

社会調査データの復元と計量歴史社会学の可能性

——労働調査資料（1945-1961）の復元を事例として——

相澤 真一・小山 裕・鄭 佳月

東京大学社会科学研究所図書室所蔵の労働調査資料に含まれる社会調査データの復元作業を事例に、過去の社会調査データをマイクロデータ・サンプルとして復元・再構成するための手続きと方法論を定式化することが本論の目的である。研究機関におけるコンピュータの普及前に実施された社会調査では、回収された個票の情報をデジタル化し、二次的な利用が可能なサンプルデータとして整備するということがほとんどなかった。それゆえ、それらを SPSS や Stata といった現在普及している統計パッケージを用いた二次分析に利用するためには、調査データのデジタル復元と再構築を行うことが前提となる。本論は、そのために必要とされる技術、手続き、方法論を豊富な具体例とともに提示することで、今後の社会調査データの（再）構成の基礎を構築するものである。

1 はじめに

過去に実施された社会調査データをデジタル復元マイクロデータ¹・サンプルへと再構築するための手続きと方法論の定式化が本論の課題である。議論の題材として取り上げるのは、筆者らが実施した4つの社会調査データの復元作業である。

現在、日本国内でも、データアーカイブの普及により、社会調査データをデータアーカイブに寄託したり、そのデータを利用して論文を公表することは極めて日常的事になった。しかしながら、研究機関におけるコンピュータの普及前に実施された社会調査では、幸運な場合でも紙の原票が残っているのみである。そのような紙の原票データについて、回収された個票の情報を（コーディングした上で）デジタル化し、二次的な利用が可能なサンプルデータとして整備することがほとんどなかった²。その

ため、それらのデータを SPSS や Stata などの統計パッケージを用いた二次分析に利用するためには、調査データのデジタル復元と再構築を行うことが前提となる。ところが、このデジタル復元と再構築の方法論の共有とその学問的可能性については、下記に示したように、大きな学問的価値があると考えられるにもかかわらず、なかなか議論されてこなかった³。本論はこの現状に一石を投じ、多くの研究者に認識を共有してもらうことにより、新たな方法論的洗練と計量歴史社会学の学問的可能性を議論していく基盤の一端となることを望んでいる。

本論は以下のような構成を取る。まず、本研究が立脚する文脈を概観した上で、復元の手順を簡潔に整理する（第1節）。次いで本論で取り上げた4つの社会調査データを紹介し（第2節）、個票の保存・復元およびデータの再構成の手順と方法論についてそれぞれ詳述する（第3節、第4節）。

1-1 社会学における二次分析と計量歴史学

本論の背景にあるのは、次の二つの研究動向である。すなわち(1)社会学における二次分析の普及、(2)計量歴史学における過去の国勢調査の復元と公開である。

社会学における二次分析の普及

社会学における二次分析は、先行研究等においてすでに収集・整理されたデータの再分析を意味し、新たな問題設定や分析手法によって、既存の理論命題を批判的に検証し、精緻化することを主たる目的としている⁴。調査設計やデータの収集・整理といった社会調査の実施に必要な膨大なコストを大きく削減できるという意味でも、また、膨大なコストをかけて収集されたデータを活用し尽くすという意味でも、二次分析は極めて費用対効果の高い分析手法である。この二次分析の対象となりうるデータには、参与観察やインタビューによって収集されたデータから、質問紙調査によって回収され、コーディング等の加工がなされたマイクロデータ・サンプルまでさまざまな選択肢がありうるが、二次分析の普及にとって決定的に重要なのは、そうした各種データへのアクセスを容易にするデータアーカイブの構築である。社会調査データアーカイブの制度化を牽引した米国において、二次分析による知見の蓄積が急速に積み重ねられていったのは偶然ではない⁵。

データアーカイブが二次分析という研究手法にとって重要な制度的基盤であるとするれば、二次分析の成果を充実させるためには、アーカイブに含まれるデータの種類を物理的に増やしていくことが決定的な前提となる。そのための方法としては、過去に行われた社会調査や今後行われる社会調査のデータセットの調査主体による寄託が最も簡易であるが、その一方で、過去

の社会調査データのデジタル復元というアプローチも可能である。IPUMSに含まれる19世紀にまで遡る過去の国勢調査(後述)や東京大学社会科学研究所附属社会調査・データアーカイブ研究センター所蔵の「新規学卒者(中卒)労働市場調査(1953年)」(苅谷・石田・菅山編2000)はその典型である。いずれにせよ史料の復元とアーカイビングということになれば、歴史学と境界を接することになる。事実、マイクロデータのデジタル復元と再構成によるデータアーカイブの充実化は、計量歴史学と歴史人口学によって牽引されてきた。

計量歴史学

過去を知るための手がかりは、文書史料から考古学的史料、口頭伝承から統計資料に至るまでさまざまに存在する。その中でも復元マイクロデータは、比較的新しいタイプの史料として計量歴史学⁶において活用されてきた。その典型は、後にIPUMSへと結実することになる1980年代末からミネソタ大学の歴史家らが取り組んできた過去の米国の国勢調査の体系的な復元作業である⁷。これは、当局等の調査主体によって、すでに集計された公開データを史料として用いていた、従来の計量歴史学の史料的限界⁸を突破することを企図した画期的なプロジェクトであった⁹。

しかしながら、計量歴史学としての歴史人口学が関心を寄せるところの住民台帳や国勢調査、宗門改帳だけが歴史研究の史料として利用しうるマイクロデータなのではない。為政者ではなく、研究者によって実施された過去の数々の社会調査もまた歴史研究の史料となる可能性を有する。先に述べたように、コンピュータの普及以前に実施された各種の社会調査は、技術的な限界ゆえに、十分に活用されたものはほと

んどない。そうした過去の有用な社会調査の原票を発掘し、収集し、復元するという作業は、マイクロデータにもとづく計量歴史学ないし計量歴史社会学¹⁰のさらなる発展にとって不可欠である。特に社会学者の手による社会調査の場合、継続的に行われ続けている国勢調査に比べると人口や家族の長期的な変化の把握には適さないものの¹¹、国勢調査よりも、調査企画者のその都度の問題関心に応じた細かい設問が含まれていることが多く、その時点での立ちいった情報を獲得できる可能性が高い。こうした性質の異なるデータを組み合わせることが可能になれば、歴史学にとっても社会学にとっても新たな研究の可能性がもたらされることは疑いえないだろう¹²。

1-2 歴史資料としてのマイクロデータとその復元

社会学における二次分析の普及と計量歴史学の進展という二つの文脈に立つ本論が提示する方法論とそれを支える理論的見通しは、社会学と歴史学の共通の基盤の構築へと向けられている。しかしながら、過去のマイクロデータの復元は、単に記載されている情報をそのままコンピュータに入力していけばよいというわけではない¹³。居住地や職業、続柄といった文字データが含まれている際には、それらをコーディングしなければならず、また、特に、他のデータとの比較可能性を考慮に入れる場合には、どのコーディング規則を採用するかが問題となりうる（参照、Ruggles, Hacker, and Sobek 1995）。さらに緻密な研究のためには、新たな変数を作成する必要が生じることもある（例えば、家族構成については Ruggles 1995 を参照のこと）。そして、コードブックを作成しなければ、他の研究者がデータを利用することはできない。し

かしながら、最も重要なことは、史料としてのマイクロデータが、それぞれに含まれている情報の種類や史料自体の成立背景等の個性を有しているという端的な事実である。それゆえ、過去のマイクロデータを、より一般的な知見を得るための科学的な分析に耐えうるものとして復元するためには、各史料に応じた構想と戦略が必要となる。これは研究者によって企画・実施された社会調査であっても例外ではない。過去の大小の社会調査の個票を書庫や研究室の片隅から発掘し、二次分析や他のデータとの比較に利用しようとするような一般化された形式を有するデータとして復活させるためには、そのための理論と方法論が要請されるのである。

しかるにマイクロデータの復元は一般に以下の手順で行われる¹⁴。すなわち、(1) 発見・収集、(2) 保存、(3) 復元の基礎フォーマットの構築とコードブックの作成、(4) 入力とコーディング、(5) データクリーニング、(6) 分析用ファイルの作成、である。

復元対象となる社会調査の個票の現存は復元作業の前提である。史料の発見と収集は、しばしば偶然的であり、社会調査データの場合は、先行研究を通じてその存在を推定することができる。本論が検討対象とする社会調査データのように、図書館職員の英断によってアーカイブ化されていたケースは、かなりの幸運に属すると言える。本論が検討対象とするデータの性格については、第2節で論じる。

史料の保存には原票自体の保存と複製による情報保存の2種類がある。社会調査の原票は、ほとんどの場合、紙媒体であり、経年劣化が不可避であるだけでなく、それまでの保存状態と紙の質によっては、触れるだけで端から崩れていく状態にまで変化している場合もありえる。そもそも効率性という観点からも復元作業中

に、その都度、原票を取り出すことは望ましくない。それゆえ、いかに復元作業に際して扱いやすい形式で原票の複写を作成するかという問題は、一つの論点たりうる。この点について論じるのが、第3節である。

復元作業の本体をなすのが復元の基礎フォーマットの構築とコードブックの作成、入力とコーディング、データクリーニングという三つの作業工程である。先に触れたように、書かれた内容をそのまま（例えば、原票と同一の形式となるように）デジタル化するだけでは、二次分析のための復元としては不十分である。特に計量歴史社会学の基礎データとして活用可能なものとして復元する場合には、現行のほとんどの質問紙型の社会調査と同様に、（多くの場合、調査票のID¹⁵と変数からなる）行列によって原票の情報を表現する必要がある。その際に原則となる指針には、次の4つがある。すなわち、(1) 情報の再現性の追求、(2) 他の社会調査データとの比較可能性の保障、(3) 正確かつ効率的な入力作業の遂行、(4) プライバシー保護のための匿名化、である。情報の再現性とプライバシー保護が典型的だが、これら諸原則は、相互に矛盾する要請を含んでいる。第4節では、そうした矛盾の調整について、実例を交えながら論じる。

2 復元したデータの概要

本論が事例とする社会調査データは、すべて東京大学社会科学研究所図書室所蔵の「労働調査資料」に含まれる調査原票である¹⁶。この調査資料の概要は、労働調査論研究会編『戦後日本の労働調査』から知ることができる¹⁷。具体的には、「No.1 都内壕舎生活者調査」（1945年）、「No.18 京浜工業地帯調査（従業員個人調

査）」と「No.20 京浜工業地帯調査（住宅調査）（1951年）」——これらは同一の調査票である——、「No.55 貧困層の形成（静岡）調査」（1952年）——正確には「被保護世帯についての生活調査」（後述）——、「No. 60 『ボーダー・ライン層』調査」（1961年）——正確には「神奈川県における民生基礎調査」（後述）——の4つの社会調査である（No. は『戦後日本の労働調査』における整理番号¹⁸、西暦は調査の主な実施年）。以下、各調査につき、復元作業に着手した順にその概要を簡単に確認しておこう。

なお、労働調査資料は、すべて東京大学社会科学研究所が実施した調査である訳ではない。むしろ、他の機関（地方自治体・研究所等）によって実施されたもののがかなりの割合を占めている。誤解を恐れずに踏み込んで言えば、労働調査論研究会によって、『戦後日本の労働調査』としてまとめられる過程を経て、東京大学社会科学研究所に所蔵されていた（あるいは実施されていた）社会調査資料群と見る方が適切であろう。

「ボーダー・ライン層」調査

本調査は、もともとは神奈川県民生部が「神奈川県における民生基礎調査」として県内の低所得世帯14,037世帯を対象に行った調査である（回収数12,551世帯）。通称『『ボーダー・ライン層』調査』と呼ばれているのは、「被保護階層とは区別された意味での低所得階層」、具体的には、「収入水準が保護基準の1.1倍の世帯」を下限とし、「市町村税の均等割だけを負担する世帯」を上限とする世帯を操作的に「ボーダー・ライン層」と概念化したことによる（労働調査論研究会編2001: 352-3）。この「ボーダー・ライン層」という用語は当時、貧困調査で用いられていた用語であった（例え

ば、日本社会福祉学会編 1958)。また、調査票も「主として神奈川県民生部が作成し、民生員が調査対象者を訪問し、調査票にもとづいて、聴きとり、記入した」ものであり、回収した調査票の整理と集計も基本的には同機関が行ったという（労働調査論研究会編 2001: 353）。にもかかわらず、回収票の中から、ボーダー・ラインを大きく超えると考えられて、分析から除外された票を除いた 11,395 ケースの原票が東京大学社会科学研究所図書室に所蔵されていた¹⁹。著者らは、これらの原票すべての撮影を行いデジタル・データにした上で、2012 年度末までに横浜市、川崎市、横須賀市を含む東半分の 6,152 世帯のデータフォーマットへの入力を行った。

貧困層の形成（静岡）調査

現在、労働調査資料「No.55 貧困層の形成（静岡）調査」として整理されているのは、もともとは、財団法人労働科学研究所が 1952 年に静岡県在住の被保護世帯を対象に実施した「被保護世帯についての生活調査」である。この原票を譲り受け、そこから当時の分析者の関心に応じた主要な情報を独自の「整理表」²⁰へと転記した上で、行われた分析が「貧困層の形成（静岡）調査」と呼ばれているものである。『戦後日本の労働調査』で、同調査の「資料保存状況」について、「転記表あり」と記されているのはそのためである（労働調査論研究会編 2001: 325）。では、そもそもの提供を受けた原票はどこにあるのか。幸運なことに、「労働調査資料」には、前述の「整理表」だけでなく、その原票が 1,000 世帯分所蔵されている²¹。内訳は、静岡県の市部が 396 世帯、町部が 154 世帯、村部が 450 世帯である。「貧困層の形成（静岡）調査」という研究企画において整理表へと

転記されたのは、この中から都市部居住と判断された 402 世帯のうち、利用不可能と判断された 8 世帯を除いた 394 世帯であるから（氏原ほか 1959: 93）、少なくとも半数以上の世帯が「都市部の貧困」という問題設定のもと、分析から除外されていたことになる。また、原票と整理表を比較すると、続柄、年齢、最終学歴といった世帯員の属性に関する情報も切り捨てられ、分析に組み込まれることがなかったことがわかる。著者らがデジタル化を行ったのは、社会科学研究所での分析に際して作成された「整理表」ではなく、労働科学研究所によって実施された調査原票の社会科学研究所図書室現存分 1,000 票である²²。

京浜工業地帯調査（従業員個人調査、住宅調査）

本調査は、神奈川県企画審議室からの依頼によって実施された「京浜工業地帯調査」の一部である（東京大学社会科学研究所の労働調査資料の No.17～No.22 に該当する）。京浜工業地帯調査は、川崎市と横浜市にある 408 ケ所の事業所²³を対象とした「事業所調査」（No.17, 回収数は 348）、同地域の事業所の従業員 22,318 名²⁴を対象とした「従業員個人調査」兼「住宅調査」（No.18, 20, 回収数は 14,594）、同地域の重化学工業において「組」を形成し作業に従事している労働者、4 グループ計 51 名²⁵に対して行われた「職場調査」（No.19, 面接調査）、川崎職業安定所の資料調査である（その意味で狭義の社会調査とはみなしえない）「職安調査」（No.21）、自前の養成機関を持つ川崎の全工場 10 社を対象とした「技能養成施設調査」（No.22）からなる²⁶。「従業員個人調査」兼「住宅調査」は、個人に対する情報量という面でもサンプルサイズという面でも京浜工業地帯調査の中で圧倒的に重要なものである。東京

大学社会科学研究所図書室書庫には、14,327票の原票が残存しており、著者らは2011年度から2012年度にかけて、これらすべての個票の撮影とデジタル化を行った。

都内壕舎生活者調査

本調査は、当時東京大学の大学院生であった邱永漢と薄信一が主催した「東大社会科学研究会」によって1945年11月に実施された²⁷。調査対象は、「壕舎」と呼ばれていた仮小屋に居住している世帯である。当時の報告書によると、最終的に490世帯に対して戸別訪問による聞き取り調査が行われた（東京帝国大学社会科学研究会編1946）。『戦後日本の労働調査』は、本調査の資料保存状況について「調査要領、面接記録ともに散逸」と結論しているが（労働調査論研究会編2001:39）、実際には295票の原票が東京大学社会科学研究所図書室の書庫に残存している²⁸。これらの個票については、すでに2004年度東京大学文学部で開講された社会学演習「社会学の方法：社会学における資料分析——調査票の二次分析を軸に」（担当：佐藤健二教授）において、デジタル化が試みられていた。しかしながら、同演習では、個票をデジタルカメラで撮影することなく、原票の白黒コピーを使用する等、技術面で問題があった。それゆえ同演習の参加者でもあった本論筆者の一人（小山）が、前述の調査票復元作業の経験を踏まえ、2011年度にあらためて個票の撮影からやり直した。

3 社会調査資料の史料的復元——データ入力前の前提作業

人口調査などの資料を社会的分析として利用可能なデータとする作業と同様に、過去の社

会調査資料をデータ化するには、資料そのものをデジタル化する作業と、デジタル化した資料を用いて利用可能なデータのフォーマットにするという2段階の作業を経ることが必要である。本節では、このうちの第1段階の資料そのものをデジタル化する作業について論じ、つづく第4節にて、デジタル化した資料を用いて分析可能なデータセットを作成していく過程について扱う。

3-1 史料の保存状況と作業工程の検討

社会調査資料をデジタル・データに復元する場合、資料が保存されている箱を開けてすぐに撮影作業が可能になることはきわめて稀である。作業を開始する前に、資料をアーカイヴしている機関のスタッフや研究者たちと共に、以下の3点について、検討を行うことが必要である。第1に「現時点で、それぞれの史料はどのように保存されているか」、第2に、「その史料はどのような形で撮影を行うことが可能か」、第3に、「その史料はどのような形で返却しなければならないか」である。この3点は相互に関連しあう論点のため、以下では、作業工程の順を追って説明しつつ、上記3点を留意点として提示する。

1940年代から60年代にかけての社会調査資料は、安価な酸性紙を用いて作成されたことも多く、資料の劣化が著しいものもある²⁹。例えば、「ボーダー・ライン層」調査では、調査原票を直接用いて手作業での集計が行われている。これらの原票は、作表を行う過程で、調査票端を油性マジックと思われる色の付いたもので分類され、そのままの状態、かなり無造作に紐で縛られていた。そして、劣化はこの紐が擦れるところから生じていた。

この資料の作業工程を検討した場合、資料を

原状復帰で返却するという原則を諦めざるを得なかった。資料を借り出して返却するにあたっては、「原状復帰」を行うことが多くの場合、鉄則である。しかしながら、「ボーダー・ライン層」調査の場合、紐をほどかなければ、作業を行うことはできないものの、この紐を切らずにほどくことは不可能であった。また、順番を並べ替えずに撮影を行えたとしても、紙自体の重さがあるところに、資料を損傷させずに紐で結びなおすことは不可能であった。そのため、「ボーダー・ライン層」調査については、この紐を切ることとし、当時の集計の通りに並べられている調査票の並び自体は維持しつつ、それを100部ずつA3の封筒に封入する処理を行った³⁰。また、保存箱に記された資料タイトルと実際の中身の間には若干の齟齬が見られたものの、この点は資料タイトルと合うように移動させることはせず、図書室スタッフに随時報告を行うとともに、「ボーダー・ライン層」調査については、箱ごとに作業記録用紙を封入し、この資料をいつ使用して撮影を行ったのか、どの市町村の資料が入っていて、原票資料が何枚入っていたのかについての記録用紙を作成し、保管されている中性紙箱の中の一番上に入れて返却を行った。

同様の問題点は他の調査票にも見られた。貧困層の形成（静岡）調査の原票はB3（つまりB4サイズの2倍）、京浜従業員調査は、A3よりも一回り大きなサイズの用紙が原票であり、これが両者ともに二つ折にして保存されていた。二つ折にされているがゆえに、そこから劣化が進んでいたり、折り方が粗雑であったために、調査票の紙の縮みや汚損が起きたりしているケースもしばしば見られた。そのため、撮影時には、紙の縮みが起きている場合は、両端を定規などで押さえながら丁寧に引き延ばしながら

ら撮影を行った。また、京浜従業員調査の傷みの激しい原票については、A3のビニルファイルに別途封入した上で撮影を行い、それらの原票がそのままの状態でも保管できる中性紙の箱を購入し、返却した。

なお、今回、作業を行った調査票の多くは、1枚で見開きあるいは両面の状態の調査票であるか、整理票に記されているものが多かったが、製本されている資料の場合、より丁寧な扱いが必要となる。例えば、京浜従業員調査における「職場調査」（No.19, 面接調査）では、面接記録が製本されていたものの、製本による保存状態が悪く、また撮影にも不都合であったため、社会科学研究所図書室と相談の上、専門業者に製本部の裁断を依頼した。労働調査資料の他の資料でも、「ソーシャル・ニーズ調査」（No.64）は6,000票ほどの原票が残されているものの、製本された調査票をどのように処理して、デジタル化を行うか、という点で大きな問題点が残されている。

3-2 資料のデジタル化およびデータの保存作業

調査票のデジタル化には、主にデジタルカメラを用いた。デジタルカメラには大雑把に分類して、一眼レフカメラとコンパクトデジタルカメラがあり、両方での撮影可能性を試した。現在の技術的な観点で言えば、一眼レフカメラ³¹が必須である。一眼レフカメラが必要な理由は2点ある。第1に、撮影した写真のテキストを作業者が読み取っていくにあたり、光を大きく取り込む一眼レフカメラの明るさと精度の高さの優位性が揺るがないからである。第2に、現在、パソコンと接続してのリモートコントロールができるカメラは基本的に一眼レフカメラだからである。そのため、たとえ出張先での



図1 一眼レフデジタルカメラを用いた作業の提示例と実際に撮影作業をしているところ

撮影を求められるような場合でも、一眼レフカメラを用いることが望ましい³²。

本研究で実際に行った撮影作業は、図1のような形である。図1左のように、パソコンとデジタルカメラを接続して撮影作業を行った。パソコン操作でシャッターを切ることができるように設定した。当初は、三脚の影を消すために、左端にあるように衝立を置いていた。その後、撮影作業専用の照明を利用することを考案し、図1右にあるように、照明を照らした上で撮影を行うようになった³³。調査票を撮影する折には、床面に対して、厳密に水平を保つ必要があるため、水準器付の三脚が望ましい。また、普通の撮影にはない形で真下の撮影を行うおとするため、そのような向きにカメラを付けられる三脚を用いるか、三脚に付けたカメラの動きを補助できる雲台を使用することが望ましい。図2(次頁)では、二台体制で撮影作業を進めており、左手前側の三脚は雲台を使用し、右奥はカメラを下に吊るすような姿勢を取れる三脚を使用している。

撮影を行ったファイルは、デジタル画像での「原票」であるため、保存には慎重を期す必要がある。本研究ではJPEG形式で得られたデジタルファイルの画像を100頁ずつのPDF冊子に変換する作業を行った³⁴。高画質画像で撮影を行った写真ファイルをPDF化するためには、動画編集と同じ程度の処理能力がコンピュータに求められる。さらに画像ファイル、PDFファイルともに大容量のファイルとなるため、バックアップも含めて複数のハードディスクドライブなどを用意しておくことが必要となる。

入力作業では、PDFファイルを見ながらの作業となるが、注意すべき事項や問題となった事項を記録するためには、原票を冊子媒体としても保管することが必要となる。そこで、資料の状況を丁寧に把握するために、カラープリンタで印刷を行い、冊子にする。大量の資料の印刷となるため、レーザープリンタを使用することが望ましい³⁵。冊子ファイルを揃え、データ入力を行うためのパソコンを中心とした機材を揃

えられたら、デジタル化した資料を用いて利用可能なデータセットを作成していく作業に移行する。

4 分析可能なデータセットの作成作業

4-1 基礎フォーマット構築の原則

社会調査データの復元の本体をなすのは、基礎フォーマットの構築とコードブックの作成、入力とコーディング、データクリーニングの三つの作業工程である。このうち最も重要なのが、入力のための基礎フォーマットの構築であることは疑いえない。自分たちで調査を設計し実施するのは異なり、他人が作成した調査票の入力フォーマットを適切な形で準備するには、試行錯誤が欠かせない。過去の社会調査は、それが先駆的なものであるならばより一層、それぞれに強い個性を持っている場合が多く、現時点でゆるやかに共有されているような統一的なフォーマットを有していないからである。

それでも次の4点は、基礎フォーマットの作成に係る原則として、念頭に置いておく必要がある。

1. 情報の再現性の追求
2. 他の社会調査データとの比較可能性の保障
3. 正確かつ効率的な入力作業の遂行
4. プライバシー保護のための匿名化

第1の基本原則として掲げられるのは、情報の再現性の追求である。復元の定義上、個票



図2 2台のカメラを用いた作業の様子

上の情報が過不足なく表現できる入力フォーマットでなければならない。第2に、他の社会調査データとの比較可能性の保障である。本論第1節でも触れたように、過去の社会調査データの適切かつ有効な利用のためには、他の社会調査との比較が不可欠である³⁶。これは主にコーディングにおいて問題となる。一見すると現行の一般的なコーディング規則を適用すれば、比較可能性は容易に保障されるように思われるかもしれない。しかしながら事態はそれほど単純ではない。例えば、職業コードは、異なる時点間の比較可能性にとって最も問題的なものの一つである（参照、Sobek 1995）。現在の日本で使用されるSSM職業小分類は、95年版以降、技能工と生産工程作業者を中心にそれ以前の75年版からのカテゴリーの大幅な統合が行われたため、データの種類によっては、古い小分類を使用した方が適切にコーディングできる可能性がある。同様に、過去の職業名の意味内容と現在の職業名の意味内容が同一とみなしうるものかどうかとも予断を許さない問題である³⁷。また続柄のコーディングについても、どのよう

にコーディングするかは、家族史や歴史人口学等の隣接分野の研究状況³⁸と次に述べる入力作業の正確性と効率性の確保という原則とを天秤にかけながら、慎重な検討を要する事柄である。

次に述べる第3と第4の原則は、先の第1の原則と緊張関係をはらんでいる。まず正確かつ効率的な入力作業の遂行である。これは、換言すれば、コーダーの負担軽減である。情報の再現性を高めるために、複雑な条件付きのコーディング規則や緻密な分類カテゴリーを設定すると、どうしてもケアレスミスを誘発させることになる。これは結果として入力完了後のデータクリーニング時の作業量を増やすことになり、かえって全体の効率を悪くする。これを回避する最も確実な方法は、コーダーの熟練度や力量に左右されることのない明瞭な入力規則の構築である。もちろんコーディングにおける例外事項の発生は日常茶飯事であるため、必要に応じて新たな値や入力規則の設定が必要となることもある。また、特に入力の初期段階では、コーダーに対するインストラクション、入力の正確性の点検と効率性の評価が不可欠であるため、監督者による適切なマネジメントが要請される。

最後に挙げられるのが、プライバシー保護のための匿名化の原則である。社会調査マイクロデータの復元は、データセットの公開を前提としている。それゆえデータセットから個人や世帯が特定できる状況は絶対に避けなければならない。すなわち、匿名性を前提としての、学問的に有意な情報への加工が要請されるのである。

以上の原則を念頭に置きながら、今回の作業ではマイクロソフトの表計算ソフト「Excel」を利用し、調査票を入力するフォーマットと入

力する選択肢一覧のガイダンス資料の2点を用意し、テスト入力を数件行った上で、本入力に入る形で進めた。次項では、その具体的な手順を再構成しながら、上記のしばしば相反する原則同士をバランスよく調整するための方策について論じる。

4-2 データの入力作業過程

佐藤香が指摘しているように、歴史資料としての社会調査データは、量的／質的という点では簡単に二分法に分けられるレベルにはない多様な意味を含んだデータである（佐藤香 2009: 26）。例えば、図3は、貧困層の形成（静岡）調査は、原票とそこに記述されている情報をまとめたものである。

これらの情報から、決して一回で完成形のデータを作ろうと考えてはならない。本研究では、入力は多段階で行い、作業の中で、完成形のデータセットのイメージを改善していきながら、決して全体像を見失わないようにデータセットを作るイメージで作業を進めていった。第1段階では、記入されているものを「そのまま電子データにする」というイメージを持って作業を行った。すなわち、性別、続柄、あるいはあらかじめ選択肢となっている最終学校の程度やその卒業の有無については、コーディング規則を作成して、そのまま数値にした。すなわち、自由記述については、そのまま入力した。例えば、図3のうち、「(10) 死亡または遺棄した生計中心者又は生活維持者」については、「夫」という続柄、「死亡あるいは遺棄」というイベント、その発生の年月については、1回目の入力で数値化した一方で、原因についてはそのままExcelのセルにテキストを入力した。同様に「(5) 保護を受けるようになった理由」についても、そのままテキストで入力を行った³⁹。

ものの一つは、くずし字であろう。特徴的な字のくずし方についての知識も必要であり、本研究では、児玉（1993）を使用して、くずし字の解読を進めた。

どのような資料であっても、特に混乱が生じやすいのは、入力作業の開始時である。5 ケースから 10 ケース程度をテスト入力してみて、入力フォーマットを作ったとしても、入力初期段階では、研究者が気づいていなかった問題や選択肢として用意していたものにあてはまらない事例が頻出する。そのため、コーダーに積極的に質問を出してもらい、一つ一つ丁寧に対応しながら、一貫性のあるルールを構築していくことが求められる。この過程では、電子ファイルから印刷した調査票にメモを取り、すぐに見直せる体制を作る、またルールを共有するために、配布資料の改訂版をこまめに配るなどの対応が求められる。

一通りデータの入力を終わったら、データクリーニングを行う。データクリーニングは、基本的に社会調査のデータセットを作成する時と同じ方法でよい。単純集計チェックと論理チェックを行っていく。また、アフターコーディングとして、地名や職業名、それぞれの設問における「その他」の具体的な回答についてのコーディングも進めていく。なお、アフターコーディングも、できれば数値の入力と同様に、複数チェックを工程に組み込むことが望ましい。例えば、京浜工業地帯従業員調査では、学歴や職業に詳細な記述があったため、コーディングルールを作り、1 度コーディングを行ったあと、2 次チェック、3 次チェックを実施した。

以上の作業において、必ず留意しなければならない点として、プライバシー情報の問題が挙げられる。「プライバシー」という言葉がなかった当時であっても、貧困調査で語られる内容

は非常に敏感なものであった。例えば、「ボーダー・ライン層」調査では、「この調査票は調査関係者以外は見えてはいけない」「どんな理由があっても調査以外の目的に利用してはいけない」と記載されている。どの調査であっても、家族の内容に非常に立ち入って尋ねており、特に、貧困層の形成（静岡）調査では、病気や死因についての情報をきわめて丁寧に尋ねているため、高度なプライバシー情報に対する配慮が求められる。また、存命中の方々もいることから、情報の取り扱いには慎重を要する。

そこで、具体的には、例えば次のような処理を行っている。「ボーダー・ライン層」調査では、「身体障害の内容」が部位までわかるようになっていて、障害の有無のみのダミー変数での入力にとどめた。また、静岡調査については、記述内容をすべて入力したもののその内容については、特別利用許可を得た者のみの閲覧・処理を行うにとどめた。そして、例えば、「就労可能かどうか」といった変数にコーディングしなおすことにより、具体的な情報がわかりにくくなる形でデータ化を行うことを心がけた⁴¹。

4-3 作業における組織マネジメント

以上の作業を通じて、社会調査資料の復元において欠かせないのは、スキルとモラルの高い作業人員の確保である。またその作業人員を、データを復元するチームとして組織マネジメントしていくことが求められる。

本論に関連する作業では、非常勤講師として教えている／教えていた学生や出身大学院の後輩たちに協力を仰ぎ、東京大学、武蔵大学、学習院大学、成蹊大学、首都大学東京の学部生、大学院生を動員して作業を進めた。これらの学生たちには、社会調査のノウハウがあることが



図4 データ入力作業の風景（左：東京大学社会科学研究所，右：武蔵大学）

望ましい。近年では、社会調査士資格が社会学系の学部学科に設けられるようになったため、この資格取得を目指す者たちに、社会調査のデータを作り上げる現場を体験させる点でも、大きな教育的効果を持つ作業であるともいえる⁴²。そのため、ただの「アルバイト」にしないことが大切である。研究発表を行う際には、学会発表のスライドのなかに作業中の写真を盛り込み、学部生にもそれを紹介して、非常に意義のある作業であることを伝えることができるとうまい。

作業の監督は丁寧に行い、質問に常に対応できる体制を整えておくことが必要である。東京大学で行った「ボーダー・ライン層」調査と貧困層の形成（静岡）調査の復元は、すべての作業日に、相澤が出勤して監督業務に就いた。武蔵大学で行われた京浜工業地帯調査（従業員個人調査、住宅調査）の復元では、研究代表である橋本健二氏（武蔵大学教授）あるいは本論の執筆者である小山、鄭に加えて中川宗人（東京大学大学院）の4名のうち、誰かが必ず監督業務に当たるようにし、4人の間でルールを共有できるようなコミュニケーションを取れる体

制を整えた。

社会調査資料を復元するチームを組織マネジメントしていく上では、データの完成形をどのようにするのか、というイメージを持つことが必要である。そして、その完成形に向けて、現在がどのような位置にあり、それぞれの組織のメンバーがそれに対して、どのような貢献ができていくのかが具体的にわかる組織体制を作っていくことが重要である。また、時代が違うことに由来するデータに対する新鮮な驚きをメンバー間でも共有し、当時のリアリティをさまざまな形で知ってってもらうことは、データへの関心および分析における背景知識の共有の両面においても役に立つ。具体的な作業の様子は、図4に示したとおりである。どちらの作業場にも進行表、具体的なルールの説明や資料などを掲示することで、学生たちに指示と進捗を共有する試みを図っている。

5 結びに代えて

本論がその方法論と理論的射程を述べてきた、過去の社会調査データの復元には、少なく

とも次の三つの効用を期待することができる。

第1に、最も基本的な点では二次分析のデータないし歴史研究の史料としての利用である。本論の冒頭で確認したように、長期にわたって継続的に実施されてきた国勢調査等に比べると、本論で検討対象としてきた社会調査資料は、長期的な変動の観察には適さないかもしれない。それでも個性的な情報が含まれていることも少なくはなく、工夫次第では、我々に新たな歴史学的・社会学的知見をもたらすに十分である。特に、今回取り上げたような、これまでデジタル化されていなかったデータの場合、統計パッケージを用いることによって容易にさまざまなクロス表を作成することができるようになることの利点は、非常に大きい。

このような例として二点紹介しよう。第1の例として、都内壕舎生活者調査の場合、当時の調査者は「最初の意図は赤坂・麻布両区によって代表される山手住宅地帯と芝・品川両区によって代表される海岸工業地帯との差異を狙ったのであるが、その期待は殆ど裏切られた。壕舎生活をする程の者は一定の条件と必要に迫られた者であって戦災前の単なる地帯別によって区別せられるべきものではなかった」とまとめているが（田沼 1946: 41-2）、実際にクロス表を作成してみると、いわゆる旧下町において、戦災後の別居家族がある世帯や電灯のない世帯が他の地域に比べて著しく多い等、戦争の影響の地域差が明確に存在していたことが明らかとなった（戦災後の別居家族あり：旧下町 32.9%、それ以外 18.0%；電灯なし：旧下町 57.0%、それ以外 8.5%）。第2の例として、貧困層の形成（静岡）調査の場合、当時、東京大学社会科学研究所で行った再集計では、当時の集計技術の限界からか、市部に集計を絞った上、「保護をうけた直前の状態」の時期を一

括し、「戦死」と「病死」のイベントが一括されて集計されている。そのため、1952年時点の調査として、戦争を原因としたものとそれ以外の事由による貧困の総量を集計できていない（氏原ほか 1959）。実際に、町村部も入力し、貧困の要因を集計したデータを作成することによって、市部の方が戦災経験者が多い一方、村部在住の女性に世帯員に戦死・戦病死が出た事例が多くみられること、女性世帯主世帯の約3割が戦争を原因とした主たる生計維持者の喪失が貧困転落の要因となっていることなどが明らかになった（相澤・小山 印刷中）。これらの知見は、現在でも推定の難しい戦争の被害の影響を検討する上で貴重なものとなりつつある。このように、現在の技術をもってして、単純なクロス集計ができるようになるだけで、新たな歴史学的・社会学的知見をもたらす大きな可能性を持っている。

第2の効用として、社会学と歴史学の共通の基盤の構築がある。社会学と歴史学はしばしば対立的に捉えられる。例えば、ゴールドソープは、歴史学に対する社会学の利点を社会調査による新たな（より質の高い）データの産出可能性に求めている。すなわち、原理的に有限の史料しか持たず、自らデータを作り出すことができない歴史学とは異なり、社会調査を通じてデータを自らの手で原理的に無限に生み出すことができる点に社会学の強みがあるというのである（Goldthorpe 1991）。このような立場からすれば、過去の不十分な環境のもとで実施された調査は、もはや現在の社会学にとって何の意味もないと判断されることになろう。しかしながら、このような社会学からの歴史学の排除は、結局のところ、社会学の課題を「非歴史的な法則の定立」へと縮減することに帰着せざるをえず、翻って、社会学的実践から自省の契機

を奪うことになる。近年では、東京大学社会科学研究所附属社会調査・データアーカイブ研究センターにある「SSJ データアーカイブ」など、データアーカイブに寄託することによって、これらのデータを広く学術的に利用価値の高い共有財産としていくことも容易にできる。このような社会学分野における過去のマイクロデータの整備は、すでに多様な歴史的マイクロデータを存分に活用している国内外の歴史人口学や社会史、経済史と社会学をつなぐ架け橋となるに違いない。

第3に、調査データの復元自体が学術史料の保存という機能を含意する。史料の保存は古事学的関心ゆえに必要とされるのではない。先に触れた、歴史史料は物理的に有限であり、それゆえ学術利用にも限界を有しているというゴールドソープの認識は端的に間違っている。たとえ史料が有限であったとしても、それにどのような解釈を施すかという方法論上の可能性は無限である。また著者らが再構成した4つの社会調査データは、佐藤健二がその必要性を強調する調査史・方法史としての社会学史という課題にとっても重要な史料となる。佐藤は、自らの課題を「理論史・学説史のみ」の社会学史だけでなく、「データをただ数列としてだけ単純にもらって、コンピュータのなかで動かして再集計してみる」だけの二次分析とも区別し（佐藤 2011: 339）、調査プロセスを総体として認識の生産として捉え直す「本源的な二次分析」の意義を説く。しかし、真に批判的な二次分析を遂行するためには、やはり調査票それ自体の分析を含む調査プロセス全体の社会的・歴史的諸条件の分析の水準にとどまることなく、そうした分析を経由した上で、「再集計」し、その結果の再検討と再解釈へと舞い戻ることが不可欠である。換言すれば、学説的・理論的検討、

調査史的批判、洗練された分析方法の三者を組み合わせた分析が、二次分析には求められるのである⁴³。その意味で過去の社会調査データの再構成には、既成のデータを利用するだけの二次分析にはない学問的可能性が含まれている。

過去に行われた社会調査データの中には、その個票が完全に散逸してしまったものも少なくない。否、すでに消失したものがはるかに多いのかもしれない。個票が残存していたとしても劣化は不可避である。社会調査データの復元は、社会学が生み出してきた知的遺産の保存と活用を含む。それゆえ同作業の重要性は、今まで以上に広く社会学学界全体に認識されるべきである。本論は、そのための礎の小さな一欠片たらんことを望み、執筆された。

【付記1】本論は、執筆者三人で全体構想を策定した上で、相澤が第3節、第4節の草稿を、小山が第1節、第2節、第5節の草稿を作成し、その後、三人で協議の上、加筆・修正・推敲を行った。

【付記2】本論で行ってきた研究作業のうち、「ボーダー・ライン層」調査と貧困層の形成（静岡）調査の復元は、「現代日本における社会的格差の世代間関係の解明——貧困の再生産に着目して」（平成22年度—平成23年度日本学術振興会特別研究員奨励費，研究代表者：相澤真一）、「戦後日本社会における貧困問題の総合的解明と社会政策の効果の検証」（平成24年度文部科学省科学研究費補助金（研究活動スタート支援），研究代表者：相澤真一）の研究成果の一部である。また、京浜工業地帯調査（従業員個人調査、住宅調査）および都内壕舎生活者調査の復元については、「戦後日本社会の形成過程に関する計量歴史社会学的研究」（平成23

年度—平成 25 年度文部科学省科学研究費補助金（基盤研究 B），研究代表者：橋本健二）の研究成果の一部である。なお、本論で行われた作業は、社会調査・データアーカイブ共同利用・共同研究拠点である東京大学社会科学研究所附属社会調査・データアーカイブ研究センターの事業の一部として研究支援を受けた。これら数多くの研究支援に対して篤く御礼申し上げる⁴⁴。

注

¹ ここでいう「マイクロデータ」とは、データの個票ごとの分析をすることが不可能な「マクロデータ」の対義語として用いている。

² 少なくとも日本国内ではごく一部の例を除き、ほとんど行われてこなかった。このような数少ない例としては、例えば、荻谷・石田・菅山編（2000）で再集計・再コーディングを行った「新規学卒者（中卒）労働市場調査」データが挙げられる。このデータは SSJ データアーカイブの調査番号 1 として既に公開されている。

³ 例えば、「新規学卒者（中卒）労働市場調査」の復元を行った石田浩のみならず、本論で言及した『戦後日本の労働調査』のなかで現存している調査について、佐藤健二や森直人も調査資料の復元を行っているものの、復元マイクロデータ・サンプルへと再構築する手順については議論の題材とはされてきていないため、方法論の共有の不在の状況が続いている。

⁴ 例えば、二次分析の先駆的な成功例と評されるコールマンやハルゼーらの教育と社会移動の研究（Coleman et al. 1966, Halsey et al. 1980）は、マイクロデータまで立ち返ることによって、集計済みのデータの使用だけでは不可能であった新たな分析視角を産出することに成功した（Smith 2006）。

⁵ 英語圏では、Hyman（1972）や Hakim（1982）

のように教科書が作られるまで、マイクロデータの二次分析は一般化している。日本でも 1998 年の SSJ データアーカイブの設置とその使い方を説明した佐藤・石田・池田編（2000）の出版、あるいは、2000 年の JGSS（日本版総合的社会調査）調査の開始とその後の逐次公開および JGSS データを授業でわかりやすく使いながら社会統計学を解説した岩井・保田（2007）などを契機に年を追うにつれて盛んになってきている。

⁶ 計量歴史学（Quantitative History）の概略については、Anderson（2007）を参照。

⁷ 同プロジェクトの文脈と概観については、Ruggles and Menard（1995）および Ruggles, Hacker, and Sobek（1995）。これはミネソタ大学人口研究所にあるデータアーカイブ IPUMS における復元作業を記したものであり、本論が示した作業の理論的射程と技術的妥当性を検討する上で、きわめて有益な題材である。なお、イギリスでも国勢調査の統合データの作成プロジェクト（I-CeM Project）が進行中であり、エセックス大学の UK Data Archive に寄託される予定である。

⁸ もちろん新たな史料の発見は、歴史研究の革新にとって必要条件でもなければ十分条件でもなく、一つの契機あるいは研究可能性の拡大にすぎないということは言うまでもない。史料それ自体ではなく、その処理、解釈のための史料批判こそが歴史学の生命線であることは揺るがない。

⁹ なお、日本では、速水による宗門改帳を利用した歴史人口学研究の他に、教育社会学において、学籍簿や企業の職員名簿を電子データとして復元することによる歴史研究が試みられてきた（広田編 2001, 井上 2003, 吉田・広田編 2004, 武石 2012 など）。本論が射程とする研究はこのような研究に対立的に位置付けられるべきものではなく、海外の計量歴史学の成果も引き受け、速水融をはじめとする日本が国際的に発信力を持つ歴史人口

学の研究との対話可能性も探りながら、社会調査資料をきっかけとした社会学の歴史研究全体の史料論と理論的射程の改善に資することを企図している。

¹⁰ 日本での「計量歴史社会学」という呼称および研究史の概観については佐藤香（2006）を参照のこと。

¹¹ ただし、SSM や JGSS が継続的に調査を続けてきたことによって、反復横断調査として扱える可能性を保田（2006）が示しており、相澤・香川（2008）をはじめとする応用例もある。一方、山口（2004）が指摘するように、ある対象者に繰り返して調査を行う縦断的調査としてのパネルデータは、社会の変化を読み解くよりは、個人の行動を読み解くことの長所がある。

¹² 社会学内部に限っても、過去の社会調査データの復元や再コーディングは、社会学の一分野としての計量歴史社会学のさらなる進展のためにも不可欠の前提である。我が国では、米国における計量歴史学の動向とは独立に、過去のSSM調査の再コードによるその活用が行われてきた。具体的には、1960年東京SSM調査の再コードにもとづいた研究として、佐藤香（2004）、1965年SSM調査の再コードにもとづく研究として、橋本編（2010）がある。

¹³ 例えば、日本の歴史人口学的研究の最重要史料群である「宗門改帳」の利用に際しては、古文書の読解能力や個々の史料を統一的に整理可能にするための適切なフォーマットの構築が不可欠であるという（参照、速水 2009: 591-2）。

¹⁴ 例えば、速水融が代表を務めた研究プロジェクト「ユーラシア社会の人口・家族構造比較史研究」では、各地の宗門改帳が「a）収集：史料のマイクロフィルムへの複写と一部の焼き付け、b）BDS化：資料のBDSシートへの転記、c）入力：コンピュータへの入力、d）クリーニング：入力あやまりな

どのチェック、e）登録：データベースシステムのテーブルに登録して実際の統計計算に役立つ形態にテーブルを加工、の5つの工程」を経て、コンピュータを用いた計量分析が可能なデータベースとして復元されたという（小野 2000）。このうちBDSは、Basic Data Sheetの略で、速水融が趣味の電車の時刻表をヒントに考案した宗門改帳用の整理シートを指す（参照、速水 2009: 11-2, 564）。本論筆者らが取り組んだ社会調査データの場合、宗門改帳の場合に必要なとされる古文書読解力のような特殊能力は不要であったため（せいぜい手書きの略字や軽いくずし字の判読程度）、ここで言われる「b）転記」の段階を省略し、原票の情報を直接に入力フォーマットへと入力することができた。この点については、原票をPDFファイルで冊子化し、入力に使用したパソコン上で適宜拡大するなどし、判読しづらい文字の検討を行うことができたことも大きい（本論第3節も参照）。

¹⁵ もちろん一つの調査票に（例えば、世帯員情報として）複数の個人に関する情報が記載されている場合、一つの調査票を複数の個人データの集合と捉え直し、世帯員と変数からなる行列を作成することも可能である。この復元方法は、特に被保護世帯や特定の地域に居住している等の任意の属性をもった世帯（例えば、被保護世帯）を単位とした社会調査の効率的な二次分析に有用である。

¹⁶ 調査票原票は、高度なプライバシー情報を含むデータのため、取扱いには、注意を払う必要がある。また、それらの資料を保管する機関においても、特別な注意が払われていることが多い。東京大学社会科学研究所の場合、これらの資料は、特別利用許可の必要な資料として扱われている。そのため、本研究を行うにあたっては、本論執筆者の一人（相澤）が、日本学術振興会特別研究員（PD）として、東京大学社会科学研究所の石田浩研究室に受け入れてもらっている折に、所内メンバーと

して特別利用許可を申請し、許可を得た上で、書庫資料の所内での閲覧を行った。閲覧許可をもらった資料は、相澤が社会科学研究所から与えられた研究スペースに移動し、撮影作業を経て、データの復元作業を行った。

なお、本論では1945年から1961年のデータを対象としたため、副題は1945-1961としたが、労働調査資料としては1967年の調査までリストアップされている。

¹⁷ ただし、後述するように同書の記述には、実際には不正確な部分が少なくない。しかしながら、この記述の不正確さや2013年時点から見る齟齬は、その後の資料保管あるいは資料発見などに起因する問題も見受けられるので、一概に労働調査論研究会に帰責する問題であるわけではない。

¹⁸ 社会科学研究所図書室書庫も同書の整理番号に従っている。この点に、『戦後日本の労働調査』がまとめられたことの大きな意義がある。つまり、社会科学研究所図書室スタッフが、この書籍を参考にしながら、資料のアーカイブを行ったため、われわれ利用者の側もこの書籍を手がかりにすることによって、資料の探索をしていくことが可能なのである。このように、過去の社会調査資料を丹念に整理する作業は、そのこと自体がのちの研究者やアーカイヴィングスタッフへの大きな手がかりにもなるのである。もちろん同書の記述については、前の注に記したような齟齬や不正確な点も見られるため、本論が行ったような批判的な再点検が不可欠であることは言うまでもない。

¹⁹ 当時、東京大学社会科学研究所は、神奈川県調査にしばしば協力していた。翌年にも、当時の母子福祉資金と世帯更生資金を貸し付けられた世帯に対して、「福祉資金の経済効果調査」（労働調査資料No.61）を5,432世帯に対して行っており、この調査の調査原票が現存していることも確認している。ただし、福祉資金の経済効果調査につい

ては、撮影作業も含めて未着手であり、研究予算の調達と作業要員の確保も含めて今後の課題である。

²⁰ 氏原ほか（1959: 94）に整理表のサンプルが掲載されている。後述するように、この整理表へと整理する段階で世帯主を含む世帯員の情報のほとんどが当時の分析から除外されることとなった。

²¹ 現時点では原票がもともと何世帯分あったのかは不明であるものの、当時の社会調査について言及した資料に基づくと1000票の可能性が高いと推測している。なぜならば、戦後日本の家計調査を整理した岩田編（1983）に所収されている平野隆之の記述によると「厚生省は1951年5月『国民生活実態調査』と称して、1年間にわたって33都道府県の被保護世帯（約1000世帯）と一般世帯（約1000世帯）の家計調査を実施する。これによって、被保護世帯の家計の全国平均値を算出することが可能となったのである。収支項目分類は前年の「社会保障総合基礎調査」のそれを踏襲したものである。この調査結果は当時さかんであった貧困研究や最低生活費研究に貴重な資料を提供することになった。籠山京『貧困と人間』（河出書房、1953年）のなかで用いられた被保護者世帯の家計実態資料や、1952年に労働科学研究所が実施した東京における最低生活費の算出などはその例として知られている」（平野 1983: 197）とあり、当時、各都道府県において、被保護世帯約1000世帯を対象とした調査が行われており、また労働科学研究所もこのような調査に関わっていた事実がここから伺えるからである。

²² この調査の再分析報告書である氏原ほか（1959）がタイトルに（二）とつけられているとおり、この調査と同時期に、貧困の分布についても再分析が行われている。この調査資料が「貧困層の分布（富山）調査」（労働調査資料No.54）である。この資料については、「調査原票および『整理票』あり」

と記されていたものの、さらに再分析が行われた形跡があり、東京大学社会科学研究所の書庫内には、報告書と対応する調査票原票を発見することはできなかった（労働調査論研究会編 2001: 318）。

²³ 「川崎市の従業員 5 人以上の全工場、横浜市鶴見区の従業員 30 人以上の全工場、神奈川区、西区の従業員 100 人以上の全工場、計 408 工場」が対象として選定され、回収数は 348 工場（回収率 85.5%）であったという（労働調査論研究会編 2001: 125-6）。

²⁴ 横浜市鶴見区、神奈川区、西区および川崎市内の 86 工場（全従業員 22,318 名）に対して、それぞれの従業員に調査票の配布、記入、回収を依頼したという。回収数は 14,594 名、回収率は 65.3%（労働調査論研究会編 2001: 130）。

²⁵ 内訳は「東日本重工横浜造船所、取付工 12 名、同ディーゼルエンジン仕上組立工 12 名、日本鋼管川崎製鉄所、大型圧延工 14 名、味の素川崎工場、中製工・精製工 13 名」である（労働調査論研究会編 2001: 136）。これらの面接記録は、製本されたものが社会科学研究所図書室に残存しており、面接記録の数も前記のものと合致していた。これらの面接記録は、調査対象者個人の No.17 従業員個人調査の調査票と一緒に保管されており、また、その一部には調査者のものと思われる加筆部が見られたため、従業員個人調査を踏まえた面接調査であったものと思われる（ただし、面接記録に「予定外」との書き込みがあった対象者が 1 名おり、その者に関しては、対応する個人調査の調査票がなかった）。

²⁶ 全 10 工場に対する面接記録が残存している。

²⁷ 同研究会の設立経緯や同調査については、邱（1993: 254-7）。

²⁸ 同研究会が次に実施した「No.2 上野・浅草浮浪者調査」についても同様である。

²⁹ 東京大学社会科学研究所では、図書室スタッフ

の尽力によって、これらの調査票をすべて中性紙の箱に入れ替えて保存する作業が行われている。このような作業が実施されていることは、おそらく社会調査資料としては極めて稀であろう。また、この作業を経ている調査票原票でも、中性紙の箱に保管される前に生じた汚損や劣化と向き合うことは、解読作業の中で不可避なものである。

³⁰ なお、A3 の中性紙封筒が特注品以外ではなかったため、予算の都合上、通常の封筒を使用せざるを得なかった。さらなる保存状況の改善については、研究予算の調達も含めた今後の課題である。

³¹ なお、今回の作業で用いたデジタルカメラは、Canon EOS Kiss X4 である。

³² なお、保存状態の比較的良好な資料についてはフラットヘッドスキャナを用いることも可能である。例えば、「ボーダー・ライン層」調査では、デジタルカメラによる撮影と A3 フラットヘッドスキャナによる作業を並行して行った。ただし、シャッターを切るのは一瞬なので、慣れてくると時間あたりの作業可能量はデジタルカメラの方が大きい。なお、フラットヘッドスキャナとして使用した機材は Brother MFC-6890CN である。

³³ なお、本研究の作業のうち、「ボーダー・ライン層」調査の調査票すべてと貧困層の形成（静岡）調査の大部分は、撮影用照明ライト購入前に画像のデジタル・データを作成した。冊子の解読についての著しい不具合はなかったものの、暗めに保存されていることには変わりなく、これを改善することについては、研究予算の調達と作業要員の確保も含めた今後の課題である。

³⁴ ここで 100 頁ずつの PDF 冊子としたが、実際に撮影する場合も 10 枚ずつ葉を挟み、100 枚撮影したところで、もう一度、紙媒体および画像ファイル双方の数の確認を行った。これにより、100 枚漏れなく撮れているかの確認を行った。また今回は片面のみの印刷であったが、裏面に何らかの

メモ書きの記載がある場合などは、そこで「調査票番号〇〇の裏面を撮影」という札を1枚撮影したうえで、裏面を撮影した。この場合は、原紙は100枚に対して、画像ファイル数が102枚、104枚…と偶数になっていることを確認しながら作業を行った。

³⁵ 資料のデジタル化およびデータの保存作業では、次のようなトラブルや問題が想定されうるので注意されたい。まず、撮影作業では、カメラの設置とピント、絞り、シャッター速度の調整などについて注意を払い、入力作業に支障のない十分に納得のいく画像が得られることを確認の上、大量に撮影する体制としていく必要がある。作業中は、大量に栞を用意しておき、10枚刻み、100枚刻み等で正確に数を数えられるようにし、作業を区切るたびに撮影漏れがないかを確認する必要がある。また、画像ファイルのPDF化では、2011年現在でメモリ2ギガ～4ギガ程度のパソコンでは処理中に強制終了してしまうことがある。8ギガ程度以上のスペックのあるパソコンを使うことが望ましい。カラー資料の冊子作りの作業にはプリントアウト出力の時間だけでなく、膨大な量のトナーを必要とするため、実施される際は、時間、予算両面に時間の余裕を見て、作業を進めるべきである。なお、保存用に高価な厚い紙を使用する時は、プリンタの設定を直す必要がある。

³⁶ 計量歴史社会学の史料論の検討を行っている森直人は、これを「『データ批判』の作法」と呼ぶ（森2005:226）。

³⁷ 例えば、IPUMSにおける職業の再コーディングでは“nurse”が“medical nurse”と“domestic servant”の双方を指示することが問題となり、世帯主との関係が雇主の場合には、後方で再コードしたという（Sobek and Dillon 1995: 72）。

³⁸ 例えば、「おじ」、「めい」、「いとこ」を「その他の親族」という一つのカテゴリーにまとめてし

まうと現在の歴史人口学で主流となっているいわゆるHammel and Laslette Model (HLM)の小分類(class)が使用できなくなる。HLMを使用することの意義の一つが国際的な世帯構成の比較可能性の担保にあるとすれば（参照、岡田2000:139）、世帯員情報を含む社会調査データでそれが十分に活用できなくなることは影響は、過小評価すべきでない。

³⁹ 貧困層の形成（静岡）調査では、全体で600コラム程度の情報量となったうえ、テキストの文章を読むことと数値を入れることを両立させながら、コーダーの熟練をはかることが難しかったため、「数値担当」、「テキスト担当」と両者を分けて多段階入力を行い、その後にデータの接合を行った。

⁴⁰ なお、単純に入力を行う場合でも、数値の入力およびその数値の解釈に置いて、河野（2012）が指摘するように、現在とは異なる意味を読み取っていく必要がある。例えば、収入についての情報を入力する場合でも、当時の金銭感覚から見て、きわめて特異な値は見分けていかなければならない。金銭感覚についての情報は、森永（2008）を参照した。また、現在では使わないような単位が用いられている場合にも、適宜、当時の資料などを確認し、情報を補いながらの入力を行った。

⁴¹ なお、この変数の作成過程の具体的な説明は石島（2013 近刊）を参照のこと。

⁴² 今回の作業過程でも、ここでの社会調査データの作成経験を生かして、自分の卒業論文において、社会調査を実施し、一連の社会調査の流れを自分でマネジメントできるような学生も見られた。

⁴³ 例えば、橋本（2013）はその好例である。菅山（2011）は、京浜工業地帯調査（従業員個人調査）の調査票それ自体の批判的分析を行った上で、日本鋼管川崎製鉄所のケースについて独自の再集計を行い、氏原（1966）の推論の問題点を指摘したが、橋本は京浜工業地帯調査（従業員個人調査）全体

をデジタル復元したデータを用いて、両者に対するより緻密な批判を行った。

⁴⁴ 本論文が成立するにあたっては、実に、多くの人たちのお世話になりました。まず、労働調査資料の存在を著者たちにご教示くださった仁田道夫先生（当時、東京大学、現：国土舘大学）に心から御礼申し上げます。作業の全工程にあたっては、東京大学社会科学研究所の前所長である末廣昭先生および現所長の石田浩先生には全面的にご協力を頂きました。心より感謝申し上げます。

以下、作業の工程ごとに、記して御礼申し上げます。なお、所属はすべて当時のものであり、出版時点とは異なるものがあります。東京大学社会科学研究所図書チームリーダーの阿食秀昭氏（2010年度）、池浦文彦氏（2011年度）をはじめとする東京大学社会科学研究所図書室スタッフの皆様には、資料の特別利用を行うにあたり、大変お世話になりました。特に、社会科学研究所図書チーム主任の清水律子氏、蛭原加奈子氏からは、労働調査資料を整理・保存する折の体験談を伺わせてもらい、作業を進める上で、大変参考になりました。撮影作業の折には、カメラの設置の工夫の仕方や雲台を借用してくださったことなどで、社会科学研究所技術専門職員の森田英嗣氏には大変お世話になりました。東京大学での入力作業用の冊子印刷および作業パソコンのメンテナンスはセンター支援室学術支援専門職員の入山浩一氏（東京大学）が全面的にサポートしてくださりました。研究費執行における事務面では、学術支援職員の稲葉めぐみ氏、谷口沙恵氏（共に東京大学）にご支援を賜りました。誠にありがとうございました。撮影作業では、茂木愛未（成蹊大学）、長谷川雅史（首都大学東京卒）、新川翔太（武蔵大学）、小川直之（武蔵大学）の4氏にご協力を賜りました。

また、撮影後の入力作業の企画では、橋本健二先生（武蔵大学）、佐藤香先生（東京大学）を中心

とする研究会にて何度も発表をさせていただき、有益なコメントを頂きました。大変お世話になってきた橋本健二先生、佐藤香先生に心より感謝を申し上げます。それだけでなく、当研究会に当初から参加されている岩井八郎（京都大学）、片瀬一男（東北学院大学）、元治恵子（明星大学）の各先生方に御礼申し上げます。

神奈川県民生基礎調査の入力では、撮影にも携わった茂木、長谷川、新川、小川の各氏のほか、久米祐太郎、小林華枝、田中信暁、森田純一（以上、当時、学習院大学学生）、伊野瀬晶子、大熊麻理枝、細田昇平、（以上、当時、成蹊大学学生）、山内明美（一橋大学大学院生）、井上紗央里、岡彩香、荻原ゆかり、勝又慶彦、川尻さとみ、塩崎寛務、鶴本玲、福本美紀、柳澤秀保、楊建修（以上、当時、武蔵大学学生）、西本恵理子（武蔵大学大学院生）（大学名、氏名五十音順）の各氏にご協力を頂きました。

貧困層の形成（静岡）調査の入力では、上述の久米（学習院大卒業後、学習院大学大学院に進学）、伊野瀬、井上、勝又、川尻、鶴本、福本、柳澤、楊、山内の各氏のほか、石丸友菜、金子純子（ともに武蔵大学）、高垣真菜（学習院大学大学院）、石島健太郎、開田奈穂美、米澤旦（以上、東京大学大学院）の各氏にご協力を頂きました。特に、本調査データの第3次チェックの問題点の整理などでは伊野瀬晶子氏の多大なる貢献なくしては本データの構築は不可能だったかもしれません。記して感謝を申し上げます。

京浜工業地帯調査（従業員個人調査）の入力では、武蔵大学の石丸、井上、勝又、金子、川尻、新川、鶴本、福本、柳澤、楊の各氏のほか、秋葉陽太、井上里穂、上野春香、菅藤晋、河津秀実、北村薫里、清水香奈子、鈴木直人、高野皓子、鶴田美咲、長濱志緒、古川太一、村上杏子（以上、武蔵大学学生）、久米（学習院大学大学院）のほか、中川宗人、開田奈穂美（東京大学大学院）の各氏にご協力を

賜りました。

都内壕舎生活者調査では、2004年度の東京大学文学部の社会学演習以来、佐藤健二先生（東京大学）にさまざまにご助言を賜りました。記して感謝申し上げます。また、撮影、入力では、武岡暢氏（東京大学大学院）にご協力を賜りました。

以上のように、大変多くの学部生・大学院生たちの尽力なくしては、本論をまとめあげることは不可能でした。ここに心から御礼申し上げます。ありがとうございました。

データ分析の報告では、萌芽段階で社会科学研究所若手メンバーでの報告機会を与えて下さり、丁寧にコメントも下さった当時、東京大学社会科学研究所助教の加島潤、加藤晋両氏および同室で

日々のうるさい作業に付き合いながら研究を聴きにいらしてくださり、コメントも下さった中澤俊輔氏（日本学術振興会特別研究員）、さらには研究成果報告会に若手の歴史に対する無知に対して温かくきわめて的確なコメントを下さった加瀬和俊先生（東京大学）に心から御礼申し上げます。

そして、このたび、丁寧に査読して下さった稲田雅也先生、土屋敦先生に篤く感謝を申し上げます。最後に、調査資料を残して下さった先達の研究者の先生方を含め、ここにお名前を記さなかった方々も含め、本調査資料とその復元にかかわったすべての方々に記して感謝申し上げます。どうもありがとうございました。

文献

- 1995年SSM調査研究会，1995→2006，『SSM産業分類・職業分類（95年版）修正版』1995年SSM調査研究会。
- 相澤真一・香川めい，2008，「後期中等教育拡大期の高卒就職者の世代内移動——JGSS累積データを用いた社会移動研究における時系列的探索分析の試み」『年報社会学論集』21: 131-42.
- 相澤真一・小山裕，印刷中，「『貧困層の形成調査（静岡県，1952年）』の概要と戦争が与えた影響」東京大学社会科学研究所 社会調査・データアーカイブセンター『社会科学研究所所蔵「労働調査資料」の二次分析』東京大学社会科学研究所。
- Anderson, Margo, 2007, "Quantitative History," Willam Outhwaite and Stephen Tuner eds., *Sage Handbook of Social Science Methodology*, London: Sage: 246-63.
- Coleman, J., E. Campbell, C. Hobson, J. McPartland, A. Mood, F. Weinfeld and R. York, 1966, *Equality of Educational Opportunity*, Washington, DC: US Government Printing Office.
- Goldthorpe, John, 1991, "The Uses of History in Sociology: A Reflection on Some Recent Tendencies," *British Journal of Sociology*, 42: 211-30.
- Hakim, Catherine, 1982, *Secondary Analysis in Social Research: A Guide to Data Sources and Methods with Examples*, London: Allen & Unwin.
- Halsey, Albert H., A.F. Heath and J.M. Ridge, 1980, *Origins and Destinations: Family, Class, and Education in Modern Britain*, Oxford: Oxford University Press.
- 橋本健二，2013，「雇用における戦前と戦後の連続性と不連続性」2012年度二次分析研究会課題公募型研究成果報告会2報告資料。
- 橋本健二編，2010，『家族と格差の戦後史——一九六〇年代日本のリアリティ』青弓社。

- Hyman, Herbert H., 1972, *Secondary Analysis of Sample Surveys: Principles, Procedures and Potentialities*, New York: Wiley.
- 速水融, 2009, 『歴史人口学研究——新しい近世歴史像』藤原書店.
- 速水融編, 2000, 『ユーラシア社会の人口・家族構造比較史研究』文部省科学研究費創成的基礎研究(平成7～11年度)最終実績報告書.
- 平野隆之, 1983, 「官庁の家計調査」岩田正美編『戦後日本の家計調査』法律文化社, 187-224.
- 広田照幸編, 2001, 『近代化過程における中等教育の機能変容に関する地域間比較研究』平成10～12年度文部省科学研究費補助金(基盤研究(B)(1))研究成果報告書.
- 井上好人, 2003, 「金沢一中卒業生からみた旧加賀藩士族の社会移動」『教育社会学研究』73: 5-24.
- 石島健太郎, 印刷中, 「二次資料における自由記述欄の再コーディング過程」東京大学社会科学研究所 社会調査・データアーカイブセンター『社会科学研究所所蔵「労働調査資料」の二次分析』東京大学社会科学研究所.
- 岩井紀子・保田時男, 2007, 『調査データ分析の基礎——JGSS データとオンライン集計の活用』有斐閣.
- 岩田正美編, 1983, 『戦後日本の家計調査』法律文化社.
- 荻谷剛彦・石田浩・菅山真次, 2000, 『学校・職安と労働市場——戦後新規学卒市場の制度化過程』東京大学出版会.
- 児玉幸多編, 1993, 『くずし字解説事典』東京堂出版.
- 河野誠哉, 2012, 「教育社会学における歴史的方法」『よくわかる教育社会学』ミネルヴァ書房 188-9.
- 邱永漢, 1993, 「わが青春の台湾(3): 全学連の『種蒔く人』」『中央公論』108(9): 246-57.
- 森直人, 2005, 「個別歴史性に定位した社会移動研究の可能性: 佐藤(粒来)香『社会移動の歴史社会学』を読む」『社会科学研究』56(5-6): 211-34.
- 森永卓郎編, 2008, 『物価の文化史事典——明治・大正・昭和・平成』展望社.
- 日本社会福祉学会, 1958, 『日本の貧困——ボーダー・ライン階層の研究』有斐閣.
- 西村幸満・石田浩, 2001, 「JGSS-2000 調査(2000年11月)職業・産業コーディングインストラクション」SSJ Data Archive Research Paper Seris 18 (SSJDA-18), 東京大学社会科学研究所附属日本社会研究情報センター.
- 岡田あおい, 2000, 「近世農民社会における世帯構成のサイクル——二本松藩2ヵ村の史料を用いて」『社会学評論』51(1): 136-52.
- 小野芳彦, 2000, 「データベース化」文部省科学研究費創設的基礎研究『ユーラシア社会の人口・家族構造比較史研究』(平成7～11年度)最終実績報告書(研究代表者: 速水融).
- 労働調査論研究会, 2001, 『戦後日本の労働調査 [復刊第3刷]』東京大学出版会.
- Ruggles, Steven, 1995, "Family Interrelationships," *Historical Methods*, 28(1): 52-8.
- Ruggles, Steven, J. David Hacker and Matthew Sobek, 1995, "General Design of the Integrated Public Use Microdata Series," *Historical Methods*, 28(1): 33-9.
- Ruggles, Steven and Russell R. Menard, 1995, "The Minnesota Historical Census Projects," *Historical Methods*, 28(1): 6-10.
- 佐藤博樹・池田謙一・石田浩編, 2000, 『社会調査の公開データ——2次分析への招待』東京大学出版会.
- 佐藤香, 2004, 『社会移動の歴史社会学——生業/職業/学校』東洋館出版社.
- , 2006, 「方法としての計量歴史社会学——階層・移動研究を中心として」『社会科学研究』57(3/4): 5-18.

- , 2009, 「歴史的資料としての社会調査データ」『年報社会学論集』22: 22-31.
- 佐藤健二, 2011, 『社会調査史のリテラシー——方法を読む社会学的想像力』新曜社.
- Smith, Emma, 2006, *Using Secondary Data in Education and Social Research*, Maidenhead: Open University Press.
- Sobek, Matthew, 1995, “The Comparability of Occupations and the Generation of Income Scores,” *Historical Methods*, 28(1): 47-51.
- Sobek, Matthew and Lisa Dillon, 1995, “Interpreting Work: Classifying Occupations in the Public Use Microdata Samples,” *Historical Methods*, 28(1): 70-3.
- 菅山真次, 2011, 『「就社」社会の誕生——ホワイトカラーからブルーカラーへ』名古屋大学出版会.
- 武石典史, 2012, 『近代東京の私立中学校——上京と立身出世の社会史』ミネルヴァ書房.
- 田辺俊介・相澤真一, 2008, 『職業・産業コーディングマニュアルと作業記録』東京大学社会科学研究所.
- 田沼肇, 1946, 「壕舎生活者の動向」, 東京帝国大学社会科学研究会 『起ちあがる人々』学生書房.
- 東京帝国大学社会科学研究会, 1946, 『起ちあがる人々』学生書房.
- 氏原正治郎, 1966, 『日本労働問題研究』東京大学出版会.
- 氏原正治郎・江口英一・高梨昌・関谷耕一, 1959, 「都市における貧困層の分布と形成に関する一資料（二）」『社会科学研究』11(2): 93-129.
- 山口一男, 2004, 「パネルデータの長所とその分析方法——常識の誤りについて」『家計経済研究』(62): 50-8.
- 保田時男, 2006, 「反復横断調査としてのJGSSの意義——調査間隔の短さがもつ利点」『社会学年誌』(47): 19-33.
- 吉田文・広田照幸編, 2004, 『職業と選抜の歴史社会学——国鉄と社会諸階層』世織書房.

(あいざわ しんいち、中京大学、s-aizawa@sass.chukyo-u.ac.jp)
(こやま ゆたか、日本学術振興会、cymytc55@gmail.com)
(ちょん かをる、東京大学大学院、chung.ka.wol@iii.u-tokyo.ac.jp)
(査読者、稲田雅也、土屋敦)

Critical Reconstruction of Social Survey Data for Historical Sociology

AIZAWA, Shinichi / KOYAMA, Yutaka / CHUNG, Kawol

In this paper, we will examine the theory and methodology to revive microdata samples of social survey conducted before the development of information technology. Microdata are unmistakably the most important source for quantitative empirical research on contemporary societies as well as previous societies. In the field of historical demography, many sources, from the Parish Register to the National Census in the 19th century, have been found and digitally reconstructed, but sociology has paid little attention to such historical data. In the libraries and archives of social sciences, however, there remain various types of microdata samples that have not been digitized and hence never fully exploited despite their value for historical-sociological secondary analysis.

Since 2010, we have conducted the digital reconstruction of microdata samples in the Documents of Japanese Labor Surveys after World War II, which were collected at the Library of Institute of Social Science of the University of Tokyo. Through this project, we have formulated basic processes and principles of reconstruction of old social survey microdata for effective use in scientific research. The restoration needs to be carried out in six steps: (1) sample collection, (2) material and digital preservation of the sample, (3) construction of the basic design of data entry and (re)coding, (4) data entry and coding, (5) cleansing of the entered data, and (6) file creation. In this process, especially at the fourth step, four principles have to be observed: (a) accurate representation of scientifically relevant information in the sample, (b) maintenance of comparability with other data, (c) accomplishment of correctness and effectiveness at data entry, and (d) complete anonymity to protect private information of each case in the sample.

These principles have opposite requirements (such as scientific accuracy and protection of privacy, and historical restoration and comparative data coding), and thus, problems can arise in the reconstruction process. We highlight the typical patterns of such problems and suggest possible ways to solve them through many concrete examples from our own experiences. The theory and methodology demonstrated through these examinations will provide a definite foundation for further discussion on the critique of sources in quantitative historical sociology.