

〈共同研究報告〉

統計分析を目的とした近世史料のデータベース化

——入力・データ利用インターフェイス

森本修馬

1 はじめに

近年のコンピュータ技術の発達により、様々な研究分野で、情報処理研究の学際研究としての役割は重要になってきている。

特に数量的研究を目的としている研究分野では、コンピュータ利用の促進と研究の発展とは切っても切れない関係にある。今回は、歴史人口学分野におけるコンピュータ利用の促進の一例として、入力インターフェイスの開発によるデータ入力の効率化について述べる。

日本の歴史人口学では、主に近世の歴史資料である宗門改帳・人別改帳（以後、あ

わせて宗門人別改帳と呼ぶ）から得られる情報をもとに研究がなされている。今回入力の対象としたのはこれらの史料である。

人口学分析のためのデータベース構築に關しては、大量のデータ入力が必要不可欠である。かつては、計算機の記憶容量の問題、入力効率の問題のため、数字によるコード化、人手によるデータ抽出などの前処理が必要であり、かなりのコストがかかっていた。昨今では、計算機の記憶容量、計算速度の向上により、かなり大規模なデータベースを構築することが可能になった。こういった機械面での発達に対して、大量の文字データの入力にコンピュータを利用

していくことはあまり考慮されていなかった。特に、機械による文字認識が不可能な歴史資料から大量の文字データを整理し、入力するにはかなりのコストがかかるという問題がある。また、歴史資料を扱うことから、入力されたデータには判断が生じざるを得ないため、元史料との比較が可能であるか、つまり、別の研究者による批判が可能であるかということが問題になってくる。本稿に紹介するインターフェイスは、この二つの問題を解決することを目標に開発されている。この開発の結果として、より情報量が豊富な、汎用性のあるデータベースを構築することができる。本稿の構

成として、まず、2章では、対象とした歴史資料についての説明を行う。3章では、実際に歴史人口学の研究者が行ってきた歴史資料の整理方法についてふれ、4章、5章、6章では、入力インターフェイスについて述べる。4章では、一般論として、入力インターフェイスを開発する際にポイントとなる事柄についてまとめおいた。5章では、入力インターフェイスを開発する際に重要となってくるデータ形式について述べる。6章では、4章で述べたポイントを実現するために、プログラムを用いてどのような工夫を行ったかの具体例を述べる。実際のデータベース構築、応用については別の機会に述べたい。

2 人別改帳と宗門改帳

この章では、日本の歴史人口学の研究に用いられている近世の歴史資料、人別改帳と宗門改帳についての説明を行う。詳細については、「速水」を参照してほしい。

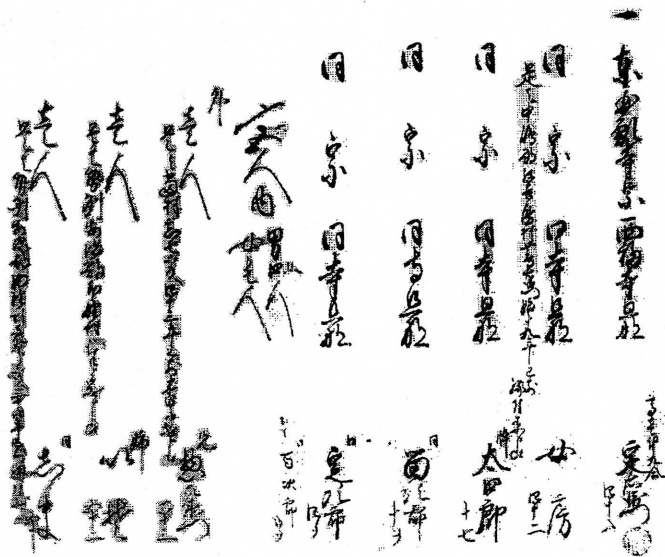
人別改帳とは、江戸時代を通じて幕府や

各藩が随時作成した領内の人・土地（持高・貸高・借高）に関する各年ごとの調査資料である。この史料には、世帯単位に、実際に領内に住んでいる者の名前、年齢、世帯内における地位などの調査時点での静態情報が記載されている。さらに、実際に世帯内に住んでいる人数が書かれた後（レ人数と呼ばれる）に村外に出稼ぎにいった者の情報などが書かれる場合もある。また、史料によっては、次の調査までの変化、死亡、出生、異動などの動態情報が記載されている場合があり、静態情報及び動態情報が同時に記載される史料は人口学の分析には貴重なものとなる。これに類するものとして、宗門改帳がある。宗門改帳は、江戸幕府のキリスト教信仰禁止にもなつて、当時の行政単位である村ごとに行われ、キリスト教信仰者でないことの且那寺の証明を報告している。人別帳との相違としては必ず宗派と且那寺が記載されることである。

例えば、図1は美濃国安八郡西条村の天

保八年（一八三七）の宗門人別改帳に記載されている世帯のうちの一つであるが、宗派「東本願寺」且那寺「西福寺」の記載の後に、筆頭者の定右衛門、四十八歳の記載が見られ、右肩にはその世帯の持高が書かれている。それに続いて順に、女房、卒の太四郎、留次郎、定次郎の世帯員が書かれ、メ人数が男女別に書かれてある。女房に関しては、出身地まで書かれてある。朱書きで百次郎と書かれてるのは、次の調査までの間に生まれた子供である。メの後には、順に、養子にいった兄の惣左衛門、伊勢に奉公にいった姉の志まと、同一世帯内に住んでいない縁者についての情報が書かれている。史料の記載について注意しなければならないのは、こういった史料は全国同一の形式で作成されたものではなかったということである。例えば、史料を用いる際に注意しなければならないこととして、「現住地主義」と呼ばれる宗門人別改帳は、実際にその地に住んでいる人と、奉公などで他の

地に出ており一緒に住んでいない人とは区別して書かれるが、「本籍地主義」と呼ばれる宗門人別改帳ではその両者の区別をつけることはできない。特に、前者の宗門人別改帳は人口学の史料としても利用価値の高いものである。



3 歴史人口学と宗門人別改帳とコンピュータ利用

この章では、歴史人口学で宗門人別改帳がどのように用いられてきたかをたどり、入力インターフェイスを開発する上で、考

一 東本願寺宗西福寺旦那	定右衛門	四十八
同宗 同寺旦那	女房	四十二
同宗 同寺旦那	太四郎	七
同宗 同寺旦那	留次郎	十
同宗 同寺旦那	定次郎	四
六 五人内 男四人	百次郎	
外 女一人	志法	
七 五人内 男七人	志法	
八 八人内 男八人	志法	
九 八人内 男八人	志法	
十 八人内 男八人	志法	
十一 八人内 男八人	志法	
十二 八人内 男八人	志法	
十三 八人内 男八人	志法	
十四 八人内 男八人	志法	
十五 八人内 男八人	志法	
十六 八人内 男八人	志法	
十七 八人内 男八人	志法	
十八 八人内 男八人	志法	
十九 八人内 男八人	志法	
二十 八人内 男八人	志法	

図1 美濃国安八郡西条村 天保8年 宗門人別改帳より (立教大学所蔵)

慮しなければならぬことを明らかにする。歴史人口学とは、人口を長期的な視点から研究する学問であり、フランスのルイ・アンリがキリスト教圏における教区簿冊を用いて行ったのが始まりである。教区簿冊は、その教区での洗礼、埋葬、結婚などの動態情報を記録している。彼は、これらの三つに分かれた動態情報を名前を手がかりに追跡し、「家族復元」を行うことによりその研究成果を挙げた。日本においては、ヨーロッパでルイ・アンリの著書にであった速水融が、宗門人別改帳を使って始めた。宗門人別改帳を用いた別の研究はそれ以前にもあったが、宗門人別改帳の整理方法としては、個標という一年一世帯単位で整理していくものであった。これを直接歴史人口学的分析に用いるとなると大変な苦勞が伴うと予想されたが、速水融はBDS (Basic Data Sheet) と呼ばれる史料整理方法を開発し、この問題を解決した。この方法(以下、BDS方式と呼ぶ)は、宗門人別改帳の記載に家番号、個人番号を付け加え、

家ごとに二十五年分を一つのシートに整理していく方法である。BDS方式の利点としては、整理面から見ると、各家、各個人の情報を二十五年という長期の期間にわたってつなげていくことにより、誤読、誤解が少なくなるということがあげられる。特に、史料作成者による誤記が少なからず存在する歴史資料を整理する場合は、この点は重要である。また、必ずしも宗門人別改帳の形式に当てはまらない情報（増減帳、付箋、人帳除きなど）をも整理していけるといった柔軟性にとんだ方法である。分析面から見れば、世帯の構造を長期にわたって把握できるという利点があり、今日までの歴史人口学における成果はBDS方式によって積み重ねられてきた。

宗門人別改帳を扱った研究でのコンピュータ利用は、計算プログラムの開発やデータベース構築までさまざまなものがあるが、「木下」にも紹介されている。例えば、宗門人別改帳の整理方法として最初にデータベースを利用した川口洋の方法がある（以下

「川口」）。この方法は、DANJIRO方式と呼ばれ、個標の形式で整理される。まず、史料開始年の記載を整理し、翌年からのデータ作成には、前年に入力されたものをコピーし、宗門人別改帳と照らしあわせて、その変化をワープロで入力していくという方法をとっている。再現性の問題は、宗門人別改帳の画像をデータベースに取り込むことにより解決している。この方法は、史料整理とデータベース化が同時に行えるという点で、データベース化までのコストを大幅に省いているが、入力インターフェイスとしてのコンピュータ利用までには至っていない。また別に、エクセル・マクロを利用して小野芳彦の試みも紹介されている（以下「小野」）。この方法は、親族関係を人類学コードに翻訳することを目標としており、その為に必要な家族関係を宗門人別改帳から、入力者が判断しながら入力していくものである。

本稿では、更なるコンピュータ利用の促進として、入力インターフェイスの開発が

歴史人口学の研究にどのようなメリットを与えるかを考える。入力面では、現在までのワープロによる単純な入力方法とは全く異なった方法をGUI（視覚的に工夫されたコンピュータとの対話手段）を用いたインターフェイスを開発することにより、情報量を減ずることのないデータ入力を可能とし、さらにデータの入力効率、データ精度の向上の面で大幅に改善した。これにより、情報量の豊富な汎用性のあるデータベースの構築が可能となった。史料整理面では、これまでの歴史人口学者が慣れ親しんでいたBDSによる整理方法をコンピュータ上で再現し、かつ、史料整理とデータベース化を同時に行っていくけるようにした。これにより、史料整理から研究・分析までの時間を大幅に短縮し、研究者はコンピュータ上で史料の整理をしながら、データベースの機能を用いた分析を行うことができる。また、逆に、データを用いて行った分析を整理された史料を見ながら批判することもできる。入力インターフェイスの開発によ

り、このような研究のレベルの向上をめざしている。

4 入力インターフェイス一般

開発した入力インターフェイスの説明に入る前に、まず、一般的に史料をデータベース化した数量的分析を行うことを目的とした研究で、入力インターフェイスを開発する際に注意すべき点を整理しておくことにする。

入力面、研究面から考えて、データベース化を目的とした入力インターフェイスを開発する上で特に留意しなければならない点を以下にあげる。

- ① 汎用性のあるデータ形式。
- ② 入力段階でのエラー削減。
- ③ 大量のデータ入力効率化。
- ④ 再現性の確保。

①の汎用性のあるデータ形式に関しては、データを目的別にできるだけ少なくなるよ

うに分類し、できるだけ単純な形式でのデータ抽出、また逆に単純な形式でのデータを入力インターフェイスでの編集が可能であるように開発するべきである。将来必要となってくるであろうデータを予想することは不可能であるので、汎用性のあるデータ形式のデータの編集、追加を容易に行えるように開発を行う必要がある。

②の入力段階でのエラー削減については、特に入力元に近い形でのデータ入力ができることが望ましい。また、よく使う言葉を選択して入力するようにする、間違いやすい入力事項はできるだけ自動化するなどの工夫が考えられる。

③の大量のデータ入力効率化のためには、コピー機能、学習機能、自動化によって、同じ事項を入力する手間をできるだけ省くことが大切である。また、①とも関連するが、入力しなければならない事項を絞り、入力時にできるだけ考えなくてもよいようにするべきである。

④の再現性の確保とは、②とも重なるが、

入力段階でのエラーを削減するために必要なことは言うまでもない。また、入力元の情報を全く失わずにデータ入力することが不可能であり、とくに研究を目的とした場合は、ある程度の妥協は必要となってくる。そのためにも入力元との比較が可能なインターフェイスが必要である。

以上が入力インターフェイスを開発する際に注意しなければならない点である。例えば、宗門人別改帳をデータ入力していくにしても、ワープロで毎年毎年の宗門人別改帳を入力していくなどといった方法では、③や②は全く考慮されておらず、入力でのコンピュータの利用としては不十分である。今回は、宗門人別改帳のデータ入力を課題に、まず、人口学の分析に必要な最小限のデータを決定し、入力するデータの形式を定める必要がある。そして、入力インターフェイスはこの形式のデータを効率よく入力していくようにGUIを用いての工夫が必要である。

5 宗門人別改帳データ入力

データ形式

この章では、人口学研究を目的とした宗門人別改帳から抽出すべきデータを決定し、コンピュータ入力していくデータの形式について述べる。

人別改帳のデータ入力に関しては、データを、家に関する情報、人に関する情報、異動に関する情報の三つに分けた。前者二つが、静態情報であり、残りの異動に関する情報は動態情報である。さらにこれらの情報を以下に定めた形式で整理することにした。傍線を引いた事項は、データの検索を行う際に用いることになる。

A 家情報

- (a) 村名 String
 (美濃国安八郡西条村、郡山下町 など)
 (b) 史料年 Integer
 (一七七三、一八六九など)

B 人情報

- (a) 村名 String
 (美濃国安八郡西条村、郡山下町 など)
 (b) 史料年 Integer

(c) 家番号 String

(四五、一Aなど)

(d) 記載順 Integer

(一五、一九など)

(e) 村役 String

(百姓代、庄屋など)

(f) 持高 String

(四八五九二合、水呑みなど)

(g) 付記 String

(馬一頭、権兵エ家来など)

(h) ^{しゅ}人数 Integer

(一〇など)

(i) 男子 Integer

(六など)

(j) 女子 Integer

(四など)

(一七七三、一八六九など)

(c) 所属家番号 String

(四五、一Aなど)

(d) 個人番号 String

(四五〇〇一、一五〇一など)

(e) 記載順 Integer

(一、八など)

(f) 宗派 String

(本願寺宗、禅宗など)

(g) 旦那寺 String

(西福寺など)

(h) 続柄相手 String

(又助、甚兵衛など)

(i) 続柄 String

(娘、女房、名子^{なご}、下男など)

(j) 名前 String

(甚八、よしなど)

(k) 年齢 Integer

(一〇、三〇など)

(l) 現住¹⁾ String

(UCHI, SOTO, SHU)

C 地文情報

- (a) 村名 String
(美濃国安八郡西条村、郡山下町など)
- (b) 史料年 Integer
(一七七三、一八六九など)
- (c) 家番号 String
(四五、一Aなど)
- (d) 個人番号 String
(一五〇〇一など)
- (e) 記載順 Integer
(一、二など)
- (f) 年月日 String
(去冬、当正月、安永七など)
- (g) 場所 String
(安八郡里村など)
- (h) 家名 String
(権八方など)
- (i) 方向 String
(く、よりなど)
- (j) 異動理由 String
(縁付遣、一年季など)

(k) 追記 String

- (読めず、年齢訂正ありなど)
- (1) 地文²⁾ Boolean
(TRUE, FALSE)

以上のデータのうち、実際に宗門人別改帳に記載されていないものとして、

- ① 史料年
- ② 記載順、現住、地文
- ③ 個人番号、家番号

があるが、①については西暦年であらわした。入力時には、元号、西暦年ともに表示したシートを予め作成し、直接入力する必要はないようにした。②については、史料に記載された位置に関する情報であり、統計分析のためには必要であると判断した。入力は、視覚的に入力することができるよ

A、家情報

- (村名、史料年、家番号、記載順、村役、持高、追記、メ人数、男子、女子)
- 美濃国安八郡西条村、1837, 1, 50, 一斗九合、5, 4, 1

うに工夫した。③については、同じ家、同じ人物を入力段階で結び付けることにより、同じデータの入力を省けるほうを優先した。入力には、誤入力を避けるために、半自動化し、検索機能を用いることにより、入力段階での負担を大幅に軽減した。

異動情報を上記のように年月日・場所・家名・方向・異動理由と細かく分けたことについては、よく使う言葉を後述の入力補助機能・データベース機能を有効に使うために行ったことである。

例えば、図1に示した美濃国安八郡西条村の宗門人別改帳からは、A家情報、B家情報、C地文情報とそれぞれ、以下のような形式のデータが入力される。各項目はカソマによって区切られているが、データの無い項目については何も入っていない。

B、人情報

(村名、史料年、所属家番号、個人番号、記載順、宗派、旦那寺、続柄相手、続柄、名前、年齢、現住)

美濃国安八郡西条村、1837、50、50-001、1、東本願寺宗、西福寺、定右衛門、48、UCHI
美濃国安八郡西条村、1837、50、50-002、2、東本願寺宗、西福寺、女房、42、UCHI
美濃国安八郡西条村、1837、50、50-003、3、東本願寺宗、西福寺、悴、太四郎、17、UCHI
美濃国安八郡西条村、1837、50、50-004、4、東本願寺宗、西福寺、悴、留次郎、10、UCHI
美濃国安八郡西条村、1837、50、50-005、5、東本願寺宗、西福寺、悴、定次郎、4、UCHI
美濃国安八郡西条村、1837、50、50-006、6、東本願寺宗、西福寺、悴、百次郎、2、SHU
美濃国安八郡西条村、1837、50、50-007、7、東本願寺宗、西福寺、兄、惣左衛門、63、SOTO
美濃国安八郡西条村、1837、50、50-008、8、東本願寺宗、西福寺、姉、いそ、52、SOTO
美濃国安八郡西条村、1837、50、50-009、9、東本願寺宗、西福寺、姉、志ま、50、SOTO

C、地文情報

(村名、史料年、家番号、個人番号、記載順、年月日、場所、家名、方向、異動理由、追記、地文)

美濃国安八郡西条村、1837、50、50-002、2、一九年已前、中島郡江吉良村、藤右衛門娘、縁
付来申候、TRUE
美濃国安八郡西条村、1837、50、50-006、6、出生、FALSE
美濃国安八郡西条村、1837、50、50-007、7、四三年已前、当村、甚七方、へ、養子遣申候、
TRUE
美濃国安八郡西条村、1837、50、50-008、8、安濃郡外野村、奉公遣申候、TRUE
美濃国安八郡西条村、1837、50、50-009、9、三四年已前、安濃郡内野村、彦太郎方、へ、縁
付遣、TRUE

6 宗門人別改帳入力インター

フェイス

以上のようなデータ形式を定めた上で、
入力プログラムの開発は、BDS作成をコ
ンピュータ上で行うことができ、かつ、プ
ログラムの援用により入力の手間を大幅に
省くことを目的とした。動作環境としては、
CPU 486、メモリー 16MB、ハードディ
スク空き容量 5MD 程度のデスクトップコ
ンピュータもしくはノート型コンピュ
ータで、Windows 95 が動作しているという状
況を想定し、開発を行った。開発言語とし
ては、GUI を用いたインターフェイスを
簡単に作れるという事で、Microsoft 社
の Visual Basic を用いて開発した。開発
にあたっては、人別帳を解読していただき、
ワープロ打ちされた大槻村の史料と、郡山
下町の人別改帳をBDSに整理していただ
いたものからの入力を行いながら、改良を
行った。

① 入力元のコンピュータ上での実現

Windows アプリケーションが容易に開発できるということで、Microsoft社の Visual Basic を使い、入力元となる宗門人別改帳、BDS をコンピュータ上で実現することを試みた。図2・図3にはそれぞれ、宗門帳画面とBDS画面をコンピュータ

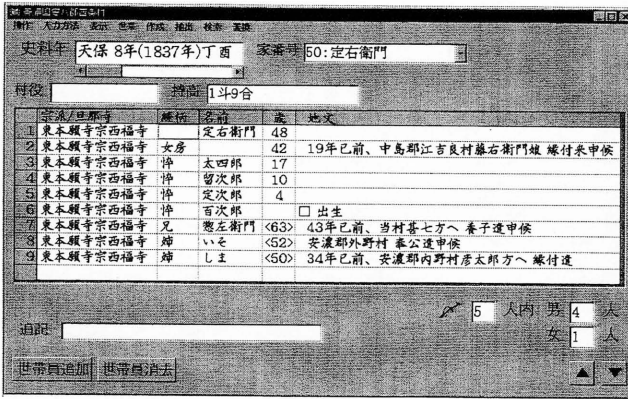


図2 宗門帳画面

タ上で実現している図である。

入力する場合は、入力元であるBDSや宗門人別改帳に書かれてある情報をコンピュータ上に再現されたBDS画面、宗門帳画面の同じ場所をクリックすると入力することができるようになっていいる。さらに、データ入力を効率よく行えるように、例え

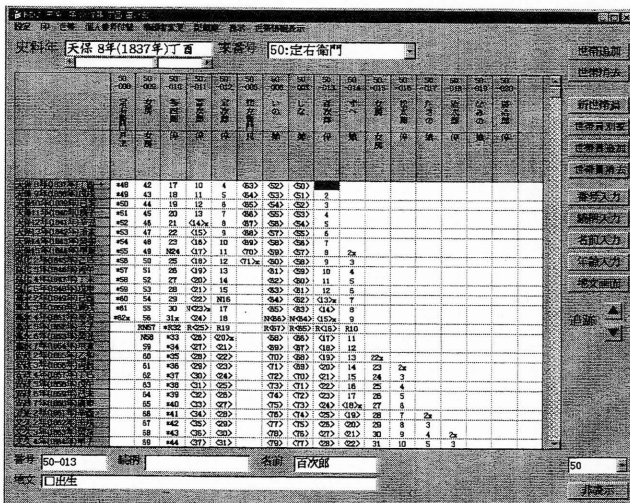


図3 BDS画面

ば、同じ人物を追加していく場合は変わらない情報(個人番号や名前)をコピーしていったり、一定の法則で変わっていく情報(年齢)をプログラムで計算して作成していく機能を加えた。こうして入力されたデータは、プログラムによって、前章で定めた形式に変換される。さらに、コンピュータ上の画面のみならず、それぞれBDSの形式や宗門帳の形式で印刷することもできる。入力ミスの校正方法はさまざまあるが、元の史料に近い形を再現することによって、最も単純な入力エラーの発見を容易にした。入力インターフェイスを開発する場合は、入力元と異なった入力になればなるほど、入力エラーが急増し、データクリーニングにかかるコストが非常に高くなることを注記しておく。

② 入力補助

入力する手間を省くようにするには、予め入力する頻度が高い言葉を選択して入力できるようにしておくことと便利である。ただ

し、この方法は、入力者が入力元から判断することによって、予め決めておいた言葉を選択するために用いてはならない。例えば、異動情報を入力する際に、予め、「出

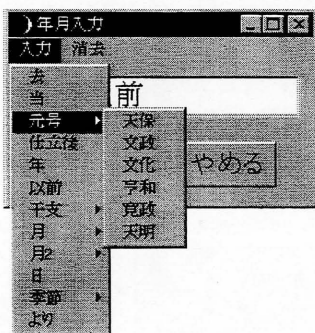


図4 元号の入力

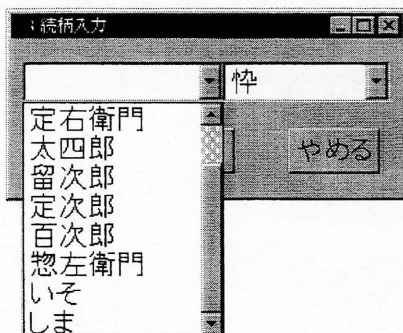


図5 続柄の入力

生」「死亡」「結婚」「養子」「奉公」といったカテゴリを決めておき、入力元に書かれてある情報をこのうちどれに当てはまるかを入力段階で判断することは避けたほうがよい。もしこの方法をとるならば、例えば、「縁付」で他家に縁付していたという情報を「結婚」と判断するか「養子」と判断するかで、縁付先でのその人の情報を得るために、長時間を費やして、史料を確認しなくてはならないか、問題が生じるたびにその場で議論を行わなくてはならないことにもなりかねない。そこで、決まりきった言葉のみ予め登録しておき、選択して入力できるように考慮した。

その一例として、異動情報には月日の情報として「天明九年」「去巳年」「四年以前」などといった表現が出てくるが、これらの言葉をそのまま入力するために、年月日の情報を入力する際には、これらの言葉を選択することができるようにした。特に、元号については、そのままの言葉をワープロ打ちする手間を省くために記載年から溯

って過去六代の元号から選択できるようにプログラムを作った(図4参照)。このようにテキストで入力されたデータの処理は、プログラムによって行われる。

また別の例としては、世帯内での地位、誰の女房であるか、誰の子であるかなどを入力するために、「誰の」に相当する部分を世帯内に同居する世帯員の名前のリストから選ぶことができるように工夫した(図5参照)。これにより、歴史人口学で重要な「家族復元」に必要な情報を確実に入力することができるようにした。

③ データベース機能

予めよく入力する言葉については、上記の入力補助機能によって入力する手間を省くことができるが、入力史料によって異なってくる言葉、つかいについては、入力を選択することはできない。ただし、一度入力したものを有効に使えるように工夫することは可能である。このプログラムで工夫したのは、入力が進むごとに、その史料でよく

使われる言葉をリストしていき、入力する際に選択できるようにしたことである。

その例として、異動情報の異動理由については、「縁付」「給取」「質物」「身受」などの言葉はできるだけそのまま残すようにした。入力が進むにつれて、これらのよく使う言葉については選択肢から選択することができるようになる(図6参照)。

さらなる利用方法としては、場所、地名情報の入力に関するものである。地名についても、入力が進むにつれて、選択肢が増

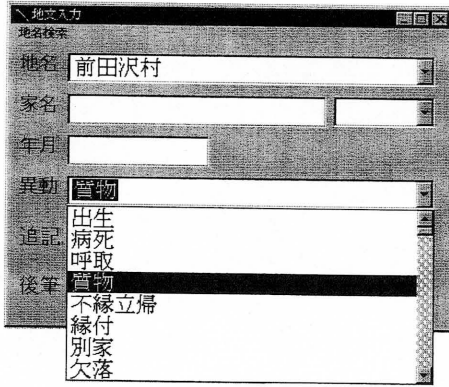


図6 異動理由の入力

えるようにしてあるが、さらに、予め近隣の村名を入力しておくことにより、入力インターフェイス上で地名を検索しながらの入力を可能にしている(図7参照)。この

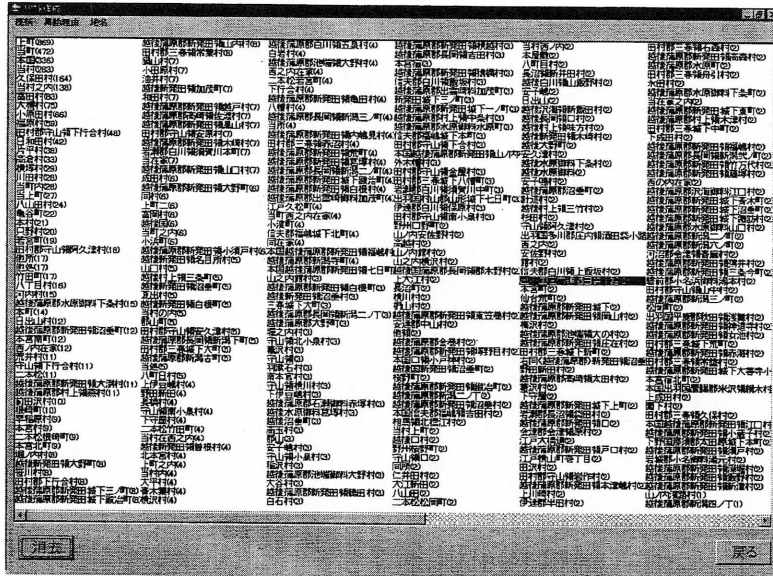


図7 地名のリスト

機能は、宗門人別改帳を解説しながら入力する際にも、大いに役に立つ機能である。

まとめ

本稿で紹介した入力インターフェイスは、文系分野でのコンピュータ利用の一例として、統計的分析を行うためのデータ入力におけるコンピュータ利用として開発を行ってみた。統計的分析を行うためには、データ入力された人口の数が重要となってくるが、今回開発したインターフェイスを用いることにより、時間当たり二〇〇人程度の入力が可能である。また、入力されたデータを用いて、すぐにでも分析を始めることもできる。このような入力インターフェイスの開発の効用を考えてみると、史料整理から分析に至るまでに

かかるコストの削減のみならず、実際に史料を用いた研究のレベルの向上にもつながるであろう。

入力インターフェイスは宗門人別改帳のデータ入力用に開発されたものであるが、ここで使われたアイディアについては、数量的分析を目的とした研究にかかわる研究者が、実際にデータ入力インターフェイスを開発しようとする場合にも応用できるものと思っている。

今後の課題としては、入力されたデータをどれだけ研究にかかしていけるかという問題が残っているが、これについては、別の機会にまとめたいと思っている。

注

(1) ^よ以前に書かれたものを現住(UCHID)、^メ以降に書かれたものを非現住(SOTO)と呼ぶ。また、お調べ時以降に入ってきた人の情報は、朱書きや貼紙(SHU)で書かれる。

(2) 異動情報には、史料作成時に書かれたものと後から書き加えられたもの(朱書、貼紙)がある。前者をTRUE、後者をFALSEとした。

参考文献

〔速水〕速水融「歴史人口学の世界」岩波書店 一九九七年。

〔木下〕木下太志「歴史人口学における人口指標算出およびグラフィック化のためのプログラム開発」『情報処理学会研究会報告 人文科学とコンピュータ』一九九六年三月。

〔小野〕小野芳彦「文科系の計算機利用Ⅱ—データ入力ユーザインターフェイス(歴史人口学の場合)—」『国際日本文化研究センター紀要 日本研究 第八集』平成五年三月。

〔川口〕川口洋「宗門改帳」データベース・システム(DANJUROの改良)』『情報処理学会研究報告 人文科学とコンピュータ』一二—一四頁 一九九二年。