

社会調査・データアーカイブ共同利用・共同研究プロジェクト

「国鉄労働組合婦人部実態調査」（1952年）の復元過程
—「データセットの整備」と「資料保存」は両立しうるか—

前田一步（聖心女子大学）

堀江和正（東京大学）

瀬戸健太郎（立教大学）

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

2024年（令和6年）8月

「国鉄労働組合婦人部実態調査」(1952年)の復元過程 ——「データセットの整備」と「資料保存」は両立しうるか——

前田一歩 (聖心女子大学)

堀江和正 (東京大学)

瀬戸健太郎 (立教大学)

1 はじめに

1.1 本稿の目的

本稿では、東京大学社会科学研究所が所蔵する「国鉄労働組合婦人部実態調査」の原票をアーカイビングし、計量的な手法による二次分析が可能なデータセットを作成する過程を論じる。本稿の主要な目的は、(1)「国鉄労働組合婦人部実態調査」のデータセットがいかにして整備されたのかを記録することで二次分析における資料批判の役割を果たすこと、(2)社会調査データをアーカイビングするための具体的な方法論を共有すること、の2点である。

本稿のプロジェクトは、東京大学社会科学研究所 課題公募型二次分析研究会「戦後福祉国家成立期の福祉・教育・生活をめぐる調査データの二次分析」の活動として行われている。本研究会は、相澤真一が佐藤香、橋本健二、森直人、渡邊大輔、石島健太郎らとともに2010年から現在まで、社会科学研究所が所蔵する「労働調査資料」を復元・分析することを目的にして活動してきた。本プロジェクトは、2021年12月に著者の前田・堀江・瀬戸らを中心にして「国鉄労働組合婦人部実態調査」のデータ化に向けて開始されたものである。なお「国鉄労働組合婦人部実態調査」は2024年8月現在SSJデータアーカイブへの寄託に向けたデータクリーニングを行っており、本稿では仮データセット完成までの過程について述べる。

1.1.1 計量歴史社会学の研究動向

日本において計量歴史社会的な研究は、現代の社会学研究を構成してきた歴史と数理というふたつの方法を出合わせる試みとして展開されてきた。この動向の過去と未来を跡づけようとする取り組みのひとつに『理論と方法』(23号, 1998年)で組まれた「特集 計量歴史社会学」の存在を指摘することができる。

ここでは説明における社会的な論理の投入と、社会学が蓄積してきたデータ処理手法の投入という、従来計量社会学が行ってきたことを歴史研究に適用する(佐藤俊樹 1998:2)立場や、計量的歴史社会学の分析は「技法によって分析の優劣が決まるのではなく、基本的には資料の質、資料と技法との上手な組み合わせ、そして何よりも、それを支える『歴史解釈』によってその分析の優劣は決まる」(蘭・中里 1998:52)という前提が確認された。

その後、社会調査データの二次分析をもとに、過去のことを解き明かそうとする計量社会学研究が蓄積されている。たとえばSSM調査を用いた戦前・戦後の産業にかんする歴史的研究(岩井 2018)や、同様にSSM調査の二次分析によって兵役の有無が職業経歴に与えた影響を明らかにしようとした渡邊勉の研究(2020)が挙げられる。

こうした計量歴史社会的な研究の蓄積のなかでは、社会調査データの発掘と復元(データ整備)それ自体をより重視する立場もある。これらの研究実践は、デジタルデータ化されないまま保管されている社会調査原票を発掘し、自らの手によってデータ化することで、二次分析研究の土台となるアーカイブの充実を図ると同時に、それを通じて過去のことを解き明かそうとする試みとしてまとめられるだろう。

そうした実践のひとつの例としては、大阪商業大学JGSSセンターによる「ビジネスエリート調査」(1960年/1970年)の復元およびデータ公開の試み(貫田ほか 2018)が挙げられる。そして、東京大学社会科学研究所が実施した「労働調査資料」と呼ばれる一連の社会調査データを復元し、分析する試みが、荻谷剛彦ほか(2000)、相澤真一ほか(2013)、佐藤香ほか(2015)を中心にして行われてきた。その過程で整備されたデータの多くが、現在ではSSJデータアーカイブに収録され二次分析が可能な状態になっている。また、復元したデータを分析した成果(橋本ほか 2013; 橋本編 2015; 渡邊ほか編 2019)もさまざま刊行されてきた。本稿で示す実践もまた、これらの試みに連なるものである¹。

1.1.2 「労働調査資料」のなかの「国鉄労働組合婦人部実態調査」

東京大学社会科学研究所では、戦後直後から1960年代終わりまでの約25年間、経済学・法学・社会学の研究者たちが、戦後日本社会における労働・社会問題を対象にする一連の労働社会調査を実施してきた。その調査実践は第I期(1945年-1950年)、第II期(1950年-1955年)、第III期(1955年-1968年)に分けられ、それぞれの時期区分のなかでさらにテーマ別に区分される(労働調査研究会編 1970: 22-3)。労働調査が扱うテーマは、基本的には(A)労働組合研究、(B)労働市場研究、(C)貧困研究、(D)中小企業および地域研究の4種類とされ、65の調査が実施された(労働調査研究会編 1970: 23)。

¹ データの発掘・復元に比重を置く計量歴史社会学が蓄積してきた技法を知るにあたり、データの復元にあたった社会学者たちが、そこで用いた技術やその過程を提示することを主要な目的として執筆された諸論考が大きな手がかりになる。先述の東京大学社会科学研究所に所蔵される「『ボーダー・ライン層』調査」「貧困層の形成(静岡)調査」「京浜工業地帯調査(従業員個人調査、住宅調査)」「都内壕舎生活者調査」の復元過程を記録した相澤ほか(2013)の論文や、「ビジネスエリート調査」のデータ整備および集計用データの公開過程を残した貫田優子ほか(2018)、そして生活時間データを含む「団地居住者実態調査」の復元についての渡邊大輔ほか(2019)の論考がそれにあたる。さらに、社会調査データではないものの、五十嵐彰(2020)は『日本統計帝国年鑑』を電子化した過程を記録している。本稿もこれらの論考と同じように、過去の歴史的資料、とりわけ社会調査データを電子化し、データ化することで二次分析可能な状態にするための過程、およびその作業に付随する問題への対応策を記録したものである。

本稿は、そのうち(A)労働組合研究に分類される、「国鉄労働組合婦人部実態調査²⁾」(1952年、労働調査資料 No. 24)の復元過程について記録するものである。とくに本調査の復元にかかわる一連の作業過程を記録することで、過去の社会調査を発掘・復元し、計量的な二次分析が可能なデータセットを整備する方法を共有することを目指す。

1.2 「国鉄労働組合婦人部実態調査」の紹介

1.2.1 調査の概要

本項では復元対象の「国鉄労働組合婦人部実態調査」の概要を述べる。

「国鉄労働組合婦人部実態調査」は、東京大学社会科学研究所講師で労働法学者の藤田若雄と国鉄労働組合婦人部長の丸沢美千代によって、1952年5月に実施された調査である。公式責任者は藤田若雄であり、藤田が国鉄労働組合婦人部による依頼を受けて調査を実施した。対象となったのは当時の国鉄女性職員の全員³⁾、約11,000人である。自記式の調査票を組合組織を通して配布し、組合本部婦人部への直接の郵送、または分会・支部・地方本部を通しての回収を求めた。研究資金の出所は国労婦人部調査費である。『戦後日本の労働調査』によると6,820票が回収され、推定回収率は60%強であった(労働調査論研究会編 1970: 155)。筆者らの復元作業の結果、表1のとおり、全国27の地方本部から6,815票を有効票として入力した⁴⁾。

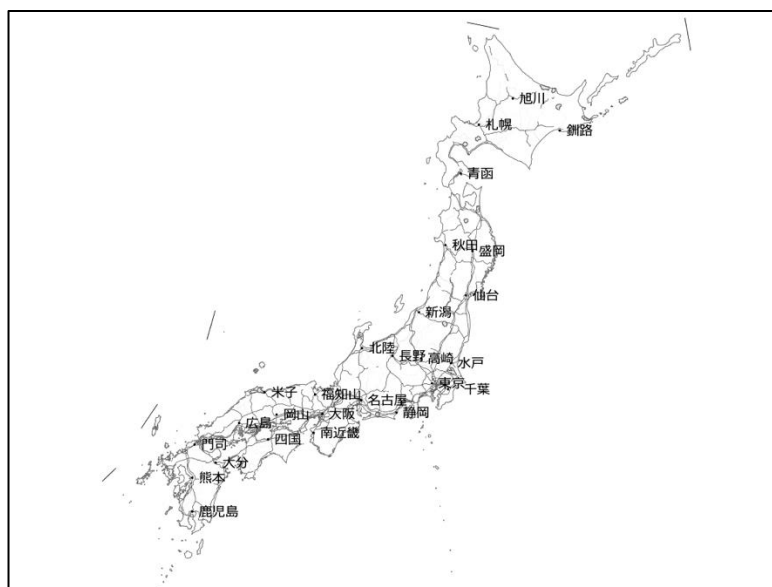
²⁾ 「国鉄労働組合婦人部実態調査」は、調査票原票の上部に記載された調査名称であり、本稿ではこの表記を採用している。なお『戦後日本の労働調査』(労働調査論研究会編 1970)では「国鉄女子労働者調査」(No. 24)と呼称されているほか、調査報告書(国鉄労働組合 1954)では「国鉄女子職員実態調査」の呼称もみられる。

³⁾ 厳密には、国鉄労働組合婦人部の組合員を対象としている。

⁴⁾ 『戦後日本の労働調査』は、調査報告書(国鉄労働組合 1954)の集計表を参照して回収数を6,820としたと考えられる。この数と、本プロジェクトの有効票数に5票の差が生じたことは、復元作業において重複票や白紙票等を入力しなかったことに起因すると考えられる。

図表 1-1 「国鉄労働組合婦人部実態調査」の対象となった 27 の地方本部⁵

| 拠点名 | 票数 | 拠点名 | 票数 | 拠点名 | 票数 |
|------|------|-------|-----|-------|------|
| A_旭川 | 105 | J_千葉 | 127 | S_岡山 | 185 |
| B_釧路 | 74 | K_新潟 | 195 | T_米子 | 138 |
| C_札幌 | 337 | L_北陸 | 150 | U_広島 | 673 |
| D_青函 | 172 | M_長野 | 176 | V_四国 | 282 |
| E_盛岡 | 134 | N_静岡 | 242 | W_門司 | 599 |
| F_仙台 | 469 | O_名古屋 | 419 | X_大分 | 70 |
| G_高崎 | 163 | P_大阪 | 532 | Y_熊本 | 73 |
| H_水戸 | 64 | Q_南近畿 | 158 | Z_鹿児島 | 111 |
| I_東京 | 1158 | R_福知山 | 3 | AA_秋田 | 6 |
| | | | | 合計 | 6815 |



図表 1-2 「国鉄労働組合婦人部実態調査」の対象となった 27 の地方本部一覧

調査項目は基本属性、就業状況、採用経緯、職歴、通勤経路、同居人（続柄、職業、収入）、収入使途、住居、家族からの援助、小卒時実家職業、職場環境（いやがらせ、辞職・配置転換勧告、私的雑用、超過勤務等）、組合・婦人部への認知・希望、政党支持からなる。原票

⁵ 後述するとおり、復元作業においてそれぞれの地方本部には北から順番にアルファベットを割り当てている。アルファベットの割り当ては下記の通りである。A_旭川、B_釧路、C_札幌、D_青函、E_盛岡、F_仙台、G_高崎、H_水戸、I_東京、J_千葉、K_新潟、L_北陸、M_長野、N_静岡、O_名古屋、P_大阪、Q_南近畿、R_福知山、S_岡山、T_米子、U_広島、V_四国、W_門司、X_大分、Y_熊本、Z_鹿児島、AA_秋田。秋田のみ保管時に除外票扱いされていたが、本復元プロジェクトでは有効票として扱っている。

は本稿末尾に掲載している。調査結果は国鉄労働組合（1954）として発表された。

1.2.2 調査当時の国鉄女性職員の置かれた状況

本項では、1952年の調査当時、国鉄に勤める女性職員が置かれた状況について確認する。国鉄の女性労働者の歴史は、第二次世界大戦を重大な契機として語られることが多い。がんらい労働集約的な特徴をもっていた国鉄は、事業の継続に大量の労働力投入が必要であるにもかかわらず、戦時期において多くの熟練労働者が軍隊の兵員、および占領地鉄道の運営要員として流出していった（林 2010: 46）。こうした状況の戦時下において、出征した男性に代わる労働力供給源として女性労働者が急増したのである。女学生の勤労働員も相まって、1943年には14,702人（3.8%）だった女性労働者が、1年後の1944年3月には49,757人（11.9%）、そして1945年3月には102,827人（22%）まで増加した（林 2010: 50）。

しかし戦後日本政府は、復員者や外地からの引揚げにより大量に増加した男性失業者対策として、女性労働者を解雇する方針をとり、1949年の労働統計局⁶推定によると失業した女性の数は250万人とされている（豊田 2007: 61）。これは女性が戦時中⁷についた職を男性に明け渡し、「女性は家庭に復帰させれば良い」とする政府の方策によるものであった（豊田 2007: 62）。日本政府は1946年に7.6万人の整理解雇を国鉄労働組合総連合会（国鉄総連）に通告したが、国鉄総連のストライキ予告によりこの整理解雇は回避されたとされる⁷。しかしながら実際には、この間も女性労働者の解雇は行われていた。そして、1949年に「行政機関職員定員法」が制定され、国鉄からは約9.4万人の整理解雇が割り当てられ、女性職員のうち約40%が解雇されたとされる（丸沢 1981b: 11）。

こうした状況に対して、1947年に国鉄労働組合婦人部（以下、国労婦人部）が結成され、成立直後だった労働基準法における女性労働者の制限（深夜残業・危険業務等の禁止、生理休暇⁸）と、それにもとづく解雇の可能性についての議論が展開された。

その中で1948年に国労婦人部長に選出された丸沢美千代は、同年に「国鉄女子職員実態調査」を実施し、つづいて1952年に本復元の対象となる「国鉄労働組合婦人部実態調査」を実施した。この過程については、つづく1.2.3で詳述する。

1.2.3 調査が実施された理由・調査体制

調査の実施においては、当時の国労婦人部長・丸沢美千代の役割が大きい。1948年8月から婦人部長を務めていた丸沢は、当時の問題意識を以下のように振り返っている。

⁶ 労働省労働統計調査局の誤りと思われる。

⁷ この当時、国家公務員や公共企業体の職員の労働基本権を制限するマッカーサー書簡を受けた1948年改正国家公務員法や公共企業体等労働関係法がなく、公務員も争議権を行使することができた。

⁸ この点は豊田（2007: 第3章）に詳しい。

私が最も懸念したのは、婦人部長として誤った指導をすることであったから、何よりも先ず、国鉄内の多岐にわたる現場労働者の実態を知りたかった。そのためには、全国各地の要請に応じてオルグに廻ったり、各地から訪れる婦人部員に聞くだけでは不十分であった。私は、広範囲な調査によって、実態とその構造・問題点とその本質を解明し、それに基づいて、婦人部の自主的な討論を経て婦人部の方針を立てることと、そのために、労働講座をもって意識を高めることを願いつづけていた。（丸沢 1981a: 172-3）

こうした問題意識のもと、丸沢は1948年にすでに国労婦人部独自の「国鉄女子職員自体調査」を実施し、国鉄当局・大蔵省との男女賃金格差解消のための交渉に活用したというが、それは「お粗末で、平面的な数字だけで、構造的には問題が出てこない」（丸沢 1981b: 17）ものであった。そのため、丸沢は国鉄労組で労働講座の講師を務めていた藤田若雄に調査協力を求めた。調査設計から実査・集計の過程は以下のようなものであった。

年中、斗争と、オルグ要請に応じて私は全国を駆け廻っていましたので、その間を縫って、知りたいことを欲ばって調査項目に挙げて、学者であられる先生と実践者の私の意見を出し合って、調査目的、項目実施方法等を決めて調査にかかりました。念願の「国鉄婦人労働調査報告」は発行まで二年を要しました。先生も極めて多忙であられ、一方私も昼間は組合業務に追われて夜間しか時間がとれませんので、調査票、結果表、関連表等を作るのに深夜までかかることがしばしばでした。集計は内閣統計局に依頼しました。分析をお願いしました先生は、非常な熱意をもたれ、私とは比較にならないご苦労がありがたかったと思います。私の任期は一年ですから、折角実施したこの調査の本報告はできなくても、中間報告だけは出さなくてはと思い、中間報告を二回出し、一九五四年にこの本報告を出すことができました。（丸沢 1981b: 17）

なお、調査票の表面にある、回答者の基本属性、就業状況、採用経緯、職歴、通勤経路、同居人（続柄、職業、収入）、収入使途、住居、家族からの援助、小卒時実家職業といった事項は、「京浜工業地帯調査（従業員個人調査）」（1951年9月実施・労働調査資料 No.18）の質問項目が流用されている。藤田は「婦人部が調査をやってくれということで、それで『京浜工業地帯の労働市場調査票』をもってきてそれに類する調査をやったりしました」（労働調査論研究会 1968: 124）と述べている。「京浜工業地帯調査（従業員個人調査）」と「国鉄労働組合婦人部実態調査」を接続する形で二次分析を行うことが可能であるという意味で、注目すべき点である。

1.2.4 当時の報告書等での論点

報告書には、以下のように調査目的が記されている。

等しく国鉄女子労働者といっても、その場にはいくつかの類型が存在し、そして、それは異った要求をもっており、その要求の解決の条件も異っている。そこで、この調査の意図するところは、組合運動の眞に組合員大衆のものとするためには、この類型を検出し、その要求の理論的な構造を明らかにし、その合理的解決の条件を検討するということであった。かような調査が労働組合にとって不必要なはずはない。なぜなら組合役員はこのような調査結果によって有効に幹部組合から大衆的組合へ自己脱皮をなしとげることができ、組合員の諸要求を運動の中に組織することができるからである。(国鉄労働組合 1954: はしがき)

調査結果について、報告書(国鉄労働組合 1954)はクロス集計とその解説という形でまとめられている。報告書の本文中で行われたクロス集計 26 点のうち、21 点は職種と各設問への回答の関係を見たものであった。本文中の集計に加えて、巻末の付表として、ほとんど全ての設問への回答実数を職名別にみたクロス表(25 点)が掲載されている。当時の集計・分析は、国鉄女性労働者の社会経済的位置や労働環境、労働組合への意識等が、職種⁹によっていかに異なっているか/類似しているかという点に集中している。

前述の調査目的にあるように、国鉄の女性労働者の「類型」の析出が目指されており、その類型としては基本的に職種が想定されていたことがわかる。また、報告書によれば集計費の不足が「調査票の関連項目の完全集計を不可能にし、基本項目の集計の使用を大半不可能にしてしまった」(国鉄労働組合 1954: はしがき)という。このように、当時の関心と調査環境の問題から、職種以外に着目した分析はほとんど未着手であった。

藤田は国労機関誌『国鉄文化』に分析の一部を発表している。そこで注目されていたのは、(1) 職場に問題が起きた時の相談先、および(2) 生家と入職経路であり、いずれも地方本部ごとのクロス表が作成されている。(1) は、「わが国の場合においては、従業員は個人的な世話まで組合委員にもち込みたい傾向^マいまだ脱却していない。だから、これを個人的な問題だからといって放っておけば上役に掌握される。そこで組合委員は組合員の個人的問題でも解決をはかってやる必要がある。ただその場合、そのことが当然に傾向^マずける、親分的、或は人身支配的な傾向を常に自己からおしのける努力を忘れてはならない」(藤田 1953a: 39)という実践的インプリケーションにつながる。(2) については、生家(=農業)と、入職経路(=縁故)が、「いずれも組合運動にマイナスの影響をあたえるものであるが故に、かような要素の多い職場では、特にその影響を最小限に喰い止めるため平常から組合の世話役活動を重要視して組合員掌握に努めなければならない」(藤田 1953b: 40)という問

⁹ 本文中のクロス集計では「電話掛」、「看護婦」、「技工 整備掛」、「事務補助」、「駅労働補助」、「其他」の分類が用いられている。巻末の付表では「課員(1 級/2 級/3 級/記入なし)」、「電話掛(電務区/駅/記入なし)」、「看護婦(病院/診療所/記入なし)」、「技工」、「整備掛」、「踏切警手」、「出札掛」、「事務掛」、「雑務手」、「駅手」、「駅務掛」、「その他」と、より細かな分類が用いられている。

題意識が背景にあった。なお藤田の論文の内容は報告書（国鉄労働組合 1954）にも収録されており、報告書作成に藤田が大きく寄与したことがうかがえる。

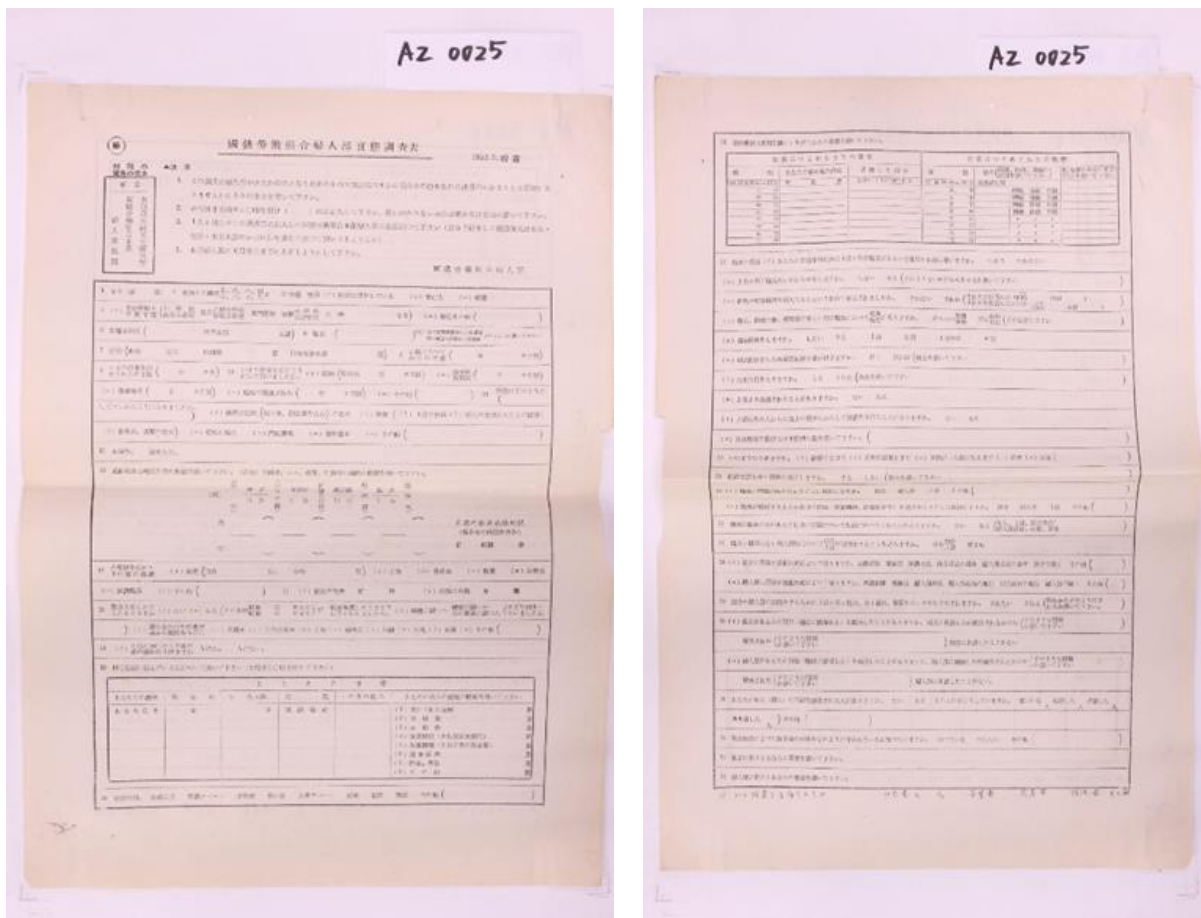
2 「国鉄労働組合婦人部実態調査」の復元の全体像

本章では、「国鉄労働組合婦人部実態調査」の復元プロセスの全体像について紹介する。計量歴史社会学におけるデータ復元の目標の一つは、過去になされた調査を後世の研究者が二次分析をすることが可能な形に整備することである。この目標からは2つの要求が導き出される。一つは、SSJデータアーカイブのような機関にデータを寄託できるレベルでの復元を行うことである。もう一つは、調査票の原回答を極力そのまま入力しておくことである。しかし、これには本章以降でもたびたび言及するように、かなりの困難を伴う。例えば、統計ソフトウェアで分析可能なデータとするためには一定程度の情報の縮約が必要であり、場合によっては第一の要求と第二の要求との間の「ジレンマ」に遭遇することもしばしばである。あえて言うならば、計量歴史社会学における復元とは、「計量分析可能なデータセットの整備」と「過去の歴史資料の保存」という、両立可能に見えて相当なジレンマを発生させる2つの要求を意識しつつ、その「折り合い」をどこでつけるか、ということが不可避免的に求められるプロジェクトなのである。このような視点からすれば、その「折り合い」を専門的見地からつけるためには、復元プロジェクトには可能な限り、社会調査経験のある研究者と、歴史社会学も含めた歴史研究者の参加が望ましいだろう¹⁰。

ただ本プロジェクトでは、「過去の歴史資料の保存」という第二の目的を、可能な限り優先させる方針を採用した。これは2つの理由による。第一に、復元に携わる研究者は当該調査の設計者ではなく、究極的な意味での「調査にあたっての仮説」や「調査対象者に対する理解」、「当時の文脈」を共有しないためである。復元に携わる研究者の問題意識は、当時の研究者のそれと一致することはありえず、一見「エラー」と見えるような回答も、当時の研究者には有意味なものだったかもしれないのである。第二に、二次分析可能なデータセットの整備にあたって調査原票の「複雑性」を極力保っておくことで、今後二次分析にあたる研究者にかかる「自分で確かめることのできない制約」を最小化できると判断したためである。

¹⁰ この点、我々が把握している計量歴史社会学における取り組みでは、必ずしも歴史研究者の参加が活発に行われているとは言い難い。本プロジェクトの場合、瀬戸が調査実査と実証分析を、前田が歴史社会学と実証分析を、堀江が歴史社会学と調査実査を経験していたが、労働史に関する専門家はいなかった。マニュアル作成や個別の質疑にあたっては、労働史や教育史に関する知識が必要になるケースが存在したが、特に複雑な回答に対しては、必ずしも適切な判断がくだせたわけではないと推測している。このことはデータの質に関する重大な問題をもたらすのが、そのような批判を甘んじて受け入れる意味でも明記しておく。

2.1 調査票の保存状態



図表 2-1 「国鉄労働組合婦人部実態調査」調査原票

1) 右上の英数字は復元プロジェクトで付与した仮コード

調査票原票の確認は 2021 年 12 月に行われた。東京大学社会科学研究所所蔵の「労働調査資料」は SSJ データアーカイブで専用の保管庫で保存されていた。「国鉄労働組合婦人部実態調査」は集計表もふくめて合計 4 箱の保存用の箱に格納されていた。相澤ほか (2013) が復元対象とした『『ボーダー・ライン層』調査』などと同様、安価な用紙が使われたと思われ、両面印刷された調査票の劣化は著しかった。また、手作業での集計記録と思いき、赤鉛筆でのエディティングのあとが残されており、これは入力時において判断割れを起こすエラー要因の一つとなりえた (対応は 4.2 で後述)。

2.2 復元の作業工程の設計方針

本プロジェクトでの復元の作業工程の設計方針について述べる。前述のように、「過去の歴史資料の保存」を最優先にした作業工程を採用することにした。作業工程の設計は、一般

的な社会調査を念頭におきつつ、相澤ほか(2013)を参考に行った。その全体を概観すれば「(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」(相澤ほか 2013: 67) となる。

相澤ほか(2013)では、具体的にどのような作業が実施されたのか、ということは比較的詳細に記述されており、工程ごとの大方針についても述べられているものの、具体的な作業工程に落とし込んでいく際の留意点や、作業工程ごとの関連構造の分析、具体的な作業手順の設計など、いわば工程設計については十分に述べられていない¹¹。そのため、一つひとつの工程ではなく、工程の連鎖としての作業フローについては、本プロジェクトにおいてあらためて設計する必要があった。本項ではこの点についての知見を共有するため、各工程の連鎖としての作業フローの設計方針について詳述することにする。

設計方針の第一は、作業工程を極力、細分化することである。このメリットは2つある。それは、(1) 社会調査の基礎科目を受講した程度の学部学生でも作業を遂行できることと、(2) 一つ一つの工程の工数計算を簡易にできるようになったことである。工数計算が容易であるために、進捗管理も可能になり、標準進捗から作業が遅れている場合に、追加の人員投入が必要か否か、また、後続作業の準備作業にいつから着手すればよいかなど、作業計画の立案も非常に容易になった。

しかし、計画を立案することは可能であっても、どのように作業工程を管理するかまで作り込まなくては、計画は往々にして画餅となってしまふ。そこで、今回の復元プロジェクトでは、WBS (Work Breakdown Structure) とガントチャートを作成して、作業工程管理を行うことにした。WBS とは以下のようなものである。

WBS は、プロジェクト全体を詳細な作業に分割する手法です。プロジェクトのゴール(開発すべき製品)を、細かな成果物に階層的に詳細化し、それぞれの成果物を開発するために必要な作業を配置していきます。WBS を作成するときには、特にプロジェクトの重要な要件について、より具体的に、成果がイメージできるまで詳細化していきます。

WBS を作成することによって、プロジェクトの目標を達成するための詳細な作業を洗い出すことができます。加えて、作成したWBS をプロジェクトに関わる人々(スポンサー、マーケット担当、プロジェクトメンバなど)によって事前に確認することで、プロジェクトのスコープ(範囲)を明確にでき、計画漏れによる後戻りを防ぐことがで

¹¹ もちろん、これは相澤ほか(2013)の実践が不十分であることを意味しない。相澤らが直面した困難の詳細は佐藤香ほか(2015)でも言及されているが、作業工程がしばしば変更され、予定通りの作業が行えないこと(「ジャングル探検」)は今回の作業でも生じた。つまり、事前設計が十分に機能しない、ということが予想されるのである。他方で、各工程で生じる異常や、それへの可能な対応の幅は、設計方針に依存することになる。

きます。また、プロジェクトを通じて WBS を管理することによって、要件の変更によるプロジェクトへの影響を管理することが可能になります。（独立行政法人 情報処理推進機構 2005: 25-7）

本プロジェクトでもこのような WBS とガントチャート（図表 2-2）を作成した。その詳細な方法は 6 章で論じるが、「誰」を「どのような工程」に投入しており、作業が「どのぐらいの期間」かかっており、その間に「どのようなリソースが空いている」かなど、作業工程ごとに必要なリソースとプロジェクト全体で保有しているリソースを可視化し、人的リソースを柔軟に最適配置できるような工夫を行うことで、作業工程管理を行った。この結果、作業者の作業重複や、指示漏れによる手持ち時間の増加といった、稼働ロスを可能な限り、小さくすることに成功したと言ってもよいだろう。同時に、大まかではあるが仕事表を作成して、人員のスキルと担当できる業務の範囲について確認し、リソース配置の最適化に活用した。仕事表とは、小池（2005: 38）でその例が挙げられているが、要員ごとの担当できる仕事の幅を記録したものである。人員一人ひとりの仕事の幅について把握しておくことで、突発的な作業が発生したときも、その作業に対処可能なスキルを持っている作業者と、そうではない作業者とを識別できるため、全体工程への影響の少ない作業組み換えが可能になった¹²。



図表 2-2 作業工程のガントチャート

¹² 大まかな仕事の幅としては、(1) カメラの機種別に単純な撮影作業が可能である、(2) カメラの機種別にチューニング作業も含めて一連の作業が可能である、(3) 調査票の基本的な入力作業が可能である、(4) 調査票について周辺情報を自分で検索した上で入力作業が可能である、(5) 複雑なケースについて一応の判定がくだせる入力作業が可能である、といった幅で確認するようにしていた。実際には、扱う機材によってもさらに細分化していたが、機材を同一メーカーで統一するなどすれば、多能工化を促進することは可能であると思われる。

設計方針の第二は、極力、マニュアル化できる作業レベルに落とし込むことである。これはいわば、作業者に対する作業手順の定義であるが、作業者と作業監督者双方に、「どこまででは自分で判断してよいか」「どこからは自分の判断を超えているか」といったことを示す、職務範囲の定義の機能も含まれる。この詳細は4章で論じられる。

設計方針の第三は、事前に作業の流れ自体はすべて計画しておくが、微妙な作業内容に関してはトライ・アンド・エラーにより修正をはかっていく、というものである¹³。第三の方針は、第一・第二の方針に一見反するようだが、導入には理由がある。一般的に社会調査であれば、原票配布以降の工程は、点検→エディティング→コーディング→入力作業と決まっているのが通例であろう。また、このような調査工程が前後することはほぼなく、エディティングガイドも事前に作り込んでおくことが多いように思われる。しかしながら、調査票の設計意図や、調査当時の調査者・回答者が共有していた文脈が復元開始時点で明確でない場合、完全なエディティングガイドを準備することはほぼ不可能であろう。つまり、入力作業においてあらかじめ全体を見渡すことが不可能な点に、一般的な社会調査との差がある¹⁴。それゆえ、本プロジェクトのエディティングガイドの作成においては、トライ・アンド・エラーにより頻繁な改訂を許容し、場合によっては作業者の裁量により改定案を作成することまで許容する、非常に柔軟な入力ルール管理を行うこととした¹⁵。

ただし、完全に動的にルールを管理すると、作業者ごとのばらつきが大きくなり、最終的なデータの質は著しく低下してしまうであろう。そこで、矛盾するようであるが、「全体の作業の流れは確定させておき」、「一つひとつの変数の入力値などについては、比較的柔軟に変更を行う」という管理方針をとることとした。つまり、「一つ一つの工程の中身を作り込む」のではなく、「全体の工程でいつ、どのように、何の作業を行うかを計画する」という視点で作業設計を行った。本プロジェクトにおける実際の作業に即して説明すれば、(1) 発

¹³ この方針は、情報システムの開発におけるウォーターフォールモデルとアジャイルモデルのハイブリッドモデルを採用したものと言える。いずれもシステム開発で利用される開発方法である。ウォーターフォールモデルとは、デジタル庁(2021)で説明されているが、計画・要件定義・設計・開発・テスト・移行の各段階を定義し、前段階の作業が完了しない限り、後段階の作業に着手できないシステム開発方法である。一般的な社会調査の工程の大半は、このウォーターフォールモデルに沿っていると言える。他方、アジャイルモデルとは、「工程を時系列に進め、原則として前工程の完了後に次工程を開始する情報システム構築作業の進め方である。設計・開発の途中で変更が少ないと見込まれる場合に用いる。工程を時系列で進めることから、計画が立てやすく、進捗の管理がしやすい」(デジタル庁 2021: 第7章 13)という開発工程のことである。

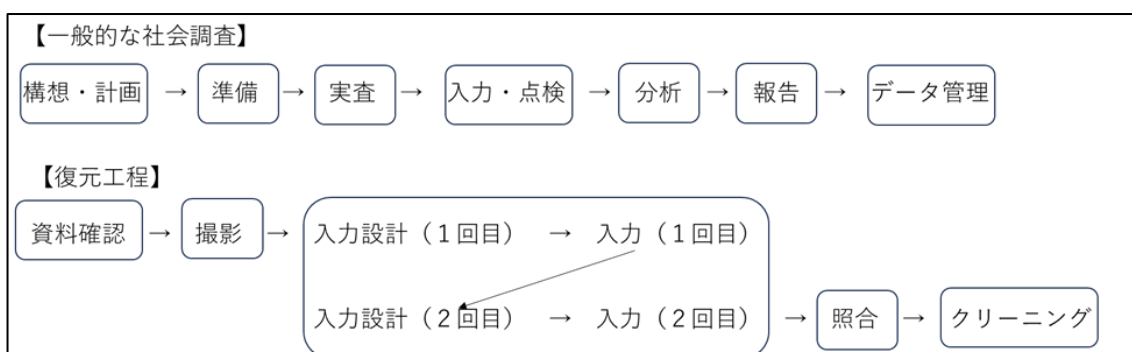
¹⁴ この点、現代の社会調査でも、特に自記式調査においては全票点検しないと予想外の記入がなされていることもあり、コーディングガイドといった入力マニュアルを事前に完全に作り込んでおくことは難しい。実際には実査が終わってからガイドが確定することが多く、その最終版もエディティング・アフターコーディングが完了したのちに作成される(轟・杉野編 2021: 137)。とはいえ、その程度は復元作業において非常に重くなる。

¹⁵ ただし、作業者の一存での判断はさせず、入力作業監督の承認を得て、マニュアル改訂を行う作業フローとした。

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という全工程の完成目標は決まっているため，それぞれの工程を完成させるのに何回の試行を繰り返せば，作業が最低限実施可能なレベルに実装することができるか，という目安を算定し，そのために必要な期間を計算した。その上で，(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には作業員 1 人あたり概ね 4 回程度のテスト，(4) 入力とコーディングは作業員 1 人あたり最低 3 日程度のトレーニング，といったように，工程の重要度と難易度を加味して試行回数を割り当てて，マニュアルや作業フローをその都度，再設計し，一応の完成度に達した時点でその工程の本番作業を始める，という流れである。

2.3 復元工程の全体像と一般的な量的調査との比較

本節では，以上で述べたような復元作業の作業工程と，一般的な社会調査の作業工程を比較する。実際の作業としては以下のようなものであった。



図表 2-3 復元作業と一般的な社会調査のフロー比較

両者の大きな違いを挙げるとすれば，入力・点検工程であろう。一般的な社会調査でも 2 回入力することはデータの整合性を保持する上で重要だが，復元工程の場合，1 回目入力と 2 回目入力とで入力設計が異なることがあるという点は大きく異なる。なぜ 2 回目で入力設計が変わりうるのか，設計変更にもかかわらずデータの整合性をどのようにして保つのか，という点については 5 章で説明するが，とくに重要なのは，「リジッドに予め設計を決めておくことができる余地はそれほど大きくない」ということを前提に，「どうすればデータの精度を保持することができるのか」ということを考慮しながら，作業設計をすることにあるだろう。

3 撮影作業

3 章では，調査票の画像撮影について論じる。当時の報告書や『戦後日本の労働調査』(労

働調査研究会編 1970) における記録では回数票数 6,820 となっているが、筆者らが確認した結果、それよりも多くの原票が東京大学社会科学研究所の書庫に保管されていた。当時除外票として扱われたと思われる調査票も保管されており、結果として報告書に記載されている数よりも多くの調査票が存在したのだと考えられる。撮影作業においては、これら除外票についても ID を付与して画像データを作成し、そのうち 6,815 票を有効票として入力対象とした。

3.1 撮影作業の準備

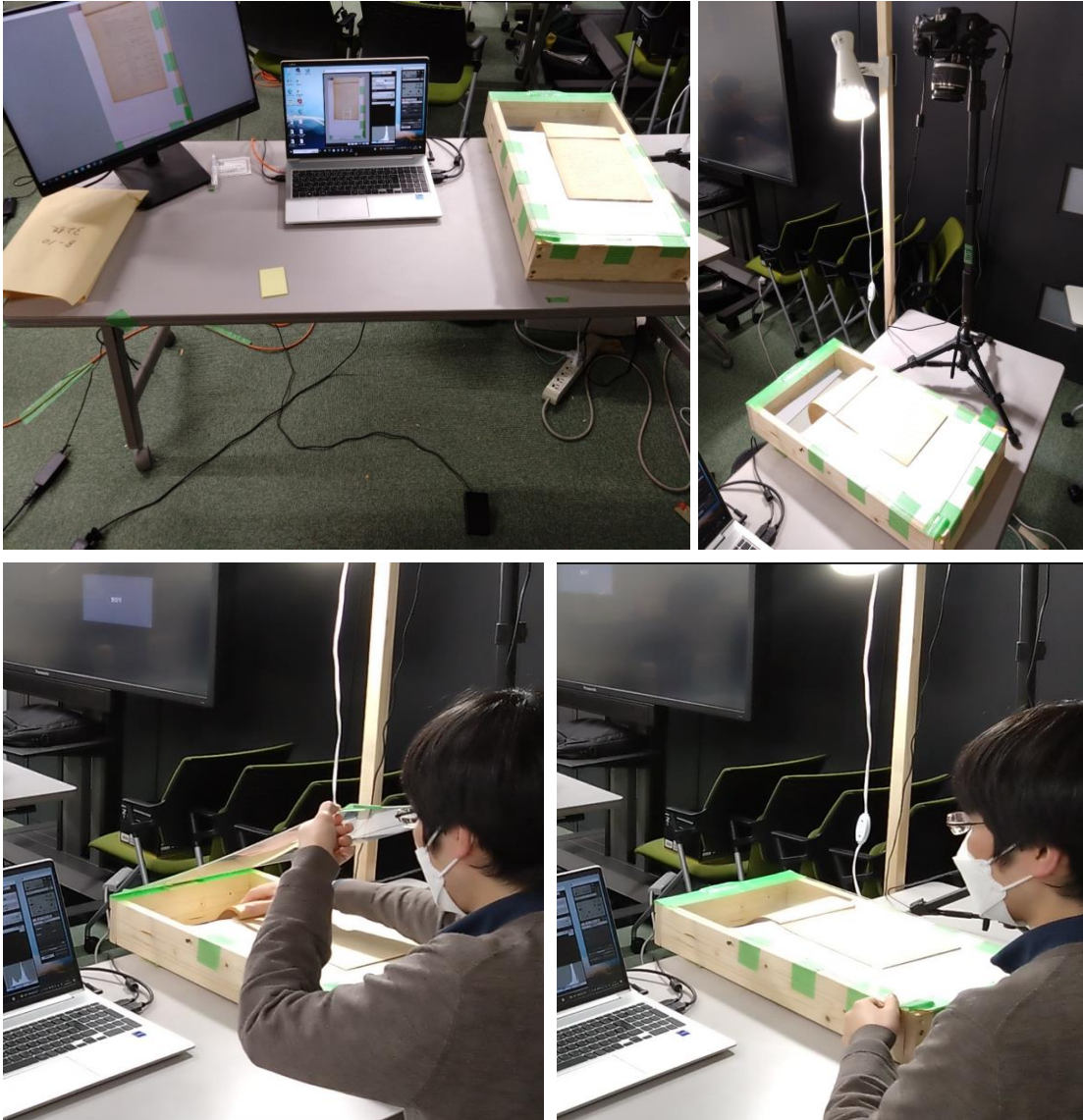
撮影作業の準備としては、ID の付与規則の作成と、撮影機材の設営の 2 種類の作業が存在する。まず ID の付与規則については、地方本部ごとにアルファベットを付与することとした。北から順に 27 の地方本部に ID となるアルファベットを付与し、つづいて個別の調査票を識別するための数値 4 桁を付与する。例えば地方本部コード「A」が割り当てられる旭川地方本部（105 票）の場合、最初の調査票は A0001、最後の調査票は A0105 となる。

撮影機材の設営については、おおまかに、デジタルカメラと、オーバーヘッド読取方式スキャナでの撮影可能性を試行し、2 種類の方法で併行して撮影を実施した。デジタルカメラで撮影を行う際には、三脚、デジタルカメラ、照明、画用紙（下敷き）、PC、カメラへの給電と PC への接続を行うコード類、ID カードが必要となる。オーバーヘッド読取方式スキャナを使用する場合は、スキャナ、PC、画用紙（下敷き）、ID カードが必要となる。

デジタルカメラの場合は、一眼レフカメラまたはミラーレス一眼カメラを使用し、作業は 2 人 1 組で実施した。1 名が PC のテザーソフトを通してシャッターの操作を行い、もう 1 名が調査票の設置・入替えをした。オーバーヘッド読取方式スキャナを使用する際には、撮影と調査票の設置を含めて、作業を 1 名で行った。

なお筆者らは本調査の復元に続き「神奈川県における社会福祉意識調査」（1964）、「東京 SSM 調査」（1960）の撮影を行い、複数ページからなる調査票に対応するとともに、1 名でも作業ができるよう撮影方法を改良した¹⁶。その際には、調査票を固定するための撮影台を作成し、また作業効率をさらに高めるために、テザー撮影ソフトの撮影キーを割り当てたフットスイッチと、画像確認用の PC モニターを導入した。フットスイッチを導入することで、1 名での撮影作業も可能になった（図表 3-1）。

¹⁶ 撮影方法の改良にあたっては、二次分析研究会メンバーであり国鉄門司鉄道局史料のデジタル化を進めていた菅沼明正氏（九州産業大学講師）の助言を得た。当該プロジェクトの詳細については九州産業大学ウェブサイトを参照（「旧国鉄史料『局報』のデジタル化に取り組んでいます」、https://www.kyusan-u.ac.jp/faculty/chiiki/news/dt_news_1128/）。



図表 3-1 調査票撮影の作業

3.2 撮影作業

撮影作業はテザーソフトを通して行う。そのため、調査票の画像撮影の工程の管理にあたっては、2種類の数字が存在することになる。写真撮影時にテザーソフトによって自動的に付与される画像のID（ファイル名）と、調査票に付与したIDである。本調査の調査票は表と裏の2枚の画像から構成されるため、画像IDを2で除すことで調査票IDを求めることができる。例えば旭川地方本部（地方本部コード=A）の場合、ID：A0001が与えられる調査票には、A0001/A0002の画像IDが与えられることになる。画像IDが奇数の場合はその画像データが調査票の表面であることを、画像IDが偶数の場合は裏面であることを示している。

調査票がすべて無回答である場合や、欠損している（調査票の印刷に不備がある）場合、

インク等で汚損されている場合¹⁷も、ID カードとともに画像を撮影した。あくまでも保存されている原票をそのままデジタル化することが、この作業工程の目的であり、調査票が無回答・欠損・汚損していることを、証拠として残そうとしたためである。

3.3 エラーチェックから印刷まで

撮影作業における最後の工程として、画像データのチェックがある。この工程で検出されるべきエラーは、大きく分けて3種類存在する。第1に画像自体が欠損していること、第2に写真の焦点がボケており、記入された文字が読めないこと、第3に調査票の折れ曲がりや作業者の指などによって、調査票の一部が隠れていることである。

目視によるエラーが検出された調査票に対して再撮影を行い、上述の画像データの命名規則にしたがったファイル名を付した上で、データの追加・差し替えを行う。エラーチェックと再撮影・画像データの差し替えを経て、画像の生データが完成することになる。これらの画像生データは、地方本部ごとにフォルダを分けて格納する。

最後に JPEG 形式で保存されているデータを PDF 形式に変換し、PDF 上で調査票 ID ごとに目次を付与し、BD 等の記録媒体に保管する。同時に画像データを、NAS およびクラウド上にアップロードすることで、PC の画面でも閲覧できるようにする（データ管理については 4.4 を参照）。

さらに、のちに続く入力作業に向けて、印刷する工程が存在する。本プロジェクトでは東京大学社会科学研究所に保管するための正本と、入力作業で使用する副本として、2部の印刷を行なった。印刷された調査票画像は、地方本部ごとにファイリングしたうえで保管・使用されることになる¹⁸。このように撮影データは、BD（保管用）、印刷しファイリングされた正本と副本、そしてクラウド上のデータと3種類準備することにより、在宅での入力作業を可能にする準備も整えた。

4 入力作業

4.1 作業の全体像と復元での特殊性

2章で論じた通り、本プロジェクトの目標は、データの SSJ データアーカイブへの寄託公開により多様な関心からの二次分析を可能にすることであった。原回答をできる限り損なうことなく入力しておくことは、将来的な分析可能性の豊かさにつながるという意味で重要である。いっぽうで、計量的に分析可能なデータとするためには、情報を一定程度縮約す

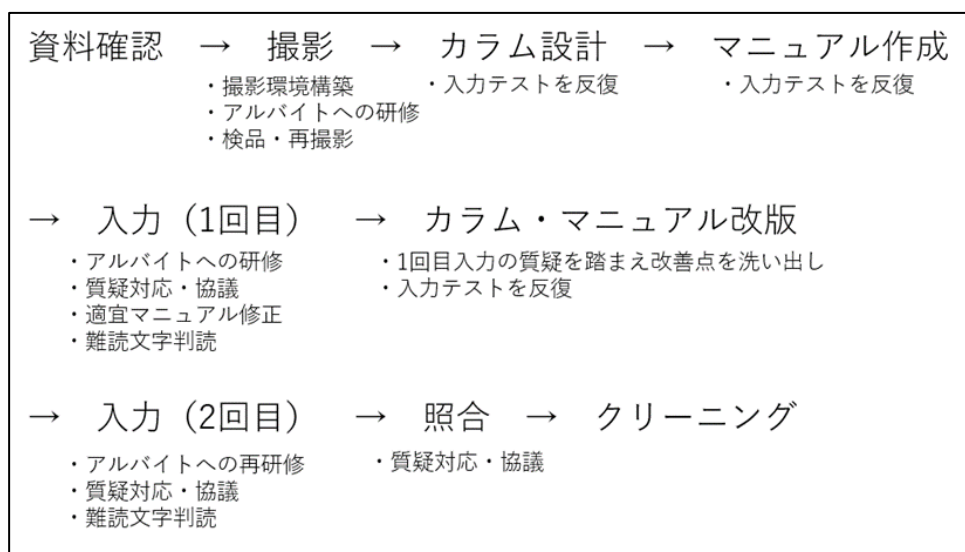
¹⁷ おそらく回答者以外の第三者によって「黒塗り」されたと思われる原票も若干数存在した。

¹⁸ なお PDF データの作成および印刷の際には、画像の色彩や明るさを調整する工程も存在する。コンピュータのソフトウェアを使用して、大量の画像データを一括して調整する際には、フォルダごと等、自動的な適用が行われるため、1つの地方本部（データフォルダのまとめり）の撮影は一台のカメラを固定して行うと利便性が高まるであろう。

ることが避けられない。復元にあたっては、このふたつの志向（「資料保存」と「データセットの整備」）の双方を意識しなければならなかった。

一般的な社会調査や、個人の関心による復元二次分析であれば、分析者の関心に沿ってデータ作成段階で情報を縮約することができる。しかし、本プロジェクトにおいては、原票の複雑性を維持したデータ作成も重要である。

本プロジェクトにおける復元作業全体のプロセスは、下図のように整理できる。



図表 4-1 復元作業全体のプロセス

このプロセスは、【問題設定→質問紙・カラム設計→プリテスト→実査→入力→クリーニング】と整理されるような、一般的な社会調査のプロセスとは異なる部分が多い。この違いは、本プロジェクトが「過去（国鉄労働組合婦人部実態調査では 1952 年）に実施され、原票で保管されていた資料」を「現在の研究環境のもとで再分析する」ための「調査票復元」を行うという性質による。

過去に実施され保管されていた調査票は、復元メンバーとは異なる第三者が調査設計をしたため設計の意図がわからない部分が多い。回答の前提となっている社会状況も、復元メンバーには共有されていない。現在の研究環境のもとで再分析するためには、統計分析ソフトで処理可能なデータを作成する必要がある。

これらの条件により、以下の点が制約になる。第一に、通常为社会調査にもまして、入力に熟練を要するため、作業者の確保・養成が難しいうえに、監督者が常時、入力作業のチェックを行う必要があり、多大な負担を要する。第二に、他者が過去に行った調査であるために、調査票の設計意図に不明確な点が残るデータ化の障壁となるほか、調査者・回答者（過去）と、復元・分析者（現在）のあいだで、社会状況の理解に断絶がある。国鉄内で共有されていた用語・認識や、当時の物価水準・進学率・世帯構成などに関する常識が、調査設計

と回答の前提となっているが、復元作業者はそうした文脈から切り離されている。それゆえ、これまでの復元プロジェクトにおいて、非定型的な回答の処理は監督者の個人的習熟に依存していたと言える。

本プロジェクトでは、こうした課題に対処するために、いくつかの工夫を行った。

第一に、通常の社会調査にもまして入力に熟練を要する点についてである。(1) 作業者の確保・養成については、マニュアルによって作業を標準化した。その際、定型的でない回答についても、どこまで作業で判断するか・どのような回答であれば監督者に相談するかなどのルールも明確化した。(2) 監督者の負担については、マニュアル・質疑対応表によって監督の判断の揺れを防止することで、複数人で監督を持ち回りにできるようにした。

第二に、他者が過去に行った調査である点についてである。(1) とりわけ非定型的な回答の処理において、本プロジェクトでは調査票に即した入力ルール・マニュアルを作成し、個人的習熟への依存を減らしつつ、作業者の恣意によるデータの揺れが生じないようにした。

(2) 調査当時の文脈との断絶については、周辺情報について当時の資料等を収集しマニュアルに掲載することで、監督・作業員間で共有した¹⁹。

4.2 カラム設計

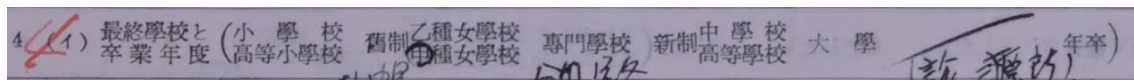
カラム設計について説明する。カラム設計は一般的な社会調査とそれほど大きく工程が異なることはない。本来、単一回答項目として準備していたものが、自記式調査で行い初票点検をしたところ、多肢選択項目として使用するほうが無難になってしまうなどといった事態は、一般的な社会調査でも起こりえる²⁰。ただし、カラム設計について特有の留意点を述べれば、(1) 欄外記入などイレギュラーだが資料的価値を持ちうる回答への対応、(2) 入力テストによる不備の発見の2点である。

まず欄外記入への対処について言及する。欄外記入には以下の図に挙げるように、回答者がなんらかの意図を持って記入したであろう項目（右端の「診療所」）と、調査集計の担当

¹⁹ 「ビジネスエリート調査」の復元を行った貫田らは、「調査実施時から長い年月が経過したデータを整備する際に直面しやすい課題」として「時代性を帯びやすい変数をどう処理するのが適切なのかという点」を挙げ、職業変数を典型として「調査時と整備時と、どちらの時代の『標準』を採用するのか。これによって、全く質の異なるデータが出来上がる」と指摘している（貫田ほか 2018: 76）。本プロジェクトにおける、マニュアルへの周辺情報の記載は、作業員が依拠する「標準」を統制することをねらっている。

²⁰ もちろん、このような場合に柔軟な集計方針の変更を許容せず、無効回答とする考え方もある。例えば、大谷ほか編（2013: 196）では、誤記入とみなして、無効回答としている。厳格に設計すれば当然このような方針となるが、一方で本プロジェクトのような復元二次分析において復元者たちは調査票の設計を担当しておらず、正確な意図は不明である。さらに1950年代の調査票となると、記入の指示などもかなり曖昧な形であり、フォーマットも現代の調査票と大きく異なる。ある設問が単一回答のものとして妥当かどうかの評価にも疑念が生じる。これらを踏まえ、データの形式的な統一性よりも資料保存を優先させたカラム設計を行っている。

者がエディティングのために記入したもの（左端の赤字「4」）の2つが存在する。これに関しては次のような対処とした。



図表 4-2 欄外記入の具体例

第一に、回答者が記入したであろう項目は、可能な限り入力するように指示をした。これには回答者本人による追記・修正や欄外記入が含まれる。調査票に当初から設置されている自由記述欄は当然のこととして、欄外記入に関してもデータ入力できるよう、各設問に追加で自由記述を入力するカラムを実装した。

第二に、集計者の集計作業の記録、例えばエディティング過程における赤鉛筆での書き込みなどは、入力しない方針とした。これについては、本人の記入ではないことから、一種の代理回答とも考えられるためである。ただし、本人記述の回答がわかりにくかったり、文字判読が困難なケースも多発していたため、実際の入力作業では、このような赤鉛筆での書き込みなどをヒントにしながら、矛盾や勘違いではないと思われるものは適宜、利用し、入力値を確定させることがあった。

以上のような対処を入力時には行ったが、このような処理は論争的な点でもある。例えば、第一の本人による追記・修正や欄外記入は、基本的に回答者の記述であるのだから、それを尊重することに問題はないと考えられるが、そもそも欄外記入をどこまで残す必要があるのか、という点は議論となりうる。加えて、第二の本人記入と考え難い記述をどこまで利用するか、という点も論争的な点であろう。少なくとも、一般的な社会調査であれば、無視するか無効票として処理することが多いと考えられる。しかし計量歴史社会学における復元に関して言えば、入力・利用せざるをえないのが実態である。それは第一に、「過去の歴史資料の保存」という要請から来るものであり、この目的を尊重する限り、欄外記入を排除することはできない。第二に、自記式調査に由来するものであると考えられるが、本来、設計者が意図していた回答方法と異なる回答が散見されたため、これらの正確な意図を掴むとともに、作業者が欄外記入から意図を汲み取って一定のコードに割り当てた回答について、その妥当性を事後的に検証できるようにするためである。これらの理由から、欄外記入についても積極的に入力し、データとして保存することを優先し、カラム設計を行った。このような方針の結果、自由回答の入力フォームを設問ごとに用意することになり、また、どのカラムにも該当しにくい自由回答を入力するカラムも別途準備したため、自由記述の入力カラム数は 167 となった。それ以外のカラム数が 326 であることを考えれば、いかに非定型データの扱いが多数に上ったか、ということがここからうかがえる。当然、このような設計は「文字読み」と呼ばれていた作業（4.7 参照）を格段に増加させるため、作業者の負担は増えることになったが、上記の性質から省略は行わなかった。

入力テストについて論じる。上記のような方針で「国鉄労働組合婦人部実態調査」のカラムを設計したため、カラム数は合計で 493 となった。両面印刷で一枚紙の調査票であることを考えると、非常に多くのカラムがあり、入力数も非常に多い。また、調査票の回答が、調査設計の想定を外れたと思われるものも多数確認されたため、本作業用としてリリースする前に入力テストを行うことで、より適切に欄外記入やイレギュラー回答を反映できるカラム設計を行った。

第一に、マニュアルとカラムの初期リリース時点では、社会調査経験のある作業者に試用のマニュアルと入力フォームで約 50 票程度の入力を繰り返してもらい、入力時の困った点をヒアリングして特定した。その点について代替案を作成し、改版した入力フォームでテストを行う、といった工程を 3 週間程度繰り返した。

このようにして作られた仮完成の入力フォームとマニュアルを用いて、1 回目入力作業前におよそ 2 週間の期間を準備して、40 人日におよぶ大規模なテスト入力を行った。この時点ではマニュアル通りの入力を行った上で、マニュアルやカラムの設計ミスを作業員から申告してもらい、質問記録として保管し、本番用マニュアルとカラム設計の参考資料とした。この工程を行うことで、本来の調査設計者の意図とは異なるものの、実際の回答の傾向を考慮して再設計すべきカラムの洗い出しを行い、マニュアルおよびカラムの再設計作業を行った。こうして修正事項の検出作業を行い、結果、3 月中旬から 1 回目入力の本番作業を行うことになった。

しかし、このような入念な準備を行ってもなお、想定していなかった回答パターンが出現し、結果的には 2 回目入力作業前に、カラムの再設計作業を行わざるをえなかった。この点は 4.6 で後述する。


4.3 マニュアル作成

4.1 で述べたように、復元作業においてはマニュアルの整備が重視された。一通りのカラム設計ができた段階で、マニュアル作成に着手し、その過程でカラムへのフィードバックも行った。マニュアル作成後は、一部の作業員によるテスト入力を行い、マニュアルの改訂を行った。本入力の開始後も、イレギュラーな事例への対応方針などを適宜追加した。

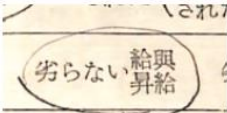
マニュアルは最終的に A4・12 ページの分量となった。全般的な作業の流れ・入力原則が 3 ページほど、残る 9 ページほどが個々の設問の入力方法である。これとは別に、全カラムについての入力規則を作成した。マニュアルは、入力規則の具体的な運用方法を直感的に理解することを可能にするとともに、代表的なイレギュラー事例への対応方法を示すよう作成した。

その際、以下のように具体的な回答例を示しつつ入力の方針を示すとともに、「どの範囲まで作業員で判断してよいか」、「どのような場合は監督者に報告・相談すべきか」についても明示した。

8 二重線や×等で選択肢を消し、特定の選択肢を意図的に残していると明確に判断できる場合、残された選択肢に○があるものとみなす。迷う場合は監督者に相談。

 ←「昇職」に○ありと判断

9 ○が複数の選択肢にまたがって書かれている場合は、またがっているすべての選択肢を選択しているとみなす。たとえば以下の例では、「劣らない」「給与」「昇給」すべてに○がついているとみなせばよい。



図表 4-3 マニュアルの例（入力の原則）

工夫のより具体的な内容を、個々の設問を事例に述べる。たとえば、調査票の問7・俸給について、マニュアルは以下の通りである。

問7
下の図の通り、「本俸」の右に書かれた数字をq7_1に、「号俸」の右に書かれた数字をq7_2に、「地域給」をq7_3に、「扶養家族手当」をq7_4に入力する。数字以外の情報が記入されている場合も、すべての情報をそのまま入力する。なお、「地域給」「扶養家族手当」が高額すぎる（おおよそ2000円以上）の場合には監督に相談すること。

7 俸給（本俸 q7_1 号俸 q7_2 地域給 q7_3 円 扶養家族手当 q7_4 円）

➢ q7_1には、原則として1桁か2桁の数字（号俸=給与区分）が入る。q7_1の位置に1桁か2桁の数字でない内容（3桁以上の数字など）が記入されていて、q7_2の位置に1桁か2桁の数字が記入されている場合、q7_2の位置に記入されている情報を号俸とみなしてq7_1に入力する。このとき、q7_1の位置に記入されていた情報はq7_2の入力欄に入力する。

| | |
|-------------------------------------|---|
| 俸給（本俸 <u>19</u> 号俸 <u> </u> 円） | q7_1は「19」、q7_2は空欄 |
| 俸給（本俸 <u>13</u> 号俸 <u>6400</u> 円） | q7_1は「13」、q7_2は「6400」 |
| 俸給（本俸 <u>14,000</u> 号俸 <u>31</u> 円） | q7_2の位置にのみ2桁の数字が記入されている →q7_1は「31」、q7_2は「14000」 (q7_1とq7_2を入れ替えて入力する) |

➢ q7_2の場所は本来、回答欄ではなく余白なのだが、この場所に回答を書いている回答者も一定数いるため、入力欄を設けている。多くの場合、q7_2は空欄となる。

➢ 金額の入力について

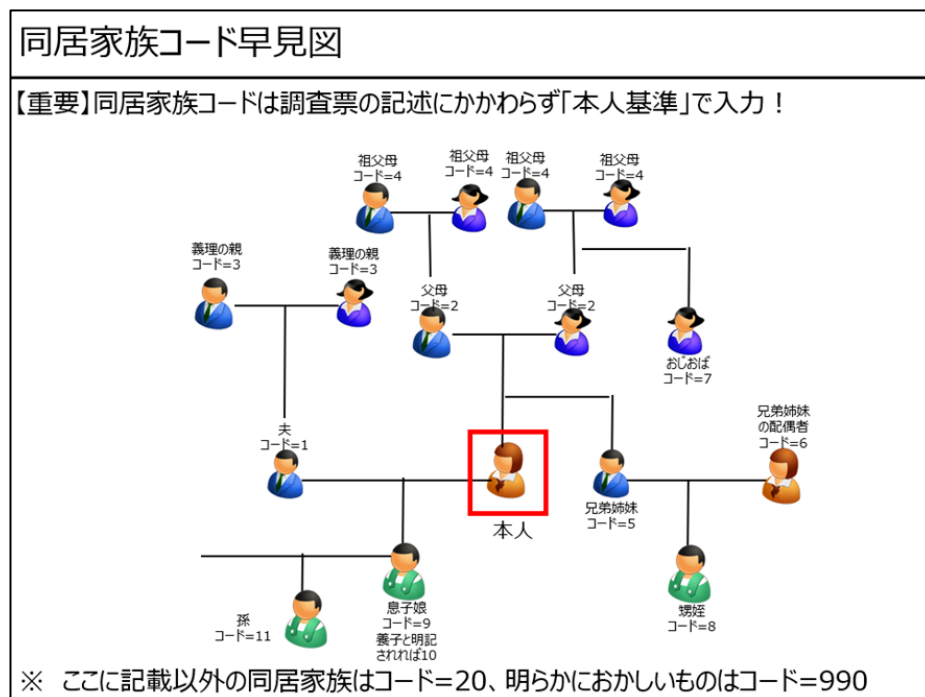
- ◇ 「円」を繰り返し記載している場合は除外して数字部分のみを入力。
- ◇ 「7000-」のように、金額の末尾にハイフンをつけている場合があるが、これは末尾であることを示す記号なので、「7000」と数字のみ入力すればよい
- ◇ 桁区切りのカンマは省いて入力する。例えば「7,000」と回答している場合は「7000」と入力
- ◇ 「銭」を含めた形で回答している場合、小数点を使って入力する。例えば「4000円00銭」と回答している場合は「4000.00」と入力
- ◇ 回答者が、「銭」の意味で点を打っている場合と、桁区切りの意味で点を打っている場合があるので、混同しないよう注意すること。
 - 「12,000」であれば桁区切りのカンマと判断できるので省いて「12000」と入力
 - 「12000.00」であれば「12000円00銭」の意味だと判断できるので「12000.00」と入力

図表 4-4 マニュアルの例（俸給）

問7の例では、(1) 調査票のどの部分への回答をどのカラムに入力すべきか、画像を用いて明確に示す、(2) 作業員で判断してよい裁量範囲を示す(号俸と本俸の入力欄の入れ替え)とともに、監督者に相談すべきケースを示す(「地域給」「扶養家族手当」が高額すぎる場合)、(3) 入力形式のばらつきを防止する(金額の入力方法)、といった工夫をしている。さらに、参考資料として調査当時の号俸表も配布している。

調査票の設計が必ずしも洗練されていないために、多くのイレギュラーな回答がみられたが、データ寄託後の多様な研究関心からの二次分析の可能性を担保すべきであるという本プロジェクトの方針からして、入力段階で情報を切り捨ててしまうことは極力避けたかった。そのため上に示したように、「q7_2 の場所は本来、回答欄ではなく余白なのだが、この場所に回答を書いている回答者も一定数いるため、入力欄を設けている」というようなカラム設計の配慮が行われている。結果として入力ルールは複雑になり、そうした入力ルールを直感的かつ個々人のばらつきなく理解できるよう、マニュアル作成には細心の注意を払った。

マニュアルの事例をもう一例とり上げたい。問19・同居世帯員について、マニュアルでは続柄の情報をコード化するための早見図を用意し、直感的な理解が可能になるよう務めた。



図表 4-5 実際に用いた家族コード早見表

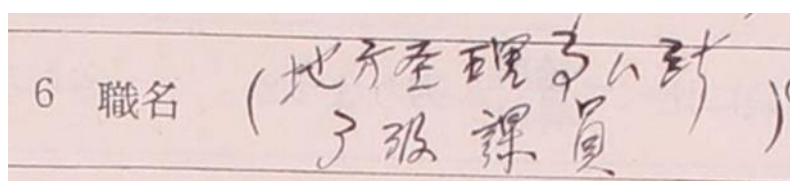
ここでは、先に述べた、調査当時(回答者)と現在(作業員)との文脈・社会状況の断絶

が課題となった。調査が行われた 1952 年時点では、現在よりも一家族の人数は多い傾向にあり、かつ幅広い続柄で同居が行われていた。しかも、ある同居家族の続柄をどのように言い表すかは、回答者によって異なっていた。以下に紹介する、家族コード入力についての説明は、このような事態に対して極力作業員自身でも判断できるようにするための工夫である。なお、作業員には中国語のネイティブスピーカーもおり、日本語・中国語では同じ漢字で別の家族関係を指す場合もあることから、とくに入力初期には注意を要した。

- 「続柄コード」は、記入されている内容をもとに、**回答者本人からの続柄を判断してコードを選択・入力する。**以下の家族コード早見図を参照。**なお、明確でない場合は憶測による判断はせず、監督に確認すること。**
- ◇ 答え方が回答者によって異なるので注意して入力する。以下で説明するように、同じ「長男」と書いていても、本人にとっての長男である場合と、親にとっての長男（＝本人にとっての兄弟）である場合がある。その意味で、「回答者本人からの続柄を判断」することが重要である。
 - ◇ たとえば兄の配偶者には「兄嫁」「嫂（あによめ）」「嫁」「義姉」など、さまざまな呼び方がある。回答一件ずつ、注意して判断する。
 - 「義姉」「義妹」「義兄」「義弟」などと書かれている場合、本人のきょうだい（＝5）ではなく、本人のきょうだいの配偶者（＝6）、あるいは本人の配偶者のきょうだい（＝その他20）である可能性が高いので、注意して判断する。判断に迷う場合は監督に確認。
 - 「養父」「養母」は父母（2）、「養女」は養子（10）とする。

図表 4-6 マニュアルの例（家族コード）

調査当時（回答者）と現在（作業員）との文脈・社会状況の断絶という点では、ごく基本的な質問項目である回答者の職名の入力すらも、当初は容易でなかった。回答者は、国労婦人部による調査、すなわち当時の国鉄の事情に精通した者による調査であることを前提として、自記式の調査票に記入をしている。その結果、当時の国鉄内で通用していた呼称・略称を、略字・くずし字で記入する、ということが珍しくなかった。



図表 4-7 略字を使用して記入された職名の例（地方経理事務所 3級課員）

当然ながら、作業員は国鉄の職名の知識を持たない。そのためマニュアルには、調査報告書の情報を参考に代表的な職名を示すとともに、当時の国鉄組織図を配布した。

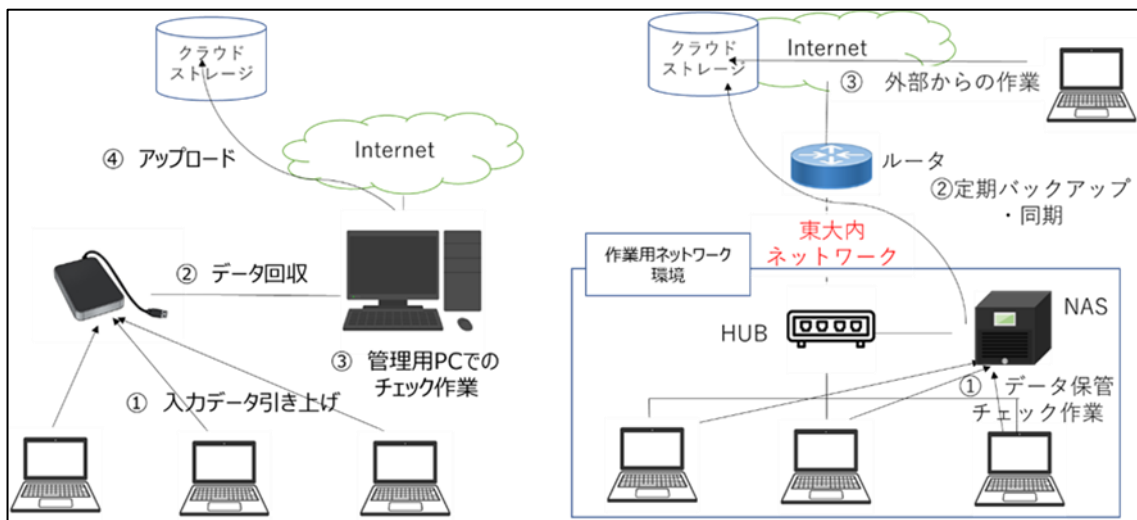
- 問6 書かれている文字を入力する。典型的な職名としては次のようなものがある
- 課員（1級課員・2級課員・3級課員）、電話掛（電務区電話掛・駅電話掛）、看護婦（病院看護婦・診療所看護婦）、技工、整備掛、踏切警手、出札掛、事務掛、雑務手、駅手、雑務掛
 - Googleドライブ「ツール一式」にある「国鉄組織図」「現業機関定員表」も参照。職場名（「電務区」など）が載っている

図表 4-8 マニュアルの例（職名）

こうした情報を示すことによって、文字の判読の手がかりにもなる。監督者が個別対応する負担を減らすためにも、このような工夫が有効であった。

4.4 入力作業体制

入力作業体制、とくにインフラ面について論じる。入力作業は東京大学社会科学研究所内に用意した作業室と、リモート作業との併用で行った。詳細なシステム構成図は図表 4-9 である。



図表 4-9 作業環境のシステム構成図

1) 左図：従前の復元体制 右図：本プロジェクトのシステム構成

このような作業体制にした理由は以下の通りである。第一に、2021年当時、新型コロナウイルスの影響があり、密集した環境での作業がためらわれたためである。第二に、書き込み可能な調査原票のコピーを準備したが、サイズ・枚数の問題で持ち出しは難しかった。第三に、従来までの作業では作業者の持ち込みPCに入力用フォームを配布し、各々が入力した結果をHDDに集約して最終的に、担当者が結合するという方式が取られるが、このような「余計な工程」が発生することにより、作業ミスが発生する可能性や、分担を事前にリジッドに決めておく必要があるため、柔軟な作業切り替えが困難になる可能性が生じる。第四に、マスターファイルの場所が分散するため、バージョン管理が困難になるだけでなく、万

一の際のバックアップ復元にも極めて面倒な工数と作業ミスが発生する余地を生み出してしまふ。そこで、東京大学社会科学研究所の担当部署と相談して以下のような作業体制をとった。

第一に、NAS (Network Attached Storage) を設置し、一つのファイルに複数名の作業者が同時アクセスできる作業環境を整備した。これにより、複数の作業で一つのエクセルファイルを同時編集することが可能になるため、無駄な作業の割り振りの手間や手持ち時間が発生せず、また、入力件数が多い地方本部に集中的に人員を投入するといった、人員配置の柔軟性を確保することができた。当然、このような作業環境を整備したため、HDD でデータを集めて統合する、といった余計な工数も省略することが可能になった。

第二に、東京大学社会科学研究所の事務用 PC を利用することで、作業者のパソコンに依存しない体制を整備した。基本的にこのような OA 環境は Windows 環境が想定されているところ、作業者には MacBook しか保有していない者も見られる。その場合、作業者に割り当てられる作業が大きく制限を受けてしまうほか、NAS に接続するためには東大側のネットワークを利用する必要があり、多数の持ち込み PC を接続させる環境では管理しきれないと判断したことによる。この結果、作業環境は Windows で統一することになり、Mac 版の作業説明をする手間が省略できる、といった副産物的な効果も生じた。

第三に、NAS の標準設定で Google Drive との同期設定²¹を取ることで、ファイルのバージョン管理をシステム側で処理させた。この結果、手作業でのバックアップが不要となったほか、NAS と Google Drive 間の同期設定を行ったことで、Google Drive 側で編集したファイルをほぼ同時に NAS 側に反映させるといったことも可能となった。このような仕組みを利用することで、東京大学に在学していない作業者に対しては、空き時間でオンライン作業をしてもらい、といった柔軟な作業投入も可能になった。また、NAS と Google Drive のシステム連携を取ることができると判明したため、さらにバックアップ体制も整備することにした。具体的には、NAS は RAID 構成という一度の HDD の故障に対しては堅牢な構成を取ることとし、一度の故障で作業データならびにスケジュールが大きく混乱しないようなハードウェア構成とした。これに加えて、外付け HDD に差分バックアップという手法を用いて、日次更新されたデータのバックアップを確保しておくことにした。これにより万が一、誤操作等でデータが消失したとしても、前の作業日までのデータを復元できる体制を整備した。

そして第四に、Google Drive ならびに外付け HDD にそれぞれ、フルバックアップデータを別々に取っておくことで、多重故障が発生してもデータが消失しにくい体制を整備した。このような多重バックアップ・同期設定を行うことにより、業務継続が可能な体制を整備し、

²¹ 細かい話であるが NAS は QNAP を導入し、Google Drive との間での同期設定は即時同期、外付け HDD への差分バックアップは週 3 回、フルバックアップは月 1 回とした。なお、バックアップのバージョンは残容量を見ながら時々に応じてチューニングしたが、基本的には 5 世代前 (フルバックアップは 1 世代前) までは保持するように設定した。

これによって、システム故障など意図せぬ要因による業務中断を防止する体制を整備した²²。

この結果、次のような作業工程管理上での成果がえられた。第一に、当初の予定通り、ファイルの結合作業が一部システムエラーが発生した場合を除いて生じなくなった。第二に、Google Drive 上でのファイル更新作業の結果が即時、NAS 上のファイルに同期されることになったため、作業結果を監督者が即時確認できる体制となった。これにより、作業者が作業室に出勤しなくとも、各自の都合にあわせてオンライン作業を行いつつ、品質をある程度、担保できる、労働力の柔軟な活用と品質の両立が可能になった。

4.5 1 回目入力作業

2022 年 3 月から、1 回目入力作業を開始した。作業者には、マニュアルに沿って監督者で初期研修（1 時間程度）を行い、数回の入力練習ののち、本入力に移行した。4.4 で述べたように、作業環境はオフライン（作業室）・オンライン（Zoom）を併用したが、それぞれの作業者が作業に習熟するまではできる限りオフラインで作業してもらい、不明点などについて相談しやすくなるよう工夫した。オフライン作業の場合、自分以外の作業者と監督者との質疑応答の内容も耳に入るため、オンライン作業よりも少ない労力で多くの情報を共有できるというメリットがあった。

入力作業にあたっては、作業者に（1）マニュアル、（2）入力規則、（3）入力用フォーマット、の 3 種類のファイルを共有した。作業者の出勤時に、その日の勤務時間などをもとに監督者が調査票を割り当て、割り当て分の入力作業が終わり次第、追加の調査票を割り当てていった。作業者の熟練度や、割り当てられた調査票の回答内容にもよるが、1 回目入力の時点でおおよそ 1 時間あたり 4 票程度の入力が可能であった。

オフラインでの作業環境は、以下のようなものであった（図表 4-10）。作業者は、割り当てられたノート PC で、入力用フォーマット（エクセルファイル）に回答を入力していく。フォーマットは基本的に地方本部ごとに 1 ファイル（票数の多い地方本部は分割）を用意し、NAS 上で共同編集ができるようにした。オフライン作業の場合は、印刷された調査票を見ながら入力を行う（調査票の印刷・ファイリングは 3.3 を参照）。文字が小さくて判読が難しい場合などは、画像データ（PDF ファイル）を拡大表示して確認を行う。机上に設置されたモニターは画像データ確認のためのものである。

²² なお、結果的にシステム障害が起きることはなかったため、このような多重バックアップ・バージョン管理を利用することはなかった。



図表 4-10 オフラインでの作業環境

入力にあたって判断に迷う点があった場合は、監督者に相談をする。相談結果は、2 回目入力・クリーニング時に引き継げるよう、印刷された調査票に赤ペンで作業者が書き込み、目印の付箋を貼ることとした。同時に監督者は、相談内容と監督者の判断について質疑記録ファイル（Google スプレッドシート）に記録した（質疑対応については 6.1 を参照）。入力作業の進捗状況は、監督者が 1 日の作業の終了時に必ず記録し、入力漏れや二重入力が生じないように注意した。進捗を可視化することは、工程管理上（6.2 参照）も重要である。

オンライン（Zoom）による作業の場合も、基本的な流れは変わらない。ファイルの共有には Google ドライブを活用し、質疑には Zoom の画面共有機能やチャット機能を用いた。オフライン環境との相違点としては、印刷した調査票を渡すことができないため、調査票の表示には各自のタブレット端末などを使用してもらった。また、外部からは NAS へのアクセスができないため、入力フォーマットは各自の PC にダウンロードしてもらい、作業終了時に Google ドライブにアップロードしてもらったものを、監督者がマージすることとした。

4.6 2 回目入力作業

2022 年 8 月には 1 回目入力作業が終了したため、2 回目入力作業に移行し、2023 年 1 月に完了した。2 回目入力作業内容は、基本的に 1 回目入力作業と同じだが、「コード」「数字」のカラムのみ入力した。「文字」のカラムには、はじめから 1 回目入力の内容を残し、あらためて入力はしないが、入力内容を読み直して、誤りがあれば修正することとした。また同一の調査票を 1 回目・2 回目ですべて同じ作業者が作業することは極力避けるようにした。

2 回目入力にあたって、1 回目入力を踏まえ、いくつかの入力ルール・カラムの変更を行った。カラムや入力ルールを大きく変更した場合、当然ながら 1 回目入力結果と 2 回目入力結果とは一致しなくなる。そこで、次のような方針をとった。

第一に、値を入力するようなカラムの再設計は極力、最小限とした。実際に行った再設計

の例として、世帯支出に「税金」カラムを追加したことが挙げられる。世帯支出の内訳のうち「その他」の欄に「税金 XX 円、雑費 YY 円」のように複数の用途の支出を並列して回答している例が多く見られ、その中でもとりわけ多数にのぼった税金について、独立のカラムを設けることが妥当と考えたためである。「税金 XX 円、雑費 YY 円」の例でいえば、XX 円と YY 円の合計を「その他」カラムに入力する、という方針も考えられるが、計算ミス等の発生防止と、実際の回答傾向を反映した暫定カラムを作ったほうがより適切であろうという判断から、「税金」カラムを新設して対応することとなった（具体的には図表 4-11 を参照）。

| あなたの収入の用途の概略を書いて下さい | | |
|---------------------|-------|---|
| (1) 家に入れる金額 | 7560 | 円 |
| (2) 食料費 | 1,000 | 円 |
| (3) 衣料費 | 500 | 円 |
| (4) 家賃間代(含む電気水道代) | 1,400 | 円 |
| (5) 教養娯楽(含む子供の教育費) | | 円 |
| (6) 借金返済 | | 円 |
| (7) 貯金、保険 | | 円 |
| (8) その他(含む税金) | 1,900 | 円 |

Handwritten notes on the right side of the form: "税金合計" (Total Tax) and "市民税年額" (Annual Municipal Tax Amount).

図表 4-11 カラム新設対応を行った項目（税金）

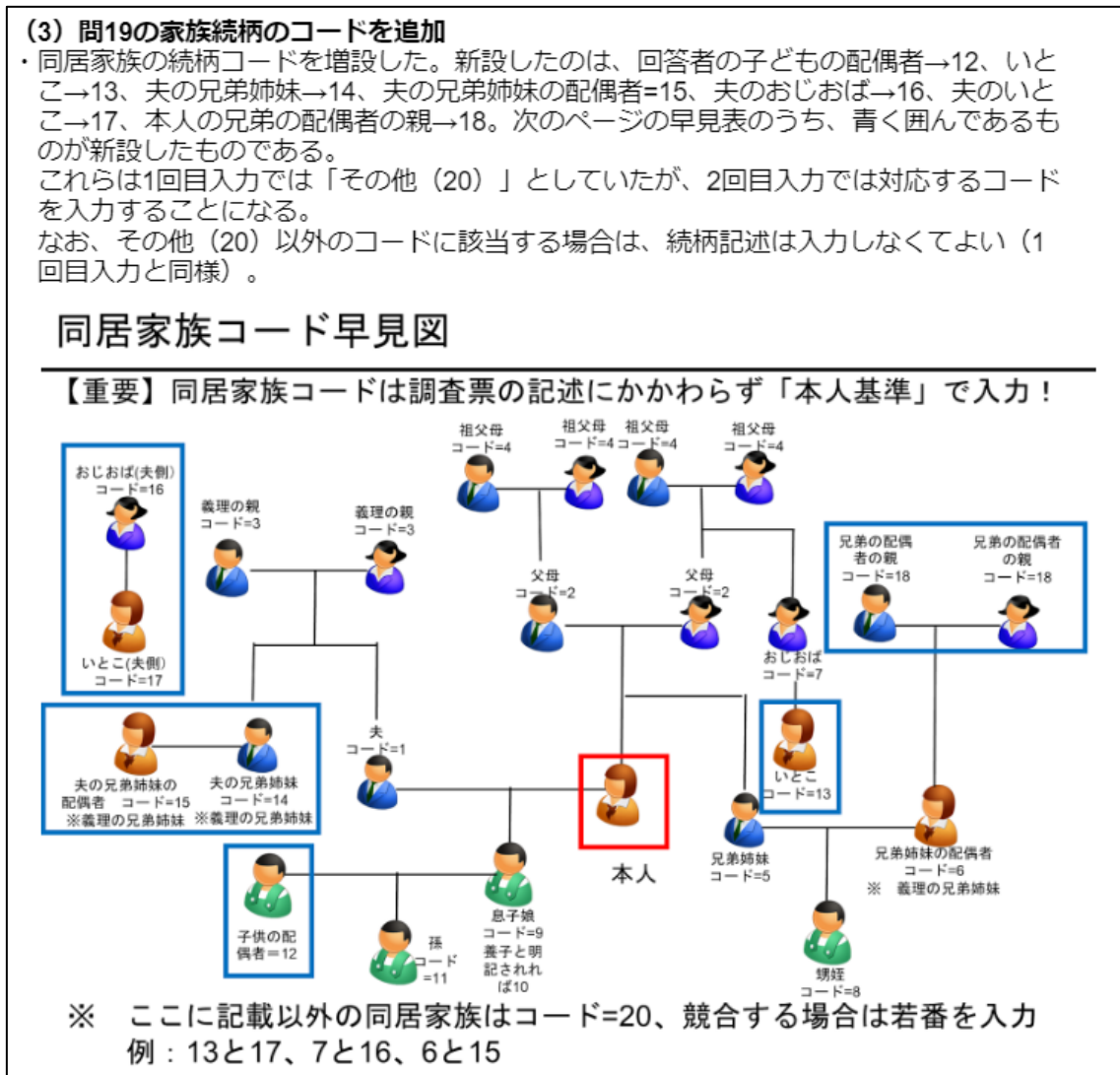
第二に、特定の設問に関して「エラーフラグ」というカラムを付け加えるという設計変更を行った。第 1 回入力作業前に入念なテストも功を奏し、カラムについて大きく再設計が必要なものは上述の「税金」のみであったが、一方で入力値に関しては想定していたコードでは対応できない場合が散見されたほか、著しい論理エラーの存在が明らかになった²³。すなわち、原票の回答を「そのまま」入力するという観点では必ずしも入力ミスではないが、一方で、データのクリーニング工程で負担が生じると予想される回答に関して、エラー検出の手間を省くためのチェックフラグを用意するといった設計変更を行った。

第三に、入力値のルールを一部変更するなどの対処を行った。例えば同居家族については、調査当時の住宅事情を反映して多種多様な続柄の同居人が存在したが、とりわけ調査対象者が女性であるために夫の親族が広く含まれることが判明した。そこで、同居家族コードを追加して対応した。このほか、数値が二通りに読める回答や、数値に幅をもたせた回答について、中央値を取るといった入力ルールを 2 回目入力では追加した。

以上のような設計変更を行った結果、1 回目入力時のカラム数は 493、2 回目入力時のカラム数は 505 となった。

²³ 例えば、調査票では本人の失業期間について尋ねているが、回答された失業期間が国鉄勤続年数から導かれる国鉄勤務期間と重複することがあった。

これら入力ルール・カラムの変更については、該当箇所について追加でマニュアルを作成し、作業への説明を行った。図表 4-12 は、先述の同居家族コードの変更に関する説明である。図表 4-5 で示した同居家族コード早見表も、この段階で修正された。



図表 4-12 同居家族コードの変更

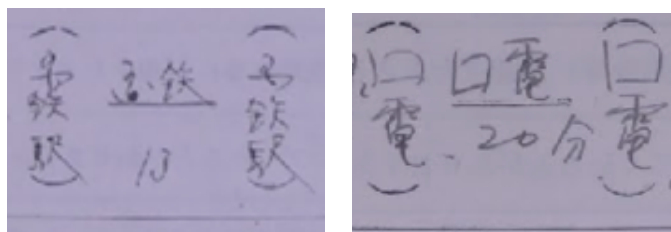
4.7 文字読み作業

復元作業において、文字の判読にはかなりの労力と工夫を要した。今回の復元のなかでは、自由記述の設問についても回答内容をそのまま文字入力した。テキストマイニング等の方法を用いた二次分析を可能にするため、そのような方針がとられた。

しかし、文字の判読・入力には複数の困難が生じた。具体的には (1) 旧字体・旧仮名遣い、(2) 国鉄内で流通していた専門用語、(3) 略字・くずし字、(4) 悪筆・誤字、が挙げられる。「国鉄」の「国」という文字ですら、入力作業を開始した直後には、判読に苦勞する

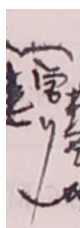
ことが珍しくなかった。入力に習熟するに従って、作業者の判読能力も向上し、また略字・くずし字に関する資料やくずし字辞典を用意したが、それでも文字判読に関する質問は多かった。質問があった場合、監督者はくずし字辞典やデータベース²⁴を検索するほか、文脈や周辺情報（地名・駅名など）を補うことで判読を試みた。なおも判読が不可能だった場合は、その箇所に「■」を入力し、後日二次分析研究会のメンバーで分担・合議する形で判読にあたった。

文字読みの困難について、特に略字と地域性の問題に関して具体例を紹介したい。例えば、「玉鉄」に見えるものは「国鉄」の略字であるし、「口電」に見えるものは「国電（国鉄電車）」の略字である（図表 4-13）。「宮ソ」は国鉄宮原操車場の略称である（図表 4-14）。国鉄鷹取工場は「広取工場」と略字で表されるが、その略字で略称「鷹工」を書けば「广工」という表記になる（図表 4-15）。「广」に類似する略字・くずし字には時期や地域性により様々なバリエーションがある。例えば図表 4-16 では、国鉄入職以前の職歴のなかで山口県の光（海軍）工場の医務部に勤務したと回答されている。このような略字や社内用語の問題は、そのまま入力するという意味では入力作業であるが、実質的には総合的判断を要するコーディング作業に近い負荷を要求されるものと言える²⁵。



図表 4-13 調査票における「国」の表記例

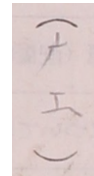
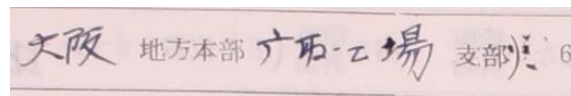
1) 左図：「国鉄」 右図：「国電」



図表 4-14 調査票における「宮原操車場（宮ソ）」の表記例

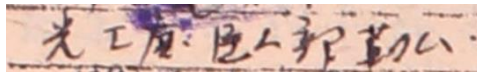
²⁴ 東京大学史料編纂所 くずし字電子辞典データベース（<https://wwwap.hi.u-tokyo.ac.jp/ships/w34/search/character>）と人文学オープンデータ共同利用センター くずし字データベース（<http://codh.rois.ac.jp/char-shape/search/>）を使用した。

²⁵ 吉村（2017: 197-200）は、単純な入力ミスの発生率は決して高くない一方、コーディングミスに関しては無視できない発生率が存在すると報告しており、特に未熟練の大学生の場合には、職業・産業を大分類レベルでコーディングさせてもミス発生率が無視できない割合になるとしている。



図表 4-15 調査票における「国鉄鷹取工場」の表記例

1) 左図：「鷹取工場」 右図：「鷹工（鷹取工場の略称）」



図表 4-16 調査票における「光工廠医務部勤務」の表記例

5 照合・再照合作業

本章では、照合・再照合作業について説明する。社会調査では、同じ調査票について、異なる作業員で異なるデータセットを作成し、双方の入力結果を突合して入力ミスを排除する「二重データ入力による検証」を採用することがある（轟ほか編 2021: 143）。本章では、このような「二重データ入力による検証」を、「照合作業」といい、照合作業を経て作成されたファイルを、さらに2回目入力データと突合して入力ミスを排除する作業を、「再照合作業」という。

5.1 照合作業

照合作業について、入力データの正確性を担保する上で、その必要性は疑うべくもない。ただし、復元作業において問題になるのは、1回目入力作業と2回目入力作業との間で、入力ミス以外の要因により、入力値に大きな差異が生じうる点である。その点についてあらかじめ、説明したうえで照合作業の詳細を論じる。

入力値が大きく異なる第一の類型は、4.6 で論じた通り、1回目入力に対応しきれなかったイレギュラー回答について、2回目入力で入力ルールの変更によって対応した結果、作業員の側に起因せず、入力値が大きく異なる、というパターンである。特に、入力値を新設した家族コードや、数値が二通りに読める回答について、入力時に中央値で処理する、といった処理は1回目入力時点で想定していなかった点であり、これらは当然ながら、照合作業で不一致を起こす。ただ、これらについては質問管理の仕組み（6.1 参照）によって遡及が可能であるほか、そもそもどのカラムで発生するか、といったことが予見できるため、照合時にあらかじめ、対処計画を用意しておくことが可能であるし、その後の処理方法も定型化されているため、工数増を招くものの、大きな負担となるわけではない。

第二に、こちらがより大きな問題であるが、1回目で担当した作業員と2回目で担当した作業員の習熟の違いにより、入力値が異なる、というパターンである。これまで述べた通り、

原票自体が相当に劣化しており、自記式調査で記入もかなり乱雑な調査票が多い。また、旧字や社内用語・略称が記入されているものもかなり多く、これらを正確に入力するのは困難を極めた。つまり、入力時期やこれまでの作業状況、資料参照の頻度により、社会調査では一般に要求されない、特殊なスキルの習熟が必要になるのだが、このような特殊な習熟の個人差により、入力値の正確性も大きく異なったのである。

そこで、1回目入力と2回目入力のデータを、照合用エクセルファイル（図表 5-1）を用意して照合作業を行った。

| 不一致数 | ID | 入力用変数名 | 変数レベル | 規則 |
|------|------|-------------|-----------------|---|
| 1 | id | ID (監督者が入力) | 文字 | |
| 2 | name | 入力者の名前 | 文字 | 照合しない 照合しない 照合しない 照合しない 照合しない 照合しない 照合しない 照合しない 照合しない 照合しない 照合しない 照合しない |
| 3 | ida | 表ID | 文字 | |
| 4 | 0 | q1 | 数字 | |
| 5 | 0 | q2 | 再観との続柄 | コード |
| 6 | 0 | q2fa | 再観との続柄 (その他・例外) | 文字 |
| 7 | 0 | q3_1 | 婚姻 | コード |
| 8 | 0 | q3_2 | 婚姻_既婚 | コード |
| 9 | 0 | q3fa | 婚姻 (その他・例外) | 文字 |
| 10 | 0 | q4_1a | 最終学校 | コード |
| 11 | 0 | q4_1afa | 最終学校 (その他・例外) | 文字 |
| 12 | 0 | q4_1bera | 卒業年度 (元号) | コード |
| 13 | 0 | q4_1byear | 卒業年度 (数字) | 数字 |
| 14 | 0 | q4_1bgrad | 卒業・中退の別 | コード |
| 15 | 0 | q4_1bfa | 卒業年度 (その他・例外) | 文字 |
| 16 | 0 | q4_2 | 検定その他 | 文字 |
| 17 | 0 | q5_1 | 所属_地方本部 | 文字 |
| 18 | 0 | q5_2 | 所属_支部 | 文字 |
| 19 | 0 | q6 | 職名 | 文字 |
| 20 | 0 | q7_1 | 係給_本係の右 | 数字 |
| 21 | 0 | q7_2 | 係給_可係の右 | 数字 |
| 22 | 0 | q7_3 | 地域給_円 | 数字 |
| 23 | 0 | q7_4 | 扶養手当_円 | 数字 |

図表 5-1 照合用エクセルファイルの構成

「1回目入力」「2回目入力」にはそれぞれ、1回目入力データと2回目入力データをそれぞれ値貼りで入力する。その結果は「結果・再入力」シートに出力される。一致したカラムは、2回目入力の数値を自動的に参照するようになっている。不一致のカラムは「X」で示され、IDごとの不一致件数が出力されるようになっている。この結果、不一致件数がゼロになるまで再入力作業を行う。また、「X」の出力箇所はセルの色が赤になるように設定しており、一見してエラー箇所を特定できるようにしている。このような照合用エクセルファイルを利用し、1回目入力と2回目入力の結果が一致するまで、作業を行うようにした。

ここで検出される不一致は大きく分けて、3パターン存在する。第一に、1回目入力が正確で2回目入力が不正確なケースである。第二に、1回目入力が不正確で、2回目入力が正確なケースである。そして第三に、1回目と2回目で入力結果が分かれ、いずれも不正確なケースである。これらについて詳述しよう。

第一のケースは単純入力ミス进行想定したものである。第二のケースは、単純入力ミスに加

え、2回目入力でのルール変更（家族コードや中央値の入力指示など）に起因するものも含まれる。こういったものについては、適切にルールを確認することで対応が可能である。

問題は第三のケースである。1回目入力と2回目入力が双方ともに不正確である、というケースも本プロジェクトでは数多く確認された。その発生原因は前述の通り、特殊的スキルの習熟によるところであると考えられるが、これらについては結局、作業者と入力監督との協議により、新しい値を入力させることとした。

5.2 再照合作業

さらに、今回の復元プロジェクトでは、再照合作業も行うように設計している。再照合作業を行った理由は以下の3点による。

第一に、入力結果の修正について、予期していない修正が行われていないかを確認するためである。具体的に言えば、カーソルキーを間違えて動かして、本来、正しい値を入力していたセルに削除処理をしてしまったなどである。これらは、再入力作業時に生じる、単純ミスを検出するという理由から実施した。

第二に、2回目入力の際に入力ルールを変更した家族コードと中央値処理について、予定している正しい値が入力されているかを確認するためである。これも実質的には3回目入力に近いとはいえ、どちらかといえば2回目入力の内容を適切に反映しているかという、単純ミスを検出するという理由が強い。

第三は、5.1で論じた、入力不一致の第三のケース対策、すなわち1回目入力と2回目入力の双方とも値が不正確である、という問題の対策である。これは結局、1回目入力と2回目入力ともに揺れが存在し、その結果、「3回目入力」を実施せざるをえなかったため、その結果についての妥当性をあらためて検証する、という目的による。

これらの理由から、再照合作業を行うことにした。ただし、再照合作業については上記2点目の理由から次のような処理を行うことにした。第一に、より習熟した段階での入力であると考えられる2回目入力と、照合作業で作成したデータ（「照合データ」とする）とを突合することにした。第二に、再照合作業を行う作業者を限定した。これは、大学院生や、社会調査法の科目を一通り受講した学部生で、ある程度以上の勤続経験のある作業者と監督者でのみ作業をすることで、判断の揺れの範囲をなるべく小さくすることを意図している。

入力ミスの検出は、図表5-1の照合用エクセルファイルとほぼ同様の内容のエクセルファイルを準備し、2回目入力データと照合データとを突合して行った。作業手順は5.1とまったく同様である。この段階でも、作業者と監督者との間で入力値の協議が必要な項目がいくつか発生したが、それらについては再度の照合は行わず、どのような理由でどのような入力値としたかの判断過程と根拠を記録しておくことで対応した。このような作業を27地方本部分、作業を繰り返し、地方本部ごとの入力確定エクセルファイルを作成し、再照合作業は完了とした。

6 監督側の体制

本章では監督側の体制について述べる。まず作業者については、二次分析研究会に所属している教員・大学院生の所属大学を中心に、東京大学・上智大学・早稲田大学・日本女子大学・一橋大学・立教大学などの学部生・研究生・大学院生で編成した。途中、人員の入れ替わりがありつつ、最盛期には29人の作業者が参加していた。

当然、このような作業者すべてが社会調査の教育を受けているわけではない。また、いかにマニュアル化しようとも、判断が必要な入力が多数、存在する。そこで作業者と監督者でマニュアルについての疑義や、個別の調査票の入力値、判読不明文字の協議などを行うことが必要になる。この点、『ボーダー・ライン層』調査の復元であれば相澤真一が1名で監督作業を担当し、「京浜工業地帯従業員調査」の復元であれば、複数名で監督をしていた。当然、複数名の監督が作業に従事するのであれば監督者間での判断根拠の揺れを最小化する必要があるし、1名の監督であっても監督側の習熟度も次第に高まるため、判断する時期によって判断の揺れが生じることが考えられる。これは、調査票の設計意図が不明であるという、復元二次分析に特有の問題であると考えられる。29名もの作業者を1名の監督で管理することは現実的に困難でもある。さらに本プロジェクトは年度予算というスケジュールで動いていたため、作業の進捗率を計測し、必要なリソースを再配分したり、先に着手できる後工程を割り出して、先に着手しておくなどにより、作業の停滞を防止する必要があった。本章では、作業者ではなく監督者側の判断の揺れをどのように最小化したか、また作業を進めるにあたって監督者はどのようにして作業進捗管理を行ったかという、監督側の体制について論じる。

作業体制は次のようなものであった。著者ら3名のうち、原則として2名、作業者が少ない場合は1名が作業室に在室して、質疑対応を行った。このほか、3名とも在室できない場合には、二次分析研究会に参加する研究者や博士課程の大学院生に、在室と作業記録の保存を依頼した。作業日は基本的に週3日、長期休暇中は週5日体制で作業を行った。結果、1回目入力作業は2022年3月中旬から8月上旬まで、2回目入力作業は2022年8月下旬から2023年1月下旬まで要した。

6.1 質問管理の仕組み

次に質問管理について論じる。前述の通り、入力に関して作業者のみならず、監督者でも判断の揺れが生じることがある。また、イレギュラー回答の入力には総合的な判断を要求されるため、「なぜそう判断したのか」ということを事後的に追跡できるようにする必要性があるほか、マニュアル作成時には想定していなかった問題が作業者からの指摘によって発見されることも想定された。社会調査の教科書を紐解いても、調査票のエディティング時に、疑問のある回答を修正することの難易度が高いことは指摘されている(轟ほか編 2021: 132)。これに加えて、「国鉄労働組合婦人部実態調査」について言及するならば、当時の時代性や

国鉄内部の事情など、調査票以外に考慮すべき要素も非常に多くなる。もちろん、くずし字に関する資料や当時の国鉄の統計、組織図を用意するなどして想定される要素にはあらかじめ対処したが、それでも予想外の問題に直面することは非常に多かった。また、調査票が自記式であることに由来する、記入の揺れも存在した。そこで、作業者からの質問に対しての監督者の判断過程の追跡や、頻出質問の割り出しを行うため、入力作業にあたっては以下のような質問管理表を作成して、作業記録の保存を行うことにした。

| 質問ID | ステータス | 質問日 | 調査票(カラム名) | 質問内容 | 回答者 | 回答内容 | 判定結果 | 備考 |
|------|-------|------|--------------|--|-----|---|------|----|
| 203 | 回答済み | 6/6 | I0395 q4 | 給与に地価補給7000円 かつ 収入に月収7000円と明記 | 瀬戸 | 21号簿7000円とみなして入力 | | |
| 204 | 回答済み | 6/6 | W0370 全体 | 下書きしてからペンでなぞっているが、一部なぞっていない | 瀬戸 | なぞっていない箇所も本人記述として入力 | | |
| 205 | 研究会送! | 6/6 | W0379 q19 | 本人世帯主だが、有職の父が同居している | 瀬戸 | とりあえず回答者申告を信じないで入力。クリーニング時に要検討 | | |
| 206 | 研究会送! | 6/6 | W0382 q21 | 鉄道引渡金は国鉄外か国鉄内の贈与か(欄内外どちらにも記入されている) | 瀬戸 | おそらく一時的な出向と考えられる上に、運用上は当時ならば国鉄内と考えても可 | | |
| 207 | 研究会送! | 6/9 | W0396 q10 | 高所得で要調査されたことになっているが、「第10号」と補足されている | 瀬戸 | 前の欄で要調査されたということなので解答者が異なるかと判断して、「その他」で | | |
| 208 | 回答済み | 6/9 | W0404 q22 | 職場で嫌がらせについて、「しない」「する」どちらにもがついているが、自由回答には「風呂場を瀬戸 | 瀬戸 | 常識に従えば嫌がらせなので、「しない」は間違ってる丸をしたと判断 | | |
| 209 | 回答済み | 6/10 | O0155 q19 | (1)夫を「戸主」と記載しているが、世帯主のOは本人の「女」についている。(2)収入が、国鉄職員 | 堀江 | (1)世帯主は取らない。(2)収入は本人16号簿(q7と一致)、夫20号簿に付 | | |
| 210 | 回答済み | 6/10 | O0167 q19 | 続柄の「養父」をどう処理すればよいか | 堀江 | 親と判断し、その旨を申し送り | | |
| 211 | 研究会送! | 6/10 | I0039 q19 | おそらく世帯員全員(子ども含む)を指して「農業」と書いてある | 堀江 | 職業はひとまず8888として、監督会議で判断。収入の「合算」のようなルールも | | |
| 212 | 回答済み | 6/10 | W0439 q13 | バス・徒歩の通勤経路が併記されている | 堀江 | 異常回答990 | | |
| 213 | 研究会送! | 6/10 | W0445 q19 | (1)年齢が一人分振れている。(2)収入の「家賃」を誰の収入とすればよいか | 堀江 | (1)「三男(弟)」は学生なので13歳と判断。長男・次男の年齢は特定不能なので | | |
| 214 | 回答予定 | 6/10 | W0427 q19 | 本人が33歳(未亡人)で、23歳の長女がいることになっている(本人はq2で三女と回答しているので、 | 堀江 | 本人および子ども4人の年齢を異常回答990とする | | |
| 215 | 回答予定 | 6/10 | W0433 q19 | (1)続柄に「3女」とあり、娘と思われる2人分の年齢が書いてあるが、続柄をどう入力するか。(2)本人 | 堀江 | (1)家族の続柄はいちおう990とした。(2)収入は異常回答99999990を入力 | | |
| 216 | 回答予定 | 6/10 | K0142 q23 | 進捗状況の「(U) 子どもが一人前になるまで」の、「子ども」が「弟」に書き換えられたうえでOがつ | 堀江 | 異常回答990とした | | |
| 217 | 回答予定 | 6/10 | L0050 q19 | 父・妹の職業が判断不能 | 堀江 | 父職業空欄、妹職業「 \bullet 」とママ入力して申し送り | | |
| 218 | 研究会送! | 6/13 | I0417 q19 | 支出に判断不明数字(1か2かまでは絞り込める) | 瀬戸 | 考えられる数字を二つないで入力 | | |
| 219 | 研究会送! | 6/13 | O0180 q19 | 世帯構成員の人数と年齢と職業の数が不一致 | 瀬戸 | 整合性の取れる群(22)まで入力 | | |
| 220 | 回答予定 | 6/13 | W0512 q19 | 支出が世帯合計支出 | 瀬戸 | 研究会で議論 | | |
| 221 | 回答済み | 6/13 | W0492 q33/34 | どちらの記述か不明 | 瀬戸 | 他の回答状況も踏まえて、どちらにも回答したとみなした | | |
| 222 | 回答済み | 6/13 | I0421 q19 | 世帯構成員の情報がまったくないが、独身寮に在住と答えている | 瀬戸 | 世帯人数=1のみを入力してあとはママ | | |

図表 6-1 質問管理表

この質問管理表についてはいくつか特徴をあげることができる。作業記録に「X日に以下のようなルールを追加」といった記録をつけることは一般的であろうが、本表ではそれに加え、「だれが・いつ・どのような質問を行い・誰が・どのように回答したか」また、「ある質問が回答済みなのか、回答保留中なのか、複数人での協議が必要なのか」といった、質問の詳細を追跡できる仕組みとステータス管理を適切に行うよう工夫した。このことのメリットは次のような点でみられた。第一に、類例の質問があった場合に以前の質問をチェックすることで、監督者に交代が生じても同様の判断を行うことが可能になる。第二に、監督者自身の判断が揺らいでいる場合、監督者同士で「なぜこの判断をしたのか」という確認を相互に行うことができる。当然、いずれかの監督者が間違った判断をしていれば、遡って入力データの修正を行うことも可能になる。第三に、質問のステータス管理を行っているため、「どの質問が難易度が高く、自分たちだけでは判断できないのか」「まだ解消していないのか」ということが可視化できるため、入力作業のハードルとなっている具体的課題をリアルタイムで共有することが可能になり、監督側の調査・検討リソースの効率的な配分が可能になった。これらを踏まえると、少なくとも自身が調査設計しておらず、設計の意図も不明であり、時代性も共有しない調査の復元に向き合う場合には、作業記録をつけておくのみならず、その判断過程まで含めて、可視化できる体制を整えることが必要と言える。言い換えれば、入力マニュアルの整備のみならず、判断過程も含めてトータルで把握することで、初めて非標本誤差への十分な対処が可能になる、というハードルが存在すると言える。この点は、近

年の社会調査において非標本誤差のうちでも、集計誤差に関する関心が薄れつつあると論じられているが（吉村 2017: 194-5）、ことに復元二次分析に関しては文字の読み取りや調査票の劣化など、機械的作業を困難にする要因があり、なおかつ、時代性を共有しないという問題から、重要な問題として浮上すると言える。

一方で運用上の課題もあった。質問管理表の記録・対応をどこまで精緻・頻繁に行うかという問題である。監督者もある程度熟練してくると、過去と同一の質問が出た場合は質問管理表を参照して回答することが可能になり、類例のある質問は記録を行わない、という対応を取っていた。データのクオリティ・コントロールという点ではこのような対応でも問題ないのかもしれない。ただ、事後的に判断の揺れを点検し、遡及的にコードを修正する可能性を想定すると、課題が残ることは否めない。また、質疑のステータス管理を行うことで、継続的に協議しなければいけない質問を可視化することはできたが、即座に協議を行える体制ではなかったため、マニュアルへの反映等のタイムラグは常に生じた。結局、これらの問題は監督側の工数として事後的な負担となるが、当時の作業体制からしてこれらをタイムリーに処理できる機構は存在しなかった。もちろん、調査票数が少なければ可能なのかもしれないが、調査規模の拡大に伴って必要となる工数は単純に比例しない（吉村 2017: 13）ため、この規模の調査の復元にはより充実した体制を整えることが望ましいかもしれない。少なくとも、6,000 票程度の調査票を復元するためには、できれば4名程度の監督者が従事できることが望ましい。

6.2 進捗管理と作業工数の計算

2.2「復元の作業工程の設計方針」が作業計画の設計についての議論であるならば、本節はいかにその計画を計画通りに履行させ、場合によっては変更したか、ということの議論である。

まず、工数の算出についてである。初期の入力テストの過程でおおよそ、1日あたりの作業量を計測した。同時に、1ヶ月ごとに作業者のシフトを編成し、作業日1日あたりの平均的な作業人数を算出する。この作業人数を作業量とで乗算することで、1日あたりないし、1ヶ月あたりのトータルの見込み作業量を割り出す。これが標準進捗率となる。

これに対して、作業については以下のように地方本部別の入力状況を管理するシートを作成した。同時に、作業日ベースでの進捗率を転記し、週当たり作業進捗率を計算した。

| | A | B | C | D | F | G | H | I |
|----|--------------------------|-------|-------|--------|-------|-------|------------|---------|
| 1 | 国鉄女子労働調査_入力作業_進捗管理表(1回目) | | | | | | | |
| 2 | 通番 | 拠点名 | 総入力件数 | 最新入力件数 | 入力開始日 | 最終入力日 | 最新入力確認(監督) | 進捗率 |
| 3 | 1 | A_旭川 | 105 | 105 | 3/9 | 3/31 | 前田 | 100.00% |
| 4 | 2 | B_釧路 | 74 | 74 | 7月7日 | 7/25 | 前田 | 100.00% |
| 5 | 3 | C_札幌 | 337 | 337 | 4/8 | 4/22 | 前田 | 100.00% |
| 6 | 4 | D_青函 | 172 | 172 | 8月1日 | 8/4 | 堀江 | 100.00% |
| 7 | 5 | E_盛岡 | 134 | 134 | 8/9 | 8/9 | 前田 | 100.00% |
| 8 | 6 | F_仙台 | 469 | 469 | 3/16 | 4/8 | 前田 | 100.00% |
| 9 | 7 | G_高崎 | 163 | 163 | 5/30 | 7/4 | 堀江 | 100.00% |
| 10 | 8 | H_水戸 | 64 | 64 | 3/23 | 3/29 | 堀江 | 100.00% |
| 11 | 9 | I_東京 | 1158 | 1158 | 4/7 | 7/25 | 堀江 | 100.00% |
| 12 | 10 | J_千葉 | 127 | 127 | 3/1 | 3/24 | 瀬戸 | 100.00% |
| 13 | 11 | K_新潟 | 195 | 195 | 4/8 | 6/24 | 堀江 | 100.00% |
| 14 | 12 | L_北陸 | 150 | 150 | 5/27 | 6/27 | 瀬戸 | 100.00% |
| 15 | 13 | M_長野 | 176 | 176 | 6/24 | 8/1 | 十文字 | 100.00% |
| 16 | 14 | N_静岡 | 242 | 242 | 3/24 | 5/27 | 前田 | 100.00% |
| 17 | 15 | O_名古屋 | 419 | 419 | 4/15 | 7/25 | 堀江 | 100.00% |
| 18 | 16 | P_大阪 | 532 | 532 | 4/14 | 7/4 | 堀江 | 100.00% |
| 19 | 17 | Q_南近畿 | 158 | 158 | 8/5 | 8/5 | 瀬戸 | 100.00% |
| 20 | 18 | R_福知山 | 3 | 3 | 3/24 | 3/24 | 瀬戸 | 100.00% |

図表 6-2 作業進捗管理表

以上のような手続きにより、1日あたりや週当たりの標準作業進捗と、実際の作業進捗率を割り出すことを行った。このような作業管理のメリットはいくつか存在する。第一に、科学的管理法と同様、動作を分解しおおよそ、どの設問にどれぐらいの時間を費やすことが妥当か、という「相場感」を客観的に計測できるため、文字判読等でいわば「泥沼」の作業状態に陥った作業員に対して、監督者から作業指示を出すことによって標準進捗へ回復させるといったことが可能になる。第二に、標準進捗がわかるため、作業が遅れている作業員を特定することができ、「なぜ遅れたか」の原因を検討することができる。これは怠業防止のみならず、作業員ごとに個別に存在するクセをヒアリングし、場合によっては監督者側の体制を改善することにも寄与した²⁶。第三に、作業全体の進捗を可視化できるため、追加の人員投入や作業日の追加といった判断を比較的、客観的に行うことができる。このように、標準作業量と作業進捗を経時的に把握できる体制を整えることで、作業改善策の検討を行うことができ、最終的には年度内での入力作業の完了という目標を後退させることなく、作業を完了することができた。

このように、標準作業量と作業進捗を経時的に把握できる体制を整えることで、作業改善策の検討を行うことができるようになるのである。これを、2.2で作成したガントチャート

²⁶ 例えば、くずし字の漢字を判読する作業は、紙媒体のくずし字辞典を用いて行っていたが、作業効率にはかなりの難が存在した。そこで、作業員からのくずし字の探し方をヒアリングし、東京大学史料編纂所の「電子くずし字字典データベース」を利用するなどの作業体制の改善を行った。このほか、過去に判読されたくずし字について、監督者に質問する前に自分で調べるといった作業員が複数いたことから、代表的な社内用語・略称パターンやくずし字のパターンを周知するといった対応を行う端緒になった。

の予定作業数と組み合わせることで、各工程にどれぐらいの人員が、いつの時期に投入されたかを可視化することができ、作業全体の工程管理を行うことが可能になる（図表 2-2 を参照）。

7 おわりに

本稿ではこれまで「国鉄労働組合婦人部実態調査」の復元作業の過程と方法について示してきた。

本プロジェクトは一方で、調査実施から 70 年以上経過したデータを復元し、SSJ データアーカイブに寄託できる品質の、二次分析可能なデータを作成することが求められた。本稿では、この目的に沿った作業の設計、作業員・監督者の体制、直面してきた困難への対処方法を論じてきた。

他方で本プロジェクトでは、資料発掘という側面、すなわち歴史的なデータを保存するという面についても、強調してきた。繰り返しになるが、本プロジェクトでは一見すると同じ方向性でありながら、細部に立ち入っていくと、ジレンマが生じる 2 つの目的を、妥協可能な水準で妥協させる、ということを行ってきた。本プロジェクトの実践がどこまで成立しているかは、今後、二次分析により多くの利用者がデータを用いることで、課題が浮かび上がっていくだろう。その意味では、データの二次利用ができるようになってからも、実践は続くのである。

最後に、本プロジェクトの持つ「二次利用可能な」「資料データ」の復元の課題と展望について簡潔に説明する。

第一に、復元に際して「正解」が存在しないことである。貫田ほか（2018: 72）では、調査当時の正確なコーディングガイドが残っていないため、当時の分析に用いられたパンチカードの結果を再現できなかったことが述べられている。この点は本プロジェクトでも妥当しており、調査報告書の結果と復元したデータが一致しないことはしばしばある。ただこれは、よるべき基準が存在しないもしくは、存在していたとしてもよりかかれぬことに起因するとも言えるだろう。この点では、基本統計量のような基礎的な項目でも、復元の妥当性をどのように担保するか、という問題がある。

第二に、繰り返しになるが、時代の文脈を共有できないという点である。この点は、回答者の問題と調査者の問題にわけられる。まず、回答者における調査票の意味通用性という問題がある。例えば、調査票にある「甲種女学校」「乙種女学校」は厳密には存在しない。前者は高等女学校、後者は実科高等女学校を指すと推測はでき、調査当時は難なく回答できたかもしれないが、現在において回答の意味を解釈するためには複雑な教育制度史に関する資料調査が必要であり、容易な問題ではない。このような、「時代を共有しない」ことによる入力のはらつきは、品質に大きな影響を与えると考えられるが、厳密な検証は困難である。

時代の文脈に関するもう一つの点は、調査者と復元者の間での認識・関心の断絶、という

問題である。貫田らは復元した「ビジネスエリート調査」の調査票について「調査者の研究関心から遠い情報ほど振るい落とされやすく、その分のバイアスが元のコード票データには映し出されているように見受けられた。今後のデータ利用者が調査者とは異なる研究関心から分析を試みる可能性を想定するならば、元のデータセットから漏れ落ちていた情報を今一度原票から拾い上げていく作業は必須であったといえよう」（貫田ほか 2018: 76）と指摘している。調査者と復元作業にあたる研究者との間には、時代の隔たりゆえに基本的な認識・研究関心の断絶が存在していると考えられる。また、調査当時は集計技術の制約も大きかった。これらの理由によって、当時は削ぎ落とされた情報も存在することは、今回の復元でもしばしばうかがえた。復元二次分析にあたってはこれらの点の自覚的な検討が必要であり、復元に際して調査者が現代の関心に沿って情報を削ぎ落とすということは、慎まれるべきであろう。二次分析は調査者と異なる関心から、異なる角度での分析が行われることに利点があるのであれば、調査者の関心から調査を再構成することは、二次分析の重要な利点を、検証不可能な方法で消失させることにほかならない。あくまで情報の削ぎ落としは、技術的問題に起因されるものに限定されるべきであろう。

第三に、組織の文脈に起因する復元可能性である。今回、本プロジェクトで取り組んだデータは国鉄という、日本最大級の企業でありなおかつ、公社組織であったため、統計資料など周辺情報は充実していた。もちろん、時代を反映して、男女別の統計が存在しない、毎年度、同じ項目が採録されているとは限らないなど、いくらかの制約事項は存在したものの、復元中に生じた社内用語の疑問などについて、まったく調べる手立てがないということは少なかった。しかしながら、「労働調査資料」に限定しても、多くの調査は企業内調査であり、第二の時代の文脈という問題と同程度か、それ以上に、組織の文脈の問題に直面せざるをえない。今回の復元対象は国鉄における調査であったため、このような「組織の文脈」を探索することは可能であったが、社史が乏しかったり、(大半の場合に妥当するであろうが)企業内史料が未公開の場合には、「組織の文脈」を探索することはできない。こうした、組織という要因による復元可能性の制約の問題は、今後検討していくべきであろう。

このように、現在、重要視されている再現可能性という観点では大きな問題を復元二次分析は内包しているかもしれない。佐藤香は、歴史学における資料批判に対応する「データ批判」が計量歴史社会学にも必要であるとしつつも、「計量歴史社会学における『データ批判』には定まったスタイルも参照すべきマニュアルもない。『データ批判』の作法あるいは方法を、これからの研究のなかで作り出していく必要がある」（佐藤香 2007: 15）と述べている。この問題提起から 15 年あまりを経た本プロジェクトにおいても、「作法あるいは方法」に確立されたものは無く、著者らは原票を前に試行錯誤を重ねざるをえなかった。この点で著者らにできることは、本稿のように復元の手続きを公開して透明性を確保することのみである。しかし、文書史料からでは明らかにできない現象にアプローチするなど、復元二次分析でしか解けないリサーチクエスションは存在する。以上の点からも、「国鉄労働組合婦人部実態調査」のデータセットが多く研究者によって分析されることを願っている。

【謝辞】「国鉄労働組合婦人部実態調査」の復元作業に際して、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から原票の特別利用許可を得ました。また、社会科学研究所からプロジェクトの遂行に関わる経費の支出を受けました。記して感謝申し上げます。

【付記】本稿の分担を右の通り示す。第1章は前田と堀江，第2章は堀江と瀬戸，第3章は前田，第4章は瀬戸と堀江，第5章と第6章は瀬戸，第7章は前田と堀江と瀬戸で執筆している。

【文献】

- 相澤真一・小山裕・鄭佳月，2013，「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」『ソシオロゴス』37: 65-89.
- 蘭信三・中里英樹，1998，「計量的歴史社会学の展開と可能性——家族史研究を中心として」『理論と方法』13(1): 41-57.
- デジタル庁，2022，『デジタル・ガバメント推進標準ガイドライン 実践ガイドブック』，(2023年12月28日取得，https://www.digital.go.jp/resources/standard_guidelines).
- 独立行政法人 情報処理推進機構，2005，『組込みソフトウェア開発におけるプロジェクトマネジメント導入の勧め』日経印刷.
- 藤田若雄，1953a，「職場における組合活動の実態——婦人部実態調査分析（一）」『国鉄文化』7(1): 36-9.
- ，1953b，「国鉄女子労働者の生家の職業および就職経路——婦人部実態調査分析（二）」『国鉄文化』7(3): 39-44.
- 橋本健二・稲田雅也・中川宗人・相澤真一・元治恵子・佐藤香・小山裕・岩井八郎・片瀬一男・久米祐太郎・開田奈穂美・米澤旦・石島健太郎，2013，『社会科学研究所所蔵「労働調査資料」の二次分析 研究成果報告書』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.
- 橋本健二編，2015，『戦後日本社会の誕生』弘文堂.
- 五十嵐彰，2020，「『日本帝国統計年鑑』を電子化したの雑感」『社会と統計』6: 55-64.
- 岩井八郎，2018，「戦前期から戦後における高学歴層の流動性と戦後階層システムの形成——SSM 調査の再分析から 1940年代を読み直す」『社会学評論』69(3): 355-72.
- 苅谷剛彦・菅山真次・石田浩編，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.
- 小池和男，2005，『仕事の経済学 第3版』東洋経済新報社.
- 国鉄労働組合，1954，『国鉄婦人労働調査報告』.
- 林采成，2010，「戦争の衝撃と国鉄の人的運営」『歴史と経済』209: 46-62.

- 丸沢美千代, 1981a, 「国労婦人部実態調査の中で」藤田起編『藤田若雄——信仰と学問』教文館, 170-4.
- , 1981b, 「国鉄労働組合婦人部創設の頃」『婦人労働と資料』3: 7-20.
- 貫田優子・孟哲男・岩井紀子・萬成博, 2018, 「日本のビジネスエリート調査（1960年・1970年）のデータ整備——データ入力からオンライン利用の準備まで」『日本版総合的社会調査共同研究地方本部 研究論文集』19: 63-94.
- 大谷信介・木下栄二・後藤範章・小松洋編, 2013, 『新・社会調査へのアプローチ——論理と方法』ミネルヴァ書房.
- 労働調査論研究会, 1968, 『労働調査論研究会中間報告（其の二）——調査参加者との面接結果および手稿』.
- 労働調査論研究会編, 1970, 『戦後日本の労働調査』東京大学出版会.
- 佐藤香, 2007, 「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.
- 佐藤香・相澤真一・中川宗人, 2015, 「歴史的資料としての社会調査データ」『歴史と向きあう社会学——資料・表象・経験』ミネルヴァ書房, 45-64.
- 佐藤俊樹, 1998, 「特集 計量歴史社会学」『理論と方法』13(1): 1-4.
- 轟亮・杉野勇・平沢和司編, 2021, 『入門・社会調査法 第4版』法律文化社.
- 豊田真穂, 2007, 『占領下の女性労働改革——保護と平等をめぐる』勁草書房.
- 渡邊大輔・相澤真一・森直人・東京大学社会科学研究所編, 2019, 『総中流の始まり——団地と生活時間の戦後史』青弓社.
- 渡邊勉, 2020, 『戦争と社会的不平等——アジア太平洋戦争の計量歴史社会学』ミネルヴァ書房.
- 吉村治正, 2017, 『社会調査における非標本誤差』東信堂.

秘

国鉄労働組合婦人部実態調査表

1952.5.現在

封筒の宛先の書方

▲注意

東京 国鉄労働組合本部 婦人部長 殿

- 1. この調査は婦人部があなたの力となるためのもので無記名ですから税金その他あなたに迷惑のかかることは絶対にありませんからありのままを書いて下さい。
2. あてはまる個所に○印を付け()には記入して下さい。
3. 1人も残らずこの調査票に記入して国鉄労働組合本部婦人部に直接送つて下さい。
4. 本部婦人部に6月末日までにとどくようにして下さい。

国鉄労働組合婦人部

1 年令(満 歳) 2 両親との続柄(長、次、三、四、五、六、七、八、女) 3 未婚(イ)結婚生活をしている(ロ)未亡人(ハ)離婚
4 (イ)最終学校と(小)小学校(高)高等学校(乙)種女学校(甲)種女学校(丙)専門学校(新)制中(高)等学校(大)学(年)卒(ロ)検定その他()
5 所属支部名() 6 職名() (例)地方経理事務所の二級課員 隣の電話・診療所の看護婦
7 俸給(本)俸() 8 国鉄に入つてからの年数()
9 いまの仕事を始めからの年数() 10 いまの仕事をどのようにして習いましたか。(イ)学校() (ロ)養成所()
(ハ)徒弟奉公() (ニ)職場で養成された() (ホ)その他() 11 国鉄にどのようにしてつとめるようになりましたか。(イ)職業安定所(紹介所、動員署を含む)の紹介(ロ)終故(1)上役の世話(2)知人の世話(あなたの関係)
(3)親兄弟、親類の紹介(ハ)学校の紹介(ニ)門前募集(ホ)廣告募集(ヘ)その他()
12 本採用。 臨時人夫。
13 通勤経路と時間を次の要領で書いて下さい。(注意)自轉車、バス、市電、私鉄等の乗物の種類を書いて下さい。
[例] 自(元)任吉(武蔵)外杉(南武)線(川)崎(徒)歩(動)務 15分 5分 15分 3分
自(宅) () () () () ()
計 時間 分
14 小学校を出たときの家の職業 (イ)農業(自作 反、小作 反)(ロ)工員(ハ)社員(ニ)商業(ホ)公務員
(ハ)国鉄職員(ト)その他() 15 (イ)戦災の有無(有 無)(ロ)引揚の有無(有 無)
16 戦後失業したときがありましたか。(イ)ない(ロ)ある(その期間昭和 年 月 から 年 月 まで) 17 戦後失業したときどうして(イ)郷里に歸つた(ロ)實家に歸つた(ロ)夫の實家に歸つた(ロ)その他()
(ロ)歸らなかつたが要兄弟から援助をうけた(ハ)退職金(ニ)失業保険金(ホ)日雇(ヘ)臨時工(ト)内職(チ)手傳(リ)商賣(ス)その他()
18 (イ)生活に困つたとき親兄弟の援助をうけますか() うける。 (ロ)うけない。
19 同じ住居に住んでいる人について書いて下さい(世帯主に○印を付けて下さい)
あなたの世帯
あなたとの続柄 男女別 年令(満) 職 業 一ヶ月の収入 あなたの収入の使途の概略を書いて下さい
あなた自身 女 才 国鉄職員 (1)家に入れる金額 (2)食料費 (3)衣料費 (4)家賃(含む電気水道代) (5)教養娯楽(含む子供の教育費) (6)借金返済 (7)貯金、保険 (8)その他
20 住居の別。 国鉄公舎 国鉄アパート 家族寮 獨身寮 公営アパート 借家 借間 自家 その他()

21 最終学校（夜間を除く）を出てからの職歴を書いて下さい。

| 国鉄につとめるまでの職歴 | | | 国鉄につとめてからの職歴 | | | |
|--------------|------------|------------|--------------|--------------------------------|-----------------------------|--|
| 期 間 | あなたの勤め先の内容 | 退職した理由 | 期 間 | 職名 (転職、降職、昇職の 區別を書いて下さい) | 理(強制か自分の都合) 由(か)を書いて下さい) | |
| 昭和15年～17年 | 町 医 者 | 女中のように使われる | 昭和18年～20年 | 看護婦見習 | | |
| 年 年 | | | 年 年 | 轉職 降職 昇職 | | |
| 年 年 | | | 年 年 | 轉職 降職 昇職 | | |
| 年 年 | | | 年 年 | 轉職 降職 昇職 | | |
| 年 年 | | | 年 年 | " " " | | |
| 年 年 | | | 年 年 | " " " | | |
| 年 年 | | | 年 年 | " " " | | |
| 年 年 | | | 年 年 | " " " | | |

22 職場の環境 (イ) あなたの分擔業務以外に上役や男子職員があなたを雑用や私用に使いますか。 つかう つかはない

(ロ) 上役や男子職員がいやがらやをしますか。 しない する (どのようないやがらせをするか書いて下さい)

(ハ) 辭職や配置轉換を婦人であるという理由で拒否されましたか。 されない された (されたが拒否した(何時) 何回) (されたが拒否しなかつた(何時) 何回)

(ニ) 職名、勤続年數、學歷等の等しい男子職員に比べて給與が劣りますか。 劣らない 給與 劣る 給與 (どの位劣りますか)

(ホ) 超過勤務をしますか。 しない する 1週 時間 1日平均 時間

(ヘ) 超過勤務をした時超過勤務手當が付きませんか。 付く 付かぬ (理由を書いて下さい)

(ト) 生理休暇をとりませんか。 とる とらぬ (理由を書いて下さい)

(チ) 上役より誘惑されたことがありますか。 ない ある

(リ) 上役以外の人から生活上の問題とからんで誘惑をうけたことがありますか。 ない ある

(ヌ) 其他職場で愉快又は不愉快な點を書いて下さい。()

23 いつまでつとめますか。(イ) 結婚するまで (ロ) 子供が出来るまで (ハ) 子供が一人前になるまで (ニ) 終身 (ホ) 其他 ()

24 結婚生活と今の勤務は兩立しますか。 する しない (理由を書いて下さい)

25 (イ) 職場に問題が起きたときどこに相談しますか。 組合 婦人部 上役 その他 ()

(ロ) 職場に關係するあなた自身の問題(配置轉換、辭職勧告等)が起きたときどこに相談しますか。 組合 婦人部 上役 その他 ()

26 職場に關係のないあなた自身の問題について相談のつてくれる人がありますか。 ない ある (友人、上役、組合役員、婦人部役員、肉親、其他)

27 職場に關係のない個人問題について組合が世話をすることを望みますか。 望む 組合 望まぬ 上役

28 (イ) 組合の問題や活動は何によつて知りますか。 国鉄新聞 家族版 国鉄文化 組合役員報告 婦人部役員報告 組合で聞く その他 ()

(ロ) 婦人部問題や活動は何によつて知りますか。 国鉄新聞 家族版 婦人部特報 婦人部役員報告 組合役員報告 婦人部で聞く その他 ()

29 組合や婦人部の活動をするために上役や男子職員、女子職員、當局からいやがらせをされますか。 されない される (誰れからどのようにされるか書いて下さい)

30 (イ) 組合があなたの問題(職場に關係ある)を解決したことがありますか。 組合に相談したが解決されなかつた (どのような問題か書いて下さい)

解決された (どのような問題か書いて下さい) 組合に相談したことがない

(ロ) 婦人部があなたの問題(職場に關係ある)を解決したことがありますか。 婦人部に相談したが解決されなかつた (どのような問題か書いて下さい)

解決された (どのような問題か書いて下さい) 婦人部に相談したことがない

31 あなたの知人(婦人)で国鉄を働かされた人がありますか。 ない ある (その人は今どうしていますか。 家にいる 結婚した 就職した 身を落した 人) その他 ()

32 現在組合によつて組合員の利益がどのように守られているか知っていますか。 知っている 知らない その他 ()

33 組合に對するあなたの希望を書いて下さい。

34 婦人部に對するあなたの希望を書いて下さい。

35 どの政黨を支持しますか。 社会党 右 労働党 共産党 改進黨 40%