

## 文科系の計算機利用 II

——データ入力ユーザインタフェース(歴史人口学の場合)

小野 芳彦

### 一 はじめに

学術研究に対する計算機支援の形態はいくつかに分類できる。第一は、計算機の高速度大量計算機能を直接利用するもの、たとえば建築物の構造解析や気象シミュレーションなどがそれにあたる。スーパーコンピュータティングや超並列計算などのキーワードがその分野では目立って使われている。第二の利用目的として、知識の集約と分配があり、データベースとネットワークというキーワードで代表できる。三番目は計算機を間接的に利用するもので、どうしても人間がやるほかない仕事の能率を計算機の補助で大幅に良くするというのがその使用目的である。これは、ユーザインタフェースというキーワードで代表される。本稿の焦点は、文科系の一分野におけるこの三番目の計算機利用に関するものである。

データベースの構築などに必要不可欠な大量データ入力は、昔は入力専門のプロに依頼するのが通常であり、データの入力以前の準備段階や後処理などを含めると膨大な時間と費用を要する作業であった。しかし、最近ではOA機器が普及し、研究者自らデータを入力したり、素人に研究室で入力させることも行なわれるようになってきた。また、印刷文字のOCR装置を使って、機械的な大量入力も多少実用の段階に入っている。

ところが、近世以前の文書を研究対象とする分野では、記されている文字の直接的処理技術がまったくないために、最近のエレクトロニクス技術の進歩による入力段階での恩恵を直接には受けられないでいる。こういう状況では、入口の入力段階のボトルネックの解消が難しいのが現状である。

歴史人口学(二章で概説)という文科系の分野もこのような困難

を抱えるひとつである。その江戸時代の宗門改帳の入力用に、ユーザインタフェースを徹底的に改善したパソコンベースの入力ソフトウェアを試作した。本稿では、その設計の際に指針とした諸原則と、その具体的な実現方式を論ずる。

## 二 宗門改帳のデータベース化

### 歴史人口学と宗門改帳

人口学とは家族や世帯を単位として、その統計的なデータから地域・社会の解析を行なう学問とでもいえばよいであろうか。時代を近代以前に広げた人口学が歴史人口学である。近代以前であっても部分的にはあるが、人口史料が残されている。近代的な意味での戸籍を作るといふ目的ではなく、宗教を個々の領民全部について確認するという目的で始められた宗門人別改帳がそれである。(正確にいうと宗門改帳と人別改帳は別の物であるが、本稿では区別せず、宗門改帳と総称する)

この宗門改帳の作成は、キリスト教信者を排斥する目的で一六三八年に始まったが、キリスト教徒の反乱の心配が皆無になっても、むしろ積極的に推進されたという。徳川幕府は、一六六五年には直轄領だけでなく全国の大名に作成指示を拡大しているし、一六七一年には毎年作成するように、さらには一七二一年には六年おきに人口を報告することを義務付けている。初期の記録が世帯の構成員の

宝曆十二年 巳二月	名主	治郎左衛門控
..... (中略) .....		
高式石八斗七升八合七勺五才本田		
一、曹洞宗大谷村圓福寺旦那	五右衛門	年四拾三
同宗右同寺旦那	右同人	女房 年三拾七
同宗右同寺旦那	右同人婿	源之丞 年式拾式
同宗右同寺旦那	右同人女子	ろく 年拾七
同宗右同寺旦那	右同人女子	つめ 年拾三
内老人数五人	内式人男三人	内式人女一人
..... (後略) .....		

図1 宗門改帳記載例

名前と続き柄だけであったのに比べて、一八〇〇年ころの記録では個人の年齢、作地の持高、家畜の種類、移動の理由などが記されるようになって、明らかに戸籍の役割を期待して作成され、またその役割を果たしていたことが分かる。同時代の世界の状況と比較しても、宗門改帳は優れた記録システムであるという。

### 宗門改帳の計算機入力

以上のような利用価値の高い古記録であるが、これを戸籍と同程度、あるいは、それ以上の人口学の資料として利用するには、適当な情報を多次元に集計することが容易な形態に変形しなければならぬ。大概の宗門改帳は調査年・宗教・世帯の順に構造化された個人に関するテキストである(図1、「4」記載の史料?『陸奥国大沼郡

大谷組桑原村宗門人別書上帳」より)。これを世帯で集計したり、個人ごとに通年集計したり、といったことが容易になるような構造に、入手を經由して計算機に入力することが具体的な作業となる。

元テキストは村の役人が達筆をふるって草書で書いた記録であり、やはり訓練を経た人でなければ読めない。ましてOCR(光学的文字認識機構)が可能になるとはとても思えない。多分、国家プロジェクト級の予算をかけないと草書古語OCRは開発できないであろう。それで、人手によってこれを入力するしかないが、どのような方法が良いかを論じるのが本稿の主旨である。

情報学の精神からいうと、全ての情報(後々利用することのないと考えられるもの、たとえば、墨の濃淡などは除く)を文字コード化し、文字テキストの形で入力するのが、情報を失わずにしておく最も良い形態ということができる。後述する川口は、これを原則とした方法を開発したとしている。ただし、文字コードとして現在の計算機の文字コード(標準的にはJIS制定の規格)を利用する場合にいくつかの問題点が発生する。これについてはごく簡単に後述する。

機械認識が難しいテキストでは、画像イメージとして(光)ファインリングするという方式が取られることがある。これは、元資料に存在する文字の位置関係や欠損などに重要な情報が含まれる場合に取られる方式である。しかし、宗門改帳の場合は記載された年齢や

移動記録の内容から情報を抽出して統計利用することに目的があるのであるから、計算機が内容を認識・処理できないこの方式は論外である。

統計利用という観点では、文字テキストは多少冗長である。つまり、記載内容が定形化しているのだから、コード化できるものは徹底的にコード化することが、後々の利用には有利である。逆にいえば、現在の文字列処理の技術はまだ未熟で、生の文字列から必要なデータを取り出すのに、かなり高度なプログラミングが必要になっているということである。この点について、川口の方法を後に検討する。一般的な現代のラテン系の言語でもそうなのだから、日本語でしかも古語ではコンピュータサイエンスの研究の恩恵が最も及びにくい分野となっているのも無理はない。

### 速水方式

日本における歴史人口学の創始者であり、宗門改帳の人口学的利用法を拓いた速水融<sup>[1][2][3]</sup>は、慶応大学から国際日本文化研究センター(現在)に至るまで一貫して宗門改帳の計算機入力とその解析を行なってきた。

速水によるこのプロジェクト(ザビエルプロジェクト)では、全ての計算機入力は、いったん紙に写されたものを經由することになっていた。これは純粋なデータ入力速度という点だけを見れば高い

宗教宗派	寺所在	旦那寺	異動理由	異動内容	村役人
曹洞宗, 曹洞宗, 曹洞宗, 曹洞宗, 0,	大谷村, 大谷村, 大谷村, 大谷村,	圓福寺, 圓福寺, 圓福寺,	結婚・婿名跡, 結婚	陸奥国大沼郡大谷組川井村出身, 1761,	, 組頭
↑馬数	↑持高 (石)	↑家屋規模* (縦横、間)	↑屋根材	↑史料作成年次	

効率の入力方式であるが、前処理に時間と人件費がかかりすぎる。計算機利用が相対的に貴重であった時代の方法と今はみなされるであろう。また、現物と入力データとの間の処理ステップ数が多くなるので、エラーが入る確率が高まり、照合を一層厳重に行なわねばならなくなる。このような欠点を考えると、入力処理をインタラクティブにする必要がある。ちなみに、一般の文科系の計算機利用においても、研究者自らがインタラクティブにデータ入力するケースは非常に多くなっていると予想される。

ザビエルプロジェクトの場合、入力前処理の複雑さを増すのは、入力データがコード

化されている、すなわち数字に変換されていることである。続き柄を例にすれば、娘で実子なら88で、男の奉公人なら881という数字に対応する、というようなデータの書き換えである。このコードはあらかじめ用意されているが、たくさんの宗門改帳を解析して得られた集大成であり、現在も時々拡充されることがあるという。コード化によって入力に数字だけになるから、入力効率は高くなるが、逆に原稿の作成と照合の手間が大きくなる。文字列で「娘」と書いてあるのと数字で「802」と書いてあるのでは、間違いに気付く確率は段違いである。ただ、速水方式のシステムの設計の時期が日本語ワードプロセッサなどの発達する以前であったため、漢字を含むデータの入力に不可能あるいは高価であり、コード化は当然の方法であった。コード方式はやむを得なかったのである。

ザビエルプロジェクトでは、宗門改帳の前に、古文書が読める専門家がデータをコード化しながら紙に整理して書き込んでいく。最初に作成する物は、世帯単位の基本データシート(BDS)である。次に、BDSをもとにして、個人単位の年齢シート(A6)と経歴シート(A7)、および、家族(夫婦)単位の世帯シート(A5)を作成する。この三表はすべて数字からなり、それを計算機に入力して種々の人口学指標を計算することになっている。さらに、これとは別に家族復元フォーム(FRF)と個人追跡シート(ITS)を同じく手作業で作成し、家族統計や出稼ぎ統計を計算する。

家番号 ↓	個人番号 ↓	名前 (ローマ字) ↓	名前 (漢字) ↓	性別 ↓	年齢 ↓	続柄 ↓	配偶 ↓
09000,	1020,	GOUEMONN	五右衛門	男	43,	戸主	有配偶
09000,	0013,	NYOUBOU	女房	女	37,	房	有配偶
09000,	1079,	GENNNOJYOU,	源之丞	男	22,	婿	有配偶
09000,	0014,	ROKU	ろくめ	女	17,	女子	有配偶
09000,	0015,	TUME	つめ	女	13,	女子	未婚
09000,	2, 3,	0, 0, 0,	0, 0,	0,	5,	520	2, 0,
人数	↑ 男 ↑ 女 ↑ 譜代 ↑ 代 ↑ 譜代 ↑ 代 ↑ 質券 ↑ 質券 ↑ 男 ↑ 女 ↑ 男 ↑ 女 ↑ 世帯 ↑ 規 ↑ 模 ↑ 世帯 ↑ 構 ↑ 造 ↑ 牛 ↑ 数						

図2 川口方式の一家族の記録内容 (文献 [4] より)

以上のように、BDSはすべての統計の情報の源泉であり、必要なすべての情報を転記しなければならぬ。宗門改帳のデータフォーマットは地域・年度などで異なることが多いため、その間の揺らぎをBDSが吸収する必要がある。このBDS方式は、この点ではかなり成功していると考えられる。

#### 川口方式

速水方式が前処理に時間と費用を要するものであるため、川口は以下のような別方式 (DANJIURO) を提案し実現している。<sup>[4][5][6]</sup>

(1) 宗門改帳の記載単位ごとに、一行 (レコード) の文字列データを作成する。記

載単位は個人であり、さらに世帯 (あるいは家屋) の情報を一戸ごとに追記する (図2)。

(2) 一行 (レコード) は固定長の文字列からなる欄 (フィールド) からなる (図2)。欄の記載順は一冊の宗門改帳で固定である。

(3) 欄に書かれる情報は、原資料の記載内容を保存する物であることを原則としている。これは、宗門改帳をデータベース化して研究者間で共有する場合に、原資料との対比ができるようにするためであるという。

(4) 宗門改帳一冊を一ファイルとし、同じ村の次の年度の宗門改帳は、作成されたファイルをコピーしてその差分だけを訂正するという方法で作成する。これによって、入力的大幅な時間短縮を図ることができる。ファイルは、定形の文字列であるから、通常の日本語ワープロやエディタで研究者の手元での編集が可能になっている。以上のような基本方針で、入力データのフォーマットが設計されているのであるが、制限や追加がいくつか存在する。記載内容以外に追加しているものとしては、(a) 家番号と個人番号、(b) 名前のローマ字読み、(c) 配偶関係と世帯構造がある。

(a) は、家や個人の同一性を識別するためのもので、既存の統計計算システムが同一性の確認をその種の識別番号に依存しているための措置である。番号は入力者が設定するが、ここに誤りの発生する可能性も小さいながら生じる。計算機側が充分賢ければ自動的な設

定も可能である。さらに、前述のBDS方式のように、時間軸での家・個人の結合を済ましておけば、本質的な役割はなくなり、省略できると考えられる。

(b)は、統計パッケージが日本語をサポートしない場合にも名前が利用できるように追加したものであろう。しかし、もしローマ字の名前をキーとして個人の検索・照合を行なうのが目的であれば、表記の統一性が大きな課題となると考えられる。<sup>(注1)</sup> 名前の文字自身に統計的な利用価値がないと判断してもよいならば、宗門改帳における女性の漢字や変体かなつづりの名前をひらがなつづりに変換して記することを川口が採用したように、<sup>(注2)</sup> 名前の読みの正体かなつづりに変換して記しても困ることはないであろう。ローマ字表記(標準つづり)が必要ならば、計算機が読みのかなから自動的に生成すれば足りるのであろう。

(c)は、現在のできあいの計算機プログラムでは抽出が難しい(あるいは面倒)と考えられる情報を、人間が付加するものである。配偶関係はそのものの文字列を記すことになっているが、世帯構造をコード化した数字で記しているのは、入力の手間の方を重視する方針のようである。

以上のようなフォーマットに従ったファイルを集めて、一カ村のデータベースを作成する。このデータベースからの種々の統計は、DBMSと前処理プログラムおよび汎用統計計算パッケージからな

るシステムが作成するように作られている。

(注1) たとえば、「五右衛門」をGoemonあるいはGouemon、Gouemon、Gouemonのどれにするか、「なほ」をNaoとするかNahoとするかなど、あらかじめ決めておかなければ判断に迷う名前が多い。さらに、一般的な現象であるが、現在はローマ字表記を统一的に表記させることが難しい状況である。昨今のワープロ・ローマ字つづりの無批判の濫用により、たとえば、「ジョ」をJoとつづり、「ン」を必ずNとするような変則的なつづり方が隆盛している。従来から見られた日本語・訓令式の揺れ以上に、ローマ字つづりの揺らぎが大きくなっていることも考慮されなければならない。

(注2) 男性の名前を漢字で取り扱うについても同様の事情があるはずである。計算機で処理する前提では、JISコード(X0208)第二水準までに含まれない漢字や異体字をどうするかが常に問題となる。史料の再現性を高めるため、異体字などを外字として固有に定義するという対処法がよく行なわれる。しかし、諸システム間で外字の互換性がまったくない現状では、データの流通性を損わないために、外字方式は避けなければならない。したがって、第二水準までの漢字に翻訳する必要がある。翻訳に関しては、採用した語のリストを公開する必要がある。

### 三 実現方式の検討

## BDSとDANJUROの比較

両方式を比較してみると、宗門改帳の入力では次のような点が問題となっていることが明らかになる。共通点は、両者とも情報のうち将来利用されないであろう末節文字列を省略していることである。<sup>(注3)</sup>

たとえば、移動理由が婚姻による場合、移動元村名（BDSでは元の家の戸主と元の名前や続柄）という中心になる情報のみを記すのであって、決して元の文（たとえば「男婿名跡縁付まし」）そのものをテキストとして記録するのではない。これは、機械の側で柔軟な情報抽出能力を保障することができない、つまり、個別に対応することはできるが、宗門改帳のどのような記述書式にも対応をあらかじめ保障することは不可能であるために、入力段階での情報抽出を作業者に任せているのである。このような情報抽出をたえ主体的でなくともよいから、計算機の側からできる限りの補助を行なうならば、大きな利益があるはずである。その形態を次の節で具体的に吟味する。

両方式の差異の方は、データの付き合わせのために必要な時間軸の視野をどれだけ確保するかという問題についての取り組みの違いとして表われている。川口が採用しているファイルコピー法では直前の年度のデータのみが入力作業者に示されている。通常の場合、それにより個人や家の結合に不都合が生じることは確かに少ない。ただ、たとえば、分家の際に孫が連れて出られるかどうかで、兄弟どちらの子であったか判明するといった場合があり、訂正の波及が

その孫の出生の時点にまで遡るようなこともたまには生じる。BDSを用いる方法は、有限ながら個人軸と時間軸を等価に一望できる点で自在性が高いと考えられる。計算機上にこの表を実現するならば、視野を事実上無限遠とすることができようであろう。

この両者の差異は、データの縮約方法の差異であるとも解釈できる。BDSの場合、個人軸方向の記載が行なわれていく段階で、個人名・旦那寺名などの縮約（同一の情報を入力しないこと）が行なわれる。これに対してDANJURO方式では、次年度ファイルのコピーという最初の段階にデフォルトとして同一化が存在し、訂正をしなければデータベース作成の段階で事実上の縮約が実行されるという二段構えになっている。データ入力の手間という観点からだけではなく、単純な誤入力防止という観点からもDANJURO方式の方が優れている。ただ惜しむらくは、年齢という最も重要なデータの縮約が果たされていないことが欠点であろう。

(注3) 川口は文献「4」の注釈でコード化による原史料の微妙な表現の画一化を重大な問題点であると捕えてはいるが、全テキストを保存するという方針を実際にとっているわけではない。

### 拡大BDS方式

以上の検討を踏まえ、両者の長所を複合したような入力支援システムの基本設計を考察すると、ほぼ必然的に次のようなものとなっ

た。

まず、差異の検討で得られたように、情報の視野を大きくするためには、BDSが採用しているように個人軸と時間軸の二次元にデータを記入していく表形式の集計が有効である。ただし、基本的に記入の順序は個人軸方向である。そして、時間軸の推移を起こす場合には、DANJURO方式のように、前年の情報の転記を計算機が行なうようにするのである。ただし、実際に時間軸推移で記載内容が変わるのは年齢以外ほとんどない。したがって、転記するのは個人年齢だけでよく、それも史料の経過年を一律に加えるだけである。

次に、必須の機能と考えられるのは、情報入力省力化を進めたり正確度を高めたりするための情報提示機能である。たとえば、家族復元などで重要と考えられる、続き柄や奉公人のバリエーションの一覧をあらかじめ用意しておいて、必要になったらその一覧を表示して選択させるようなことである。具体的には、メニューとして画面に表示してそこから選択するといった形式が最も一般的である。さらに、情報指示機能もできるかぎり直接的（あるいは選択的）な手法を用いるべきである。たとえば、親子関係を指示するにあたって、個人番号をキーボードから打たせるのではなく、対象となる個人が書かれている場所をマウスでクリックする（あるいはカーソルを合せる）ようにするといったことである。

最後に、データ規模についての検討が必要である。DANJUROでは一冊（一カ村）の宗門改帳をひとつのファイルで扱うようになってきている。BDSは一世帯が一枚のシートになっている。ただし、紙の大きさの関係で二十五人二十五年間が一枚になっているので、それを超える場合は、複数の紙を使用することになる。実際の集計作業を検討すれば、世帯ごとにファイルを取り替えるのは作業効率からいって得策ではないことがわかる。したがって、計算機の中に複数の世帯を入れた大きなBDSシート（拡大BDS）を作成するという方針が両者の特徴を取り入れたものとなる。結果として、最大で約二百五十年、述べ数万人程度が収集できることが必要である。

### スプレッドシート言語

上記のような機能を満たすものを最初からプログラムとして個々に作成するのは、マンパワーが通常の文科系の研究者の許容範囲を超えてしまっている。できれば既存の計算機ソフトウェアが望ましい。そういうものとしてデータ集計システムであるスプレッドシート言語（表計算言語）がある。周知のように、表状態にデータを入力し、統計や財務などの各種の計算をし、グラフを簡単に出力することができる汎用ソフトウェアである。<sup>[7]</sup>

スプレッドシート言語が関数主体のものであるにも関わらず、新



しいタイプのプログラミング言語という範疇に入れられるのは、セルの中に他セルを参照する計算式を書き込むことでデータ依存関係を作り出し、ソースとなるセルの数値を書き換えることで、一度作りに上げた表を再利用することができる、つまり実際はプログラムとして働くようになっていくからである。スプレッドシートのこの特徴は、本システムの設計に本質的ではないが、相互に関連するデータの新しいタイプの集計方法（後述）を開発するのに必要な機能であるという、別の意味での重要性があった。

スプレッドシート言語のビジネス需要が高まるにつれ、単なる計算式をユーザーが書くだけではなく、計算式中で利用する関数や表・データを操作するコマンドをユーザーが作成するといった高度な利用法の要望が高くなった。その実現の一例はマクロ機能である。マクロ機能と名付けてはいるものの、実体は手続き型のプログラミング言語であり、別の連続するセルに計算あるいは処理の手順をあらかじめ書いておき、関数あるいはコマンドとして適宜実行するというものである。

マクロ機能は、ふたつの点で有効である。ひとつは、細かな操作の集積を一動作のマクロで実行させることができる点である。この宗門改帳の集計のように計算機の素人が操作する場合、操作数を減らすことは、慣れ・誤操作防止・操作時間短縮など多くの利点を生む。もうひとつは、共通する計算をマクロで共有することで、消費

するスペースを節約することができる点である。一カ村という単位はかなり大きく、年度ごとの個人データにいくつものセルを消費すると、場合によってはパーソナルコンピュータでは手に負えない大きさとになってしまいう可能性もある。

以上のような条件を満たすものとしてマイクロソフト・エクセル (Microsoft Excel version 2.2)<sup>[8]</sup> を選定し、パーソナルコンピュータであるマッキントッシュ上で試作を行なった。エクセル・マッキントッシュ版はマッキントッシュが提供する画面操作インタフェースをマクロ機能として提供しており、後述のメニューのユーザー定義機能などで、計算機初心者に親切な設計を可能にしている点が優れている。

### 基本機能

**環境** 作業環境を確定するため、マクロを収めたファイル「宗門

帳マクロ」をオープンする。これによって、エクセル・システムが起動され、また、宗門改帳専用の集計環境を設定することができる。エクセル・システムが本来提示するメニューに追加して、作業者が可能な操作をメニュー表示する。最初は、既存のVBDS（本システムが作る集計表、前節で拡張BDSといっていたもの、村基本データシートのアクロニム）の追加

編集と、新規の拡張BDSの作成である。作成・編集が開始されると、マクロにまとめられた操作が指定できるようになる。

#### 項目の自由な追加

新規作成を指定した場合は、エクセルのワークシートを新規に作成するとともに、ヘッダーや欄の幅などの設定をして空のVBD S (図3の1) を作成する。<sup>(注4)</sup>ヘッダーは

図に示した通り、時間軸に西暦、和暦(元号+年号)、干支の三目盛、個人軸に名、性、続き柄、宗派と旦那寺の四項目である。ただし、宗派と旦那寺の項目は、宗派、旦那寺、寺の所在地のように複数に分けても良いし、別の項目を加えても良い。さらに、なくても良い。これらは、通常世帯の戸主と同じであり、経年による変化も非常に少ないという性質を満たすと考えられるので、列を同一の基準で扱うようにプログラムしているため、このような自由な追加変更を可能にしている。

#### 計算式の効用

空のVBD Sには、ひき続いて、史料開始年のラベルと空の一世帯が作られる。史料開始年は、初期化と同時に一時的に表示される入力ボックス(図3の2)の指示に従って、西暦のみを入力する。和暦と干支は西暦からシステムが計算する。ただし、思い違いなどで史料の開始年を誤った場合でも対処できるように、計算結果の値をセットしているのではなく、

計算式をそのセルに記入しているのである。このことによって、該当の西暦年を書き換えるだけで和暦と干支が自動的に変更されるようにすることができる。また、この方式によって矛盾したデータを入力することを防ぎ、記述の誤りの発見を、入力場でできるようにしている。

#### 単純集計の自動化

続いて成員がまだ誰もいない世帯を作成する(図4)。世帯の情報は、初期値では世帯員の男女別小計および全員の合計のみである。この後に家屋の情報や家畜・石高などを追加してもよい。この後追加される世帯には、これらの行が自動的に設定される。世帯員の三つの合計が当該年度にあらかじめりと書かれている。これは、スプレッドシートの計算式をあらかじめその年度セルに記入してあるからである。西暦と和暦の関係と同様、世帯員の合計のような単純な計算はシステムにやらせる方が確実である。

#### 表の動的変更

世帯員追加メニューを選択して、世帯の合計行の前に戸主およびその世帯員を挿入する。このようなことは紙ではできない計算機独特のダイナミックな編集である。行挿入はスプレッドシート言語にとっては通常の機能であるが、それによって計算式の指定範囲が狂わないように細心のプログラミン

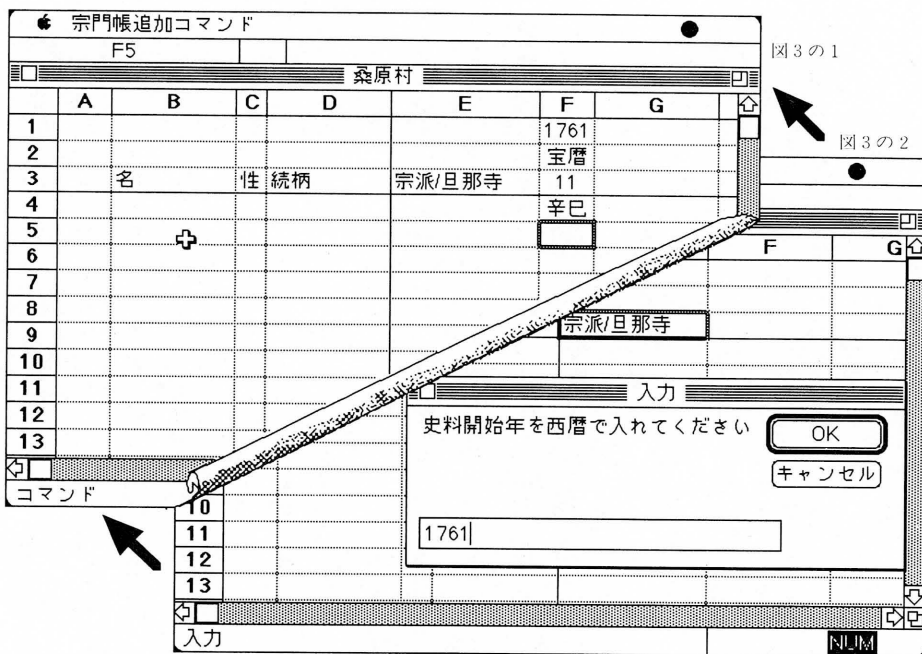


図3 空のVBSDの作成

宗門帳追加コマンド

F5

=宗門帳マクロ集計(\$C4:\$C5,\$C5)-1

桑原村

	A	B	C	D	E	F	G
1						1761	
2						宝暦	
3		名	性	続柄	宗派/旦那寺	11	
4						辛巳	
5		小計	男			0	
6		小計	女			0	
7		合計				0	
8							
9							
10							
11							
12							
13							

コマンド

図4 空の世帯の追加

グを要する。たとえば、成員の男女別小計を計算する式は、成員の先頭から最後まで<sup>(注5)</sup>の性別のデータの男および女という文字を数えることで計算する。挿入が先頭や最後で起きた場合に範囲に含まれないことを防ぐには、範囲を上下に一行ずつ余分にとっておくといった工夫が必要である(図4、F5セルの内容参照)。

高級メニュー 世帯員の追加のデータは、図5のような一連のデータ入力ボックス(エクセルではダイアログ・ボックスと呼ぶ)によって必要なデータを入力し、所定の位置に正しく設定されるようにしている。データ入力ボックスは、マッキントッシュが提供する種々のデータ入力法をユーザーが適宜利用できるようにしたものであり、ひとことで説明すれば高級なメニューといえるものである。

文字列入力ボックス 世帯員の名前を入力するのは、任意の文字列を入力できる文字列入力ボックスである(図5の1、「五右衛門」と書きこまれた名ボックス)。これは単なるセルであり、特殊な援助を行なうことはしない場合のスタイルである。

数値入力ボックス 史料に初出時の世帯員の年齢を入力するのは、数値入力ボックスである(図5の1、「43」と書きこまれた

年齢ボックス)。数値のみ入力できる。初出は出生時であることが多いと考えられるので、あらかじめ2(数え年)を設定してある。ちなみに、入力忘れて本来の年齢が設定されずに2が設定されても、その数値を表の該当位置で書き変えるだけで訂正が可能になっている。

排他的ボタン たとえば性別(図5の1)の場合「男」か「女」のどれかであり、排他的なボタン(一方をオンにすれば他方がオフになるもの)というスタイルをとるのがふさわしい。特に性別の場合、正確にこのどれかの値でなければ、前述の男女別の世帯員の小計が正確でなくなる。もし自由に文字で入力させた場合、同義である「M」か「F」などとなる可能性があるが、この方法ではそういうあいまいさが排除される。

チェックボックス たとえば、続き柄で嫡子であるかどうかといった情報は、記載されている場合もない場合もある。記載されている場合にはそれを捨てないで取って置くために、付加情報としてチェックマークを付けるというスタイルが、この種の情報を入力するのにふさわしい。

リストボックス 例では、奉公人などの続き柄を入力するのに利用しているのがリストボックスである。可能な選択肢をリストの形で提示しており、その該当する欄を選択すること

(1)

名  年齢

性別  男  女

戸主との続柄:

戸主本人  
 誰かの血縁者  
 同居人

図5の1 ダイアログボックス(戸主スイッチ)

世帯別追加コマンド

	A	B	C	D	E	F	G
1						1781	
2						宝福	
3		色	住	親類	赤羽巨摩寺	11	
4						幸巳	
5		五右衛門	男	戸主		43	
6		小計	男			1	
7		小計	女			0	
8		合計				1	
9							
10							
11							
12							
13							
14							
15							

図5の2 世帯員(戸主)追加

(2-1)

名  年齢

性別  男  女

戸主との続柄:

戸主本人  
 誰かの血縁者  
 同居人

(2-2)

指定方法:

両親  
 配偶者  
 その他

甥姪  
 兄弟  
 叔父(母)  
 曾孫

(2-3)

入力

配偶者の行の名前セルをマウスで選択

図5の3 ダイアログボックス(配偶者指定)

世帯別追加コマンド

	A	B	C	D	E	F	G
1						1781	
2						宝福	
3		色	住	親類	赤羽巨摩寺	11	
4						幸巳	
5		五右衛門	男	戸主		43	
6		安男	女	妻		37	
7		小計	男			1	
8		小計	女			1	
9		合計				2	
10							
11							
12							
13							
14							
15							

図5の4 世帯員(女房)追加

図5 世帯員の追加のメニュー系列

で文字列を入力する。典型的なメニュー形式である。選択肢が多くて画面を圧迫する場合は、一部分だけを表示して、残りをスクロールさせるようにすることもできる。

リストボックス付き文字列入力ボックス 文字列入力ボックスにリストボックスを付随させたスタイルである。任意の文字列の入力であるが、典型的な例が存在する場合に、それを候補としてリストボックスに並べておいて、そこにあればそれを指定することで入力が省力化されるという効果がある。さらに逆に、本来はリストボックスであるが、ごく稀にしかない選択肢を列挙することが困難であったり実用的でなかったりするとき、その稀な場合に対処するためにこのスタイルを取ることがある。

スイッチ 画面上に表示された入力ボックスで、全てのパラメータ入力を終えたことを確認するOKボタンと、全体の処理を止めるキャンセルボタンがある(図5の1)。

関係の記述法 家族間の関係である続き柄は、独特の親族名称で記述されているが、計算機がその用語を理解して親子関係を正確に再現することは難しい。しかし、逆に親子・婚姻関係が与えられれば、家族間の戸主を中心とした続き柄を合成することはかなりの程度可能である。したがって、本システムでは血縁

者の場合は原則として配偶者か両親を指定させることにしている。

アドレスの利用 スプレッドシート言語は、関係を表わす機能を直接もっているわけではない。そのため、マクロ関数を使って工夫をする。マクロ関数に限らず、スプレッドシートの関数は引数としてセルのアドレス(セルの位置を行と列の番号で表わしたもの)をとることができる。マクロ関数のプログラムの中では、引数として与えられたアドレスから、そのセルに書かれている数値や文字列などの情報以外に、そのセルの回りの情報を引き出すことも可能である。たとえば、引数が世帯員の名前のセルを指しているも、同行の隣接したセルにあるその性別や続き柄の情報も引き出すことができる。具体的な手法は次のようなものである。プログラムでは、関係が婚姻(配偶)で相手(引数)の続き柄が「戸主」で性別が男であれば「妻」を返し、性別が女であれば「夫」を返す。相手の続き柄が「嫡子」で性別が男であれば「嫁」を返し、性別が女であれば「婿」を返す。「戸主」と「嫡子」以外の場合は「○○女房」「××婿」という具合に名前を付けて関係を返す。以上のように、続き柄セルにはマクロ関数が書かれているが、その計算値は続き柄を表わす文字列になっている。画面に表

示されるのは後者の続き柄文字列である。親子関係の場合  
は、このような処理が非常に複雑になっている。

セルの直接指定 セルに書かれるマクロ関数の引数がアドレスで  
あると述べた。その場合、マクロ関数の名前やアドレスを  
作業者がタイプしなければならぬとしたら、続き柄をタ  
イプする以上に手間がかかって、このようなことをする意  
味がなくなってしまう。そこで、対象となるセルをマウス  
で直接指示することをデータ入力ボックスの処理の中で要  
求し(図5の3)、引数のアドレスを得て、それからマク  
ロを呼び出す関数式を合成して該当するセルに書き込むこ  
とにした。このようにすることで、家族関係を直接的に指  
定するスタイルを実現できたと考えている。

備考事項の記述 世帯員が追加される場合には、その理由が宗門  
改帳に付記されている。BDSでは該当セルにコメントマーク  
をつけ、欄外にそれら付書を転載している。

メモ機能 エクセルにはメモと呼ばれる注記を各セルにつける機  
能がある。メモは通常は表示されない。VBDSではこの  
メモ領域に付書の内容を要点だけ転載することになっている。  
ただし、メモ領域は日本語で百文字強しか容量がないので、  
超える場合には一般セルの空き領域へ内容を移すことにし

ている。メモ領域の書き込みや読み出しは専用のコマンド  
マクロで対処する。

以上のように、世帯追加・世帯員追加のコマンドをメニューバー  
から選んで実行することで史料初年度のデータを入力する。

次年のデータ作成 史料次年のデータ入力はDANJUROのよ  
うにデータのコピーをとることで行なう。ただし、単純な値の  
コピーではなく、前回のセルへの参照とする。こうしておく  
と、前回のデータを修正した際にそれ以降の年のデータも同  
時に修正される。このようにすることが親切な場合もあれば迷  
惑な場合もあり、功罪は半ばするが、前回年と同じであること  
をシステムが明らかにすることができると良いと考えている。  
この延長として、個々人の年齢は単なる参照ではなく経過年数  
(前回調査年との差)を加える式として記入することになって  
いる。このため、年齢の照合だけ行なえばよく、通常は変更の  
必要がないはずである。

VBDSの見せ方 BDSでは、年次が進むにつれ、分家・転  
出・除籍・死亡などで世帯の構成員の間に隙間ができてきて、  
一覧性が悪くなってくる。VBDSでも事情は同じであるが、

計算機上にあるという柔軟性を發揮して、一覽性を回復させることができる。

行の縮減 エクセルには行の表示高さや列の表示幅を変える機能がある。特に、それらを0にすると、行や列が画面から見えなくなるといふ効果があり、これを一覽性の回復に利用することができる。メニューバーから一世帯あるいは全世界について、世帯表示整理を選ぶと、その年度にはもう存在しない世帯員の行を高さ0にするコマンドマクロを実行する。全員再表示を選ぶとすべての列の高さを標準に戻す。

もうひとつの時系列データ 現代では起こらなくなっているが、江戸時代は改名がかなりあった。旦那寺も結婚当初は親元の旦那寺であったものが世帯の旦那寺に変わるといったことが起こる。これらは、いずれも毎年変化するものではないが、時系列として扱うべきデータである。BDSは年齢と世帯諸情報だけを調査年毎に転記しているが、名前や旦那寺はそういう扱いをしない(DANJUROはすべて調査年毎に転記する)。VBDSでは、セルの消費をおさえるため、DANJUROの形式にすることが憚られた。そこで、変化のほとんどない時系列を表現する別の方式を考案した。

可変表示マクロ エクセルは、関数の引数に任意個のデータをつ

ないだリスト構造を指定することができるようになってい  
る。このことを利用して、変化のあった年(西暦)とその  
内容を組にしたリストを引数に与えてあるマクロ関数を設  
定した。具体的な計算はたとえば以下の通り。仮に生年一  
五四二年幼名「竹千代」、結婚(一五五七年)して「元康」、  
その後一五六〇年「家康」とする。編集・確認を行なっ  
ている年度が一五五九年ならば「元康」、一五六〇年ならば  
「元康改め家康」、一五六一年ならば「家康」という文字列  
を引数のリストの中を検索して作り出すような関数を作成  
した。

また、改名があれば、メニューバーから改名コマンドマクロを  
起動して、名前欄に名前ではなく変更年とセットになったリス  
トを引数に入れ、当該マクロを呼び出す式を記入する。

(注4) VBDSはBDSとは軸を反転して、時間軸を水平方向にと  
っている。これは、スプレッドシートでは行数より列数の制限がき  
ついのが普通であり、個人数は延べ一万を超える可能性があるが、  
年数は三百年を超えることはなく、個人軸を垂直方向にとるのが最  
適であることが一つの理由である。また、人名を縦書きにすること  
がスプレッドシートでは難しいため、視野を広くするには個人軸を  
垂直方向にして個人名を横書きにする必要があることがもう一つの  
理由である。

(注5) 文字列が一致するセルを直接数えるという組込みの関数は、



集計には必須であるにもかかわらず、エクセルのこのバージョンには未だそれがない。同等の機能を持つ関数は、データベース操作関数として実現されているが、その場合、データベースとしての特定の体裁を集計範囲や照合文字列に整えさせねばならないため、任意の範囲について適用することができない。本システムではマクロで自作している。

#### 四 おわりに

以上のように、計算機に不慣れな人のための入力サポートシステムの一例を作成した。設計思想では、ベースとしたスプレッドシートに存在する細かい機能はできるだけ作業者にみえないようにした。このため、すべての操作をあらかじめマクロでプログラムして対応しておく必要がある、全体の完成には日数がかかっている。しかし、今後はこの経験を生かすことができ、どういった点を注ぐべきかも判明しているので、作成にかかる時間は大幅に短縮されるはずである。本稿を書いている現段階では、未だ実用に供していないので、作成したシステムの評価は行なっていない。

今回のシステムの設計は、徹底的に作成者の補助をするという観点で、既存の汎用プログラムがどのような力をもっているかを試すという実験でもあった。ユーザーインタフェースとしてユニークでしかも初心者向きといわれる機能を満載したマッキントッシュならでのサポートも大きかった。筆者が開発した(特に文科系向きの)データハンドリング方法は、ベースシステムの開発者の予想の

範囲を超えた新しい利用法ではないかとの期待もあり、これを詳しく公開した次第である。

なお、本システムは無償で利用可能である。興味をもたれた方は筆者までご連絡されたい。

#### 参考文献

- [1] 速水融「戸口」赤松他編『日本古文書学講座』第七巻近世編 II 所収 雄山閣 一九七九年。
- [2] 速水融『近世濃尾地方の人口・経済・社会』創文社 一九九二年。
- [3] ローレル・L・コーネル 速水融「宗門改帳——日本の人口記録」仲井美由紀訳 速水他編『徳川社会からの展望——発展・構造・国際関係——』第六章所収 同文館 一九八九年。(原著 Laurel L. Cornell and Akira Hayami, 'The Shamon Aratame Cho Japan's Population Registers, Journal of Family History, 11-4, pp. 311-328, 1986.)
- [4] 川口洋、中山和彦「宗門改帳」データベース(DANJUR O)の開発「一九八九年情報学シンポジウム講演論文集」一二五—一三四頁 一九八九年。
- [5] 川口洋「江戸時代における人口分析の方法——奥会津地域における「宗門改人別家別帳」のデータベース化を事例として——」『歴史地理学』一五一 一六一—一三三頁 一九九〇年。
- [6] 川口洋「宗門改帳」データベース・システム(DANJUR

○の改良」『情報処理学会研究会報告 人文科学とコンピューター』一三一—一四頁 一九九二年。

〔7〕 小野芳彦 「文科系の計算機利用Ⅰ——選挙制度のシミュレーション」『日本研究』第六集 一九九二年。

〔8〕 マイクロソフト社 『Microsoft Excel 関数・マクロリファレンス』 マイクロソフト社 一九八九年。