

# 市民の声の政策形成への活用に関する一考察

## A study on the use of the citizens' voice in the policy formation

金井 茂樹<sup>1</sup>  
Shigeki KANAI

<sup>1</sup> 行政広報アドバイザー Public Relations Adviser

**Abstract** In progress of decentralization, it has become more important for local governments to analyze the diverse voice of citizens. The voice of citizens is useful for discovering potential needs of the citizens. This study aims to clarify the policy issues of the local governments and the structure of the voice of citizens by using text mining and statistical analysis.

キーワード 広聴, 市民の声, 政策価値, 自由回答質問と選択肢型質問, テキストマイニング

### 1 はじめに

自治体は半世紀以上にわたり様々な手法を活用して市民の“生の声”を聴く努力を行ってきた。しかし、自治体の広聴活動は行政施策に特段の影響をもたらすことなく、首長や行政の「聴く姿勢」をアピールすることにとどまり、世論調査や住民意識調査も政策形成や行政評価に十分に活用されていないとの批判もなされてきた（上野, 2000, p.112）。とくに、広聴活動によって得られた市民の声は地域の課題発見に結びつく可能性をもつ情報として、その重要性は認識されているが、市民の声への迅速な対応という側面が重視され、政策形成への活用に関しては多様で変化に富む市民の声の代表性の問題、政策課題抽出に伴う困難性、抽出過程における恣意的判断の介入のおそれなどが常に問われ、その分析手法や政策への活用方法が確立しているとはいえない。個々の市民の声から政策形成に有用な知見が必ずしも得られるわけではないが、市民の声を無視した政策形成はありえず、その分析・活用は多くの自治体にとって重要な課題になっている。

そこで、本稿では自治体世論調査・意識調査で多用される選択肢型質問や人口統計学的変数に基づく、主に質的な数値型データと、自由回答質問により収集された非定型で非構造的なテキスト型の自由回答データの両者に対して、自然言語処理やデータマイニングなどの複合技術であるテキストマイニング手法による分析を行うことで、市民の声から具体的な政策形成の手がかりとなる客観的な情報を効率的に抽出するひとつの枠組み（フレームワーク）を提示することを目的とする。

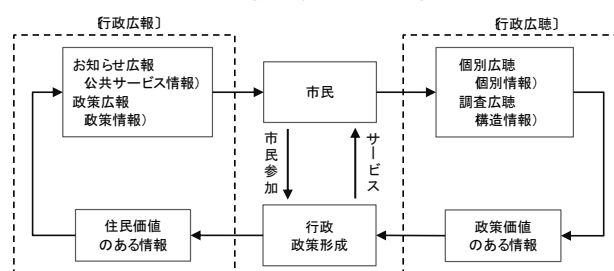
## 2 市民の声の分析の現状

### 2.1 行政広報広聴

筆者は以前、戦後 PR (Public Relations) の導入以降の行政広報広聴に関する先行研究の検討において、行政広報広聴の基礎的枠組みに関する検討を行った（金井, 2016）。そこでは、行政広報の本質を“情報開発と価値創造による行政と市民との信頼関係の構築・維持”と捉えるとともに、行政広報広聴の 4 つの理念（真実

性、双方向性、網羅性、並行性）とともに、対象とする情報を広聴による「個別情報」「構造情報」と広報による「公共サービス情報」「政策情報」に分類して基礎的枠組みを構築した<sup>1)</sup>。この基礎フレームは、行政がその理念のもとに多様なチャンネルを通して個別・構造情報の集約・調整と公共サービス・政策情報の提供・公開という情報循環を戦略的かつ継続的に実践することにより情報開発・価値創造を実現し、住民との信頼関係を構築するというものである（図 1）。この枠組みでは、広聴活動による「政策価値」と広報活動による「住民価値」という“二つの価値”を創出することが重要である。政策価値とは広聴活動によって得られた情報の分析から導き出される政策形成に有用な価値であり、住民価値とは住民が真に必要な自治体の政策・施策・事務事業に関連した価値である。つまり、市民の声を反映した政策価値をいかに創出するかが広聴活動の課題なのである。

図 1 行政広報広聴の基礎的枠組み



出所：筆者作成

### 2.2 行政広聴の手法と情報特性

次に、行政が市民の声を収集する手法とそれによって得られる情報の特性について述べる。

行政の広聴活動は、実務上その機能と形式から個別広聴・集団広聴・調査広聴に分類<sup>2)</sup>されることが多い。ここでは広聴活動の枠組みとして実務上の分類と行政側の情報収集態度（受動的・能動的）（辻, 1962, 阿部, 1998）を用いて分類整理を行った（表 1）<sup>3)</sup>。

受動的活動から得られた情報は、一定期間において“集まる”データであり、市民の個別的な意見や要望、苦情が中心であり、そのほとんどが非構造的なテキス

ト型データである。最近では SNS（Social Networking Service）を通じて発信された市民の声が増加しているが、こうした情報もテキスト型データであることが多い。それらの活用には個別性・代表制の問題への配慮が必要である。他方、能動的活動から得られた情報は、主として調査によって行政が主体的に“集めた”情報であり、設定した質問群に対するその調査実施時点での市民の意見構造や分布を表すものである。この場合、主に数値型データを対象とするが、テキスト型データの収集も頻繁に行われている。世論調査や意識調査は調査票の質問に対する回答のみを集める点では収集できる情報に制約があるものの、科学的手法として多くの自治体で活用されている。ただし、昨今の回収率の低下により調査品質やその信頼性は危機的状況に陥っている。

表 1 行政広聴活動の枠組み

| 収集態度<br>情報特性 | 受動的活動<br>住民が主体          |                   | 能動的活動<br>自治体が主体             |
|--------------|-------------------------|-------------------|-----------------------------|
|              | 個別広聴                    | 集団広聴              | 調査広聴                        |
| 実務上分類        |                         |                   |                             |
| 手 法          | 面談・手紙・<br>電話・メール・SNS など | 対話集会・<br>市民モニターなど | 世論調査・意識調査<br>ウェブ調査など        |
| 性質           | “集める”データ<br>一定期間のデータ    |                   | “集める”データ<br>一時点のデータ         |
| 主な内容         | 個別的な<br>意見・要望・苦情など      |                   | 提示した質問群<br>に対する意見構造分布       |
| データ形式        | 主に非構造的な<br>テキスト型        |                   | 主に数値型<br>一部自由回答などの<br>テキスト型 |
| 傾 向          | 情報の量的増加                 |                   | 情報の質的低下                     |
| 問 題 点        | 意見の個別性・代表性              |                   | 調査票の制約                      |

出所：筆者作成

### 2.3 自治体による市民の声の分析

表 1 に示したように、行政はさまざまな手法を用いて市民の声を収集してきた。しかし、数値型データの分析と比較して、テキスト型データについては十分な分析が行われてきたとはいえない。市民の声には、市民が強く伝えたいメッセージや地域に暮らす人たちに共通する経験や意識などが記され、市民のニーズ、ライフスタイル、問題意識などを知る手がかりになる可能性がある。これからの行政経営にあって、複雑・多様化する市民の声の分析はこれまで以上に強く求められている。

それでは、行政は市民の声をどのように分析しているのであろうか。ここでは、1980 年代から最近までの自治体における市民の声の分析状況について概観する。

#### (1) 1980 年代および 1990 年代の市民の声の分析

まず、1980 年代および 1990 年代における東京都による広聴行政研究の報告書を取りあげ、当時の市民の声の分析状況と課題について整理する<sup>4)</sup>。1980 年代の報告書においては、自治体に対する住民の意見・要望が多様化するなかで、広聴活動によって集まる年間数万件の市民の声が行政施策に有効に活用されていない

状況を踏まえて、今後どのように市民の声を施策に反映させるかという広聴手法の開発の必要性が課題として提起された。また、1990 年代においては、市民の声への迅速、公平な対応と行政経営への対応のために採用された多様な広聴手法の状況が報告されるとともに、市民の声の収集、蓄積、分析、政策への反映といった広聴活動の各フェーズにおける課題が指摘された（東京都、1997, p.18）。この翌年の報告書では、市民の声の分類に関する全庁的な統一項目の未作成、データベースの未整備、全庁的な情報共有の未実現などの指摘とともに、市民の声を計画策定過程や予算編成過程に反映させるためのシステムの枠組みが示された。また、市民の声を政策形成に活かす具体的な展開として「明確な事実から重要な情報を発見し、政策情報を創造していくこと」の重要性が指摘された（東京都、1998, p.17）。ただし市民の声の具体的な分析手法までは論じられてはいない。

#### (2) 2000 年代以降における市民の声の分析

2000 年以降、情報化の進展により自治体においても全庁的に情報端末・ネットワークが導入され、テキスト型データの電子的な保存・蓄積が可能になった。宮城県や福岡市、浜松市などいくつかの先進的な自治体において市民の声データベースが構築され、データの蓄積と全庁的な情報共有が始まった。しかし、2002 年に行われた政令市（12 市）を対象にした調査では、ほぼすべての政令市が蓄積された市民の声データを政策に反映させるための分析手法を持ち得ていない実態が報告された<sup>5)</sup>（仙台都市総合研究機構、2003）。

2012 年には神奈川県による全国都道府県の広聴担当課を対象に行われた「広聴のあり方に関する調査」の結果が公表されている<sup>6)</sup>。そこでは、都道府県の約 70%が市民の声をデータベース化していることや、先進的な市民の声の分析・活用を実践している大阪府・徳島県の事例が紹介され、データベース化による情報共有とデータ分析が進みつつあることが示された。また、データを政策に結び付けるための具体的な分析手法として KJ 法や KIT-PT 法、テキストマイニングの活用が指摘されるなど、市民の声の戦略的な活用の必要性が強調された<sup>7)</sup>。

#### (3) 基礎自治体における市民の声の分析

足立区（東京 23 区）は毎年『区民の声報告書』と『足立区政に関する世論調査報告書』を公表している。ここでは、市民の声の分析事例として 26 年度版の報告書を取りあげる。足立区においては「区民の声」は区政に関する意見・要望を区の事業や施策に反映させていく制度と位置づけられていることから、その分析においては個々の意見・要望からの政策価値のある情報の抽出が目的とされていると考えられる。『平成 26 年度版区民の声報告書』では、集まった市民の声の概要とともに、それらが施策別に 70 項目に分類整理され、件数の多い上位 10 位までについてはその内容と区の対応、行政サービスや事務の改善に結びついた具体的な事例が掲載されている。ただし、区民からの個別的な意見・要望に対する回答が重視され、集積データとしての意見の分析はこの報告書には示されていない。また、『第 43 回足立区政に関する世論調査報告書』（平成 26 年度版）では「区政についてのご意見、ご

要望」という質問によって得られた自由回答データの分析結果が報告されている。そこでは自由回答の内容が主観的に 25 の施策項目に分類整理され、その項目別件数と代表的な意見が掲載されているが、この報告書でも集積データとしての意見に対する分析は示されていない。

このように、“市民の声の分析手法”は 1980 年以前から検討されてきた行政広聴の課題であるが、情報通信技術が発達した現在においてもデータベースの構築や政策・施策別の分類分析が中心であり、市民の声の分析手法のさらなる進展が求められている。

## 2.4 市民の声のテキストマイニング分析

このように、自治体において収集した市民の声については分類分析が中心に行われているが、市民の声を対象にテキストマイニング分析<sup>9)</sup>を用いた事例研究もいくつか報告されている。

たとえば、近藤（2003）は、浜松市の「市民の声システム」に蓄積されたデータ（1,455 件）についてテキストマイニング分析によりキーワードを抽出し、それにもとづいて実務担当者でディスカッションを行い、政策形成のための情報としての活用可能性を示した。また仙台市都市総合研究機構は、仙台市における市長への手紙・メールから得られたデータと仙台市民を対象に行った確率標本に基づく市民意識調査によって得られたデータの両者に対してテキストマイニング分析を適用した結果を報告している。この報告では、両データの同質性の検証を通して、「そのまま市民意識調査の代替とはなりえないものの、住民ニーズを把握する情報、政策形成に有用な情報になりえるものである」ことが示されている（仙台都市総合研究機構, 2003, p. iii, 竹村, 2004）。

これらの研究は、表 1 に示した受動的活動によって得られたテキスト型データに対してテキストマイニング分析を適用することにより、政策形成に有用な知識の発見ができるのか、という問題意識に基づくものであり、市民の声の分析手法を持ち得ていない現状にあって、その活用可能性を示した点では意義がある。しかし、近藤および仙台都市総合研究機構の研究はともに、受動的活動によって得られた“集まった市民の声”を分析対象にしていることから回答者の性別や年齢といった属性情報には欠落が多く、また手紙・はがきといった紙媒体からのデータ入力時においては担当者による事後の編集や分類といった主観的判断が入らざるを得ないといった短所がある。

筆者は、これら先行研究の市民の声に対するテキストマイニング分析の有用性を踏まえて、集まったデータではなく、能動的活動である市民意識調査によって“計画的に集めた”テキスト型データに対してテキストマイニングおよびクラスター分析を適用し、4 つの政策課題クラスターに分類するなど、そのデータ構造を明らかにした（金井, 2015）。この研究は、住民意識調査で得られたテキスト型データと回答者の人口統計学的変数を手がかりにクラスター分析による類型化を行ったものであり、そこでは自由回答データと選択肢型質問から得た質的な数値データとを関連させた分析までは行っていない。そこで本稿では、この研究

を進めて、調査目的やデータの取得方法が明示されている市民意識調査によって得られた自由回答データと質的な数値データとを関連させて分析を行うこととした。

## 3 分析

### 3.1 調査の概要

本稿では、A 自治体が実施した住民意識調査を事例として取りあげて分析を行う<sup>9)</sup>。調査の概要と分析対象とした調査項目は、以下のとおりである。

調査名 A 自治体世論調査

測定方法 郵送配付・訪問回収・郵送回収の混用

抽出方法 住民基本台帳からの単純無作為抽出法

標本の大きさ 2,000 名

回答数（率） 1,512 件（75.6%）

#### (1) A 自治体の現況に関する調査項目（選択肢型）

この調査では、A 自治体の現況評価として 18 の政策項目と総合評価項目が設定された。その調査項目は、「買い物」「交通」「道路・歩道」「公園・水辺」「緑化」「景観」「バリアフリー」「ごみ」「リサイクル」「治安・安全性」「災害への備え」「近所づきあい」「文化施設」「スポーツ施設」「熟年者の生活環境」「子育て環境」「健康相談・健康診査」「病院・診療所」といった政策項目の利便性や整備状況などに関するものであり、その回答尺度は「満足」「やや満足」「ふつう」「やや不満」「不満」「わからない」の 6 段階である。総合評価は A 自治体の現況をトータルとして評価したものであり、回答尺度は「わからない」を除いた 5 段階とした。

#### (2) 自由回答に関する調査項目

自由回答に関する質問は、「あなたが、A 自治体で暮らしているなかで、A 自治体の『良さ』または『改善してほしい』と感じるのは、どのようなことですか。子育て支援、熟年者施策、生活環境、産業振興、災害対策、まちづくりなど、どのような分野でもかまいません。できるだけ具体的にお書きください。」というものである。ここでは回収後の分析を考慮し、意図的に、「良さ」「改善してほしい」のそれぞれについての回答欄を設定した（以下「良い」「改善」とする）。

以下では①～③に示す順序で分析を進める。まず、①現況評価に関する選択肢型質問によって得られた質的な数値データから A 自治体の政策項目のなかで評価の高い項目と低い項目を抽出する。次に、②「良い」および「改善」欄に記入された自由回答データから年齢区分ごとの特徴的な語句を抽出する。さらに、③①で抽出した A 自治体の政策項目（評価の高いもの・低いもの）に焦点をあてて、それらの項目に対する評価（満足・やや満足・やや不満・不満）ごとに特徴ある語句を抽出し分析を行う。つまり、本稿は選択肢型質問で得られた質的な数値型データの分析から A 自治体の特徴的な政策項目を抽出し、それに焦点をあててテキスト型データと選択肢型質問を関連させた分析を行うという手順を用いる。分析対象のデータは表 2 のとおりである。

表2 分析対象とするデータ

| 自由回答データ       | 質的データ         |      |
|---------------|---------------|------|
| 「良い」に記入されたデータ | 「改善」に記入されたデータ |      |
| 総合評価          | 分析対象          | —    |
| 政策項目評価        | —             | 分析対象 |

### 3.2 政策課題の抽出

住民意識調査による行政課題の抽出については、有馬ほか（2005）による報告がある。そこでは、単純集計とクロス集計、CS ポートフォリオ分析とともに対応分析（コレスポンデンス分析）が地域の類似性や行政課題の構造分析としての可能性をもつことが示されている<sup>10)</sup>。ここでは、A 自治体の 18 項目の政策項目に対する 6 段階の評価の集計表である大きさが（18 行×6 列）の 2 元データ表に対して対応分析を適用した。その結果、A 自治体の政策項目に対して高い評価を得た項目は、「公園・水辺の整備」および「緑化の推進」であり、これらは住民からみた場合の A 自治体の競争優位の源泉になっている可能性を示すものと考えられる。他方、評価の低い項目は「災害への備え」「街のバリアフリー」「地域の治安・安全性」であり、これらの評価の低い項目は A 自治体にとって解決すべき政策課題として捉えることができる。図 2 は対応分析を行った結果の同時布置図であり、図の左から右へ満足層から不満層が布置されている。それぞれの項目に対応する円の大きさは回答頻度（人数）を表している<sup>11)</sup>。

図2 A自治体の現況とその評価の同時布置図

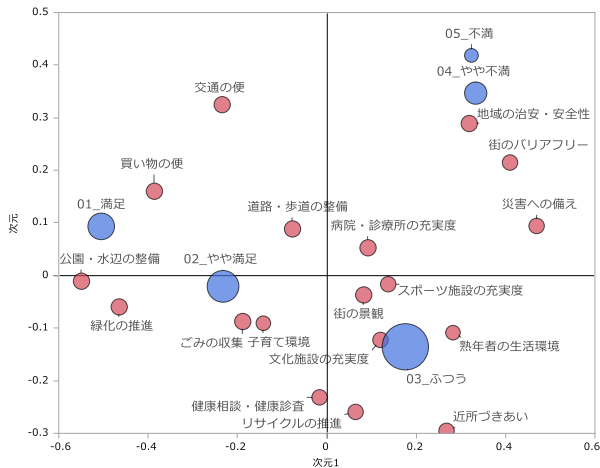


表3 固有値と累積寄与率

| 次元<br>(成分) | 固有値<br>(成分スコアの分散) | 割合(%) | 累積(%) |
|------------|-------------------|-------|-------|
| 1          | 0.083             | 70.1  | 70.1  |
| 2          | 0.030             | 25.4  | 95.5  |
| 3          | 0.004             | 3.5   | 99.0  |
| 4          | 0.001             | 1.0   | 100   |

固有値の合計（総変動）= 0.118

### 3.3 自由回答の特徴と分析対象データ

この住民意識調査の全回答者 1,512 人のうち自由回答への回答者数は「良い」=670 人（全体の 44.3%）,

「改善」=735 人（全体の 48.6%）であり、「改善」に対する記入率が高い。性別の記入率は男性 47.6%、女性 52.4%であり、女性のほうが記入率が高くなっている。年齢別の記入率は 40 代が 53.4%で最も高く、次いで 30 代（49.8%）、60 代（49.1%）、50 代（49.1%）の順であり、各年代ともに約半数の回答者が記入している。

#### (1) 「良い」に記入された自由回答の特徴

まず、「良い」に記入された自由回答文の特徴を述べる。この自由回答データに分かち書き処理を行った結果、出現頻度が「1」以上、つまり全構成要素数（総語句数）が 12,287（語）となり、異なり構成要素数が 1,896（語）となった。これが全体に占める割合（異なり構成要素率）が、 $1,896 \div 12,287 = 15.4$ （%）になる。この異なり構成要素率は、回答者が記述した語句の重複の程度を示すひとつの指標と考えられ、値が小さいほど同じ語句が繰り返し使われたことになる。

ここでは、分析対象とする語句を限定するために閾値を「5」と指定し、出現頻度が「4」以下の構成要素（語句）を分析対象から除外した。なお、自由回答文のなかの記号、句読点、助詞など不要な語句を削除するとともに、同義語と判断したものについては構成要素の置換といった辞書編集操作を加えた<sup>12)</sup>。頻度 5 以上の全構成要素数（2,579 語）に対して異なり構成要素数が 114 語であり、その割合が 4.4%にあたる。図 3 は出現頻度 5 以上の語句の分布であり、分析に用いた 114 語のなかで出現頻度が高い語句群は「子育て」160 回、「公園」156 回、「緑」153 回、「支援」131 回、「充実」107 回である。逆に、出現頻度が低い語句群は「カルチャー」「役所」「仕事」「物価」「花火大会」など 5 回になっている。

図3 114 語の構成要素の頻度分布（頻度 5 以上）

| 頻度の水準    | その頻度以上の総構成要素数 | その頻度以上の異なり構成要素数 | その頻度以上の異なり構成要素率 |
|----------|---------------|-----------------|-----------------|
| 1 頻度 5   | 2579          | 114             | 4.4             |
| 2 頻度 6   | 2489          | 96              | 3.9             |
| 3 頻度 7   | 2393          | 80              | 3.3             |
| 4 頻度 8   | 2337          | 72              | 3.1             |
| 5 頻度 9   | 2249          | 61              | 2.7             |
| 6 頻度 10  | 2213          | 57              | 2.6             |
| 7 頻度 11  | 2183          | 54              | 2.5             |
| 8 頻度 12  | 2139          | 50              | 2.3             |
| 9 頻度 13  | 2115          | 48              | 2.3             |
| 10 頻度 14 | 2050          | 43              | 2.1             |
| 11 頻度 15 | 1980          | 38              | 1.9             |
| 12 頻度 16 | 1935          | 35              | 1.8             |
| 13 頻度 18 | 1903          | 33              | 1.7             |
| 14 頻度 19 | 1885          | 32              | 1.7             |
| 15 頻度 20 | 1828          | 29              | 1.6             |
| 16 頻度 22 | 1788          | 27              | 1.5             |
| 17 頻度 24 | 1766          | 26              | 1.5             |
| 18 頻度 25 | 1742          | 25              | 1.4             |
| 19 頻度 26 | 1717          | 24              | 1.4             |

#### (2) 「改善」に記入された自由回答の特徴

次に、「改善」に記入された自由回答文の特徴を述べる。分かち書き処理を行った結果、この改善に關す

る意見は出現頻度が「1」以上、つまり全構成要素数が 23,992（語）となり、異なり構成要素数が 3,875（語）となった<sup>13)</sup>。これは「良い」の約 2 倍の構成要素数であり、この調査では「良い」よりも「改善」に多く回答が記入されたことがわかる。異なり構成要素率は、 $3,875 \div 23,992 = 16.2$ （%）である。この値は「良い」の構成要素率（15.4%）よりも大きいことから、「良い」に比べその内容が多様であると考えられる。

ここでも分析対象とする語句を限定するために閾値を「5」と指定し、出現頻度が「4」以下の構成要素を分析対象から除外した。「改善」についても回答文における語句に対して削除、置換の辞書編集操作を行った<sup>14)</sup>。頻度 5 以上の全構成要素数（2,788 語）に対して異なり構成要素数は 98 語であり、その割合は 3.5% になる。つまり、5 回以上利用された 98 種の異なる構成要素（語句）が合わせて延べ 2,788（語）あるということである。図 3 は出現頻度 5 以上の語句の分布であり、分析に用いた 98 語のなかで出現頻度が高い語句群は「子ども」107 回、「自転車」93 回、「災害対策」88 回、「バス」75 回、「交通ルール」73 回である。逆に、出現頻度が低い語句群は「リサイクル」「呼び込み」「防犯カメラ」5 回、「学力」「景観」など 6 回となっている。

図 4 98 語の構成要素の頻度分布（頻度 5 以上）



### 3.4 テキストマイニング分析

本稿で用いるテキストマイニング分析とは、自由回答文を分かち書き処理によって分割した単語や語句といった構成要素群と、回答者の人口統計学的変数や選択肢型質問などの質的変数とを組み合わせ得られる 2 元データ表に対して対応分析やクラスター化法による類型化を適用することをいう。この分析手順の特徴は、調査設計時に調査目的に合わせて構造化された選択肢型質問と市民の自由な意見（テキスト型データ）という非構造的な自由回答質問との間の関連性を可能な限り客観的に分析することにある。

#### (1) 年代別の特徴

まず、「良い」と「改善」について、年代別に使用された語句の特徴を明らかにするために、[構成要素×質的変数] = [114 語の構成要素×年齢区分（7 階級）] の大きさが（114 行×7 列）の 2 元データ表および [構成要素×質的変数] = [98 語の構成要素×年齢区分（7 段階）] の大きさが（98 行×7 列）の 2 元データ表に対応分析を適用した。年齢区分は 18-19 歳、20 代から 60 代までの各年代、70 歳以上の 7 段階である。その結果の同時布置図が図 5（良い）と図 6（改善）である。それぞれの固有値と累積寄与率から「良い」については、はじめの 2 成分が全情報（総変動）の約 53.4% を占め、「改善」については、はじめの 2 成分で全情報（総変動）の約 50.8% を占める（表 4 および表 6）。年齢区分ごとの特徴ある語句を抽出したもの、すなわち世代間の差を表したものが表 5 および表 7 である。

図 5 「良い」に関する語句と年齢区分の同時布置図

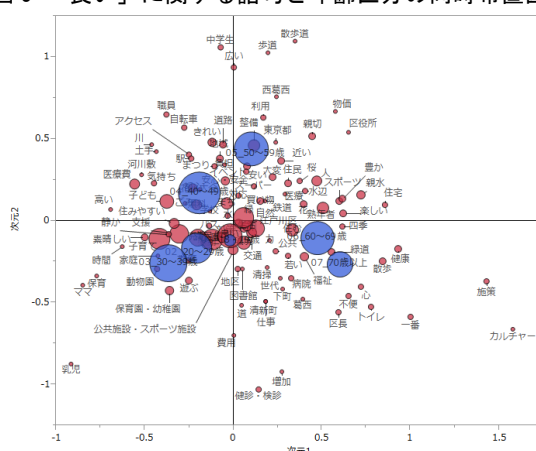


表 4 固有値と累積寄与率

| 次元<br>(成分) | 固有値<br>(成分スコアの分散) | 割合(%) | 累積(%) |
|------------|-------------------|-------|-------|
| 1          | 0.110             | 33.8  | 33.8  |
| 2          | 0.063             | 19.5  | 53.4  |
| 3          | 0.050             | 15.4  | 68.8  |
| 4          | 0.043             | 13.4  | 82.2  |
| 5          | 0.033             | 10.2  | 92.4  |
| 6          | 0.024             | 7.6   | 100   |

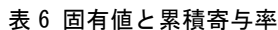
固有値の合計（総変動） = 0.326

表 5 年代別の特徴ある語句（良い）

| 18・19   | 20 代        | 30 代    | 40 代 | 50 代 | 60 代  | 70 以上 |
|---------|-------------|---------|------|------|-------|-------|
| 保育園・幼稚園 | 便利          | 子育て     | 医療費  | 整備   | 楽しい   | 熟年者   |
|         | 制度          | 支援      | 自転車  | 親切   | 散歩    | 健診・検診 |
|         | 自然          | 保育園・幼稚園 | まち   | 利用   | 四季    | 施策    |
|         | 公共施設・スポーツ施設 | 遊ぶ      |      | 中学生  | 住む    | 健康    |
|         | 地区          | 公園      |      | 散歩道  | 豊か    | 一番    |
|         | 交通          | 健診・検診   |      | 東京都  | 不便    | 親切    |
|         |             | 子ども     |      | 歩道   | カルチャー | 緑道    |
|         |             | 充実      |      |      | トイレ   | 住宅    |



図6 「改善」に関する語句と年齢区分の同時布置図



固有値の合計（総変動） = 0.161

| 18・19  | 20代       | 30代     | 40代      | 50代   | 60代  | 70以上  |
|--------|-----------|---------|----------|-------|------|-------|
| 悪い     | 東西線       | ごみ      | 給食費補助    | 学校教育  | 避難場所 | 生活環境  |
| A自治体   | 子育て       | 保育園・幼稚園 | 公園       | 整備    | 歩道   | 老人ホーム |
| バス     | 駐輪・駐輪場    | 喫煙      | 防犯対策     | 自転車   | 良い   | 熟年者   |
| 改善     | スポーツ施設    | ポイ捨て    | モラル      | 防犯カメラ | プール  | バス    |
| 子育て    | ショッピングモール | 緑道      | 健康・健診・検診 | 税金    | 高齢化  | 介護    |
| 行政サービス | 鉄道        | イメージ    | 道        | 商店街   | 住宅対策 | 病院    |
|        | 環境        | 支援      | 子ども      |       | 都営   | ペット   |
|        |           | 危険な所・時  |          |       | 介護   | 商店街   |

## ①治安・安全性に対する評価別の特徴

図 7 構成要素と治安・安全性への評価の同時布置図



固有値の合計（総変動） = 0.301

| 満足   | やや満足   | やや不満  | 不満    |
|------|--------|-------|-------|
| プール  | 学力     | 高齢化   | 安心・安全 |
| 文化   | 駐輪・駐輪場 | まちづくり | 緑道    |
| 生活環境 | 保育     | パトロール | 防犯対策  |
| トラック | 景観     | 支援    | 呼び込み  |
| 向上   | 向上     | 介護    | マナー   |
|      | ボランティア | 歩行者   | 区画整理  |
|      |        | 緑     | 道が狭い  |
|      |        |       | 夜間    |
|      |        |       | 景観    |
|      |        |       | 駅周辺   |

表 9 では、A 自治体の治安・安全性に対する不満層（やや不満、不満）が記入した語句として「防犯対策」「呼び込み」「マナー」「夜間」「駅周辺」「まちづくり」「パトロール」「支援」「歩行者」などが確認できる。この不満層の回答のなかに、前述した A 自治体が高い評価を得た政策項目である「公園・水辺の整備」や「緑化の推進」に関連する「緑道」や「緑」といった語句が見られることが特徴的である。

## ②治安・安全性に対する市民の不安

ここで、これら特徴的な語句が自由回答原文において、どのような文脈で使用されているのかについて確認を行った。これは、市民と行政とではそれらの語句を異なる意味・文脈で使用している可能性があり、行政は市民が使用する語句の意味レベルでの共有が不可欠だからである。また、これは地域住民個々人の状況や立場を常に考慮しながら市民と同じ目線での広報広聴の実践を要求する広報広聴の並行性の理念（補注 1）に基づくものである。

ここでは、自由回答原文の確認から、現況評価が低かった治安・安全性に対する市民の声から具体的に以下の仮説発見的なアプローチによる推論を進めた。すなわち、①夜間の駅周辺で飲食店への呼び込みなどが多く、安全への不安があり、その対策として駅周辺のパトロールや防犯カメラの設置がもとめられていること、②喫煙、歩きタバコ、ポイ捨てなど喫煙者のマナーに対する不安があり、それによって子どもをはじめ喫煙しない人への受動喫煙の影響が懸念されていることであり、これらは治安・安全性に関する地域住民がもつ具体的な不安・懸念事項であると考えられる。また、「緑」「緑道」を手がかりに自由回答原文の確認を検索機能を用いて行った結果、A 自治体の 18 の評価項目のなかで高い評価を得た緑化政策に関連する改善要望を確認することができた。すなわち、多くの公園や緑道が整備されていることに対しては、住民は一定の評価をしているものの、狭い歩道にも植栽するなど安全性の点からの不安の声があり、また夜間の公園での中高生の集まりやごみの放置による火災や事件などを不安視していることが、仮説発見的に捉えることが可能である。これは、地域住民は A 自治体の緑化政策・施策に対して一定の評価を与えている一方で、緑化管理の不十分さや緑化に伴う安全性の低下に懸念をもっていることを示すものである。

以上の結果は、市民の声の収集方法から分析に至る一連の過程が明らかな点において客観性をもつものであり、政策形成に有用な価値を創出するための手がかりとなる可能性を有すると考えられる。

## (3) 高い評価を得た政策項目に関する分析

次に、A 自治体の現況への総合評価と高い評価を得た政策項目である「公園・水辺の整備」や「緑化」に関連した語句の関係を明らかにする。〔構成要素×質的変数〕＝〔114 語の構成要素×現況に対する総合評価（4 階級）〕の大きさが（114 行×4 列）の 2 元データ表に対応分析を適用した（ここでは回答選択肢「ふつう」を除外）。図 8 は構成要素と総合評価の同時布置図である。累積寄与率から、はじめの 2 成分で全情報（総変動）の約 79.5%を占める（表 10）。また、評価別の特徴ある語句を抽出したものが表 11 である。

図 8 「良い」に関する語句と総合評価の同時布置図

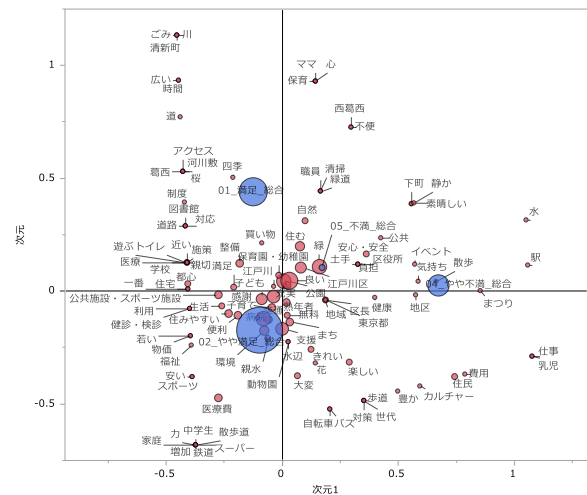


表 10 固有値と累積寄与率

| 次元<br>(成分) | 固有値<br>(成分スコアの分散) | 割合(%) | 累積(%) |
|------------|-------------------|-------|-------|
| 1          | 0.070             | 41.4  | 41.4  |
| 2          | 0.064             | 38.1  | 79.5  |
| 3          | 0.034             | 20.5  | 100   |

固有値の合計（総変動）＝ 0.169

表 11 総合評価における特徴ある語句

| 満足  | やや満足 | やや不満 | 不満  |
|-----|------|------|-----|
| 広い  | 医療費  | 住民   | 四季  |
| 清新町 | スーパー | 駅    | 気持ち |
| ごみ  | 親水   | 散歩   |     |
| ママ  |      | 緑    |     |
| 心   |      | 乳児   |     |
| 川   |      | まつり  |     |
| 道   |      |      |     |
| 保育  |      |      |     |

表 11 では、A 自治体の現況に対する満足層（満足・やや満足）のみならず、不満層（やや不満・不満）においても前述した評価の高い「公園・水辺の整備」や「緑化」に関連した語句を確認することができる。ここで、不満層に見られる「緑」という語句の自由回答原文における回答の特徴を確認した。その結果、不満層の回答のなかに「緑や水辺も多く、歩いていても自転車に乗っても気持ちの良い街」「町に緑や花が多く美しい環境」「緑が多く心がなごむ」など 23 件の回答すべてが A 自治体の緑に関連した政策に対して一定の評価を与えていることが明らかになった。こうした傾向は、不満層による A 自治体の現況に対する低い総合評価の要因は緑化関連政策以外の事柄にあることを示唆したものといえる。

## 3.5 考察

以上にみたように、自由回答質問と選択肢型質問から得たデータを関連させて分析することにより、新たな視点からの分析結果が得られることが明らかになった。本稿では、A 自治体の政策課題である治安・安全

性および高い評価を得た「公園・水辺の整備」や「緑化」に焦点をあてて分析を行ったが、今回対象としたA自治体世論調査の現況に関する18項目のすべてについて同様の手順で進めることが可能である。また、調査回答者の居住地区や子どもの有無といった人口統計学的情報や永住意向や行政への関心度など他の質問で得られた質的な数値データと組み合わせた分析も可能である。つまり、多様な視点からの分析によって、これまで把握できなかった情報源の異なるテキスト型データの構造を把握することが可能になる。

以上の分析はテキストマイニングを前提としなくても、類似した結果を導出することも可能であろう。しかし、そのためには、今回の分析事例では分析者は「良い」に対する回答者670人および「改善」に対する回答者735人全ての意見に対して、あらかじめ用意したコード表にしたがって分類整理するコーディング処理が必要となる。コーディング処理においては分析者・編集者の熟練度や調査内容によって分類結果に偏りが生じることが知られている。

今回の事例で示したように、テキストマイニング分析とその周辺技法の活用により、自由回答から重要語句を選出、編集し絞り込むこと、さらに質的な変数（選択肢型質問や人口統計学的変数）との併用による特徴抽出と類型化、それらを手がかりにした自由回答原文に遡及した詳細分析までを一貫して行う客観的かつ効率的な市民の声の分析が可能になる。

#### 4 おわりに

自治体の政策形成能力が重視されるなかで、自治体が地域の課題を発見し、その問題解決を図っていくためには、これまで以上に住民個々人の意見・要望といった非構造的なテキスト型データの分析が不可欠である。本稿では、自治体の意識調査による数値データ（選択肢型質問から得た質的な数値データ）の分析による政策課題の抽出に加えて、市民の声というテキスト型データの構成要素群（語句群）と選択肢型質問や人口統計学的変数からなる質的なデータに対してテキストマイニング分析を行った。個別の政策項目に対する具体的な市民の声（語句群）を抽出し、それを手がかりに自由回答原文を検討することで、従前の単なる分類分析とは異なる新たな政策価値を創出するための手がかりを得る仮説発見型の推論手順を示すとともに、効率的なテキスト型データ分析の可能性を示すことができた。

市民の生の声のテキストマイニング分析は、市民意識調査のような計画的に“集めた”データの分析にとどまらず、受動的活動によって“集まったあるいは集まってくるデータ”や、今後の行政経営において活用可能性をもつミドルデータやビッグデータ解析の基礎になるものと考えられる。この意味でも、自治体にとってのテキスト型データの分析は、これまで以上に実務レベルでの客観的かつ効率的な分析手法として期待される。とくに、分析の質の向上と政策価値の創出に寄与する有効な手法のひとつといえる。しかし、テキスト型データの分析から政策価値を創出する適切な手がかりを常に得ることができるわけではない。有効な

手がかりを得ることができるか否かは、データの収集方法を前提とする議論であるという側面もある。今後、周到に準備された調査実施環境下で集めたデータと集まってくるデータをいかに分析するか、その分析手法の継続的な検討が必要不可欠である。

最後に、実務の現場におけるテキストマイニングの活用について提案しておきたい。前述のとおり、「市民の声」の分析手法は確立しているとはいえず、多くの自治体で試行錯誤のなかで分析を行っている。人事異動や組織改正のなかで、継続的な分析を実践していくためには、分析の「枠組み」をもつことが有効である。そこで、実務レベルにおけるテキスト型データの分析の「枠組み」の構築を提案したい。これは、市民意識調査によって得られた自由回答データをはじめとするテキスト型データの分析によって得られた一定の知見や洞察を基礎にしながら、面談、手紙、メール、対話集会などの受動的な広聴活動で得られた意見（集まるあるいは集まったデータ）を随時追加修正していくというものである。つまり、ある一時点で得られた市民の声の構造を基礎にして、受動的活動で一定期間に得られるデータを順次位置付けて分析を進めるという方法である。もちろん、この枠組み自体の修正は定期的に必要にはなるが、集まるデータを市民の声という総体のなかで迅速に位置づけることが可能となり、分析の効率化と得られた情報の質の向上を期待することができる。

#### 補注

- 1) 行政広報広聴の理念の概要は以下のとおりである。真実性とは、行政が提供する情報が真実であると同時に、住民から真実の反応（意見、要望、苦情、問合せ等）を得なければならないことを要求する理念であり、行政活動の実態を正確に外部者に示すという現代行政に求められる透明性確保のために最も重視されるものである。双方向性とは、広報と広聴が双方向かつ一体として実践されるべきことを要求するものであり、網羅性とは障害者や高齢者、外国人を含む多様なニーズをもつすべての住民に対する広報広聴の実践を要求するものである。並行性とは、相手を人格的存在として認めることであり、地域住民一人ひとりが置かれている状況や立場を常に考慮しながら情報保有者の論理に陥ることなく、住民と同じ目線で広報広聴を実践すべきことを要求するものである。
- 2) 個別広聴とは、個々の市民からその不平、苦情、意見、要望、要求などを直接聴取するものであり、もっとも基本的な役割を果たす活動である。集団広聴とは、タウンミーティングや市民集会に代表されるように、行政と住民間のみならず、住民間の相互の対話によって声を集める活動である。タウンミーティングなどへの参加者は自主的活動者であり、かつ人数も限定的という側面があるが、個別広聴に比べ住民相互の対話を交えたコミュニケーションといえる。調査広聴とは、統計的手法を活用した調査による情報収集活動である。
- 3) 阿部(1998)は、政策形成において「問題を認識するための情報」をいかに取得するかということの重要性を述べ、その情報を情報収集の源泉が組織内部か外部なのか、収集主体の態度が能動的か受動的か、という二つの基準によってマトリクスを提示している。なお、集団広聴については、懇談会等の会場設



- 定という点では自治体の能動的な活動であるともいえるが、情報の収集段階においては住民が主体的に意見・要望を発言するという点では受動的な性格が強いといえる。
- 4) 東京都は1980年代に三つの広聴行政に関する報告書を公表した(東京都生活文化局都民広聴部,1981,1982,1983)。また、1990年代においては、東京都が学識経験者と都職員による「広聴システム研究会」を設置して、二つの報告書を公表した(東京都,1997,1998)。
  - 5) この調査は、仙台都市総合研究機構が平成14年10月に12政令市を対象に実施したもので、その調査内容は広聴制度、市民ニーズの把握方法、政策立案過程などであった。
  - 6) これは、岩手、宮城、福島を除く44都道府県の広聴担当課を対象に、広聴の取組み状況や先進事例、課題などに関する調査である(神奈川県政策研究・大学連携センター,2012)。
  - 7) KIT-PT法は住民の意見・要望・苦情・提案をカードを活用して整理することによって体系化・政策化し、合意形成を目指す技術である。詳しくは浦野(2008)を参照されたい。
  - 8) テキストマイニングとは、自然言語処理やデータマイニングなど多様な技術を組み合わせた複合技術である。そのため、計算機科学、言語学、統計学をはじめ多方面の研究領域からの横断的なアプローチが試みられている。
  - 9) 分析に使用したソフトウェアはJMPおよびJMPスクリプト(SAS社)およびWordMiner(日本電子計算株式会社、富士通エフ・アイ・ピー・システムズ株式会社)である。WordMinerは産学協同によって開発されたソフトウェアである。詳しくはテキスト・マイニング研究会ウェブサイトを参照されたい。  
<http://wordminer.org/> [2016.4.20 accessed]
  - 10) 対応分析は、選択肢型質問の回答尺度に対して成分スコア(座標値)という数量を付与することに特徴がある。例えば、回答尺度「満足」「やや満足」「ふつう」「やや不満」「不満」のそれぞれに付与の座標値が成分スコアであり、図2では横軸が第1成分スコアに、縦軸が第2成分スコアに対応する。またそれぞれの分散が第1固有値、第2固有値となる。つまり、成分スコアの変動の大きさ(回答尺度間の変動の様子)をこれによって評価することができる。対応分析法およびその特性を利用したクラスター化法の詳細についてはテキスト・マイニング研究会ウェブサイト参照されたい。  
<http://wordminer.org/> [2016.4.20 accessed]
  - 11) 累積寄与率から、はじめの2成分で全情報(総変動)の約95.5%を占める、すなわち全情報の約95%がこの2次元空間内で説明されることになる。
  - 12) 自由回答原文をWordMinerにある検索機能(例:コンコーダンス機能など)によって確認したうえで、たとえば「駅」や「駅前」を「駅周辺」へ、「火災」「地震」「津波」を「災害」、「カメラ」を「防犯カメラ」へと置換編集を行っている。
  - 13) 分かち書き処理とは切れ目のない文章を単語・語句の単位で区切ることをいう。本稿ではWordMinerにならって、この区切られた単語・語句の要素一つ一つを「構成要素」と呼ぶ。
  - 14) ここでは、「駅」や「駅前」を「駅周辺」へ、「火災」「地震」「津波」を「災害」、「カメラ」を「防犯カメラ」へと置換編集を行っている。
  - 4) 浦野秀一(2008)「自治体職員に求められる“聴く”技術の広聴マインド」『月刊ガバナンス』No.82,ぎょうせい。
  - 5) 大隅昇,L.L.パール,A.モリノウ,K.M.ワーウィック,馬場康維(1994)『記述的多変量解析法』日科技連出版社。
  - 6) 大隅昇,横原東(2004)「テキスト・マイニングが目指すもの～最近の動向、そしていま何を必要とするか～」『マーケティングジャーナル』23(3),日本マーケティング協会。
  - 7) 大隅昇,保田明夫(2004)「テキスト型データのマイニング:定性調査におけるテキスト・マイニングをどう考えるか」『理論と方法』19(2),数理社会学会。
  - 8) 大隅昇(2015)「対応分析法・数量化法Ⅲ類の考え方」<http://wordminer.org/uptodateinfo/3028/> [2016.4.20 accessed]
  - 9) 金井茂樹(2015)「テキストマイニングによる『市民の声』の分析:自治体世論調査における自由記述データを事例にして」『自治体学』自治体学会。
  - 10) 金井茂樹(2016)「自治体における信頼関係構築の脆弱性に関する一考察ー横浜市「広報よこはま(放射線特集号)」を事例としてー」『広報研究』日本広報学会。
  - 11) 神奈川県政策研究・大学連携センター(2012)『平成23年度調査研究報告書 広聴のあり方に関する調査・研究』神奈川県。
  - 12) 菰田文男,那須川哲也(2014)『ビッグデータを活かす技術戦略としてのテキストマイニング』中央経済社。
  - 13) 近藤田津(2003)「市民の声を市政に活かすためのIT活用の研究」『行政&ADP』7月号,行政情報システム研究所。
  - 14) 仙台都市総合研究機構(2003)『「市民の声」の活用法に関する調査研究』2003SURF研究報告。
  - 15) 竹村弘子(2004)「『市民の声』の整理・分析による政策への反映方策について」『みやぎ政策の風』Mar(1),宮城県。
  - 16) 辻清明(1962)「都市の広報活動(1)」『都市問題』53(8),都市問題研究会。
  - 17) 東京都生活文化局都民広聴部(1981)『東京都広聴行政研究会報告書ーこれからの広聴活動と住民相談ー』東京都生活文化局。
  - 18) 東京都生活文化局都民広聴部(1982)『東京都広聴行政研究会報告書ー住民からみた役所の窓口 問題点とその改善についてー』東京都生活文化局。
  - 19) 東京都生活文化局都民広聴部(1983)『東京都広聴行政研究会報告書』東京都生活文化局。
  - 20) 東京都(1997)『もっと活かそう都民の声 広聴システム研究会報告書』東京都。
  - 21) 東京都(1998)『もっと活かそう都民の声(Ⅱ)ー計画・予算・事業へ都民の声を反映するための“回路”設定ー平成9年度広聴システム研究会報告書』東京都。
  - 22) 那須川哲哉(2006)『テキストマイニングを使う技術/作る技術』東京電機大学出版会。
  - 23) 保田明夫(2003)「テキストマイニングの技術と適用性」『薬学図書館』48(4),日本薬学図書館協議会。
  - 24) Sten-Erik Clausen(1998), *APPLIED CORRESPONDENCE ANALYSIS An Introduction*, Sage Publications, Inc(藤本一男訳・解説(2015)『対応分析入門 原理から応用まで』オーム社。)

## 参考文献

- 1) 阿部孝夫(1998)『政策形成と地域経営』学陽書房。
- 2) 有馬昌宏,繆青,川向肇(2005)「JMPを活用した住民意識調査データに基づく行政課題の構造に関する計量的抽出法の検討」『SAS Forum ユーザー会学術総会 2005 論文集』。
- 3) 上野征洋(2000)「自治体広報広聴の現状と課題ー変化から新たな時代の展望へー」『判例地方自治』(202),ぎょうせい。