

布施グループ・夕張調査データのテキスト化 における「入力ルール」生成過程の記録

— 社会的質的データの定型性と再現性をめぐって —

A record on the process of generating data input rules for an investigation by Fuse group on the coal mine city, Yubari

— on the definitive and reproducibility of Sociological Qualitative Data —

庄司知恵子

SORDプロジェクトでは、平成18年度から、「社会調査史の博物館」構築に向けて、1970年代に「北海道大学・生活社会学研究会」（代表 布施鉄治）によって行われた炭鉱都市・夕張をフィールドとする実証的社会調査において収集された資料のデータベース化、調査資料のデータベース化、炭鉱労働者調査データのデジタル化を行ってきた。筆者は、「社会調査史の博物館」構築における一連の作業のうち、〈調査データのデジタル・アーカイブ化〉において、質的データのテキスト化作業を行っている。本稿は、質的データのテキスト化作業における「入力ルール」の生成過程を記述することを第一の目的とする。その上で、質的データの入力作業が有する機能と、テキスト化を目指した質的調査のあり方について述べる。

1. はじめに

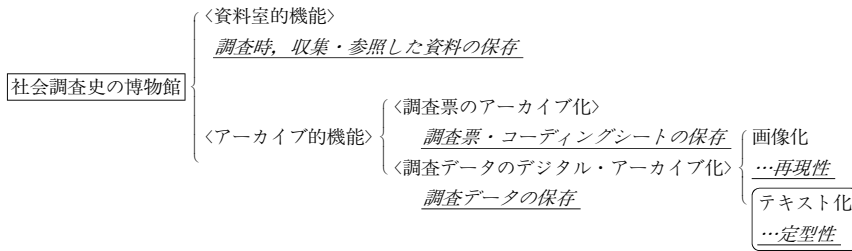
本稿は、本誌掲載論文である齊藤論文「社会的質的データのデジタル・アーカイブ方法論・序説——北海道布施鉄治グループによる炭都・夕張調査に即して——」の「3.2 コーディングシートのテキスト化」について、コーディングシートのテキスト化作業に携わった側からの視点を盛り込み、作業過程を記述することを目的とする。特に、作業過程において生じた入力の困難性をいかにして解決してきたのか、その試行錯誤に着目し、質的データのデジタル・アーカイブ化におけるテキスト化作業の「入力ルール」生成過程を記述する。その上で、さいごに、質的データの入力

作業からみえてきた入力作業が担う役割について述べる⁽¹⁾。その意味で、質的データのテキスト化作業の指南として読んでいただきたい。

2. 「社会調査史の博物館」構築における本作業の位置づけ

2.1 「社会調査史の博物館」構築作業

札幌学院大学社会情報学部 SORD プロジェクトでは、平成18年度から「『社会調査史の博物館』としてのリージョン拠点データアーカイブの構築」（平成18年～21年度文部科学省研究費補助金基盤研究（B））をテーマとした研究活動を行っている。その一環として、この一年半、1970年代に「北海道大学・



【図1】「社会調査史の博物館」における作業

生活社会学研究会」(代表 布施鉄治)によって行われた炭鉱都市・夕張をフィールドとする実証的社会調査(以下、「夕張調査」とする)において収集された資料のデータベース化、調査資料のデータベース化、炭鉱労働者調査データのデジタル化を行い、「社会調査史の博物館」構築に向け、作業を行ってきた。調査関連資料をそれぞれ適した方法で保管し、二次分析への道筋を作ることが「社会調査史の博物館」の目指すところである。これまで社会調査の分野において量的データのアーカイブ化・二次分析の利用は行われてきた。本プロジェクトの目指すところは、質的調査・質的データの特殊性に根ざしたアーカイブ化・二次分析の利用であり、その方法論の確立にある。

本プロジェクトが構築を目指す「社会調査史の博物館」が有する機能について簡単に説明する。詳細については本誌掲載齊藤論文を参照されたい。「社会調査史の博物館」が有する機能には、大きく分けてふたつある。ひとつは、調査チームが調査時に収集・参照した資料の保管である「資料室的機能」と、もうひとつは、調査チームが調査時において作成した資料である調査票・フィールドノート・学会発表原稿などの保管を目的とした「アーカイブ的機能」である。このうち、調査票の「アーカイブ的機能」は、調査票自体を保存する「調査票のアーカイブ化」と、調査で得られた調査データの保存である「調査データのデジタル・アーカイブ化」がある。夕張調査

の場合、「調査票のアーカイブ化」においては、調査票自体の保存と、調査票をもとに作成されたコーディングシートの保存が行われた。「調査データのデジタル・アーカイブ化」においては、コーディングシートをもとに、データの「再現性」を目指したコーディングシートの画像化と、データの「定型性」を担保するための調査データのテキスト化が行われた。

以上が夕張調査に即した「社会調査史の博物館」の大枠の作業である。

2.2 質的調査・質的データのテキスト化作業の位置づけ

さて、夕張調査に即した「社会調査史の博物館」の一連の作業のうち、筆者が携わっている作業は、コーディングシートに記された調査データをテキスト化する作業である([図1]の枠で囲った部分)。この作業は、筆者と他二名の大学院生(以下、「入力チーム」)⁽²⁾により行われている。現在、「夕張市民(の)生活実態調査」における世帯を対象としたA票(「基礎表」)の入力作業が終了した。

われわれ入力チームが行ったテキスト化の作業は、他の作業に比べ、質的調査・質的データの特殊性と、それに伴う困難性をもっとも顕著に現れる作業といえる。調査時、収集・参照された資料の保存を目指した「資料室的機能」における資料の整理は、実物の資料を何らかの客観的で明確な分類枠組みのもと整理することが可能である。調査票本体の整理

を目指した〈調査票のアーカイブ化〉においては、炭鉱・階層・票の種類ごとに整理することが可能であり、この分類もまた、客観的で明確な枠組みが用意されている。〈調査データのデジタル・アーカイブ化〉の画像化の作業も、実物をそのまま画像に落とし、炭鉱・階層・票ごとにファイル名をつけて保存すれば良い。これらの作業は、客観的で明確な指標のもと整理可能であり、結果、誰にとっても共有可能な資料・データとして保存可能である。

しかしながら、調査データのテキスト化作業においては、次の点で、困難な状況が存在する。齊藤論文でも書かれていたように、質的調査・質的データにおいては、量的データのようにフォーマット化された回答が用意されているわけではない。コーディングシート上に設けられている欄からはみ出している記述、欄外に書かれている記述、単位で応えるべき内容に対し、文章で書かざるを得ない回答、図で表現されている記述など、ときに、(調査票が)「意図せざる回答」が出現する。これら記述は、現場や当事者のリアリティを描き出す上で貴重な記述であり、質的調査においてこそ得られる醍醐味ではあるものの、テキスト化作業においては、処理に困る「やっかいな記述」である。これらデータを処理しながら、コーディングシート上の記述を忠実に再現していくことが質的データの入力に求められている作業である。

われわれ入力チームが行った作業は、単に流れ作業でデータを入力していくというのではなく、記述ひとつひとつの内容を読み解き、その内容に適した入力方法を探索し、質的データの「入力ルール」を生成していく作業であったといえる。

3. 「入力ルール」の生成過程

以下、入力作業の流れを追い、「入力ルール」の生成過程を記述していく。

3.1 事前に決めたルール「欄内記述」/「欄外記述」

作業開始前、とりあえず2つの入力ルールを決定した。ひとつは、デジタル化する上で基本的なものである。入力されたデータの二次分析を考えた場合、様々な利用者、利用方法が想定される。そのため、入力データは汎用可能な形として残す必要がある。機種依存文字を利用しない、英数字は半角を利用するなどといった、最低限のルールを用意した。

ふたつめとして、コーディングシート上に記述されたデータを理解する際のルールである。コーディングシート上のデータは、フォーマット化された回答が用意されていない質的データであるという性質上、多様な記述のありようが想定される。これら記述を分類することを目的としたシンプルなルールを事前に用意し、入力作業を行なうことが望ましい。要するに、事前に記述を理解するルールを決め、作業自体をフォーマット化することにより、フォーマット化されていない回答の煩雑製を回避するのである。

そこで、入力チームでは「あくまでも原票に忠実に」という合言葉のもとに、当該記述がコーディングシート上に用意されている欄の内か外か、という視覚的な判断から記述の分類をした。その分類のもと、「欄内記述」は「設問の回答」として理解し、事前に用意された「入力フォーム」(File Maker Pro8を使用)に入力、「欄外記述」は「自由記述」として理解し、テキスト形式で保存(当初、「秀丸」を利用)するというルールを作った。このルールに従い、「ファイルメーカー」と「秀丸」という二つのファイル形式を往復する形で、入力作業を進めることとした。

それぞれ3シートほど入力してみると、入力作業に「違和感」を覚え、効率よく作業を進めることが難しいという意見が入力チーム内で発せられた。その最たる理由は、コーディングシート上の記述に対して事前に用意して

いた分類と、それに従った入力ルールが作業遂行上、適していないというものであった。どのような点で適していないのか？

当該記述を視覚的な判断のもと、「欄内記述」「欄外記述」というように機械的かつシンプルな形で分類し、「欄内記述」=「設問に関連する回答」→「ファイルメーカーに入力」,「欄外記述」=「設問に関連しない回答」=「自由記述」→「秀丸に入力」というようにルールを定めたわけだが、実際作業を進めていくと、目の前に現れる記述は思っていた以上に多様であることに気づかされる。それぞれ記述されたデータには、記述である以上、「内容」が存在する。入力作業を進める中で、われわれは記述の内容に触れることになる。機械的かつシンプルに記述を分類し、入力していく作業は、記述の内容を無視する形で展開され、否応なく、内容把握の思考の筋との入力ルールとのズレを感じさせるものであり、入力チームに対し、「果たしてこの入力方法で良いのだろうか」という疑問を常に投げかけるものであった。

コーディングシート上に記述されたデータの分類においては、欄の内か外かといった機械的かつ視覚的な処理によるものではなく、当該記述の内容を、前後の設問ないしは調査票全体における位置づけから理解し、分類することが重要であり、その分類に従い入力方法を模索し、入力ルールを定める必要がある。以下、具体的な例から、記述の分類において内容を重視する必要性をみていく。

[図2]は、欄の内に収まっている「欄内記述」であり、一見して、その設問にかかわる回答として理解し、ファイルメーカーの当該設問箇所に入力していけば良いように思われる。しかし、よくよく見てみると、他設問の欄にまたがって記述されていることがわかる。ファイルメーカーへの入力場所は、果たして当該設問箇所であらうのだろうか。また、[図3]のように記述が欄の内から始まり、

欄の外に続いている回答もあった。このような記述は、欄外の記述が欄内からそのまま続いているのか、それとも一度、欄内で文章が切られ、欄外の記述は新たな、もしくは補足的な文章なのか、内容を読み解かなければその判断は難しい。

「欄外記述」の内容を読んでみると、設問への関連がみられるものがあった。これら記述は、記述に対して設問から矢印や線が引っ張られており、客観的に関連が示されていると想定できる場合もあるが[図4]、矢印や線がなく設問の近くに記述されており、内容を読み解くことによって、設問との関連性が認められる場合もあった[図5]。また、[図6]のように、設問への関連はみられるもののテキスト形式では入力困難な記述が存在した。他にもシートを目にした途端、まさに「茫然自失」となってしまうような膨大な量の記述も存在した[図7]。この記述の場合は、設問との関連性を確認するよりも先に、そもそも読み始める場所はどこなのか、項目の区切りはどこにあるのか、一瞥しただけでは皆目検討がつかず、記述を読み解き、判断する作業が求められる。

当初、「あくまでも原票に忠実に」という合言葉のもと、「欄内記述」=「設問に関連する回答」→「ファイルメーカーに入力」,「欄外記述」=「設問に関連しない回答」=「自由記述」「秀丸に入力」とするルールを定め、入力作業をしてきた。この判断は一見して原票に忠実のように思われるが、それは視覚的な意味での忠実さでしかない。また、記述に含まれた内容を重視しない作業は、記述入力に「違和感」を覚えさせ、結果として、入力チームの作業の手は止まってしまった。質的データであるが故の多様性による煩雑さを防ぎ、機械的な処理に徹するために定めたシンプルなルールであったが、そのルールが反対に、作業の効率性を奪ってしまう結果となったのである。「原票に忠実に」再現するといった場合、

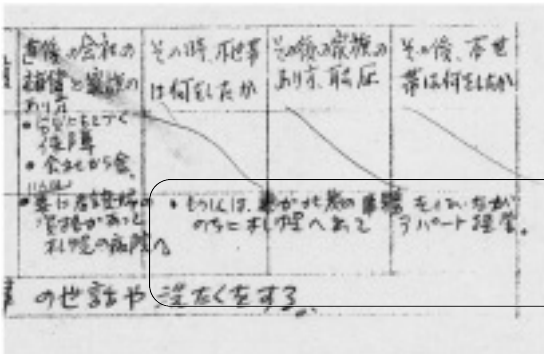
果たしてわれわれ入力チームは、どのような点を忠実に表せばよいのであろうか。

3.2 内容をくみとり記述を類型化

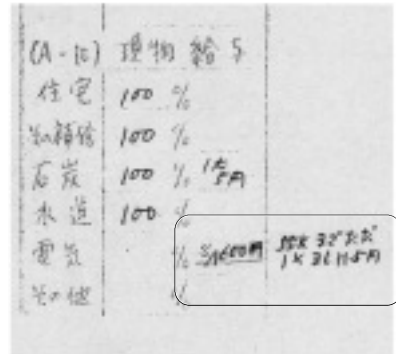
そこで、入力チームは、記述に対して新たな分類方法を考え、それに従ったルールを作成することにした。

先に示したように、「欄外記述」であっても

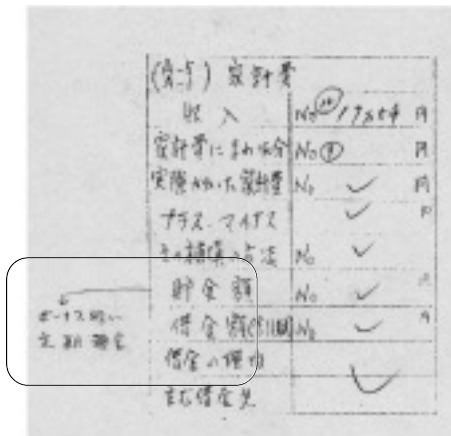
設問にかかわる内容が含まれている記述が存在する。この場合、「設問に関連する回答」として「欄内記述」と同様に扱い、ファイルメーカーの入力フォームに入れ込んでいくことにした。しかし、入力作業を進めてみると、設問にかかわる記述といっても、大問レベル⁽³⁾でのかかわり [図8] と小問レベル⁽⁴⁾でのかかわり [図4] [図5] といったように、違う



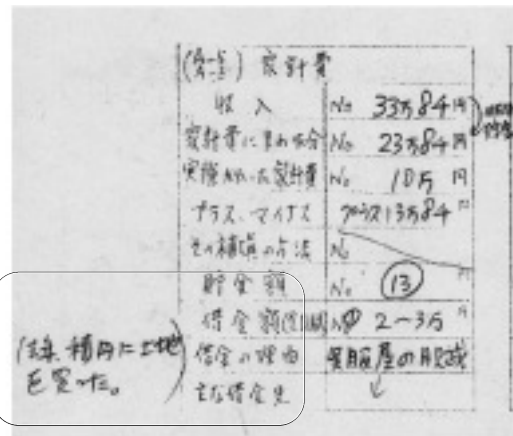
[図2] 欄をまたがって書かれている記述



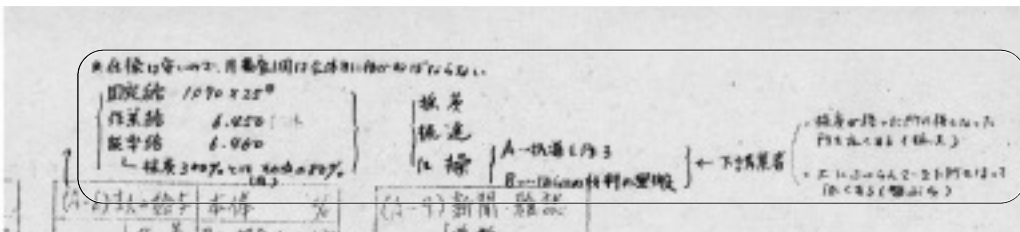
[図3] 欄をはみだして書かれている記述



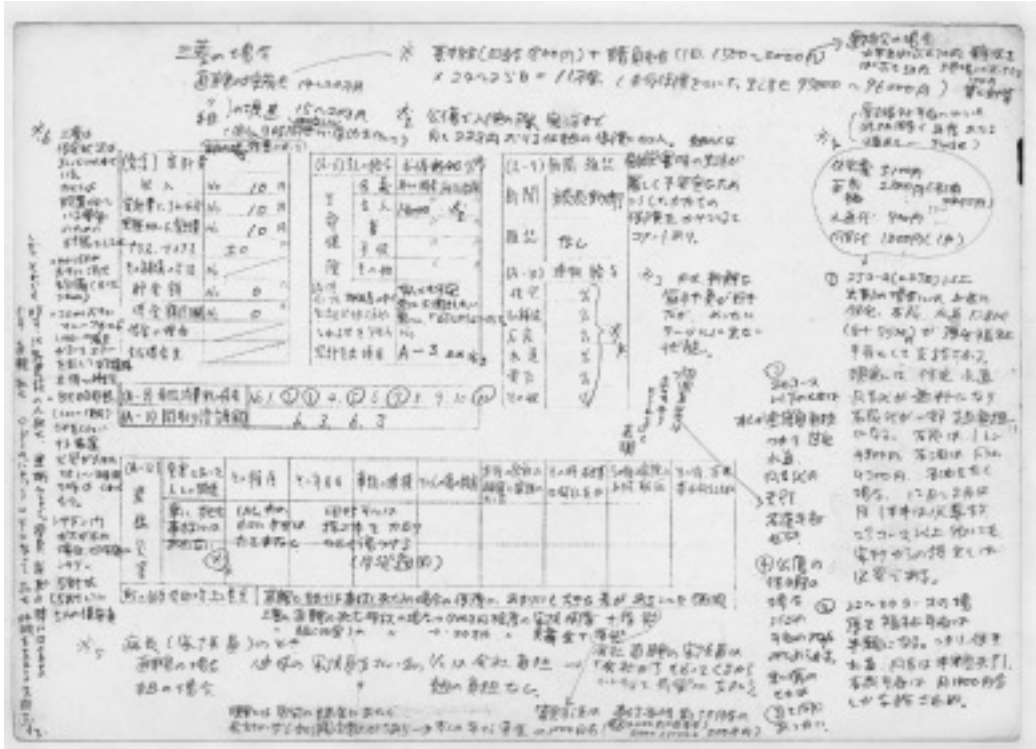
[図4] 設問への関連が認められる記述①



[図5] 設問への関連が認められる記述②



[図6] 入力困難な記述



【図7】膨大な量の記述

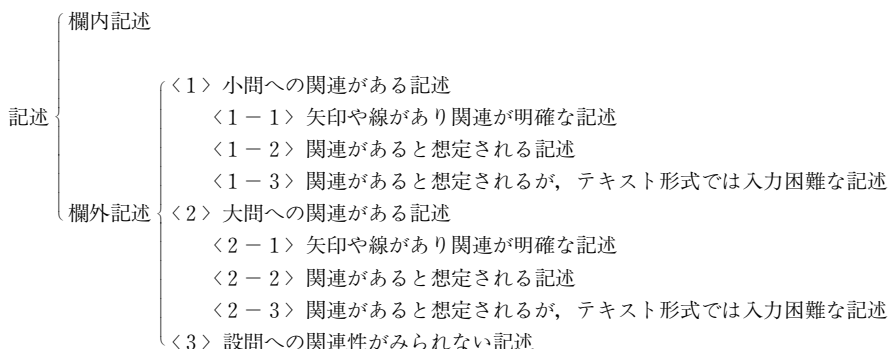
規模	その人の得意分野	直後の会社の 業績と家賃の 推移	その時、その人 は何をしていたか	その後、家賃の 推移、取戻	その後、その人 は何をしていたか
大規模		30月入居 (8月付家賃) 家賃17000	家賃11月 5月用家賃17000 (16,200円)		

責任上では、誰の何をどのどのどの所もある
家賃で買った人とは違う人。

【図8】大問レベルでの関連が認められる記述

レベルでのかかわりが存在すること、また、欄外記述の多くが、設問となんらかのかかわりを持った記述であることがわかってきた。要するに、これまで機械的に「欄内記述はファイルメーカーへ」「欄外記述は秀丸へ」といった、二つのファイルを往復する作業があまり意味を持たない、ということが判明したので

ある。
この状況を鑑み、入力チームは入力作業をいったん止め、これまでの作業を見直すことにした。第一の見直しは、記述を理解する際に用いる分類ルールの再構成である。当該記述を、設問へのかかわり方のレベル（大問／小問）を軸として分類し、第二に「ファイル



【図9】 記述の分類

メーカー」と「秀丸」を用いて行っていた往復入力を見直し、ファイルメーカーのみで入力するというファイル一元化の試みである。

記述を理解する際の分類ルールの再構成においては、入力者それぞれが担当するコーディングシートの中から事例を出し合い、「この場合、どのように理解する必要があるか」といった共同討議を何度となく行った。結果、「欄内記述」「欄外記述」という大分類のもと、「欄外記述」には中分類として、<1> 小問への関連がある、<2> 大問への関連がある、<3> 設問への関連がない、という3点、更に<1>と<2>は、当該記述と設問が、矢印や線で結ばれており、客観的に関連ありと考えられる記述<1-1><2-1>、内容を読み解いた結果、設問に関連があると想定される記述<1-2><2-2>、設問に関連は認められるものの、ファイルメーカー上においても、テキスト形式でも入力困難な記述<1-3><2-3>、というように分類された【図9】。

尚、「欄内記述」においても「やっかいな記述」が存在したのは前述の通りである。これら記述に関しても入力ルールを決めた。欄をまたがって書かれている記述は【図2】、記述が始まったところの設問に対する回答として理解し、ファイルメーカーの当該設問箇所に入力するというようにした。欄から始まった記述が欄外まで伸びている場合【図3】、記述

の切れ目が問題となるが、一文で伸びている場合は、欄内記述として理解した。ひとつの文が欄内に収まり、それに続くようにもうひとつの文が欄外にある場合、内容の切れ目によって、欄内からつながっている記述として理解するか、それとも欄外記述として扱うか、疑問の余地が残る。その場合は、入力チーム内で話し合いをもち、欄内記述として分類できればそのまま入力、欄外記述であれば記述分類上の場合に該当するかということを決め、入力を行なっていった。

これら作業を通してみてきたことは、われわれ入力チームが示さなければならない「原票の忠実さ」とは、単に字面を映し出す作業により表現されるものではなく、調査者が、被調査者の回答から読み取った方法や調査者と被調査者との往復によって作られたデータが含むうるリアリティというものを、記述の内容を分析することによって読み解き、入力ルールを作り上げ、再現していくことなのではないだろうか、ということであった。

3.3 ファイルの一元化

さて、記述の分類が成立したことにより、ファイルメーカーのフォームも変更された。変更点を記述の分類【図9】に従いながらみていこう。尚、入力フォームについては、齊藤論文【図8】を参照されたい。

まずは、<1>の小問への関連がみられる記

述は、そのままファイルメーカー上の当該設問箇所に入力すれば良いが、以下の補足が必要である。〈1-1〉に関しては、設問と記述を結ぶ矢印や線が、客観的な関連性を示しているものとして理解し、原則として躊躇なく当該設問箇所に入力することとした。〈1-2〉に関しては、若干の曖昧さが残らざるを得ない。というのも、入力者側である筆者たちの判断により、設問と記述の関連性を認めているからである。果たして、この記述が真に設問に関連しているものなのかどうか疑問は残るが、それを知っているのは、コーディングシート記入者であり、その記入者でさえも忘れていく可能性が高い。よって、このような記述に出会った場合は、入力チームで関連性があるかないかを話し合い、入力の是非を決めた。小問への関連が認められなかった場合は、他分類に落とし、処理に従った。〈1-3〉に関しては、いくら関連がみられても、ファイルメーカー上には入力不可能であり、またテキスト形式での入力は、その記述の広がりや再現性を再現実ることはおろか、記述が含むうるリアリティを捨象してしまう可能性もある。保留とし、後に適した形での入力方法を考えることにした。

〈2〉の大問への関連がみられる記述に関しては、新たにファイルメーカー上、大問自由記述欄を設問の最後に設け、〈1〉を判断する際用いた同様の方法で分類を判断し、入力作業を行なった。〈3〉に関しては、コーディングシートが表裏二枚一セットであったことから、表面／裏面自由記述として扱い、新たにそれぞれの記述を入力する箇所をファイルメーカー上に設け、入力を行なった。

このように記述の内容をもとに、設問との関連性から記述を分類し、入力フォームを変更した結果、効率よく作業が進むようになった。作業開始当初、1シート40分から1時間ほど要していた入力時間が、[図9]の分類に従い入力を行なった結果、1シート20分から

30分ほどで入力可能になった。また、表／裏の自由記述欄を設けたことで、その都度、入力されていない記述のチェックにつながり、入力に際しての取りこぼしミスも少なくなった。

さて、次なる課題は、〈1-3〉〈2-3〉といった入力困難な記述の入力方法検討である。これら記述は、そんなにたくさんあるわけではないが、だからといって入力しなくても良いという類の記述ではない。「チャート」による記述は、調査者が文章で表現するよりも「チャート」で表現したほうが好ましいという考えのもと選択した記述であろう。同様に、「図表」や「計算式」も、文章で表現するよりも、このような形で表現したほうが、被調査者の意図をくみ取ることにつながり、リアリティが伝わると考えた結果によるものであろう。その点を考えれば、単なるベタ打ちでは、被調査者から聞き、それを他に伝えようとした調査者の意図が失われてしまう。できる限り原票に忠実に入力し、データとして残していく必要がある。さて、それはいかにして可能であろうか。

以下に述べる点は、われわれ入力チームが反省すべきところなのだが、入力当初、画像化されたファイルの存在を知っていたが、このファイルは、最終的に判読不能な文字を読み取る際に利用すれば良い、というぐらいにしか考えていなかった。つまり、自分たちの作業が、画像ファイルと照らし合わせることでどのような効能を発揮しうるのか、自分たちの作業が「定形性」を、画像が「再現性」を担保しているということは、全く考えずに作業を進めていたのである。そのおかげとっては語弊があるかもしれないが、「チャート」や筆算、図表といった記述を、なんとかして原票に忠実な形で再現すること重要であり、後でみた人が、すぐその内容を理解できるような形で入力することが大切であると考えた。次なる入力方法をデータの定型

性及び可能な限りの再現性を重視して検討し結果、最も適しているとして行き着いたのがMicrosoft Word上での入力であった。

大括弧や中括弧、斜めに延びる長い矢印、紙面上縦に延びる筆算式、記述を囲んだ大きな丸、吹き出し、線を使って書かれた間取りなど、これら記述の多くは、入力に際して幅広い空間を必要とする。Microsoft Word上で入力することにより、記述が必要とする幅広い空間を確保でき、オートシェープ機能を使うことにより、長い矢印や丸など、再現が難しいと考えられていた記述も再現可能となった。そうはいつても、原票の図表やチャートに含まれる調査者、被調査者によって作りあげられたデータの息遣いとでもいべきものは、無機質な記号による表現からは感じ取ることは難しい。また、Microsoft Word上表現されたデータには、入力者側の主観も入らざるを得ない。二次分析の際には、これらファイルを画像ファイルと照らし合わせて利用することが重要である。

以上の作業により、残された記述である〈1-3〉〈2-3〉の入力方法が定まった。入力方法見直し当初、入力作業に際して、ファイルメーカー一元化を検討していたが、結果としては、ファイルメーカーとMicrosoft Wordのファイル二元化となってしまった。

そもそも、何故一元化であることが望ましいと入力チームは考えたのか。その理由は、第一に作業効率のアップであり、第二に、二次分析等で利用する者にとって、ファイルメーカー一本で情報を把握できたほうが望ましいと考えたからであった。作業の効率性があがったことは前述した通りである後者を反映する方法として、ファイルメーカー上に、表面／裏面注釈欄を設けた。この注釈欄は、当該シートに「チャート」「図表」「計算式」がある場合、「あり」をチェックすることになっている。この「あり」のチェックは、利用者に対する道標的機能と画像ファイルの確

認を促す作用を持っている。つまり、当該シートには、「チャート」「図表」「計算式」といったファイルメーカー上入力困難な記述が存在する。これら記述は、別ファイルとして保存(Microsoft Word)しているの、必要に応じて別ファイルをみてほしい、その際、画像ファイルとの照合も忘れないでほしい、といった案内である。他にも、[図10]に記されているような文章では表現しがたい記述についても(主人・妻・子どもに渡ってひかかれている大括弧)、画像ファイルとの照らし合わせが必要になると考えられる。この場合、コーディングシート上の記述は全て主人の欄に入力し、入力が難しい括弧が存在することをMicrosoft Word上に「コメント」として残した。「チャート」「図表」と同様に、注釈欄の「あり」をチェックし、当該シートには「コメント」と入力した⁽⁵⁾。その存在を注釈に「コメント」として示し、画像ファイルとの照合を促している。

以上の作業を通して、ファイルメーカー上でのデータ一元管理が可能となったといえる。これら作業を通して感じたことは、質的

(A-6) 収入給与	本俸	%
生命保険	名義	月の掛金別々給与
	主人	{ 夫: 30万合計で 月7000円、満期時 取100万 子: 2万、50万+50万 計200万
	子供	
	その他	
(A-7) 収入給与の中、生活とどう分けてるか	山にいれ、手帳	
しゆせどうけり	No	
家計支出項目	食心のみ以外。	

[図10] 入力し難い括弧の存在

データのテキスト化において、入力困難な状況が予想されても、最初からファイル一元化で入力することを試みるべきである。どうしても入力し難い記述に出会った場合に、初めて、ファイルの二元化を考えるべきである。質的データといった場合、私たちは多様な「自由記述」の存在を連想し、その整理の難しさを思う。そのために、質的データのテキスト化を諦めてきたのが、これまでの状況であったのかもしれない。本作業においては、フォーマット化された回答が用意されている量的データの整理に対抗するかのように、当初は機械的な処理方法を講じた。

しかし、その作業はすぐに破綻する。なぜならば、自分たちの入力作業には、本来質的データが持っている内容と、そのデータが含むうるリアリティを捨象してしまう危険性があるということに気づいたからである。確かに、量的データに比べれば、質的データにおける回答のありようは多様であり、煩雑であることは否めない。しかし、内容を分析していくことにより、[図9]のように記述は分類可能であり、それに従いルールを定めて入力していけば、データの一元管理は可能なのである。質的データである以上、質的データが持つ長所を活かし、定型性と可能な限りの再現性を目指し、入力作業を講じる必要がある。

また、以上に示してきたことは、調査票およびコーディングシートの作り方の問題にもつながる。本作業は、夕張調査のデータをもとに行われた。当時、調査メンバーは、このような形でデータがデジタル化がされるとは考えていなかっただろう。よって、調査票・コーディングシートの記述の仕方が、入力作業に適したものではなかったといえる（ただし、これは批判すべき点ではない）。

今後、質的データのデジタル化を希望し、調査データを残していきたいと考えるならば、調査票やコーディングシート作成段階において、入力作業を想定しながら、欄を設け

ていく必要があり、調査データも入力に適した形で記述していくことが望ましい。

3.4 追隨的に生み出されていったルール

作業を振り返ってみると、「入力ルール」は、記述の定型性と可能な限りの再現性を検討することによって生成されていった。それはやはり、フォーマット化されていない質的データの特异性によるものであろう。また、記述の分類を、記述の内容を読み解きながら進めていくことは、調査者と被調査者の往復によって作り上げられた調査データが含むうるリアリティというものを描き出す作業につながっていたからともいえる。

データのデジタル化に関する他基本的な側面は、これら作業に追隨していく形で生成されていった。以下、生成されたルールについて、いくつか記していく。

作業開始当初、機種依存文字を利用しない、英数字は半角利用といった、基本的な点について最低限のルールだけ用意していた。記述の分類が成立し、それに従ったルールで作業を進めていくことで、作業効率が上がったと同時に、基本的な点にも目が届くようになった。

例えば、頻繁に登場する記号の存在である（cm, CO₂ など）。機種依存文字を利用しないと決めつつも、「原票に忠実に」という合言葉が頭から離れず、当初ファイルメーカーに入力する際、機種依存文字で入力してしまっていた。これら記述は、「CO₂」であれば「CO2」というようにベタ打ちで入力してくようにした。入力チームのメンバーが、機種依存文字で入力した場合、公開時にどのような障害が生じるのか、といったような入力作業の最終地点に関する予測がまったくつかなかったため、安易に入力を進めてしまっていた点に反省が残る。

コーディングシート上、選択肢を入力する箇所において、調査時用いられたカード上の

記号をそのまま入力したものがみられた (①や㊸など). これらは明らかに機種依存文字であり, 他入力方法を検討しなければならない. 単純に「1」「A」として入力してしまうと, 他記述に用いられている英数字と混同してしまう可能性がある. 例えば, 家計費のところの貯金額は, 選択肢を選んだ上で, 額も記入され, コーディングシート上, 「No.㊸ 28万円」というような記述となっている (「No.」はもともとコーディングシート上に印刷されている). この場合, ファイルメーカー上, 「3 28万円」と入力してしまうと, 「3」が何を示しているのかわからない. そこで, コーディングシート上の No. を利用し「No. 3 28万円」というようにファイルメーカー上入力した. 数字同士の間には全角スペースの多用を心がけ, 選択肢の数字と額としての数字の区別を行なった.

調査者によっては, 記述において文章の流れを説明するために, ①②など利用している人がいる. この場合も, ①②に関して, 他入力方法を検討する必要がある. 当初 () を用いて, (1)(2)としてはどうかという案が出たが, () も, もとものの記述で利用されている場合があり, 混同してしまう可能性がある. 文章中に用いられた①②に関しては, [1][2]とすることにした. もとものの記述と重複しない記号を探し出すのに苦労した.

判読不能な文字に関しては, 画像ファイルを確認し, 解読を心がけたが, それでもわからない場合は入力チームで相談しあい, 尚且つ判明しない場合は, 「□」として入力した. また, 誤字脱字も何個か見受けられたが, そのまま記入した.

以上示した点も, 3.3の最後の部分で述べたように, データのテキスト化を考慮する場合の調査票作成にかかわってくる点であろう.

4. 入力作業が有する機能と入力者の役割

以上示してきたように, われわれは質的データのテキスト化における「入力ルール」を生成してきた. 入力ルール生成過程を通して, 以下の点がみえてきた.

入力作業開始当初, われわれ入力チームは, 本作業を, コーディングシート上のアナログな文字群を, 機械的なルールをもとにデジタル化するという単純作業として理解していた. 言い方は悪いが, 「入力さえすれば良い」という理解であったといえる. しかしながら, 作業を開始してみると, 思いのほか作業は困難を極めるものであった. 作業中生じた困難性を, 内容をもとに記述を分類し, 入力ルールを再構成したり, 入力フォームの変更を行ったりして, 解決してきたわけだが, その過程で私たちは, 入力作業を通して担わなければならない自分たちの役割について考えるようになっていた.

われわれが行なっているデータの入力作業は, 質的データの文字群をテキスト化することにより, データを保存するということが第一の目的である. この点は, 量的データの保存とさほど変わりはない. さらに, 質的データであるが故の, 第二の目的が登場する. それは, 調査者と被調査者との往復によって作り上げられた調査データの内容とデータが含むうるリアリティを, 適切な形で二次分析利用者に, 大きく言えば後世に伝えていくことである. これらの作業は, 単に字面を入力フォームに移行するということだけでは成立しない. 記述内容をひとつひとつ紐解き, 内容に適した記述分類, 入力ルールの作成が重要な意味を持つ. データ入力方法・保存方法にしても, 利用者の立場を考え, 利用しやすい方法を検討することが求められる.

以上, 作業を通してみえてきたわれわれの役割は, データ入力者であると同時に, データが存在するに至った調査者と被調査者との

往復による営みを利用者および後世に伝えていくという橋渡しの役割を担っているのである。

しかしながら、若干の疑問は残る。二次分析の利用において、誰が、どのような形で、どのような目的でデータを利用するのかということが不確定なまま入力作業を進めなければならなかった。特に本作業で用いたデータは、夕張という地域の歴史を描き出す上で重要な資料になりうる。よって、データの利用に関しては、一般住民も想定される。この場合、われわれの入力作業が、利用者全体にとって共有可能かどうかという点については、疑問が残らざるを得ない⁽⁶⁾。

さいごに、付録的ではあるが、質的データ入力からみえてきた、質的調査者への調査心得を述べたい。われわれの作業から、質的データのテキスト化作業は、決して不可能なものではないことがわかった。但し、データのテキスト化を望むのであれば、3.3、3.4のさいごの部分に述べたように、調査票を入力に適した形で作成していく必要がある。また、〈1-2〉〈2-2〉のような線や矢印などの存在によって客観的な関連性が認められない記述に関しては、入力において、どうしても入力者の主観が入ってしまう。それを避けるためにも、調査者は設問への関連を明確に表示する必要があるだろう。〈1-3〉〈2-3〉のような File Maker でもテキスト形式でも入力困難な記述については、現場や当事者のリアリティが読み取れる記述であるので、大いに採用すべきであるが、入力方法を念頭に入れた上で取り組むべきである。

以上、簡単ではあるが、質的調査者への調査心得を述べ、むすびとしたい。

【注】

- (1) 筆者は、平成19年4月から札幌学院大学社会情報学部 SORD プロジェクト「『社会調査史の博物館』としてのリージョン拠点データアーカイブの構築」(平成18年~21年度文部科学省研究費補助金基盤研究(B))のリサーチ・アシスタントを務めている。本稿は本プロジェクトの研究成果の一部である。このような貴重な機会を与えて頂き、札幌学院大学社会情報学部に感謝申し上げたい。
- (2) 入力チームは、平成19年7月に結成された。メンバーは、筆者のほか、吉野航一(北海道大学大学院博士課程)、寺沢重法(同修士課程)、計3名である。本稿の内容は、吉野氏、寺沢氏との共同討議・共同作業が重要な意味を持っていることを、ここに記しておく。
- (3) 大間レベルでのかかわりとは、[図3]の図を用いて説明すると、現物給与全体に対して述べられている記述のことを言う。
- (4) 小間レベルでのかかわりとは、[図3]の図を用いて説明すると、現物給与の水道の所だけに対して述べられている記述のことを言う。
- (5) [図10]内の括弧は、おそらく家族保険の加入を示していると思われるが、この判断は入力者の主観によるところが大きく、「コメント」には、記述に対する判断は入力せずただ「~に括弧あり」という記述にとどめている。
- (6) 市民への公開を考えた場合 Microsoft Word は、一般の人でも利用しやすいソフトだが、ファイルメーカーの場合は、多少数居の高さが気になる点である。