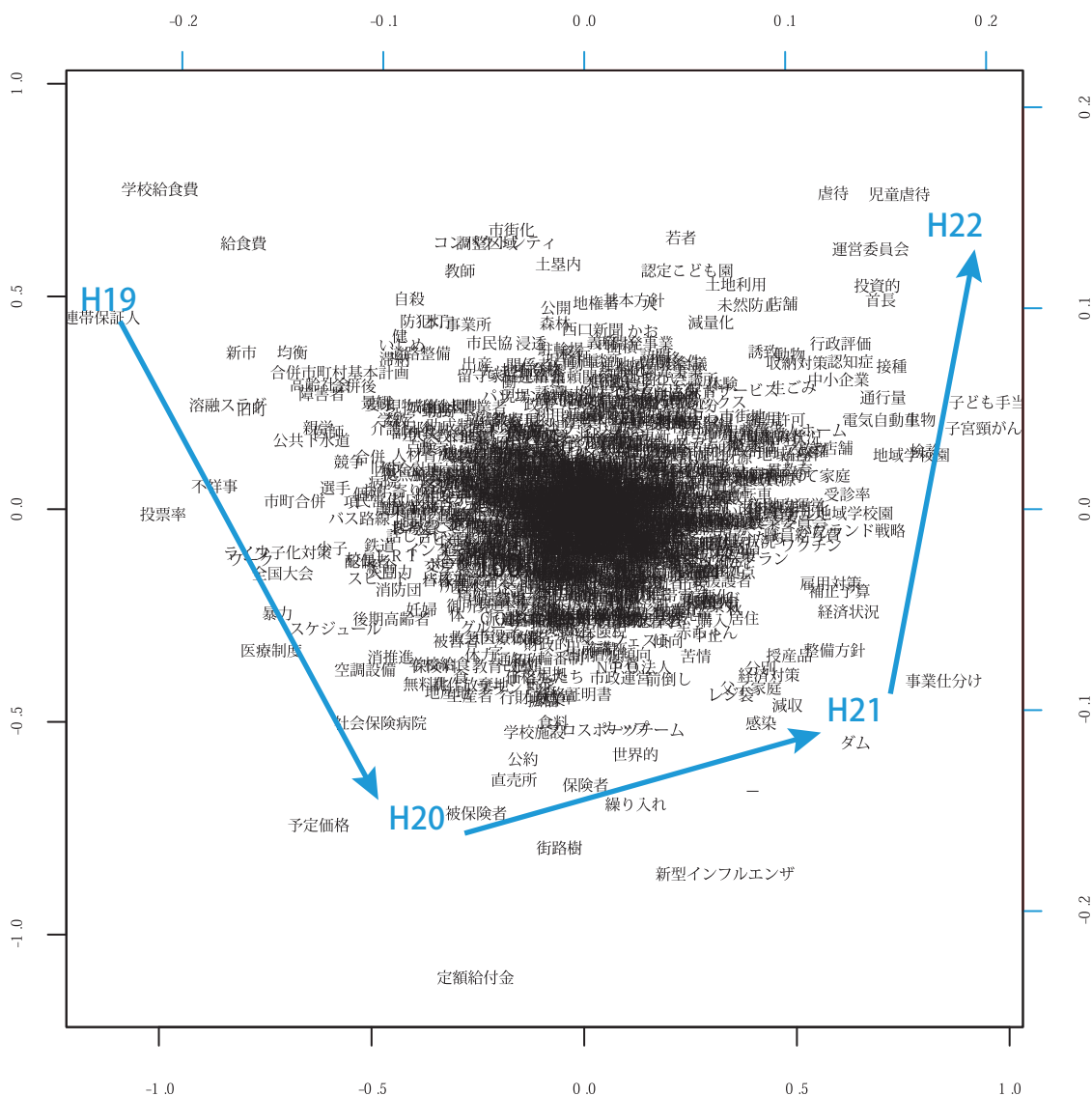


図4 発言年と複合名詞頻度の対応分析

「新型インフルエンザ」「感染」、平成 21 年度には「ダム」「事業仕分け」「レジ袋」、平成 21～22 年にかけて「経済状況」「雇用対策」「補正予算」、平成 22 年度には「児童虐待」「子ども手当」「子宮頸がん」「接種」「認定こども園」などが特徴的な語として出現していることが分かる。

3.3. 発言者肩書き別の分布

同様に総出現頻度が 50 以上の複合名詞 1,393 語を対象に、発言者肩書きと出現頻度の対応

分析を行った。ただし、2 章で示した肩書きの 6 分類のうち「その他」は発言数が非常に少ないため、本分析から除外した。寄与率は第 1～3 軸に対して、それぞれ 59.77%、23.53%、10.29% である。第 1,2 軸の散布図を図 5 に示す。

質問する立場の「議員」とそれに答弁する立場の「市長」「行政」が左下にプロットされており、概ね同じ語を使用していることが推測される。一方、横軸で「議員」「市長」「行政」のまとまりと「議長」が分離され、縦軸で「委員

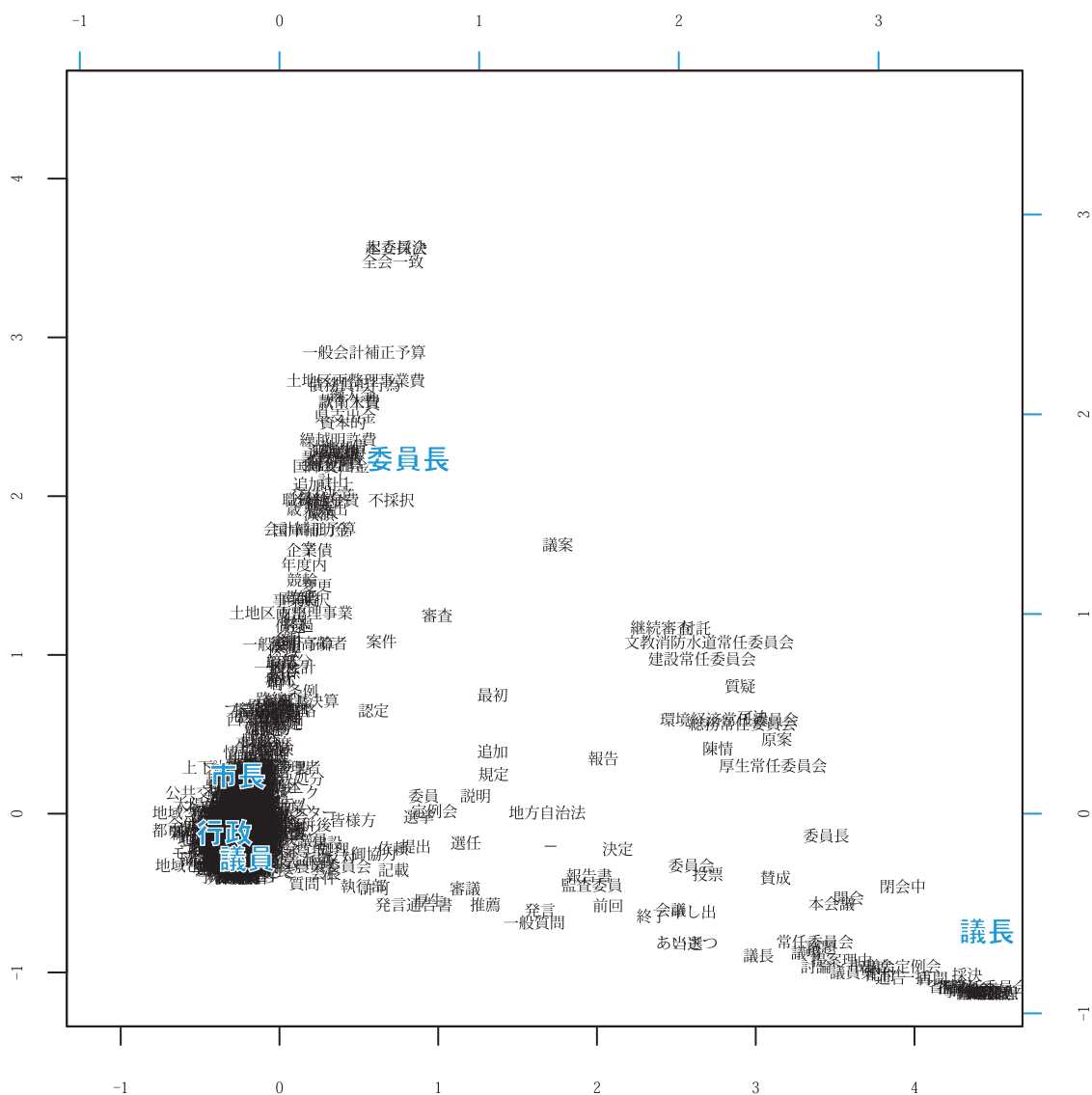


図5 発言者肩書きと複合名詞頻度の対応分析

長」が分離されていることが分かる。これは「議長」「委員長」はそれぞれ異なる特徴的な語を発言していることを意味する。例えば「議長」の付近には「採決」「定例会」「通告」「閉会中」「開会」「賛成」など、議会の運営にかかわる語がプロットされている。「委員長」の付近には、「全会一致」「起立採決」「本委員会」「一般会計補正予算」「不採択」等委員会での審議の報告にかかわる語がプロットされている。また、「議長」と「委員会」とを結ぶ線分上にプロットさ

れた委員会名（「文教消防水道常任委員会」「建設常任委員会」「厚生常任委員会」等）のように、それぞれのグループの中間の位置には、双方が発言する語を見ることができる。

3.4. 議員による発言の分布

発言者肩書き別の分析のよって、発言者のうち、議長と委員長は議員、市長、行政職員とは傾向の異なる語を発言していることを捉えた。

議会では議員の質問に対して、市長や行政職

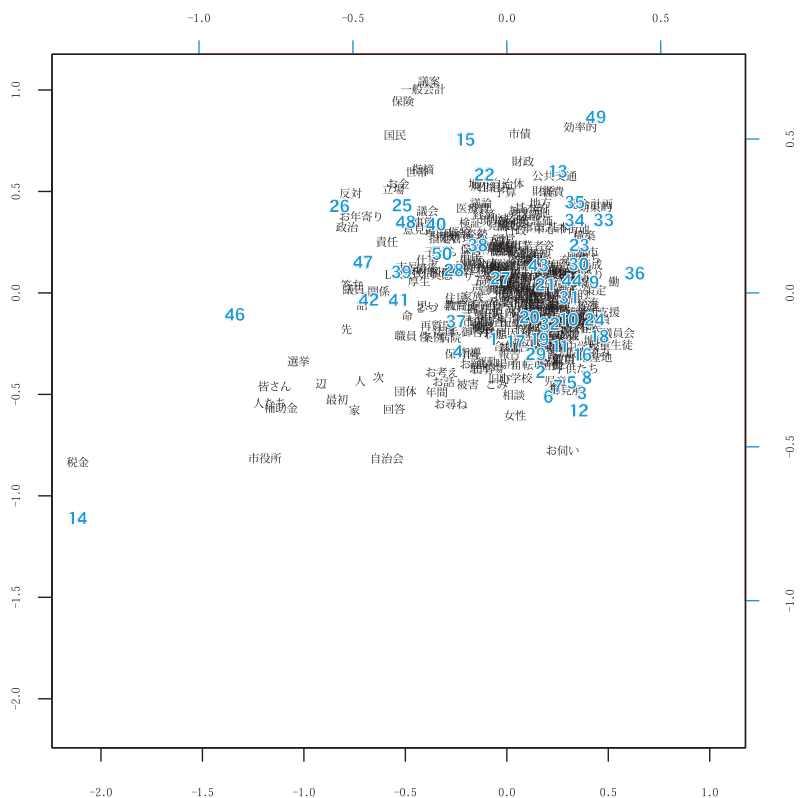


図6 議員と複合名詞頻度の対応分析

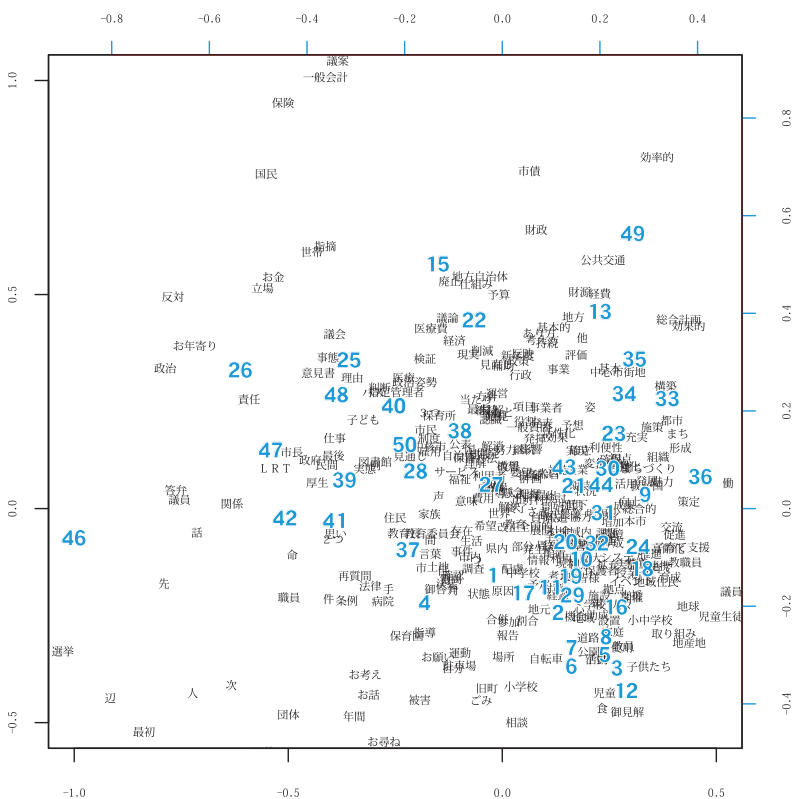


図7 議員と複合名詞頻度の対応分析 (拡大)

員が答弁する。このため、市長や行政職員の用いる語は、議員の質問に依存する。そこで本節では、議員に焦点を当て、議員個人と複合名詞との対応分析を行う。

図6に議員の発言における出現頻度が100回以上の複合名詞335語を対象とした対応分析の結果（第1,2軸の散布図）を示す。どの議員が何を発言したのかを個別具体的に見ることは本論文の範囲を超えるため、図には議員の個人名ではなく議員番号を示している。対応分析の寄与率は第1～3軸に対して、それぞれ37.62%、25.42%、21.76%である。議員と語のプロットがともに中心部分に密集しており、質疑において地域の課題等についてどの議員も同じ単語を繰り返し用いて発言されていることが推測される。より詳しく見るために図7には、図6の中心部分を拡大したものを示す。まず、図6を見ると「税金」が左端に、「財政」「市債」は上部にプロットされていることわかる。また図7を見ると、「小学校」「小中学校」「児童」などが右下に、「都市」「まちづくり」などが右部中央に、「LRT」「図書館」などが左部中央にプロットされていることがわかる。

議員（もしくは会派）ごとに重視する課題や取り組みに違いがあり、質問に取り上げる回数が異なっていると考えられる。図6,7にはこのような議員（会派）別の関心領域の微細な差異が、単語の使用頻度の差として現れている。

4 出現頻度の比較

4.1. 議会ホームページのトピックとの比較

宇都宮市議会会議録検索のウェブページには、「最近の話題」としてトピック語が列挙され、その語を含む発言の検索結果へリンクが張られている。表4にこのトピック語、および、これに完全一致する複合名詞の出現頻度（年度別）

表4：ホームページに掲載されたトピックの出現頻度

トピック語	H19	H20	H21	H22	ALL
個人情報	12	18	17	10	57
情報公開	27	8	6	3	44
合併	148	71	53	39	311
行政評価	7	2	17	27	53
新交通システム	8	18	3	20	49
電子入札	0	2	0	0	2
バリアフリー	8	6	2	5	21
支援費制度	該当なし				
Oー157	記号を含むため対象外				
ダイオキシン	0	0	5	1	6
斎場	9	10	0	5	24
住民基本台帳ネットワーク	該当なし				
最終処分場	16	3	3	3	25
中央卸売市場	6	7	2	4	19
大谷	地名を含むため対象外				
チャレンジショップ	1	0	0	1	2
中心商店街	6	14	4	14	38
競輪場	17	37	13	26	93
ろまんちっく村	固有名詞のため対象外				
市営住宅	14	31	26	24	95
都市開発	1	0	0	0	1
城址公園	61	26	23	29	139
区画整理	14	24	3	5	46
再開発事業	36	12	26	46	120
地籍調査	6	12	11	0	29
学校給食	25	43	17	9	94
救急救命士	2	4	0	0	6
湯西川ダム	地名を含むため対象外				
水道料金	3	5	10	15	33
通信指令	0	0	0	0	0

を示す。

トピック語のうち、記号を含む「Oー157」、地名・固有名詞を含む「大谷」「湯西川ダム」「ろまんちっく村」は本研究の複合名詞出現頻度とは比較できないため対象外とする。また、本表の数値はトピック語と完全一致する複合名詞の頻度であり、より長い複合名詞の一部になっているものはカウントされていない。「通信指令」の表4における頻度は0であるが、実際には会議録中に、より長い複合名詞「通信指令塔」「通信指令システム」が計4回出現している。

「支援費制度」および「住民基本台帳ネットワーク」は当該期間の会議録に一度も出現しない語であった。ただし、「支援費制度」ではなく「支援費」含む「担い手育成支援費」「住宅

表5：高崎市議会と宇都宮市議会における
名詞出現頻度の順位の比較（上位60位）

両市の 出現順位に近い		宇都宮市の 順位が高い		高崎市の 順位が高い	
地域 質問 事業 市民 計画 教育 施設 学校 制度 支援 議案	都市 行政 対策 生活 取り組み 活動 調査 医療 関係 管理 報告	市 整備 委員 実施 市長 予算 環境 推進 まち 利用 議員 検討 交通 活用 設置 国 地区	確保 お答え 公共 企業 児童 中心 補助 向上 効果 情報 連携 子供 会計 見解 導入 経済	答弁 年度 設備 子ども 状況 合併 社会 お願い 一般 対応 住民 高齢 基本 質疑	職員 課題 条例 センター 現状 財政 保険 建設 施策 保護 負担 要望 期待

復興支援費」「子育て支援費」「自立支援費」であれば、合計15回出現している。また、「住民基本台帳ネットワーク」ではなく「住民基本台帳」を含む複合名詞（例えば「住民基本台帳オンラインシステム」「住民基本台帳法」「住民基本台帳事務費」）は計22回出現している。

出現頻度の低いトピック語の中で「中心商店街」（38回）より「中心市街地」（695回）、「新交通システム」（49回）より「LRT」（244回）の方がそれぞれ出現頻度が高い。まったく同一のものを指す語ではないものの、会議録の情報を幅広く提供するためには、出現頻度の高い語の方が適している。また、「ダイオキシン」「チャレンジショップ」などのトピック語も出現頻度が非常に少ない。これらは、重要な話題であるにもかかわらず市議会では議論されていないということを明らかにするには役立つものの、最近の話題として挙げるには適さない。議会における最近の話題に関する語をウェブページで提示する場合、会議録中の出現頻度（実際に議論が集中している語）や検索システムの検索履歴（利用者の関心が高い語）などに基づいて、適切に選択することが望まれる。

4.2. 高崎市議会会議録との比較

増田(2012)の表1、2に示されている平成19年度から22年度における高崎市市議会の名詞の出現頻度の合計値と本研究における名詞出現頻度（表2）を比較する。増田(2012)では、名詞とサ変名詞それぞれの出現頻度上位30語（計60語）を示している。これを1つにまとめて降順に並べ替えた上で、本研究の表2の上位60語と比較する。高崎市の60語と宇都宮市の60語の計120語における異なり語数は82語である。これらを両市の出現順位に近い語（±10位以内）、宇都宮の出現順位が11位以上高い語、高崎市の出現順位が11位以上高い語の3つに分類した。これを表5に示す。

出現の順位がその分野についての議論の活発さを近似していると考えれば、「教育」「生活」「医療」はどちらの市でも同程度に重視されている。宇都宮市の順位が高い語のうち特徴的なものには「環境」「交通」「企業」「情報」「経済」等がある。一方、高崎市の順位が高い語のうち特徴的なものには「住民」「高齢」「合併」「保険」「建設」等がある。これらはすべて出現頻度の多い語であるが、地域や期間によって、出現順位（議論に占める割合）に違いがあることがわかる。また、高崎市は「子ども」の順位が高いが、宇都宮市では「児童」「子供」が上位にあり、議会ごとの表現の違いもこの結果から読み取ることができる。他都市との比較の方法については今後検討を進めたい。

5 まとめ

本論文では、地方議会では何が話題になっているのかを捉えるために、ウェブに公開された会議録を対象としたテキストマイニングを試みた。宇都宮市議会会議録から抜き出した発言内容を形態素解析し、年度別、発言者肩書き別、

議員別に複合名詞の頻度を計算した。「何が話題になっているか」を一言でまとめることはできないが、対応分析（第1軸と第2軸）のグラフを利用して可視化することで、議会における話題の特徴を捉えることができた。

年度別の対応分析から、年度を問わず出現する語が多い一方、その年に起きた事件や生じた課題によって、出現する単語に違いがあることを示した。肩書き別の対応分析からは、発言される単語の傾向は、「委員長」、「議長」、「議員、市長、行政（職員）」の3つに大別されることを示した。これは役割による使用語彙の違いである。議員別の対応分析から、質疑や議事の基本語彙や重要な課題に関わる語はどの議員も用いている一方、分布を詳細に観察することで議員ごとの使用語彙の違い（関心のある分野に違い）を示した。

議会における発言内容について知ろうとしたときに、議会を傍聴したり、会議録をすべて読むことは時間的コストの観点から容易ではない。議会における注目の出来事や重要な審議は新聞等で取り上げられるが、議会での議論の全体像を知る機会は非常に少ない。地域において常に議論の対象となる諸課題に関する語は年度、発言者（議員）を問わず繰り返し出現しており、報道されずとも、議会ではこれらの課題についての議論が常に交わされていることがうかがえる。その上で、議論の集中度、関心の度合いを可視化するテキストマイニングの手法（単語出現頻度に対する統計分析を利用した手法）を、市民が議会への関心を高める取り組みの一助としたい。

今後は、分類カテゴリと単語との結びつきをより適切に表現するための可視化方法についてさらに検討していきたい。また、全国の自治体から収集した地方議会会議録コーパスを利用し

た地域横断的な比較とその可視化方法についても検討を進めたい。

謝辞

本研究の一部は、科研費（No.2230086 および No.25370524）の助成を受けた。

参考文献

- 葦原史敏, 木村泰知, 荒木健治 (2012) 「節の分類情報を用いた地方議会会議録における要求・要望表現抽出」電子情報通信学会 言語理解とコミュニケーション研究会 第2回テキストマイニング・シンポジウム, pp. 1-6
- 上田翔, 八木田浩史 (2012) 「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), pp.283-288
- 乙武北斗, 高丸圭一, 渋谷英潔, 木村泰知, 森辰則 (2013) 「地方議会会議録コーパスの学際的応用を目的とした n-gram データの構築およびウェブ UI の試作」言語処理学会第19回年次大会発表論文集, pp.733-736
- 齋藤誠, 大城卓, 菅原晃平, 永井隆広, 渋谷英潔, 木村泰知, 森辰則 (2011) 「地方議会会議録の収集とコーパスの構築」言語処理学会第17回年次大会発表論文集, P2-21
- 高丸圭一, 木村泰知 (2010) 「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, 第10号, pp.74-86
- 高丸圭一, 渋谷英潔, 木村泰知 (2011) 「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, 第11号, pp.47-72

高丸圭一 (2011)「規模の異なる自治体における
地方議会会議録の整文の比較」社会言語科学
会第 27 回研究大会, pp.256-259

高丸圭一 (2013)「形態素 N-gram を用いた地方
議会会議録における地域差の分析手法の検討
—ひらがなで構成された文末の 4-gram に着
目して—」明海日本語, 18, pp.1-10

藤岡亮介, 渡邊俊彦, 植崎博司 (2010)「自然言
語処理技術を活用した議会議事録の要約支援
方法について」バイオメディカル・ファジィ・
システム学会誌, 12(2), pp.33-46,

星野卓也ら (団体名: 宇都宮共和大学 4 年高丸
ゼミ) (2011)「議会会議録を利用した議員の

活動の見える化と市民と議員との協働支援」
まちづくり提案集 2011 (「大学生によるまち
づくり提案」発表会資料), pp.105-114

[http://www.city.utsunomiya.tochigi.jp/dbps_data/_material/_files/000/000/023/319/gikai.pdf]

増田正 (2012)「地方議会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, 15(1), pp.17-31

松本直樹 (2008)「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
議録分析をもとに」日本図書館情報学会誌
54(1), pp.39-56

表 2: 対象全期間における名詞頻度 (上位 100 語)

名詞	頻度	名詞	頻度
(1)事業	8451	(51)情報	1608
(2)地域	7248	(52)答弁	1608
(3)市民	5786	(53)職員	1604
(4)市	5138	(54)連携	1604
(5)質問	4881	(55)子供	1583
(6)整備	4129	(56)関係	1554
(7)支援	3985	(57)会計	1538
(8)計画	3955	(58)見解	1513
(9)教育	3439	(59)導入	1485
(10)委員	3325	(60)経済	1475
(11)実施	3268	(61)基本	1466
(12)市長	3259	(62)センター	1438
(13)制度	3200	(63)保険	1435
(14)施設	3199	(64)人	1430
(15)予算	3029	(65)報告	1407
(16)議案	3007	(66)保育	1389
(17)学校	2988	(67)農業	1389
(18)環境	2960	(68)負担	1373
(19)対策	2878	(69)課題	1370
(20)推進	2843	(70)説明	1362
(21)まち	2773	(71)県	1333
(22)利用	2769	(72)計上	1311
(23)状況	2712	(73)運営	1299
(24)議員	2662	(74)住民	1298
(25)検討	2528	(75)意見	1271
(26)行政	2528	(76)自治	1268
(27)取り組み	2516	(77)サービス	1264
(28)交通	2480	(78)充実	1257
(29)都市	2385	(79)補正	1247
(30)生活	2318	(80)促進	1219
(31)活動	2205	(81)財政	1212
(32)活用	2199	(82)指導	1205
(33)社会	2191	(83)策定	1204
(34)設置	2158	(84)保護	1204
(35)対応	2141	(85)全国	1195
(36)国	2109	(86)家庭	1188
(37)地区	2005	(87)施策	1171
(38)確保	1992	(88)提案	1170
(39)お答え	1989	(89)市街地	1149
(40)公共	1963	(90)介護	1144
(41)企業	1896	(91)道路	1135
(42)児童	1808	(92)認識	1121
(43)調査	1770	(93)一般	1107
(44)高齢	1725	(94)総合	1078
(45)医療	1724	(95)土地	1072
(46)中心	1703	(96)福祉	1071
(47)管理	1685	(97)税	1058
(48)補助	1648	(98)団体	1053
(49)向上	1621	(99)体制	1042
(50)効果	1611	(100)会議	1024

表 3: 対象全期間における複合単語頻度 (上位 100 語)

複合名詞	頻度	複合名詞	頻度
(1)市民	3850	(51)増加	864
(2)地域	3086	(52)行政	825
(3)議案	2932	(53)具体的	813
(4)市長	2851	(54)追加計上	810
(5)実施	2744	(55)地区	795
(6)取り組み	2378	(56)予算	794
(7)本市	2371	(57)職員	787
(8)質問	2322	(58)開催	782
(9)検討	2189	(59)再質問	773
(10)事業	2118	(60)一つ	736
(11)整備	2081	(61)御異議	735
(12)国	2076	(62)お伺い	735
(13)お答え	1972	(63)提案	733
(14)活用	1916	(64)魅力	732
(15)推進	1914	(65)効果	727
(16)対応	1849	(66)活性化	725
(17)まちづくり	1849	(67)見直し	723
(18)議員	1759	(68)期待	708
(19)確保	1734	(69)国民	708
(20)設置	1637	(70)まち	698
(21)支援	1583	(71)企業	696
(22)状況	1573	(72)中心市街地	695
(23)御質問	1485	(73)都市	692
(24)市	1463	(74)社会	676
(25)連携	1382	(75)減額	663
(26)向上	1318	(76)強化	659
(27)見解	1264	(77)理解	655
(28)利用	1210	(78)提出	654
(29)充実	1177	(79)委員長	654
(30)課題	1155	(80)活動	649
(31)学校	1152	(81)方々	639
(32)導入	1152	(82)意見	639
(33)報告	1118	(83)対象	637
(34)人	1110	(84)役割	632
(35)認識	1063	(85)原案	627
(36)答弁	1049	(86)陳情	624
(37)制度	1017	(87)要望	616
(38)施設	1013	(88)決定	615
(39)計画	1000	(89)補正	607
(40)高齢者	996	(90)生活	604
(41)お願い	984	(91)現状	603
(42)子供	972	(92)拡大	601
(43)県	971	(93)目的	599
(44)策定	968	(94)評価	594
(45)対策	955	(95)全国	590
(46)調査	948	(96)保護者	587
(47)施策	928	(97)情報	583
(48)環境	926	(98)可決	582
(49)説明	903	(99)安心	580
(50)実現	882	(100)月	578