

令和 5 年 4 月 6 日現在

機関番号：12601

研究種目：挑戦的研究（萌芽）

研究期間：2019～2022

課題番号：19K21720

研究課題名（和文）大規模社会調査における欄外記入文のデータ化と分析手法の探究

研究課題名（英文）Exploring Analysis Methods for Free Text Questions in Large-Scale Social Surveys

研究代表者

橋本 摂子（Hashimoto, Setsuko）

東京大学・大学院総合文化研究科・准教授

研究者番号：70323813

交付決定額（研究期間全体）：（直接経費） 4,900,000円

研究成果の概要（和文）：本研究は、大規模社会調査の回収済み調査票において、通常のコードデータ分析では分析対象とならない自由記述の重要性に着目し、調査データとしての整備・記録および再分析手法の開発を探索する探索的研究である。具体的には、2011年に福島大学が実施した「双葉地方の住民を対象とした災害復興基本調査」の回収された調査票原票を対象に、データ保存および二次分析がおこなわれた。実績としては、1. 双葉調査の調査票原本13,600票のデジタル化によるアーカイブ作業、2. 自由記述文のデジタル化と個人情報保護にもとづいた匿名加工作業、3. 原票PDFを含めた電子データのデータアーカイブセンターへの寄託が挙げられる。

研究成果の学術的意義や社会的意義

本研究では、東日本大震災の原発事故においてもっとも過酷な経験を強いられた双葉地方の人々の自由記述を記録することで、大規模社会調査の回収原票を歴史史料とみなす視点をひらいた。また、社会調査研究からみても、これだけ大規模な社会調査の自由記述をそのままテキスト化した調査は例がない。コードデータのデータ寄託は数多いが、質的調査のデータ寄託事例はまだ少なく、社会学全体をみても質的データのアーカイブ化事業は端緒にすぎたばかりである。こうした状況のなか、本研究ではコンプライアンスと研究倫理に従って慎重に手順を踏み、原資料のアーカイブ化を進めた先駆的事例となる。

研究成果の概要（英文）：This research is an exploratory study focusing on the importance of free-text questions in collected questionnaires of a large-scale social survey, which are not subject to analysis by ordinary code data analysis. Specifically, main purpose of this research is to preserve and record of the original questionnaires collected from the "Disaster recovery field survey for residents of Futaba district" conducted by Fukushima University at Great East Japan Earthquake in 2011. The achievements include: 1) digitization and archiving of 13,600 original survey forms of the Futaba survey, 2) anonymization of free-text questions in the original questionnaires, 3) deposit of electronic data toward secondary analysis use for academics at the Center for Social Research and Data Archives of Tokyo University.

研究分野：社会学、社会調査

キーワード：東日本大震災 社会調査 自由記述データ 双葉地方 データ・アーカイブ

## 1. 研究開始当初の背景

本研究は、「双葉地方の住民を対象にした災害復興実態調査（2011）」（以下「双葉調査 2011」と表記）の回収済み調査票をめぐって、その震災史料としての価値を鑑み、デジタルアーカイブ化による原票保存の必要性から始められた探索的研究として位置付けられる。双葉調査 2011 は、福島第一原子力発電所事故による大規模災害の中心地となった福島県双葉群 8 町村（広野町・楢葉町・富岡町・川内村・大熊町・双葉町・浪江町・葛尾村）の住民全世帯を対象に、事故から半年を経た 2011 年 9-10 月にかけて福島大学行政政策学類（内：災害復興研究所）によって実施された調査票調査である。周知のように、同年 3 月 11 日から始まった原子力発電所事故は未曾有の原子力災害へと発展した。事故直後から双葉地方にはほぼ全域にわたり強制避難命令が発令され、ほどなくして深刻な放射能汚染が判明、住民の立ち入りが無期限に禁じられる事態となった。そのため、調査時点における双葉地方の人々は、その多くが住居と職と故郷とを同時に失い、家族や地域の人々と分断され、短期間での転居（転院、転所）と移動を強いられる、非常に切迫した状況下にあった。

双葉調査 2011 はこうした時期に 8 町村すべての自治体からの協力をえて実現した世帯全数調査（発送数 28,184 票、有効回答数 13,634 票、回収率 48.4%）である。調査対象者は、東日本大震災の被災者として、おそらくもっとも過酷な経験を強いられた層と重なるだろう。結果として双葉調査 2011 は、終わりのみえない避難生活、家族の離散、金銭的負担、心身の疲弊、故郷の環境汚染、見知らぬ場所での孤立、被ばく恐怖と将来の健康悪化への不安、帰還に向けた今後への希望と絶望、復興への想いなど、双葉地方の人々が直面した被災後のリアリティを写し取る貴重な記録となり、関係者の間で原票保存の必要性が共有された。

こうした経緯に加え、本研究は社会調査と経験社会学の文脈において重要な問題提起を担う。双葉調査 2011 の大きな特徴として、回収された調査票にコードデータには回収されえない膨大な量の自由記述回答と欄外記入文がみられるためである。調査票が自由記述で埋められた主な原因は、震災、原発事故、強制避難がもたらした回答者の過酷な状況、準備不足から来る調査票の不備、そして調査スキームと回答者の現実との間に生じた深刻な認識ギャップが挙げられる。従来の大規模社会調査において、分析対象となるデータは主に統計的手法に適したコードデータを中心としてきた。しかし双葉調査 2011 の自由記述文は、通常の質問紙調査の範疇を超え、さらに調査実施者の思惑をも超えた、回答者のリアリティを伝える貴重な一次資料であるだけでなく、前例のない未曾有の災害に対し、あらかじめ全員に共有される認識スキームを前提しなくてはならない質問紙調査自体の限界を開示しうる新たな次元のデータとなりうる。

## 2. 研究の目的

以上を踏まえ、本研究の目的と意義は以下の 3 点にまとめられる。

- 1) 本研究は広義には災害社会学、被災地研究の一環として位置づけられる。断片的かつ無秩序な調査データを再度精査し、歴史史料として再構築する。
- 2) 大規模社会調査における自由記述回答と欄外記入という、これまでデータとみなされてこなかったランダムな記述に情報価値を見出すことで、量的データと質的データを横断し、質問紙調査の意味づけを問い直す。
- 3) 原資料のアーカイブ化作業そのものに、すでに社会的意義が含まれる。震災によってすでに多くが離散した、かつての双葉地方の人々の、事故当時の声を正確に記録し保存することで、次世代への記憶の継承となる。

もとより調査データの適切な保存と公開は、すべての社会調査において調査実施者の担うべき社会的責務の一つであるが、本事業は今後要請が高まるであろう震災調査データの整備・保存

事業における先駆的試行例として位置づけられる。

### 3. 研究の方法

以上の目的のもと、本研究では統計的分析手法からは除外される自由記述文や欄外記入文の重要性に着目し、原票自体の電子化とともに、コードデータとリンクしたテキストデータとして自由記述をデータ化し、量的データと質的データの統合・整備をおこなう。コードデータだけでは見えない被災者の現実を描き出そうとするこの探索的研究は、双葉調査 2011 の原票の電子化と並行して、自由記述欄のテキスト化およびデータクリーニング事業を通じて進められた。

本研究の方法は大きく、次の三つの段階に分けられる。

- 1) 双葉調査の調査票原本約 13,600 票全票のデジタル化作業 (1, 2 年目)
- 2) 欄外記入文の読み込みとカテゴリー化、分析手法の探索 (2, 3 年目)
- 3) 自由記述文のテキスト化と個人情報配慮の観点からの匿名化加工作業およびデータアーカイブセンターへの寄託 (3, 4 年目)

1) のデジタル化作業では、原票のアーカイブ保存がデータの移管と二次利用にあたるため、福島大学・東京大学双方における倫理審査をへた上で、双葉地方 8 町村の各自治体に保存事業の内容を伝え、オプトアウトを実施した。それと並行し、2) では原票から欄外記入文を文書資料として読み込み、属性コードデータを参照しながら、既存の質的データ分析手法の適用可能性を探りつつ、パターン化、コード化、カテゴリー化等、多様な側面から情報圧縮・分析・統合の方途を探った。3) では、福島大学およびデータ寄託先である東大社研データアーカイブセンターとの連携体制のもと、匿名化の基本方針を設定し、原票を参照しつつデータクリーニングをおこなった。年齢、職業、病名、入居施設名等の具体的情報はカテゴリー化あるいは匿名化する一方、被災者の広域避難経路が追跡できるよう地域名については極力匿名化しないなどの方針が共有された。

### 4. 研究成果

本研究において構築・整備されたデータは、学術的二次利用のための公開とともに、原史料の保存という、時に相反する二つの目的のもとアーカイブセンターに寄託される。そのため、個人情報の加工度にあわせてデータの階層を 4 層に分類し、それぞれの層において公開の方針を定めた。分類表は以下となる。

双葉調査 2011 データの 4 階層

	コードデータ	自由回答	欄外記入文	匿名性	データ公開
第 1 層	○	×	×	○	○
第 2 層	○	○ (匿名化)	×	○	○
第 3 層	○	○ (未加工)	×	×	△
第 4 層	○	○	○	×	×

第 1 層は通常の調査票調査によるデータと同様にコードデータのみで構成される層であり、個人を識別する情報は含まれない。個人情報が含まれる自由回答のテキストデータについては、一般公開のため個人識別につながる情報を削除した「匿名化」バージョンと、情報の正確な保存のみを目的とした「未加工」バージョンの 2 つの版を作成し、第 2 層をコードデータ+匿名化テキストデータ、第 3 層をコードデータ+未加工テキストデータとした。さらに原票そのものを PDF 化したデータが第 4 層となる。第 4 層には、自由記述回答とともに原票に書き込まれた不規則な欄外記入文が含まれる。第 4 層の原票画像 (PDF) については原則非公開とし、個人情報未加工の第 3 層とともに、アーカイブ機関への寄託後も、閲覧許可は福島大学の個別判断に依拠するという基本方針が決められた。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 橋本摂子	4. 巻 34
2. 論文標題 社会調査データの保存と公開：「双葉地方の住民を対象にした災害復興実態調査（2011）」デジタルアーカイブ化事業の概要	5. 発行年 2022年
3. 雑誌名 行政社会論集	6. 最初と最後の頁 63-82
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 橋本摂子	4. 巻 第8回
2. 論文標題 震災の記録としての社会調査 「双葉調査（2011）」のデジタルアーカイブ化を事例として	5. 発行年 2023年
3. 雑誌名 震災問題研究交流会研究報告書	6. 最初と最後の頁 78-82
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 橋本摂子
2. 発表標題 震災の記録としての社会調査：「双葉調査（2011）」のデジタルアーカイブ化を事例として
3. 学会等名 震災問題研究交流会（第8回）
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

本研究によって公開と保存用に再構築された「双葉調査2011」のデータは東京大学社会科学研究所附属社会調査・データアーカイブセンター（SSJDA）に寄託される。2023年4月現在、センターにてデータクリーニング中であり、その後匿名化されたデータ階層が一般公開される予定である。

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------