

文化資源学の創設をめざして

For Constructing the Cultural Resources Study

原 純輔

今、新國先生のお話の中で、私の名前も出てまいりました。先程、田中先生ともお話ししましていろいろのことを思い出しました。学部創設のころからいろいろと関係があったのですが、あまりお役にもたてず、今日は何かそのことを責め立てられているような気がしています。その罪滅ぼしという意味で、いくつか話題提供をさせていただきます。具体的には、お配りしました資料に沿いながら、1.「文化資源学」プロジェクト、2.「非定型データ」の処理・分析法、3.調査原票の保存と利用、という3点についてお話をしたいと思います。



原 純 輔 氏

1.「文化資源学」プロジェクト

さて、「文化資源学の創設について」というような、非常に大仰な題を掲げたのですが、中身は、私どもが今、自分のところでやっておりますプロジェクトの紹介を中心にして、少しお話をしてみたいと考えております。

私は現在、東北大学大学院文学研究科の、4つの専攻のうちの1つである人間科学専攻というところに所属しております。この専攻で、文部科学省（文部省）から教育研究拠点形成支援経費という特別予算の配分を受けまして、「文化資源学プロジェクト」というものを進めております。これは昨年度から開始されたもので、一応3年間にわたって実施す

ることになっています。御存知のように、こういうお金を下さいというときには、なにやら長ったらしい、なんとかに関する研究というのをつけるのが普通です。このプロジェクトにももちろんそういうものが付いているのですが、私も覚えきれないのと、ちょっと読みあげても恥ずかしいような企画なものですから、一応略号として「文化資源プロジェクト」という名前で私自身は推進しているということです。

この経費の趣旨はよくわからないのですが、いわゆる COE ということ盛んに文部科学省の方がいってございまして、それと関連があるのだと思います。「形成支援経費」というのだから、何かをつくるぞといえいい

のかなということで、応募をして、何とかうまい具合に対象になりました。全国でだいたい30専攻くらいを対象にしているそうで、金額は私どもの場合は年約1800万円ほどです。これには、一般の研究費、設備備品費の他に、ポスドク対策として学位取得者を非常勤研究員（講師）に採用する費用とか、あるいは外国の研究者を呼ぶ費用とか、そんなものも含められておりますので、使いようによっては非常に有効なのではないかというふうに考えております。

さて、御存知のように東北大学というのは東大ほどは長くないとはいえ、非常に長い歴史を持つ大学です。それぞれの研究室には、非常に大量の研究資料というのでしょうか、たとえば調査データであるとか、実験データであるとか、こういうものが蓄積されております。ただし、蓄積というべきか退蔵というべきか死蔵というべきか、この辺が怪しいと思うのですけれども。

具体的にはどんなものかといいますと、人間科学専攻には4つの研究室（講座）がありまして、それぞれが多少違った資料を持っているのです。

1つは、ここにおられる高橋先生のご出身講座である社会学です。ここでは農村調査、とりわけ東北地方を中心とする農村調査というものがずっと長いこと行われていて、主に聞き取り調査が中心だと思うのですけれども、これらの資料が蓄積されております。これらの調査がどういう形で進められたか、研究室としてどういうデータがあるかというようなことが曖昧になっています。それぞれの先生方に古い記憶を呼び起こしてもらいつつ、家捜しでもしないと実際のところはわからない状態です。

それから私が属しておりますのは、行動科学ということなのですが、これはまあ実態は社会学でありまして、社会学の中でも数理的方法や計量的方法を中心にする社会学とい

うふうに位置づけています。歴史はそんなに古くないのですが、いろいろなかたちでの統計分析を目的とした統計調査がそれなりに蓄積されています。まあ、割合質のよいものが多いと自負しております。

次に心理学です。これは非常に長い歴史を持っておりまして、一方で心理学実験の資料であるとか、他方で社会心理学の分野では、主に学校を対象としたものが多いと思うのですが、多くの調査資料が蓄積されています。

それから4番目の研究室は、人間文化学という、何だかわけのわからない名前なのですが、実体は2つの分野で構成されております。1つは文化人類学で、非常に様々な、東南アジア、あるいは東北アジアを中心とした、さまざまな調査のフィールドノート類をもっております。もう1つは宗教学あるいは宗教社会学の分野なのですが、いろいろ活発に調査等を行っており、その中でも非常にこの研究室が自慢にしておりますのは、葬祭儀礼をビデオで撮影したものや、あるいは8ミリフィルムに撮ったもので、これは相当大量に存在しています。

それからもう1つ、これは人間科学専攻ではないのですが、人間科学的な意味で非常に貴重だと考えておりますのは、国語学の研究室というのが他の専攻にありまして、ここでは東北地方中心とした方言調査というものを非常に精力的にやっております。このデータが非常に大量に存在しております。

こういうものが、先程退蔵とか死蔵とかという悪口をいいましたけれども、そういう状態で結局使われなくなるのは非常にもったいないということを、私は前から思っておりますので、このプロジェクトを始めてみたわけです。どんなことをやるのかというと、レジュメにも書いておきましたが、それぞれの研究室で蓄積されてきた研究資料の相互利用を促進をしたいということです。

更に、ただ単に相互利用を促進するだけで

はなくて、私はそのことによってそれぞれの領域における調査であるとか、実験の方法の改善を図りたいというふうに考えています。具体的にいいますと、たとえば、同じ調査といいましても、一方に私どもがやっておりますような統計調査があるのですが、もう一方で心理学などでやっております、学校を中心にした調査があります。その心理学の調査データは、なかなかおもしろい項目を含んでいるにもかかわらず、いわゆる社会学の方から見ると使いでがない。それは何故かという、回答者の背景に関する情報がデータに含まれていないのです。学校で調査をやるということになると、重要であることがわかっていても調べられないということもあるのですが、重要だということをわかっていないのではないかなという気もしまして、相互に利用する中でこういう点の改善を図りたいというふうに考えているわけです。

具体的には、1番目にはデータの整備を進めるということ、それから2番目には、相互利用をするといいましてもそう簡単ではないわけで、どうやっていくかという方法論の研究が中心になります。その上で、この経費の趣旨であります、何かをつくるという意味での新しい総合的学問分野というものを創設するための取り組みということです。

新しい総合的学問分野ということになりますと、名前をつけなくては駄目なわけで、そこで「文化資源学」という名前をつけたわけです。しかし、これは御存知の方も多いと思いますし、それから後で佐藤健二さんの方から話があるかと思うのですが、実は東大で既に文化資源学という独立の専攻が発足しております。私がこの名前を考えたときに、それを知らなかったというわけではありません。知っていたにもかかわらず、こういう名前をつけたのには事情があります。

実は、こういう経費の計画だとか要求というのは、ある日突然上の方から申請用紙とい

うのが回ってきて、しかもそれを1週間後くらいに計画を出せというような話になるのです。普段ですとそんなものは出せないというふうに断るのですが、この時は研究科長から、COEを目指すのだから各専攻は必ず1つは出すようにというような命令がありまして、仕方なくというか慌てて出しました。その時に名前をいろいろ考えたのですが、どうも思い浮かばなくて、まあ東大にあるけれどもそれとは違うことを言えばいいだろうというので、この文化資源学という名前を拝借したのです。

それで、こちらで今日お話しをするようにというふうに言われて、佐藤健二さんも来るとし、ちょっと格好が悪いなと思ひまして、何人かの人たちと別な名前はないだろうかと思ひに考えました。しかし、どうもよいアイデアがないものですから、まあとりあえず「仮称文化資源学」ということで、このまま使わせていただいたわけです。

先程相互利用というふうに言ったのですが、相互利用というのは2つの側面があると思うのです。1つは各研究室、あるいは領域単位で世代間の相互利用というのですか、世代間の継承というのでしょうか、この側面。それからもう1つは、分野、領域を超えた人たちでの相互利用ということです。

私どもも、もちろんその両方が目標ではあるのですが、現実には前者、つまり世代間の継承ということの方が割合簡単なわけで、これが先行しております。後者については、もしかしたら今日は後者についてお話しすればいいのかもしれませんが、そこまではなかなかまだ手が回りきっていないというのが現実です。

具体的には、まずデータの整理をやることから手をつけ始めました。またその一環として、各研究室が所蔵している資料には、どんなものがどれくらいあるのかというリストづくりと、一部資料のデジタルデータ化、これ

がまず今のところ進んでいます。この各研究室が所蔵している資料のリストづくりというのは、最初からきちんと計画的にやっておられるような先生方から見ると、アホみたいだと思われるかもしれませんが、とにかくこれが長い歴史を持つ研究室では大変なことで、たとえば心理学なんかの例でいうと、去年1年かけて、ようやく戦前の分が終わって、今年は戦後をやりますというような、そういう状態です。

これができたのは、ちょっと余談になるのですが、先程申し上げましたように、この経費で非常勤の研究員を雇用できたからです。各年度に1人というふうに言われていたのですが、去年は発足の年で、お金が来たのは9月だったでしょうか、文部省の方も慌てていたと見えて消化に困ったらしく、今年は特別に5人どうですかというようなことをいつてきたのです。それで5人、ちょうど各分野に1人ずつ採用しまして、専らその非常勤研究員にリストづくりをさせたものですから、割合スムーズに仕事が進行したという、そういう事情がございます。

そういうリストにもとづく、一部資料のデジタルデータ化ということなのですが、これはいろいろな形で進めています。そうした中で、これは期待通りの成果が上がったなと思いますのは、心理学の分野なのです。心理学にはいろいろな分野がありますが、いわゆる生理的な現象である脳波だとか、心拍数だとか、そういうようなものを測定して進める実験心理学の分野では、かなり以前から、おおよそ27,8年前くらいからではないかというのですが、専用のデータレコーダーが使われてきました。つまり、測定して紙に筆記するのではなくて、専用のテープに自動的に記録されるわけです。このテープが大量といってよいくらい所蔵されているのです。

データレコーダーは非常に高価なものだったようです。ただこれが、だいたいどこでも

そうなのですけれども、機械が古くなってくるとか、あるいは新しいコンピュータ関連のいろいろなものが出てくることによって、半ば放置状態になるのです。ところが、今度お金が来たのを機会に、まず故障して使えなかったデータレコーダーを修理して、このテープを再生できないだろうかということを考えまして、やりましたら、完全ではないのですけれども、ある程度テープが再生できるようになりました。さらにこのデータをコンピュータに取り込むということによって、そこからグラフをつくるとか、表をつくるというようなことが可能になった。

これを何に使っているとかというと、研究資料というよりは教材に使うのです。つまり、実験心理学の歴史というのを、ただ教科書に記載されたことを教え込むのではなく、身近でやられた実験データの解析例を目の前に示すことによって、非常に教育の成果が上がったとのことなのです。私には教育の成果が上がったかどうかわからないのですが、見せてもらいまして、30年近く前の実験結果が目の前に再現されるのに非常な感動をおぼえました。

2. 「非定型データ」の処理・分析法

さて、このプロジェクト全体について、まだいろいろといい足りないこともあるような気がするのですが、あまり時間をそれほどばかりに使うわけにはいきませんので、2番目のお話に入らせていただきます。

2番目のお話は、非定型データの処理・分析法というテーマです。「非定型データ」というのは私がよく取り上げるテーマで、またあの話かというふうに思われる方もあろうかと思います。ここまでは非常に一般的な形で文化資源学プロジェクトについてお話してきましたが、私の専門の1つであります非定型データの処理・分析法との関係で、私自身はこのプロジェクトに積極的にかかわってみようかと考えておりますので、その話をさせて

いただきたい思います。

「非定形データ」とは何かといいますと、典型的なのは、調査における自由回答です。それから佐藤健二先生なんか、主に分析といますか、対象にしておられるようなものは、みんなこれに入るのではないかと思います。自由回答の例を考えてみますと、一般の調査では、回答者の回答というのは選択肢が示されまして、そのことによって、回答のベクトルというのでしょうか、これを調査者の側で限定するわけです。自由回答の場合にはそれが無限定であったり、あるいは、限定されていても非常に緩い。「何々について」だともう無限定なのですから、「何々のこういうことについて少し話して下さい」みたいな、やや限定した形をとることもあります。そういう形で進められるのです。私自身は、そういうものに関心があるのですが、その関心というのはややちょっとひねくれております。というのは、そのデータの内容と同じくらい、あるいはそれ以上にその処理・分析法ということに実は関心をもっています。

処理・分析法とはどういうことかといいますと、私自身は、実証研究をやってきた人間ですけれども、基本的に、統計的な分析を自分の研究方法としてずっと用いてきました。ですから、自由回答にしる何にしる、非定形データを最終的には統計分析、具体的には多変量解析でしょうか、これに持ち込みたいというふうに考えています。それ以外の方法よりこの方法が優れているとか、統計分析にすべきだとか、そんな考えは全然ありません。それは私は選択の問題、趣味の問題だというふうに考えておりますが、とにかく統計分析に持ち込みたいというふうに考えております。

この統計分析をやるためには、実はデータについて一定の制約、要請があるわけでした、個体×変数という形のデータ行列の存在が前提なのです。個体は調査（測定）の単位ですから、個人の場合もあれば世帯の場合もある。

あるいは地域みたいなものもある。

個体の数は必ず複数ですが、変数の方は、たとえば何かの平均値を求めたいという場合であれば、1でいい。相関変数を求めるのであれば2、多変量解析をやる場合には普通はもうちょっとたくさんある。調査データの場合、たった1つの分析をやるわけではないですから、普通はたくさんの変数が並んでいる。こういうデータが必要なわけで、そして、それぞれのセルには特定の個体をもつ特定の変数の値というものが、一義的に与えられている必要があるのです。かつ、これは矩形の行列であって、ある個体は変数が100個くらいあるのだけれども、あるものは2個とかでは駄目なわけです。もちろん、例外的な欠損値は許されるでしょうが、基本的に矩形のデータ行列の存在が前提になるわけです。

私がいくら多変量解析だとか統計分析だとかといっても、これがなくては何もできません。選択回答式の調査の場合には、調査結果をコンピュータ入力すればデータ行列となるように、普通は調査票が設計されています。しかし、自由回答の場合には、そのままではデータ行列にならないわけで、そのデータ行列をどういうふうに作っていくのかというのが、私自身の非常に強い関心の対象です。

それで、今、非定形データの典型として自由回答をあげたのですが、レジュメにも書いておいたように、私自身が非定形データと呼んでおりますものには、大きくいいますと2つのタイプがあります。1つは、自由回答がありますとか、あるいは、作文であるとか、演説であるとか、誰かの手記であるとかというような、文章の形をしたデータ。あるいはその他、私自身はそこまでちょっと手が回らないのですけれども、絵だとか音声だとか、そういうような形のものも含まれると思います。

それから、これの仲間なのですから、皆さんがちょっとお気づきにならないものも

あります。それは、本来は選択回答にすべきであるにもかかわらず、分類カテゴリーが細かすぎて（多すぎて）、選択肢のリストが作れない、あるいはそのリストを見せても、回答者がうまく選択できないようなものについても、自由回答の形で回答を求めることがある。

たとえば職業の場合がそうです。職業は、通常は1桁くらいのカテゴリー数に分類して用いることが多いのですが、私どもの場合、288くらいに分類して使うこともある。その288職種の見やすいリストを作るのはもちろん無理ですし、仮に作ることができたとして「あなたの職業はこのうちのどれですか」といっても、回答者は困ると思うのです。これは自由回答で、「あなたはどのような仕事をやっているのですか」とたずねて回答を細かく記録し、それをあとでコーディングしてやるという手順をとります。

今述べましたものは本当に形をなしていないものですから、「不定形データ」と呼んでいるのですが、もう1つ、「定形不定長データ」と呼ぶものがあります。

OHPで示したのは、実際の調査票の一部で、職業経歴に関する質問の部分です。私どもの調査の場合には職業経歴をどういうふうに見つけるかといいますと、最初についた仕事から現在までを切れ目なく、勤務先・従業上の地位・仕事内容・職位が変わるたびに、全部細かく記録します。この調査票でいえば、一段一段がそれに対応しています。

そうすると、ここにおられる若い先生方が、もし定年までこの大学におられるとすれば、たった1段で済んでしまうのに対して、同じ大学教師でも、私のようにいろいろな大学を変った場合には、何段も要することになります（ちなみに、講師・助教授・教授などの職位は、職業分類上は同一のコードが当てられています）。私どもが95年にやった調査では、16段という猛者もありました。話を簡

単にするために各段を1つの変数と考え、回答者によって変数の数が異なってきます。要するに長さが決まらないわけです。そういうものも、ちょっと拡張のしすぎかなとも思うのですけれども、非定形データに含めています。

ここからデータ行列をどうやってつくるかということなのですが、一番単純な、これまでやられている方法は、たとえばこの不定形のデータ（自由回答）でいいますと、まず全部の回答をよく読むのです。それでウンと唸って、「うん、これは5つくらいのタイプに分けられる」というふうに決めて、これはAタイプ、これはBタイプというふうに分類していくというのが一番素朴でよくやられている方法だと思います。

それから不定長データ、たとえば職業経歴の場合でいいますと、最初に就いた職業と、最後の、現在の職業と、この2つで経歴を代表させて、初職×現職というクロス集計表を作って分析するということをよくやります。そうしますと不定型データの場合は、データ行列でいうと、どのタイプかということの1変数。それから、不定長の場合でいうと、初職と現職という2変数で統一されますから矩形のデータ行列になるわけなのです。いくら何でもこれでは、非常にいろいろな情報が入っていると思われる、せっかく集めてきたデータのかなりの部分、あるいは大半を切り捨てることになってしまいます。だからそういう豊かな情報を切り捨てないかたちで、さっきのデータ行列をどう作っていくかというのが私の研究課題なのです。

あまり自慢できる成果とはいえませんが、私の書きました文章を御紹介しておきますので、お読みいただければ有難いと思います。

1番目は、私が実際にやった職業経歴の分析、2番目は、世界各国から集めた子供の作文の分析、それから3番目は問題提起みたいなものです。

原純輔, 1979. 「職業経歴の分析」. 富永健一編

『日本の階層構造』東京大学出版会: 198-231.

原純輔, 1987. 「子どもは日本をどうみているか」.

辻村明・古畑和孝・鮑戸弘編『世界は日本を
どう見ているか』日本評論社: 205-218.

原純輔, 1988. 「非定型データの処理・分析」.

海野道郎・原純輔・和田修一編『数理社会学
の展開』数理社会学研究会: 461-471.

さて、多くは今述べた程度のことしかやられていないのですが、実はそこまでやればよい方でありまして、その情報が、非定形だというだけで、こういうデータ行列のようなかたちで蓄積することができないというだけで、他の調査結果とは切り離されて死蔵されてしまう。たとえば、皆さんも、もしかしたら耳が痛いのではないかと思うのですが、調査票の末尾に、この調査について何か御意見や御希望がありましたら自由にお書き下さいという設問がよく付け加えられています。そういうふうに書いてもらっても、それは捨ててしまうか、分析も何もしないというのが非常に多いような気がします。これもみかたを変えれば非定形部分の切り捨てです。

私のかかわった別の例をあげておきます。厚生省が1965年以来でしょうか、10年に1回ずつ、原子爆弾の被爆者実態調査というのを実施していたのです。東大の福武直先生との関係で、85年の調査にだけ私は協力しました。これは正直にいいまして、科学的調査というよりは政治的調査でありまして、いろいろな力関係から、被爆者手帳を持っている人全員を対象にした全数調査をやりました。郵送調査であるにもかかわらず回収率が90数パーセントでした。これは、実は被爆者団体協議会という団体が、後ろの方で回収に協力するようにという圧力をかけた結果なのですけれども、36万人になります。

この中で、被爆して死亡した家族の思い出と、それから被爆者の立場からの意見という

のを、B5版1ページくらいずつを充てて書いてもらいました。驚いたことに、ほとんど全員が書いています。

本当に悲惨で感動させられるものなのですが、厚生省は、これについてどう扱っているのかというと、今の身体の状態はどうだとか、どこで被爆したかみたいな、前半の選択式の部分についての簡単な報告書を出しただけで、残りの部分については一切目をつむったままです。厚生省としてはやりたくなかった調査なのです。

私は腹に据えかねたものですから、私が当時横浜国大にいた関係で、神奈川県の実答者の分だけでも詳しく分析してやろうと思い、調査票を借り出しました。いまだに返却の要求がありません。それくらい、軽視されている調査なのです。

残念ながら分析はできておりません。しかし、すぐに利用するのは難しいかもしれないけれども、何とかうまく蓄積して後で使えるようにできないかということで、いろいろな方法を検討してみました。その可能性の1つとして、このあとでお話します調査原票というものを、そのままデジタルデータといいますか、画像データに変換して、蓄積して利用するという方法があるのではないかと考えるに至ったわけです。

他の方法でもいろいろ検討したのですが、すぐに思いつくのは、文章をテキストファイル化することでしょう。被爆者調査についてもやってみました。テキストファイル化しますと、文章データの分析法みたいなものの研究は非常に盛んですから、さまざまな分析プログラムが提供されておりまして、こういうものを使えば、いろいろなことができると思うのです。

ところが、皆さんのレベルではないのですが、ごく普通の社会学者のレベルでテキストデータを扱おうとすると、いろいろ使いにくいというか、問題があります。たとえば、今

の被爆者調査の例でいいますと、一方でそういうテキストデータがあるのですが、もう一方で既にコード化された選択回答のデータがあるわけです。そして、多分おもしろい分析は、そういうテキスト化されたデータとコード化されたいろいろな属性とか何とかを組み合わせるとのことだと思うのです。テキストファイルとコード化されたデータを、どう繋いでどう蓄積していくか、またどう利用していくかという技術的な問題が生じます。コード化されたものがデータ行列の形をしていますが、テキスト部分は長短いろいろのものが混在している。それを、たとえば一般の社会学者が、SPSSとかSASのレベルのプログラムを使って分析できるかという、そのままではできないのです。そんな問題点もあるということだと思います。

それで、解決の先延ばしではあるのですが、とりあえず1つの可能性として私が考えたのは、デジタルデータ、画像データとしての蓄積なのです。ここでは、非定形データの処理・分析という側面からそういう方法があるのではないかというお話をしたのですが、調査原票を保存するということの重要性を、別な側面から痛感したという経験を私は持っておりまして、その辺から次のお話を始めたいと思います。

3. 調査原票の保存と利用

「社会階層と社会移動全国調査」通称SSM調査と呼んでいる調査に私はずっとかかわっております。この調査については名前くらいはお聞きになったことがあると思います。1955年以来10年に1回ずつ、社会学研究者が続けてきた職業階層の問題を中心にした調査です。一番新しいものは1995年。一番古いものは1955年で、この間40年の時間の経過があるわけですが、たとえば集計や分析の方法をとってみても、非常に大きくこの間に変わったわけです。

たとえば、1955年は、基本的には手集計で進められました。道具としてホールソートカードという手集計用のカードが使われたと思います。B6版くらいの大きさでしょうか、このカードに回答を転記するのですが、周辺には多くの穴が開けられています。次に、回答にしたがって紙の特定部分に切れ目を入れて切り落とす。そうすると、穴の残っている部分とそうでない部分とができてきます。たとえば、ある質問の回答がAであったら穴が残ったままであるけれども、回答がBであったら残っていないというように、針金をその穴に通して持ち上げ、カードを振ると、切ったものはストンと落ちる。切らなかったものは残る。それで数を数えるというわけです。

現物を持ってくれば良かったのですが、興味がお有りの方は福武先生が書かれた『社会調査』（1958年）を読んでいただきますと出ていたように思います。1955年当時、小学生でありまして、さすがにこの調査に私は参加してはおりません。

2回目、その10年後になりますと、この時は、コンピュータも使えなくなかったのですが、基本的に手集計だったと思います。ただ、パンチカードは既に使われていました。カードソートマシーンという郵便葉書の自動分類機の元祖みたいな機械を使います。パンチカードをその機械にセットし、特定の1コラムを指定してスタートボタンを押すと、コラムのパンチ内容を読みとって、自動的にカードを分類してくれるわけです。最初の頃は分けられたカードを人手で数えていたのですが、その後、機械的に何枚というのがちゃんとカウンターに出るようになって、ずいぶん楽になりました。

しかし、単純集計でコードが0から9までの1桁だったら、これは非常に便利なのですが、2桁になったらどうするかということをやっと考えていただきたいのです。まず十の位に関して区分けをする。次に、区分けさ

れたそれぞれのカードをもう1回マシンにセットし、一の位について区分けをしてからその数を数える。クロス集計をしたり相関係数を求めるということになると、2つの組み合わせですから大変なことなのです。まあそういう時代です。

第3回の1975年になりますと、ようやく各旧制帝大を中心にした大型計算機センターが整備されてまいりまして、そこで汎用のコンピュータが使えるようになった。それから75年の非常に大きいトピックは、今、非常によく使われております、SPSSというプログラムパッケージが導入されて、これを社会学者でも使えるようになったことです。それまでは手計算の世界ですから、普通の社会学者が多変量解析をやるなどというのは本当に想像もできない時代だったのです。それができるようになった。

1986年に、次のような簡単な文章を私は書いております。

原純輔, 1986. 「数理社会学・計量社会学と電子計算機」. 『東北社会学年報』15巻, 東北社会学会: 23-41.

これは社会学の世界にコンピュータがどういうようにして導入されてきたかという、その歴史を書いたものです。その中で、私だけではないのですが、何人かの人間がSPSSを日本に導入するために、いかに苦闘したかということが書かれておりますので、一度お読みいただくと、SPSSが使われる場合も感謝の念が高まるのではないかと思います。

それから第4回になりますと、1985年ですから、パソコンもかなり普及してきたのではないと思われるかもしれません。ただ、たとえば先程のSPSSレベルのプログラムになりますと、これはまだパソコン用ソフトにはなっていないですから、やはり分析は計算機センターに行き、磁気テープを回して汎

用コンピュータで計算するという時代であったようです。

95年、第5回になると、これはもうパソコン全盛時代です。データ媒体はフロッピーディスクが中心です。過去のデータを分析するために磁気テープを学生にみせたことがあるのですが、それが何であるかということもわからないという者が、現在ではかなりおります。それから、各地の計算機センターでさえ磁気テープ装置が消えております。東北大学の場合でいうと、今1台しかないのです。ですから大急ぎでそれをフロッピーディスクだとか何かに移し替えるという作業を迫られております。SSM調査について詳しくは、次の文献をお読み下さい。

原純輔, 2000. 「SSM調査（社会階層と社会移動全国調査）について」. 原純輔編『日本の階層システム1—近代化と社会階層』東京大学出版会: xvi-xviii.

こうした長い経過の途中の、第3回（1975年）のSSM調査が終わった時のことです。その分析をやって報告等もして、「さて」と一息ついたときに、過去の第1回、第2回のデータも同じようにSPSSなどを使って分析ができれば非常によいのだがという話になりました。

ところが、そのデータがコンピュータで使えるような形にはもちろんなっていません。それからコーディング自体も、たとえば職業でいいますと、先程のカードソートマシンでの苦労話と関連しているのですが、簡単に分類できる程度のカテゴリー数に抑えられている。これに対して、第3回の調査では288に職業を分類しました。同時に行った別の調査結果にもとづいて、それらすべてに職業威信スコアを与え、職業を量的な変数として扱うことが可能になり、重回帰分析やパス解析が行われるようになりました。

前の調査についても共通の分析をやるためにはどうしても、もう一回調査票をコーディングし直して、75年と同じレベルのデータをつくる必要があるのではないかという話になりました。1980年から81年にかけて、これは富永健一先生が中心になっていたのですが、その作業をしました。大学院生クラスの学生に手伝ってもらい、今田高俊さんや私などが指導して、夏と冬に一週間くらいずつ、乗鞍と伊豆の温泉にこもってやりました。

そうはいってもなかなかわかりにくいと思いますので、1955年の調査票のコピーを持ってまいりました。これは字が小さいので見えづらいと思うのですが、職歴の部分で、サインペンで書いてありますのは、80年から81年にかけて再コーディングをした結果です。仕事の中身をこの3桁、055とか057とかというように再コードしていきます。当時ですから、兵役などが混じっておりまして、そういう経歴をどう扱うかみたいなやっかいな問題もありました。いろいろ書いてありますのは、何回も何回もチェックしたその結果です。私はこのころ職業のコーディングばかりをやっておりまして、町を歩いている人を見ると、コード番号が思い浮かんできました。農耕作業者076とか、事務員は055とか……。大事なことは、これは自由回答だったからできたということです。

それから第4回、5回についても同じやり方でコーディングをやりましたから、55年から95年まで40年間にわたるデータが蓄積されて、時系列の分析が可能になった。

たとえば、今、社会的に大変話題を呼んでいますけれども、日本社会が閉鎖化状態を強めているのではないかという、東大の佐藤俊樹さんの分析などは、戦後40年にわたるデータがあればこそできるようになったといえると思うのです。

戦後40年だけではなく、さらにさかのぼ

ることも可能になります。1つだけ例をご紹介しますと、ここにお示ししたグラフには、25歳から34歳の人びとに関する階層の閉鎖性の推移が、1935年から10年間隔で描かれています。1935年と45年については、55年調査の職歴データを使って計算しています。つまり、1935年に25歳から34歳であった人びとは、55年には45歳から54歳になっているわけですから、55年調査における45歳から54歳の対象者の20年前の職業をみればよいわけです。1945年についても同様です。もちろん、この間には戦争と敗戦があって、1935年であれば調査対象になった多くの人びとが、すでに亡くなっています。ですから、さかのぼるといっても限界はあるのですが、まったく不可能というわけでもありません。

私たちの分析は、調査票が捨てられずにあったからこそ可能になったのです。そしてまた、分析の視点が変わるとか、あるいは新しい分析方法が導入されるというような、そういう展開が今後あるとしたならば、この調査票がもう一回コーディングされる可能性もあると思うのです。そのためには調査原票の保存ということが非常に望ましいというように判断されます。すべての調査とはいいませんけれども、やはり重要なものについては保存しておいた方がよいと思うのです。

ただし、ここに大問題があります。調査票用紙の劣化です。紙全体の劣化だけでなく、このSSM調査の場合、地点でいうと全国300地点にわたっており、各地点ごとに調査票を重ねて封筒に入れておいたせいで、封筒に含まれているワラか何かではないかと思うのですが、一番上と一番下の封筒に直接接触しているページが黄色く変色して非常に読みにくい。それから調査票をとめておくステープラーの腐食も進んでいます。

そこで、調査原票の画像データファイルとしての保存を考え、作業を開始したわけです。

1955年のデータから始めまして、これについてはスキャナーを用いた入力作業は一応終わりました。55年から始めた一番の理由は、劣化が最も進行しているということです。55年の調査票は1人分16ページから構成されております。対象者数が2014人ということですから、掛算しますと32,224ページ、30cm×25cm×40cmという箱で12箱になります。

これを画像ファイル化したのですが、ソフトウェアとして何を使うかが問題となりました。このごろ画像データでアルバムなどを作るためのソフトがいろいろあります。しかし、われわれのような大量データには向いていません。最終的に、野村総研が開発しましたパワーバインダーという事務用のソフトなのですが、これを使うことにしました。単純にいうと各ページをファイルとして取っておけばよいはずなのですが、1人分をまとめて取り出すとか、あるいは検索をするというようなことになりますと、統一的なファイリングという考え方が必要なわけで、野村総研のソフトは、この点が非常に柔軟にできています。

野村総研と画像ファイリングというのは何か変だと思われるかもしれませんが、これは書類のファイリングを電子化したいという、企業などの要請に応じて作られたもので、なぜ画像データかというと、テキストファイルだけでは駄目だからなのです。というのは契約書等の事務書類では、印影（ハンコ）が重要になってきます。これを含めた形でファイリングしようとする、どうしても画像データが必要なのです。

とにかくそれを使いまして、専用のファイルを作ってやりますと、何とCD1枚に収まります。もちろん、パワーバインダーが使えない場合もありますから、TIF形式のファイルで書き出すということもやっております。この場合にはCD10枚になります。

これまで調査票などというものは、ごく一

部の人間しか見なかったわけで、CDにしてもそんなに公開して配るといような予定はないのですが、それにしても見る人の数が増えるものですから、プライバシーの保護ということに気がつかいました。今回の作業では、調査対象者の氏名だけは付箋紙みたいなものを貼って、隠して入力しています。なお所用時間は、最初は1人分16ページで20分くらいかかったのですが、今は頼んでいる学生諸君も慣れてきまして、5分から6分でいくそうです。

他にもいろいろな問題がありました。今日は簡単に1、2点だけをお話したいと思います。何ととっても、もう45年くらい前のものですから、鉛筆で回答が記入にされているのですが、薄れて画面上での判読が困難だということが大問題です。それで、あまり極端に薄いものについては、鉛筆で文字をなぞって再入力をやっております。ただし、そこまでしていいかどうかはやや議論のあるところでしょう。

ただ、調査中に回答を聞いて調査員が記録するわけですから、必ずしも丁寧に書かれているわけではありません。なぞるといっても、どういう字が書いてあるかということがわからないと、なぞりようがないわけで、判読に非常に苦労することが少なくありません。

そこには癖だとか、あるいは走り書きだけということではなくて、それ以上の本質的な問題があります。ちゃんと書かれていても、調査当時の職業だとか社会、あるいは職歴や親の職業の場合には、もっと古い時代の職業や社会についての知識がないと判読できないことが結構多いのです。はっきり書かれていても、「字の間違いでないのですか」といい出す。

字の間違いでもそのままなぞるようにと指示しているのですが、学生諸君にしたら、間違った字をそのままなぞるのは、なんとなく不本意なのです。たとえば、「青果物小間物

行商」という記述に先日出くわしました。これは、小間物と行商というのはわかったらしいのですが、それと青果物が対になっているということが、学生はどうしても想像できなくて、何と読むのだろうと議論しているのです。私どもから前の世代になると、リヤカーか何かに青果物だとか小間物を載せて、一軒一軒行商して回るおばさんみたいなものが頭に浮かぶのですけれども、今の学生、あるいはその後の世代の人にとっては、それ自体がわからないものですから、間違じゃないかといわれたり、悪いときには字を直されて、せっかくの情報が改ざんされてしまう。そういう可能性もあるのです。

それから、もう1点、やや弁明めいたことがあります。とにかく計画を1週間くらいででっち上げさせられて、そしてお金が来て、さあそれから何をやりましょうかというような話だったわけです。年度も終わりに近づいてきて、慌てて作業を開始したのですが、実はこういうのをやっている人たちはいくらでもいるわけで、出版社で古書の復刻をやるとか、あるいは官庁で公文書を保存するとか、あるいは、今、大学の図書館なんかでもサービスの一環として所蔵物を画像ファイルにするとかというような経験をしているわけですから、その辺から事前にきちんとしたヒアリングをすれば、苦労もずっと少なかったかとは思うのです。

ただ、まだ年度がありますので、もう一度そういう専門家からのヒアリングをやりながら、改良を加えたいと思っています。「この調子でCDのコレクションを作って、それを抱きながら引退するのだ」といって、学生たちにはあきれられております。

友人にいわせると、それは原さんの趣味の世界だということになります。貴重な情報をあとの時代に伝えるのだと、大上段を振りかぶってはいるのですけれども、趣味的な要素も否定できません。まあ、楽しみながらそう

いう苦勞をしているということをお伝えして、あまりに散漫なお話で申し訳なかったのですが、一応これで終わらせていただきます。

司会（高橋）：原先生、どうもありがとうございました。

現在進めておられるプロジェクトの話から始まって、非定形データの問題、調査原票の再生作業のお話しなど、いろいろな観点を交えていただきました。特に私のようなフロッピーディスク以降しか知らない若手の社会学者としては、手集計作業のご苦勞などは感動的に伺いました。早速皆様の方から、ご意見、コメント等を頂戴したいと思います。

田中：田中です。大変おもしろくて、私もお話しを伺っていて、いろいろあれこれと気づいたことがあるのですが。

そのうちの1つで、確かに紙の問題、いろいろな資料の紙の問題というのは非常に大きな問題だと思うのです。真空パックにする装置は今、たぶん購入可能なレベルにあるかと思うのですが、真空パックは食べ物用にもっぱら使っていますけれども、精神的食べ物にも使えます。真空パックに入れておくと、材質が変わるのは酸化のためが多いかと思うのです。少なくともその部分は防げるかと思っています。

原：ただ、ちょっと素人でわからないのですが、好きなときに取り出して見るというのは、かなり慎重にならざるを得ませんね。

田中：今、私が言いましたのは、データ化してコンピューター入力するまでの話です。始終使うというのには難しいかと思いますが、でも、データ化するまでの保存としてとりあえず真空パックに入れて置いておいてというのは、あるいは有効かなという気がするのですけれども。

原：実は、私の研究室は、建物の裏に青葉山というのがあって、青葉山から伏流水が建物の下を、音をたてて流れているようなところ

なものですから、湿気が非常に多いのです。今回の作業をやった調査票は、基本的に東大で保存していたので、比較的乾燥しております。こういう作業ができるぐらいのかたちで、今まで残っていたというのは、ある意味で非常な幸運もあったと思うのです。

ですから、そういう意味では、当時そういうものがあれば良かったのでしょうかけれども、現実には先程ご紹介したような状態になっていたわけです。

田中：いまからでも。

原：そうですね。

田中：それから、また別のことなのですが、実は不定型データに類するものは、かなり前、25年にもなるかと思うのですが、扱ったことがあります。

実際の仕事はそこにいる新國さんがやってくれたのですが、研究論文の中のセンテンスを検索するという、そういう仕事です。センテンスは不定形なものです。そのセンテンスの検索は大変役に立ちました。

たとえば福井大学の院生諸君を対象に実験してみたのです。ある分野の総合報告がどのぐらいの期間で書けるだろうかという実験なのですが、データベースを使うとかなり早く総合報告が書けるようになったのです。他とどこが違うのかといいますと、まず論文の中にあるセンテンスを十幾つぐらいのカテゴリーで分類します。タイトルだとか、結論の部分だとか、あるいは課題意識のセンテンスだとか、それだけでは使い物にならないのですが、それともう1つ、検索するときには、今度はその中にあるターム、いろいろな言葉、それから文を分類したカテゴリーと、この二次元のデータで検索する。

そういう質の違った二次元のデータで検索する。しかし、データとして入れるときには一次元の分類で入れるというような方法も、1つの方法です。不定形データを扱う方法として役に立つのではないかと思います。もっ

ともその仕事は特定研究のお金が切れるとともに、金の切れ目が縁の切れ目で終わりましたけれども、今だとソフトも進歩しましたから、いちいち手で入力しなくてもできるだろうとは思いますが、まあそういうことにちょっと気がつきました。

原：今回のテーマと直接関係があるかどうか分からないのですが、非定形のデータの分析では、今田中先生がおっしゃったように基本的にはワードを単位とすることが多いと思います。言葉を取り出すという作業はある種客観性を持っているものですから、それを自動的にやるというような研究というのは結構進んでいるわけです。一方で、社会学者あるいは社会心理学者もやっていますけれども、一方で文献学やマーケティングリサーチなどでも研究が盛んです。ただ、私自身は、「非定形データの処理・分析」という文章を読んでいただくといいと思うのですが、今進んでいる、ワードを取り出して何とかというのには、あまり賛成ではないのです。ということかと言いますと、社会学という学問では、価値だとか意味だとか、非常につまえにくいものが主題になっていることが多いのです。それをワード、たとえば単語なんかで代表させられるかということを考えると、もうちょっと大づかみに価値だとか意味みたいなものを取り出していくにはどうしたらいいかという、その辺の方法の研究が、もう一方で必要だと思います。そういう方向で非定形のデータを処理して分析しましょうみたいなのが、私自身の主張なものですから……。

田中：それには私も全く賛成なのです。もう1つの部分として、私の場合には、いくつかのカテゴリーをつけ加える。これがポイントと思うのです。だからワードだけというのは一次元のデータで処理しようとするもので、それはちょっと無理じゃないかと思います。1つのセンテンス全体を表現するような何かのものが、もう1つ必要になってくると思い

ます。それがセンテンスを分類するカテゴリーとして私たちが使ったものだったと思います。

原：センテンスを単位とした分析というのは結構あるのです。ただ、作文の分析なんかでいうと、やっぱりひと固まりのものであるから、これをなんとか、ひと固まりのままなんとかできないかというような、そういう発想で私はやっているのです。

田中：それは、センテンスだとかパラグラフ、全体というものを特徴づけるようなものを考えてというのは、センテンスの自身の分類用に必要じゃないかと思います。

原：それも1つの次元ではなくて、1つのものを多次元的にそうやって、ある種の評価をして、それで今度は多変量解析に持っていきみたいなものは、基本的に私の発想です。

中澤：この学部の中澤と申します。高橋さんと同じ所属です。今の歴史のお話を大変感動しながら聴いておりまして、SPSSを使うときはこれから感謝しながら使いたいと思います。

お話を聞いていて3つ大きなテーマがあったと思うのですが、それぞれ1つずつ質問させていただきたいのですけれども、まず1番最初の文化資源学に関するお話しですと、分野を超えてというのは実際にはなかなか難しいというお話ではありましたが、もしイメージとして、分野を超えてデータベース化するときに、何かフォーマットみたいなものを考えていらっしゃるのであれば、そういうことについて教えていただきたいと思います。

これは、たとえば心理学の場合だと、サンプリングなんかの項目がありませんとか、いろいろな欠落があると思いますけれども、そこら辺をどういうふうに考えていらっしゃるのかということをお聞かせいただきたいと思います。

それから2つ目に、非定形データの処理分析についてなのですが、やっぱり今ま

では情報処理技術の限界があって、必ず矩形のデータでないと定型化したデータだと言えないということがあったと思うのです。ただ、情報技術の制約がかなりとれてくると、必ずしも矩形がなくても分析できるという可能性がひょっとしたら開けるかもしれないと話を聞いていて思ったのですが、たとえばSPSSの場合だと、バイト数の制限があって1つの変数の中に一定の長さの文字列しか入れないということがありましたけれども、これなかだと今の技術では解決できるはずですので、そうすると、ある程度の長さのテキストの文章でも入れ込めるかもしれない。

そういうふうに、むしろソフトウェアの側を考え直してみたいことが必要なのかもしれないというふうに思ったので、そこら辺のお考えがあればお尋ねします。

それから3つ目に、調査原票の保存についてなのですが、これはプライバシーについて問題になるところがあると思うのです。先程名前については付箋紙を貼って隠したというふうにおっしゃっていましたが、それでも見る人が見れば誰かがわかってしまうとか、その調査の事情を知っている人であれば、この調査票は筆跡からして誰だとか、いろいろな属性から推定して誰だということがわかってしまうと思うのです。

だからそういうものをOCR等で、そのままかたちとして後世に残してしまっていていいかという問題が、これから社会学会なんかで非常に大きな問題になると思うのです。

たまたま先々週に都市社会学会がありまして、磯村英一先生が残した売春婦の調査の話が出ていました。都市社会学会では、それをOCR化せずに、テキストとして入力しただけで残すということで、原票自体は電子化せずにそのまま封筒の中に入れておくということにしたのだそうです。

そういうことも含めて、どういう基準を社会学会として考えていったら良いのかという

ことについて伺いたいと思うのですが、以上3点です。

原：いずれもお答えするのが難しいですね。

まず第1点目なのですが、私はフォーマットを統一するとかというのは、あまり生産的ではないというのが基本的な発想なのです。それよりもどういうフォーマットで入っていたかという、その情報がきちんと整理されていることが大事だと思うのです。

今度、SORDに提供する予定の「青少年の性行動全国調査」というデータがあります。これは非常に貴重なものなのですが、第1回調査のデータは、磁気テープはあるのですが、それがどういうかたちで入っていたかという、その情報がないものですから、どうしてもこれを復元できないし、提供もできないということになっています。これは、磁気テープの例です。ですから、フォーマットは分野によっていろいろと違いますので、データベースみたいな、リストづくりみたいのだったら統一してもいいのでしょうけれども、私はやっぱりデータそれ自体を使うというようなことを考えています。具体的には、私は行動科学というところに属していて、心理学と社会学の中間ぐらいに位置づけられることもあるわけです。その立場から、社会学で考えられてきた個人の意識に関する命題を、心理学の理論と実験データや調査データの中で少し検証できないだろうか。そういうようなことを考えているのです。ですから、無理に統一する必要もないし、それもなかなか難しいのではないかというのが今の実感です。

それから、非定形データの場合は、確かにおっしゃるように、ソフトウェアの進歩の問題があると思います。ただ、統計分析そのものというふうに考えてみますと、たとえば統計的な手法でも確率モデルみたいなものを使って、生存確率だとかなんとか、そういうような事柄を扱うような分析も、今は盛んなのですが、そういうものが必ずしも矩形である必

要はないでしょう。けれども、一般に使われているような多変量解析の技法というのは、それはソフトの問題というよりは、統計分析の考え方の問題で、やはりさっき言ったデータ行列の存在というのが前提になるわけです。ソフトのそういう問題は確かに否定できないという気はします。

それから、ついでなのでちょっとお話ししておきたいことがあります。ここの学部ではコンピューターというか、情報教育というのはどういうふうにしてやっておられるのかわからないのですが、多くの大学では、SASなりSPSSなりというような統計分析プログラムのパッケージを使って分析法を学ばせたり、コンピューターの操作を学ばせるというのが基本だと思うのです。私はそれが正しいと思います。正しいと思うのですが、ただ、今言ったような非定形のデータであるとか、あるいは過去のデータを扱おうとしたときに、それだけで良いのかということを感じます。つまりデータそれ自体にさわって修正したり構成を変えたりという作業は、もちろんSPSSだとかSASだとかでできないわけではありませんけれども、結構煩雑です。私自身は基本的にフォートランを未だにを使って、そういう操作をいろいろやっているわけなのですが、まあフォートランとは言いませんけれども、そういうファイルに直接さわられるための技法なりというものを、せめて大学院レベルの一部の学生に、伝達していく必要があるのではないか。私だけでなく、東大の盛山和夫氏も、もう数年して引退モードに入ったら、フォートランかベーシックの授業をやるんだと言っております。

それから3番目の問題は、これは大問題になります。大問題なのですが、私は、基本的には調査をやった人間か、それと同等の位置にある人間だけが調査原票にアクセスできるという原則が大事なような気がするのです。それを超えて原票を公開することについては、

私自身は疑問に思っています。ですから今度の作業も、公開を目的にしているわけではありません。紙が劣化している、だからやるのだというのが基本的な発想です。

それから売春婦の調査というのは、私も聞いております。いろいろと資料を送っていたいて、もし都立大に私がまだいたら、こんなことをやらされていたかなと思って、ひそかに胸をなで下ろしました。私は現物を見ていないのでわからないのですが、多分、原爆被爆者の思い出の記録と、かなり似ているものがあると思うのです。それでも、果たしてテキストファイルが作れるのか。さっき言った文字の問題があります。先程紹介したSSM調査の例の、これは自由回答そのものではありませんけれども、調査員の記録はまだこれでもましなのです。回答者の字は、ホントに読めるものもあれば、判読に苦労するようなものもあるわけです。

それで、佐藤先生ならばそういう情報をとおっしゃるかもしれません。確かに、原爆の被爆者の人々は言いたいことがいっぱいあるわけです。そういうことが脇のページとか、欄外に書かれている。あるいは文字も、間違い字がたくさん出ている。出ているのですけれども、なんとなくそこから、言いたいことがたくさんあるとか、あるいは非常に衝撃を受けた思い出だというのがわかるのです。そういうものをテキストファイルにするというのがどこまで意味があるのか。保存することは重要だ、ではテキストファイルだ、はい万々歳というわけではないということが、実際にこういうものにあたると出てくるわけです。そういうことで、1つはプライバシーの問題は大問題なのだけれども、あまり一般論として議論をすると、かえって議論の幅を狭めてしまうような気がするのです。

それからテキストファイル化という話があるのだけれども、それだってそう簡単ではないよという、その2点ぐらい。お答えになっ

ているかどうかわからないのですが、

中澤：私も調査票を見たときのアナログ情報の大変さみたいなものはすごくわかるのですが、やっぱり古いデータになると、国立大学でも相当歴史のあるところでしか残っていないので、アクセス性という意味ではかなり制限された人しかそれを見れないという状況が一方にありますよね。

その古い調査票を1枚1枚めくって、アナログ的に伝わってくる感動を受け取れる、経験できる人というのは実は限られているので、そこら辺を逆に、もっとアクセス可能にするような……。

原：情報の階層性みたいな感じで……。

中澤：そんなところも私大にいたりするのだ。

さっきの話と全然逆の主張になっちゃうのですけれども、そういうものをもっとアクセス可能にしていくという方法論が必要なのかなということをお話しを聞いていて思いました。

千葉：北大工学部の田中譲先生にミームメディアという概念を学びました。ミームメディアというのは、外在化された知識の媒体で交換だとか、複写だとか、改編と淘汰だとか統一的効率的に行うメディアアーキテクチャです。そこでは、たえず複製されることが大事なことで強調されるように思います。先程の調査表も、世代を超えて継承させるというのは、ある一定の期間ごとに複製されるところが大切なのではないかという感想を持ちました。

原：そうですね、CDだって劣化するわけですからね。

千葉：ただ保存しておけば良いというものではなくて、絶えず複製していく、そのことによって文化を繋いでいくという要素があるように思っています。絶えず複製して使っていくということに積極的な意味があるのではないかという気がしています。

原：私は冷たいようですが、選別というのがかなり重要だとおもいます。実際にすべてについて複製してとっておくということはなかなか難しいわけです。そうすると確かに捨てたものについて後になって、「しまった」と思うことがあるのですけれども、やっぱりそれぞれの時点で、それなりの責任を持って選別をして、重要と判断したものについてはそういうかたちでとっておくということが必要だと思います。すべて蓄積するということは、ちょっと現実的ではないという気がするのです。

心理学の、例をご紹介しましたけれども、すごい量なのです。しかも、大半はゴミなのです。ゴミなのですが、中に非常に有用なものがあるわけです。

それができたというのは、さっきの質問とも関係するのですが、そのテープに実験をやった先生が非常に的確なラベルを貼っておかれていた、それだったのです。そういう意味で作成した人のある種パーソナリティーみたいなものが非常に影響するなという、そういう印象を持っています。

千葉：その時点で継承すべきことを複製して、意味がないと判断されたものは捨てられるということを繰り返して淘汰の働きをしていると思います。

原：そうですね。やっぱり使われなくなったというのは、1つは後の時点からみて意味がないということ。もう1つは、立派な成果なのだけれども、既にいろいろな文献や何かに載せられたり、それからもう実際に研究の中に取り込まれているという、そのいずれかなのでしょね。

佐藤和洋：本学の佐藤和洋といいます。データベースについてお聞きします。今のお話の中で非定形データ或いは不定形データという言葉が使用されていますが、データベースコミュニティの分野では非定形（型）データ処理という表現はあるのですが、非定形データ

という言葉はほとんど使われていないと思います。ここでの非定形データという意味は、不定長或いは可変長データとも取れます。データベースの分野では非定型なデータ処理という意味ですが、基本的にデータベース内のデータはある種の定型な構造で格納管理されていますので、ここでお話されたデータについてもデータベースとして管理できるように思います。データベース管理システムというソフトウェアを使用されては如何かと思うのです。どうしてデータベース管理システムのようなシステムを使用されなかったのだろうか、ということが気になっています。

データベース管理システムの世界に落としてあげれば、色々なビュー（view）でデータベースのデータをみる事ができます。ここで示されたデータ表現も一つのビューかもしれませんが、他の見方も可能となります。他のビュー表現もあるということであれば、積極的にデータベース管理システムというソフトウェアシステムを使用されてはどうでしょうか？ そうすることによって、社会学での調査研究データという財産が世代間を越えて、分野を越えて活用されていくのではないかと考えられます。これからの社会調査の分野においても、手書きデータや音声データなどマルチメディアデータも扱う必要が出てきているかと思っていますので、データベース管理の重要性が増すのではないかと考えています。また、データベース管理の技術とともにデータ検索および情報検索の技術も進展しています。これらを活用していくことでこれまでとは違った形での社会調査データに対する取り組み方がでてくるようにも思えます。

何を質問したいのかといいますと、データベース管理システムをどうして使用されないのかな、ということです。使用されたらもっとメリットがあるように思えるのですが。

原：2つぐらいお答えがあるのですが、1つは私自身もデータベース管理システムという

言葉も知っていますし、実際に使ったこともあるのですが、それはかなり古典的なものでしたので、先生がおっしゃったような、自由度の高いものは知りませんでした。

もう1つは、私自身は基本的にそういうものを公開してというようなことは、あまり考えていないのです。ですから、ちょっと非定形データの話は、話のつながりとしてまずいなというふうに自分でも思ったのですが、あまり自動処理ソフトなどには手を伸ばさずに、もうちょっとシコシコという感じでやっているのです。

実際、田中先生とお話したときも、伝わらなかったかもしれないのですが、私が実際やっている分析そのものは、コンピューターを使ってやるようなものでもないし、私自身はやれないものだと思っています。というのは申しましたように、価値だとか意味みたいな、抽象的な概念について判断をしていくわけです。その判断そのものの合理性というか、客観性を高めるためにはどうしたらいいかということは考えるのですけれども、基本的には分析者が目で見えて判断してやる。無形のものについて判断してやるという、そういうものなものですから、基本的にそういうデータベースとかなんとかというものとは、話がつながらない。そこまで考えがいかなかったのだらうと思います。

佐藤和洋：とりあえず昔のものを知っているものですから、確かに昔の物は大変だったと思います。

原：そういうようなものを扱うというような

発想とは、なかなか昔のイメージがつかないもので、良い勉強をさせていただきました。

田中：文化資源学を使わずに、学術資源学という言葉にすると良いのではないかとという感じがするのですが。

原：そうですね。「学術資源学」はいいですね。ではこれからそれでいきましょうか。

田中：聞いていると、大学の中のいろいろな資料を扱っておられますから、文化資源学とおっしゃったら、もっと広く社会全体の文化資源全体を対象とされるのかと思ったもので。

原：私どもの専攻でやっている仕事は、今のところは学術資源の中でも、いわゆる調査・実験データが中心です。ただ、どこまでできるかわからないのですけれども、新しい研究科をつくるとか組織をつくるということになると、他の、いわゆる哲史文という分野の協力も必要なわけで、そういう理由からで、人間科学専攻に特化した名前をつけるのではなくて、もうちょっと広くという意味で「文化」という名をつけたのですが、「学術資源学」というのはちょっと思いつかなかったもので、これからもしかしたら先生からお借りしたというクレジット付きで使わせていただくかもしれません。

田中さん：借り賃はいりません。

司会：どうもありがとうございました。時間が来ましたので、これで午前中の予定を終わりにしたいと思います。明日も討論の時間をとってあります。原先生、ありがとうございました。