

KH コーダーによる住民意見の計量テキスト分析
—公共政策学科 WEB 調査自由記述欄分析を事例として

田島恵美
社会共生学部 公共政策学科 准教授
専門分野:社会学・社会調査法

キーワード:計量テキスト分析・KH コーダー・住民意見・自由記述欄

1. はじめに

公共政策に関する学問とは、「公共的問題を解決するための、解決の方向性と具体的手段を導き出し実施する過程、およびその公共的問題に関する諸要素を取り扱う科学」(首藤, 2021:9)であり、恣意的な判断ではなく、合理的な根拠に基づく判断が重要であるとされている(首藤, 2021: 10)。

地域における公共の課題を明らかにするためには、住民など関連する個人・団体の意見を適宜取りまとめ、分析していく必要があり、現地での実地調査が必要とされる。こうした調査方法は、聞き取り調査や公開された文書による質的な分析と質問紙による大規模な量的調査がある。地域の課題を明確化するためには、地域で問題を抱える人の状況を実際に見聞きすることの重要性は言うまでもないが、事前に情報を整理し、それに基づいて効果的に聞き取り調査や参与観察を行うことが望まれる。そのため現地でのフィールドワークに先立ち、先行研究の整理から課題の概要を把握するだけでなく、WEBを利用した調査を行い、独自に地域の課題を把握することも重要である。公共政策学科では、そのような意図のもとWEB調査を行い、報告書としてまとめているが、ここではその自由記述部分に関してテキストマイニングを用いて再度分析を試みることにする。データ数の制約はあるが、KH コーダーを利用した自由記述欄分析を行い、課題把握の手法の確認、利点・課題を学生が実際に利用することを考慮しながら検討していく。

2. 公共政策学科のWEB調査と自由記述欄

公共政策学科では、フィールドワークを必修科目としており、聞き取り調査や観察法を利用して地域の課題を把握し、それに対し具体策を提案できる力を養うようなプログラムを構築している。その一環として、WEB調査を実施している。WEB調査は、厳密なサンプリングができないため、それをもって地域課題の存在を検証できるわけではないが、仮説探索のひとつの方法であり、聞き取り調査や観察法などと併用することで、地域の実情の把握には有効な方法である。WEB調査の目的は2点ある。一つは、フィールドワークにおもむく自治体での課題に関する見取り図を得ることである。独自の調査に基づくデータは、地域での生活に関する最近の状況を知ることが可能になるとともに、当該自治体による課題の再認識や、新たな課題発見につながることも期待している。もう1点は、リアルなデータによる実際的な分析を体験することで、課題の把握について、主観にとらわれず、様々な角度から分析する能力を養成することである。

2022年(令和3年)度のWEB調査では、基本属性のほか、生活満足度と定住意欲、さらにSDGs及び公共の課題について重視していること29テーマについてたずねる項目を設けている。

WEB調査は2021年8月4日～6日、調査委託業者のモニタ会員対象として、学生が

フィールドワークを行う対象である7自治体の住民を指定して、サンプルサイズ各自治体100、合計700で行っている。

また調査の最後には、居住自治体への意見を自由に記述する欄を設けており、合計177件、各自治体に対して少ないところで14件、多いところで35件の意見が寄せられている。これは居住自治体への意見であるので、自治体ごとの分析が最優先され、自治体ごとに寄せられた意見を集約することでまとめがなされている。しかし、より多くのデータが集まった場合はこのような方法では限界があり、分析の裏付けのためにも、別の角度からの方法が望まれる。こうした方法として計量テキスト分析による自由記述欄分析の手法がある。

ここでは、計量テキスト分析の方法を紹介し、それに基づいて自由記述欄を分析し、そこでの結果から明らかになったこと、利用することの利点、課題について示すこととする。

3. 計量テキスト分析

計量テキスト分析はもともと質問紙調査における自由回答項目のコーディングに関するものとして始まったが、その後インタビューデータの分析も加わり、現在ではテキスト型データ一般を扱う方法となっている(樋口, 2020: 12)。具体的な手順・方法に違いはあるが、今回扱う自由回答欄における意見のような短いものだけでなく、長時間のインタビュー記録などにも用いることが可能である(樋口, 2020: 13)。また先行研究として、新聞記事にもとづく社会意識探索(樋口, 2020: 65-77)やTwitterでのツイートデータから緊急事態宣言解除がもたらした影響を分析した事例(峰滝, 2020)があるように、表明されたさまざまな言説からその背景にある意識を析出することが可能である。

計量テキスト分析を行うツールは、無料のものから有料のものまで様々あるが、ここではKHコーダーを用いることとする。KHコーダーは、樋口耕一が開発したもので、単に計量的分析で終わるだけでなく、元の文章を検索・閲覧する機能により、質的な分析も行えるという利点がある。分析が主観に偏っていないかを確認するとともに、計量的分析で示された手がかりをもとに、自由記述欄を読むことで、さらに理解が深まることが考えられる。無料であるので、学生でも手軽に利用できるという利点もある。他にも無料のソフトは存在しているが、KHコーダーには製作指針として、

「段階1 Correlationalアプローチにならない、多変量解析を用いることで、分析者のもつ理論や問題意識の影響を極力受けけない形で、データを要約・提示する

段階2 Dictionary-basedアプローチにならない、コーディングルールを作成することで明示的に理論仮説の検証や問題意識の追求を行う」(樋口, 2020: 19)という二つの段階からなる「接合アプローチ」としているという特徴がある。これにより、分析者の主観によらない分析、問題意識に基づく仮説にかかわる分析双方を行うことが可能となっている。こうしたツールを使えることは、住民の意識の把握、さらに課題を

めぐる現状の把握には有用であると考えられる。そこで、現在入手できる KH コーダー Ver 3. Beta01 を用いて、分析を行っていくこととする。

4. 計量テキスト分析の手順

計量テキスト分析の手順は、以下のような段階を踏む。(図1) 各段階で実際に行ったことは次で説明する

(1) データの作成

調査データから、必要な部分を抽出したファイルを作成する。

(2) データのクリーニング

誤字・脱字や個人情報のチェックを行う。

(3) データの整形・読み込み

(4) 前処理

テキストのチェックは、文字化けや長すぎる 行の存在をチェックし、自動で修正できるものは修正してくれる。さらに複合語のチェックでは自治会やごみ袋など一語として扱ったほうがよいと思われる語の候補をあげてくれるので、リストをチェックし、一語として扱うものを登録する。

その後形態素解析を行い、品詞ごとに切り分ける。

そこで出力された抽出語リストをチェックし、語の切り出し方のおかしいところを変更し、ごみ・ゴミなどの表記ゆれの吸収する

(5) 分析

上の前処理を終えたのち、分析を行う。

図1 計量テキスト分析の手順



5. 計量テキスト分析の準備

計量テキスト分析の準備段階として、上の(1)~(4)の段階について行ったことをまとめておく。

(1) データの作成

初めにデータの作成を行う。今回の調査結果はエクセルファイルで提供されているが、そこから関連のありそうな、居住自治体、年齢・性別・定住意欲及び自由記述の回答を抜き出してファイルを作成する。¹

(2) データのクリーニング²

次のそのファイルの自由記述欄を見て、まず、誤字脱字の修正、個人情報や個人を示す固有名詞を削除、不要な改行記号や記述記号の削除、半角文字の全角への変換を行う。

10 件の意見で、数字や IT などの用語で半角が見つかったので、それを修正した。固有名詞として俳優の名前と知事の名前があげられていたが、俳優の名前は前後のコメントなく単独で記載されているため居住する自治体の意見としては適当ではないと判断して削除し、知事の名前は公共性が高いと判断してそのまま残した。またこの際に表記ゆれなどで気が付いたことがあれば、リスト化しておくことにする。今回は、1 人称をワタシとカタカナで表記している例が見受けられたので、漢字に変換し、リストに加えた。

(3) データの整形・読み込み

今回の自由記述欄は一つだけ設けられており、そのなかに自治体に対するよい点と改善点が混在している。有償のソフトではこうした分析を同時に行えるが、学生が教材として手軽に利用することを考え、あらかじめファイルを分けておくこととした。そのため「よいと思う」「満足している」「対策してほしい」「不便」「心配」などの言葉を手がかりに（金井, 2015:44）、良い点と改善点の 2 つのファイルを作成した。一つの自由記述の中に良い点と改善点を併記しているものについてはそれぞれの部分を分割して、最小限の文字編集を行い、格納した。その結果、よい点を挙げた記述は、44 件、改善点・要望をあげた記述は 145 件であった。

この時点で一度 KH コーダーを起動し、新規プロジェクトとしてファイルを読み込んだ。

(4) 前処理

前処理は、テキストのチェック・複合語のチェック・語の強制抽出の設定・形態素解析・抽出語のチェック・語の切り出し方の変更と表記ゆれの吸収の各段階からなる。最初にテキストのチェックを行う。これは文字化けや長すぎる行が混在している場合に警告を出し、自動で修正できるものは修正してくれる。今回は、そのようなエラーは出なかった。次に複合語のチェックでは、まず「Term Extract」を利用して複合語を検出し、自治会やゴミ袋などひとかたまりにしたほうがよい語のテキストファイルを作り、「語の取捨選択」のメニューにより強制抽出する語として設定した。ここで一度「前処理の実行」を行った。

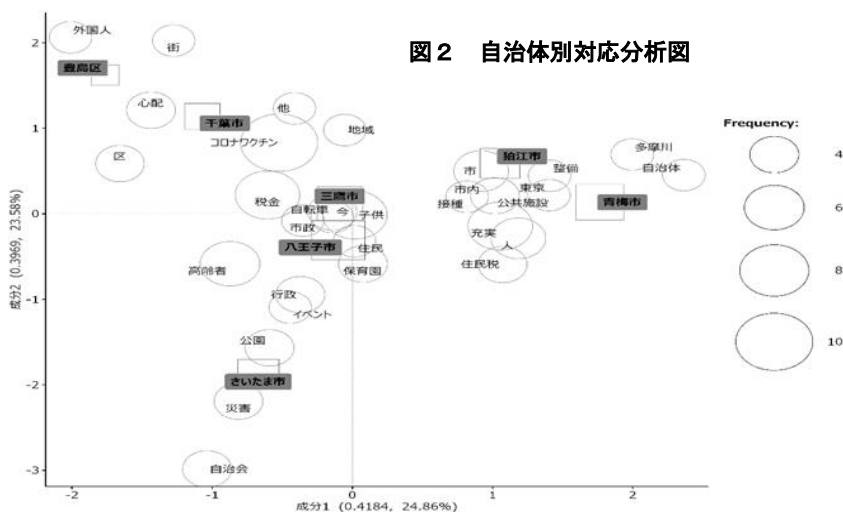
処理が終わったところで抽出語リストを出し、語の切り出しでさらに追加が必要か、また表記ゆれについて検討を行った。表記ゆれとは一つの事柄についての様々な表記法である。今回特に目立ったのは、「コロナワクチン」「コロナウイルスのワクチン」といった例や、「ごみ袋」「ゴミ袋」といった表記である。こちらも無料で公開されて

いるプラングインを利用し、うまくいかない部分については同一ルールでエクセル上の置換機能も併用した。

ここから、繰り返し、気になる表記をチェックし、再度前処理を行うことを繰り返して、クリーニングを続けた。

6. 計量テキスト分析による、自治体別課題・要望の分析

居住している自治体の良い点については、44件であり、その多くが住みやすい・満足しているなど理由が具体的でない表明が多いため、課題・要望を中心に分析を行うこととした。まず自治体別の課題・要望の全体像を見るため対応分析図(図2)と自治体を外部変数とした共起ネットワーク図(図3)を作成した。対応分析は、自治体という外部変数と関連の高い用語をプロットすることができる。原点付近には共通性のある用語が、原点から離れた点には、自治体ごとに特徴的な語が配置されている。また、自治体と語の共起ネットワーク図は、画面上では色が付き、線の太さも判別しやすいので関係の強弱も見ることができる。どちらの図からも子供、保育園、税金など対象の自治体で共通してみられる要因と、青梅市、狛江市に特徴的な多摩川に関するもの、千葉市と豊島区に特徴的な外国人に関するもの、さいたま市に特徴的な自治会に関するものなどその自治体での特徴的な要因を見ることができる。さらにソフト上では、こうしたキーワードをクリックすることで実際に書かれている文章を読むことができるので、どのような内容なのか整理しやすいという利点がある。今回は件数が少ないので、偏りはあるだろうが、フィールドワーク先に関する多数の意を読む際には、こうした見取り図を持つことと同時に、その場で実際に書かれている内容を確認する作業は、課題の把握のために有効であると考えられる。



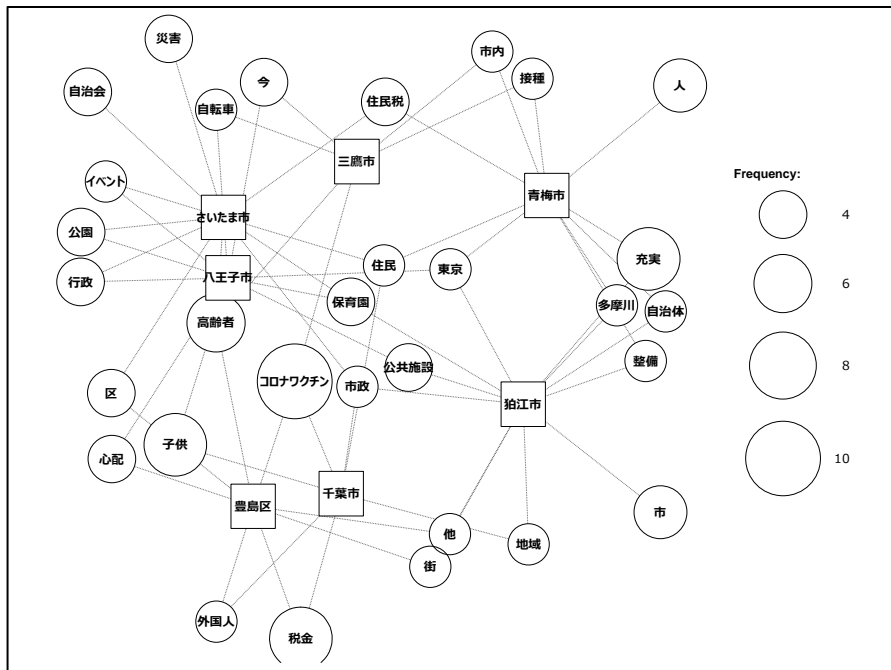
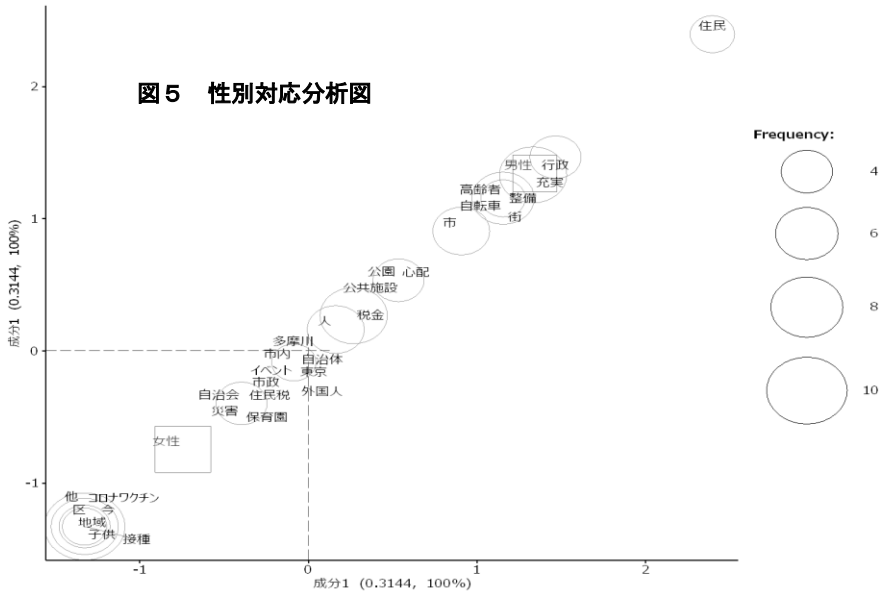
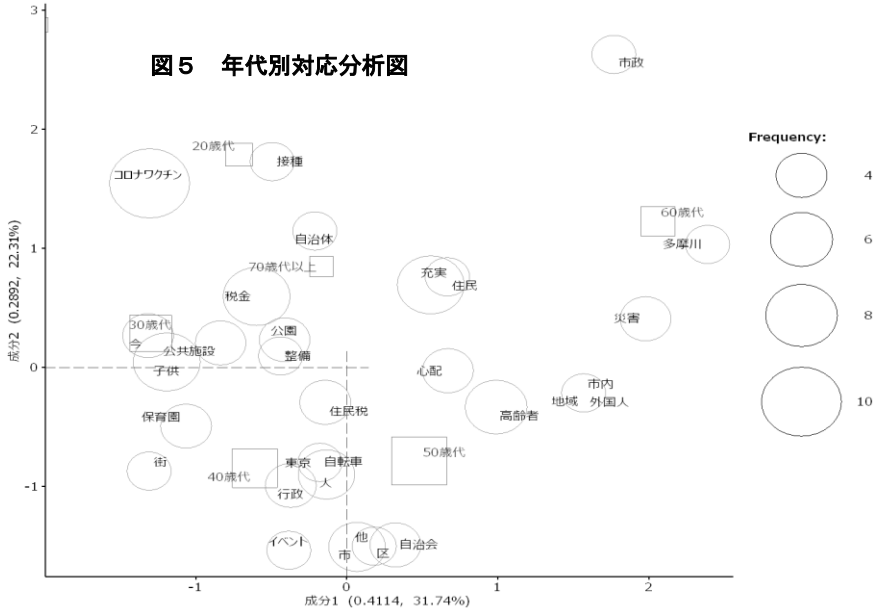


図4 共起ネットワーク図

次に属性との関連を見ておく。年齢・性別、定住意欲との関連を見ていく。定住意欲は5段階でとっているが、ぜひ住み続けたい、どちらかといえば住み続けたいを合わせて住み続けたいと記録し、機会があれば他地域に移住したい、すぐにも他地域に移住したいを住み続けたくないと記録している。これの3つの属性と記述との対応分析が以下の図5～図7である。年代別にみると20代ではコロナワクチンへの要望が多いことがわかる(図5)。調査時期は、高齢者の接種から若年層に拡大されつつあった時期であり、自治体ごとに進行状況は異なるが、順番が比較的あとになりがちな若年層の要望があらわれている。30代40代は子供や保育園の要望があるなど、年代別の異なるニーズを見ることができる。性別でみたものが、図6である。コロナワクチンの接種や子供に関する要望は女性に特徴的であり、住民や行政に関する要望は男性に特徴的であることがわかる。また、記録した定住意欲との関係を見たものが図7である。住み続けたいと考える人に特徴的な要望と他地域に移住したい人の要望の特徴を見ることができる。意見の内容をみると自治会活動のわずらわしさや保育園の状況などが不満として寄せられており、それが他地域に移住したいと考える要因の一つであるといえると思う。このように属性との対応分析は、個別の意見を整理していく中で見落としてしまうような、個々の意見の背景になる要因を可視化し

てくれる。今回はサンプルが少ないのでここで現れた関係を一般化することはできないが、自由に記述された意見の見取り図を得るためには対応分析が非常に有効なツールであることはいえる。



材に、前処理の手順を演習する方が、効果的に利用できるようになるように思える。今後さらに、住民意見・住民意識を量的側面から把握し、質的な分析とあわせてより実態に迫れるような方法の検討を重ねていきたい。

【謝辞】

住民意見・住民意識把握方法としての計量テキスト分析に関する基礎的な研究については、令和3年度大正大学学術研究助成金「分散型エネルギーインフラプロジェクトと住宅用太陽光発電システムユーザー」（研究代表者：田島恵美）の助成を得て実施した。

またここで用いられたデータは、大正大学社会共生学部公共政策学科で行った WEB 調査によるものである。データの利用を許可して下さった公共政策学科と教材用に整形して下さった大正大学社会共生学部公共政策学科の本田先生に感謝いたします。

【註】

1. 今回は学生への教材として本田先生がこの情報を切り出して作成したファイルを加工して作成している。この段階で「特になし」というような記述が削除されている。
2. テキストマイニングは、様々な分野で行われているが、データをチェックし、成型する作業をクレンジングという表現で表している分野も多い。ここでは社会調査法でこれまで用いられてきたクリーニングという用語を用いている。作業については用語で変わるわけではないが、1回では終わらず、前処理後再度切り出しや表記ゆれを吸収するなど繰り返し元データに対し作業を加える点では、従来のクリーニングよりも手間がかかる印象である。

【文献】

- 樋口耕一(2020)『社会調査のための計量テキスト分析 第2版』ナカニシヤ出版
- 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 28 巻 2 号
(DOI : https://doi.org/10.34547/jichitaigaku.28.2_42)
- 峰滝和典(2020)「Twitter データを利用した「新型コロナウイルス」関連語句の分析」 商経学叢 第 67 巻第 1 号
- 首藤正治(2021)「これからの地域における公共政策」高橋正弘・首藤正治編『公共政策基礎ゼミナール』大正大学出版会

