

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 16 日現在

機関番号：34504

研究種目：挑戦的萌芽研究

研究期間：2011～2013

課題番号：23653142

研究課題名(和文) 新聞読者投稿欄の集積と世論変遷の計量テキスト分析

研究課題名(英文) Construction of a readers' contributions database and quantitative text analysis on Japanese opinions

研究代表者

中野 康人 (NAKANO, Yasuto)

関西学院大学・社会学部・教授

研究者番号：50319927

交付決定額(研究期間全体)：(直接経費) 2,900,000円、(間接経費) 870,000円

研究成果の概要(和文)：本研究の成果は、(1)新しい社会調査の手段として大手日刊新聞社の読者投稿欄の記事をデータベース化、(2)そのデータベースを利用した投稿者の属性や掲載年月日などと内容との関連の分析、(3)日本語テキストを統計解析環境Rで分析するためのツールの整備、である。読者投稿欄は、投稿者の主張だけでなく、年齢や職業等が明記されている。媒体による選別があるとはいえ、このデータは、約60年にわたる「世論の覗き窓」として活用できる。

研究成果の概要(英文)：The results of this research project are (1) construction a database of readers' contributions in daily Japanese newspapers, (2) analysis on the data base to find changes of Japanese opinions and (3) text analysis tools for 'R'. The database, that contains contributors' age, occupation, address and their opinion, could be a source of new social researches.

研究分野：社会科学

科研費の分科・細目：社会学・社会学

キーワード：社会調査法 計量テキスト分析

1. 研究開始当初の背景

これまで、計量的社会調査のデータ獲得方法の主流であった調査票調査（アンケート調査）は、近年さまざまな困難に直面している（盛山 2008 など）。プライバシー概念の浸透や、情報保護関連法案を盾にした住民基本台帳閲覧の制限などが原因となり、標本調査の実施環境は急速に悪化している。現在、データ収集の根本的な部分から「新たな調査手法」の必要性が高まっている。

2. 研究の目的

本研究の目的は、悪化する社会調査環境の現状を踏まえ、新たな社会調査の手段として、日刊新聞の読者投稿欄に注目する。読者投稿記事は毎日一定数が蓄積され、投稿者の意見表明に加えて、住所、氏名、年齢、職業が付記されている。このテキストから、既存の調査票調査と同様に、社会現象にかかわる変数を抽出し、意味や関係や概要を分析する「計量テキスト分析」の手法を確立しながら、以下のことに取り組む。

【目的 1】投稿記事のデータベース化
大手日刊新聞（朝日、読売、毎日）の読者投稿欄の記事を、投稿者の年齢、職業、住所とともにデータベース化する。

(2) 【目的 2】投稿内容の計量テキスト分析
投稿者の職業、年齢、住所と投稿内容との関係、投稿内容の時系列的变化、新聞社ごとの投稿内容の違いについて分析を行う。

(3) 【目的 3】テキスト分析ツールの開発
フリーな統計分析環境 R を利用した、テキスト分析パッケージの開発を行う。

本研究の特色は、読者投稿記事を投稿者情報と同時に網羅的にデータベース化することにより、世相を反映する継続的なデータを構築できること、そしてその新聞記事テキストを量的に分析すること、にある。前者は研究の素材として、後者は研究の方法論として、「新たな社会調査」の礎となることが期待される。構築されたデータベースは、関係法令が許す限りにおいて公開する予定であるので、世論研究や職業研究に貢献できる素材になる。また、計量テキスト分析の手法は、自由記述の分析等、既存の社会調査にも応用できる技法であり、フリーの分析環境が公開できれば、社会調査環境に大きく貢献できるものとなる。新聞記事テキストを社会調査のデータ源として使うということは、これまでいわゆる質的調査の一環として行われてきた。例えば、見田（1963）は新聞の人生相談記事に表象される当時の日本人の「不幸」を分析しようと試みた。一方、新聞記事を対象にして量的に調査・分析をすることは、内容分析

の名のもと蓄積がある。しかし、そこで分析の対象となるのは、記事の長さや面積、単純な言葉の数など、形式的なものが多かった。本研究で確立を目指す調査・分析手法は、こうした既存の調査・分析手法の欠点を克服し、新たな社会調査のフィールドを開く試みである。

3. 研究の方法

本研究では、以下の三点の研究目的を三年間の研究期間の中で達成していくことを目標とした。

(1) 【目的 1】投稿記事のデータベース化

【方法】

電子化されていない過去の記事をデータベースに入力すると同時に、随時発行される電子化された最新の投稿記事をデータベースに入力する。

(2) 【目的 2】投稿内容の計量テキスト分析

【方法】

投稿記事を計量テキスト分析にかけ、内容を量的に抽出する。そこから、記事内容の意味空間の分析や、投稿者の属性との関係の分析を行う。さらに、既存の調査票調査データによる世論調査結果と、計量テキスト分析の結果を比較し、両者の互換性が高まるような、テキストの抽出・分析の方法を検討する。

(3) 【目的 3】テキスト分析パッケージの開発

【方法】

データ分析の際に統計分析環境 R を利用する。その中で、テキストデータを整形し、分析する手法を一つのパッケージとしてまとめ、最終的には一般に公開する。

以上の計画は、既存のテキストデータ（新聞記事）や調査票調査データを用いるので、データそのものは必ず入手できるものである。データが整備できれば、如何なる形であれ分析結果の公表まで実現できるものと考えられる。本研究は、基本的に研究代表者が個人で行うものであり、社会学者による社会調査法の研究ではあるが、言語学や情報科学分野の学会でも研究成果の発表を行う。学際的な交流を行うことによって、革新的な調査手法の確立を目指した。

4. 研究成果

本研究の成果は、以下の通りである。

(1) 【目的 1】投稿記事のデータベース化

【成果】

朝日新聞の読者投稿欄「声」と読売新聞の

読者投稿欄「気流」について、1980年代からさかのぼって、1940年代までの記事をすべて抽出し、データベース化した。また、研究開始時にデータベース化されていなかった2010年以降の記事については、朝日新聞、読売新聞、毎日新聞の記事をデータベース化した。結果として、最長で2012年から1946年までの66年間にわたる記事が蓄積された。記事データは、投稿者の意見表明である記事内容だけでなく、投稿者の属性(氏名、住所、年齢、職業)と掲載年月日を含むものである。記事内容、投稿者の諸属性、掲載年月日を、それぞれ変数としたデータベースが構築されたことになる。このデータは、世論の一側面を測るためのデータベースとして有意義なものとなる。投稿者の属性と投稿内容との関連をみることにより、既存の調査票調査の分析と同じように、社会現象にかかわる変数を抽出し、その意味や関連を分析することができるようになる。新しい職業カテゴリーの登場とその定着、名前の変化や同一言説にかかわる関連語句の変化など、一般的な調査票調査ではアプローチし難い内容について、実証的に分析が可能となる。データ化にあたっては、手作業による入力に頼ったが、古い新聞記事は保存状態・スキャン状態が芳しくなく、文字の判別が困難な記事が多くあった。そのため、作業効率が著しく低下し、必ずしも研究開始時に予定していたすべての作業が完了したわけではない。しかし今後も、網羅的なデータベースを構築すべくデータの蓄積・整備作業を継続していく予定である。

(2)【目的2】投稿内容の計量テキスト分析

【成果】

整備されたデータベースを利用して、世論変遷の分析を行った。例えば、第二次世界大戦時に日本と関係の深かったいくつかの国について、その国名が出てくる記事の特徴が時間の流れの中でどのような変化があるのか、またどのような投稿者が記事を書いているのか、分析を行った。時系列的に見て、安定して語られる内容もあれば、近年大きく変化している内容も確認された。こうした研究成果を、国内外の学会・研究会で報告した。特に、NIOD(NIOD instituut voor oorlogs-; holocaust- en genocidestudies, Amsterdam, the Netherlands)での研究発表後(学会発表)同研究所との共同研究(戦前戦後の日本における世論変遷の計量テキスト分析)が開始されることとなった。

(3)【目的3】テキスト分析パッケージの開発

【成果】

本研究では、既存のテキスト分析用のツールを援用しつつ、統計解析環境Rで利用できるパッケージの構築をめざした。特に、

日本語の計量テキスト分析に欠かすことのできない形態素解析器について、複数の解析エンジンをRから制御する方法を模索した。その研究成果の一部を、R user conference UseR!2013(Albacete, Spain)(学会発表)や統計数理研究所R研究集会(東京)(学会発表)など、国内外の学会・研究会で報告した。本研究で整備したデータは日本語テキストであるが、ネパール語やオランダ語など、いくつかの外国語テキストについても、分析環境を確認することができた(学会発表)。本研究で整備されたパッケージは、現時点では一般公開する段階に至っていないが、今後も開発を継続し、CRANなどでの公開をめざしていく。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 2 件)

中野康人, 2012, 「景観評価と地域移動-安曇野市景観意識調査の分析-」, 『地域ブランド研究』7:19-31, 査読無.

中野康人, 2011, 「社会学部生・学部教育へのアドバイス-関西学院大学社会学部卒業生調査の分析(6)-」, 『関西学院大学社会学部紀要』113:1-11, 査読無.

〔学会発表〕(計 6 件)

Yasuto NAKANO, "Quantitative Text Analysis with R", *Workshop on the quantitative analysis on text data at NIOD*, 2014.03.21, NIOD, Amsterdam, the Netherlands.

Yasuto NAKANO, "Analytic Dataset of Readers' Contributions on Japanese Daily Newspapers", *Workshop on the quantitative analysis on text data at NIOD*, 2014.03.20, NIOD, Amsterdam, the Netherlands.

中野康人, 「日刊新聞読者投稿欄の計量テキスト分析」, 統計数理研究所R研究集会, 2013.11.30, 統計数理研究所(東京).

Yasuto NAKANO, "Quantitative text analysis of readers' contributions on Japanese daily newspapers," *R user conference UseR!2013*, 2013.07.11, Universidad de Castilla-La Mancha, UCL, Albacete, Spain.

Yasuto NAKANO, “Quantitative Analysis on Political Values and Activities in Nepal and Other Asian Countries,” *International Conference on Social Science Methodology: With Special Reference to Social Movements*, 2011.11.17, Kathmandu, Nepal.

中野康人, 「社会調査とデータ・アーカイブ」, 関西社会学会, 2011.05.29, 甲南女子大学(兵庫).

〔その他〕

ホームページ等

<http://www.soc-nakano.net/>

6. 研究組織

(1) 研究代表者

中野 康人 (NAKANO, Yasuto)

関西学院大学・社会学部・教授

研究者番号: 50319927