

宗門人別改帳のデータベース化とその応用

森本 修馬

京都大学大学院理学研究科

統計的分析を目的としたデータベース構築には大量のデータ入力、および、データ抽出が必要不可欠である。分析にいたるまでのデータ抽出に、計算機の利用が最大限に行われている分野も有るが、歴史史料を扱う分野では、なかなか困難である場合が多かった。歴史人口学の分野では、宗門改帳、人別改帳のデータベース化、入力インターフェイスについて、さまざまな方式が試みられてきた。特に、分析にいたるまでの過程で改良の余地が残されているものとして、人別改帳の入力の効率化、再現性の向上という課題が有る。著者は、歴史人口学・社会学における分析のために必要最小限のデータを決定し、研究者による批判が可能な入力インターフェイスを開発した。ここでは、人別改帳からデータ抽出にいたるまでの過程を紹介する。

The creation of a database of “Shumon aratame cho” and its application for statistical analysis.

Shuma Morimoto

Department of mathematics, Faculty of Science, Kyoto University.

The creation of a database for statistical analysis requires a great amount of data entry and extraction. Some fields of research make full use of calculators for the process from data. Historical demographers have used many varying data interface methods for making data database out of early modern religious and population registers (Shumon Ninbetsu Aratame Cho). In the process from data entry to analysis there remains considerable room for improvement regarding the issues of efficient data entry and the ability to reproduce the original data from the database. I have developed a data entry interface that will provide the minimum data necessary for critical analysis by historical demographers and sociologists. I introduce and explain the process from population register to data extraction.

0. 歴史人口学と宗門改帳、人別改帳

歴史人口学とは、近代以前の社会をその構成要素である人口・家族の統計的分析により考えていく研究分野である。日本における近世の人口史料としては、「人別改帳」、全国民が仏教徒である事を村・町ごとに証明させた文章である「宗門改帳」がある。典型的な人別改帳・宗門改帳（以後、人別帳と呼ぶ）には、名前、年齢、続柄、異動理由が記載されており、人口研究を行う上で世界的に見ても最良の史料である。その他、補足史料として、「五人組帳」、「過去帳」、「増減帳」、「人別送り」などがある。（cf. [コーネル, 速水]

1. 人別帳のデータ整理

人別帳のデータの整理方法としては、日本の歴史人口学の祖である速水融の BDS(Basic Data Sheet)による方法が最初のものである。速水の方法（以下、BDS 方式と呼ぶ）は、人別帳の記載を家番号、個人番号を付け加え、家ごとに 25 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、入力面から見ると、各家、各個人の情報を 25 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。史料整理の面から見ると、必ずしも、人別帳の調査時点に記載された以外の情報（付箋、朱書き など）及び、人別帳以外からの情報（増減帳、人帳除きなど）をも吸収することができるという利点がある。また、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、いわば、空間的な把握ができるという利点があり、今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。

川口洋は、人別帳の整理方法として、最初にデータベースを利用した（cf. [川口 1, 2, 3]）。この方法は、DANJURO 方式と呼ばれる。DANJURO 方式では、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、人別帳と照らしあわせて、その変化を入力するという方法をとっており、入力する手間を大幅に省いている。人別帳のデータベース化の方法については、速水方式、速水・木下方式、川口方式、VBDS 方式がある。前 3 者については詳しい調査ができなかったので、VBDS 方式を検討することによって、入力インターフェイス、データクリーニング、研究者による史料批判という面から、人別帳のデータベース化について考えたい。

2. VBDS 方式について

VBDS 方式とは、BDS に整理されたデータを、計算機上でプログラムにより制御し、入力していくものである。人別帳のデータベース化をユーザーインターフェイスの面から考えられた最初の方法であり、今まで、紙の上での史料整理か、ワープロでの史料整理しか考えられていなかった歴史人口学の分野では画期的なものであった（cf. [小野]）。著者は、この入力方法を分析、入力の立場から検討してみたところ、以下に挙げる改良点があるこ

とがわかった。

- a) BDS を作成し、BDS をもとに家族の復元をしておかないと、入力が困難である。
- b) 入力には、BDS に書かれた言葉から入力者による抽出が必要で（例えば、縁付を養子とするか、結婚とするかなど）、入力者に依存する任意性が生じる可能性がある。
- c) 言葉を入力するのではなく、関係式を入力しなければならないので（例えば、倅の入力には、人別帳に書かれていない母親の入力をもしなければならない。）、データの输入が困難な場合がある。
- d) 入力されたデータと史料との双方向性が確保されていなかった。

データクリーンの面から考えてみることにする。入力されたデータに存在する不整合な点は以下のように分類できる。

- a) 史料を作成した村役人の間違い。
- b) 古文書解読段階での間違い。
- c) 入力段階での間違い。

である。a) については、データベースを構築する際には訂正を残すべきで、b) については、史料の整理方法を向上させることにより、少なくすることを考えなければならない。BDS による整理方法は、この種の間違いを発見しやすくするという点で優れている。c) については、さらに細かく、以下のように分類できる。

- c1) BDS の読み違い。
- c2) 操作ミス。
- c3) 判断の違い。

である。c1) については、入力者が BDS に習熟することからある程度解消されるものと期待してよい。c2) については、入力ソフトウェアのインターフェイスの改善が進んだとしても、ある程度生じることを考慮しなければならない。c3) については、特に、規模の大きな史料の入力には、顕著に表れてくる。データベース化を考える際には、特に、c2) と c3) を分離し、できるだけ、c3) の生じる可能性を少なくし、データエラーは c2) に収束していくように配慮しなければならない。そのために、c3) に関する判断をできるだけプログラム化し、研究者が批判できるかたちで公開することが、肝要である。

著者は、以上の問題点を考慮に入れて、宗門帳からの直接入力が可能で、任意性の生じる判断のほとんどをプログラム化し、誤入力をできるだけ避けるために、入力者による抽出を避け、入力段階でのデータ抽出を最小限にするシステムを確立した。さらに、史料との照らし合わせが可能であるように、入力されたデータを宗門帳の記載に近いかたちに印刷できるようにした。

4. 人別帳入力インターフェイス

入力インターフェイスとしては、容易に Windows のプログラムを開発できるということから、Microsoft 社の Visual Basic を用いて開発した。インターフェイスとしては、毎年、世帯ごとに、個人の情報を追加していくウインドウ（宗門帳画面）、ある1年の家と全筆頭者の情報のリストのウインドウ（世帯画面）、BDS のウインドウ（BDS 画面）の3つからなる。第1の画面は宗門帳の形式に近いものとし、宗門帳・人別帳などの元史料からの直接入力を容易にしている。第2の画面は、ある1年の全筆頭者の情報のリストを表示することで家の検索を容易にするために設けた。第3の画面は入力されたデータを繋ぎあわせてみることによって、誤入力の発見を容易にしようとするためのものである。また、BDS からの入力もこの画面を用いる事により簡単に行える。データの保存方法、アクセス方法については、年、家番号、個人番号など検索のために必要最小限のデータをメモリーに読み込み、その他のデータを1つのファイルに保存し、ランダムアクセスすることにより、大量のデータをパフォーマンスのよくないコンピュータでも扱えるようにした。（例えば、メモリー 16MB のマシンで、一度に、50000 人年分のデータを快適に扱える。）分析のためのデータ抽出を考慮に入れ、入力プログラムで入力されるデータを以下のものに絞った。

家情報	a) 史料年	(1773,1869 など)
	b) 家番号	(45,1A など)
	c) 記載順	(15, 19 など)
	d) 村役	(百姓代、庄屋 など)
	e) 持高	(48592 合、水呑 など)
	f) 付記	(馬 1 頭、権兵エ家来 など)
	g) 総人数	(10 など)
	h) 男子	(6 など)
	i) 女子	(4 など)

人情報	a) 史料年	(1773,1869 など)
	b) 所属家番号	(45,1A など)
	c) 個人番号	(45-001, 1-501 など)
	d) 宗派	(本願寺宗、禅宗 など)
	e) 旦那寺	(西福寺 など)
	f) 続柄	(女房、又助娘、源八後家 など)
	g) 名前	(甚八、よし など)
	h) 年齢	(10,30 など)

h) 現住* (TRUE, FALSE)

地文情報** a) 史料年 (1773, 1869 など)

b) 家番号 (45, 1A など)

c) 個人番号 (15-001 など)

d) 記載順 (1, 2 など)

e) 年月日 (去冬, 当正月, 安永 7 など)

f) 場所 (安八郡里村 など)

g) 続柄 (彦六娘 など)

h) 異動理由 (縁付遣、一年季 など)

i) 後筆*** (TRUE, FALSE)

* 現住とは、 $\wedge$ 以前に書かれたものを現住、 $\wedge$ 以降に書かれたものを非現住と呼ぶ。

現住地主義の人別帳(cf. [山口],[木下])には、 $\wedge$ 以降にも、移動に関する重要な情報が書かれてある。

** 地文情報には、史料作成時に書かれたもの、後から書き加えられたもの(朱書き、貼紙など)がある。前者を FALSE、後者を TRUE にした。後筆の情報は、史料作成時点(各地域により何月に行われるかがおおそ決まっている)の静的な情報のみを扱うのではなく、お調べとお調べの間に起こった動的な情報を扱う場合、重要になる。

以上のデータのうち、実際に人別帳に記載されていないものとして、

a) 記載順、現住、後筆。

b) 個人番号、家番号。

があるが、a) については、史料に記載された位置に関する情報であり、宗門帳との双方向性を保つために必要であると判断した。入力には、視覚的に入力することができるように工夫した。b) 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、同じデータを入力する手間を省けるほうを優先し、誤入力を避けるために、半自動化し、その上、検索機能を付け加えることにより、入力段階での負担を大幅に軽減した。

入力方法としては、1 年分のデータの入力がすむと、翌年のデータを前年のデータをもとに生成するという方式を採用し、さらに、各年毎に入力されたデータを長期にわたってつなげて確認できる BDS 方式を取り入れた。この方法により、史料解読からデータベース化への時間が大幅に短縮できるものと予想される。分析のためのデータ抽出プログラムとしては、抽出が困難と思われる結婚関係のデータ抽出プログラムを作成し、未婚率の計算をしてみた (cf. [浜野、黒須、森本])。

5. 今後の課題

現在のところ、実際に必要となるデータが得られるかどうかのテストが終わり、システムの評価についてはこれからである。入力されたデータを使い、ユーラシア人口・家族史プロジェクト（研究代表者 速水融）における国際比較、また、日本の歴史人口学的指標の地域差などの研究を行っていきたい。

[コーネル、速水] ローレル・L・コーネル、速水融 「宗門改帳－日本の人口記録」 中井美由紀訳 速水他編「徳川社会からの展望－発展・構造・国際関係－」第6章所収 同文館 1989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」 1996年3月。

[山口] 山口喜一編著 「人口分析入門」 古今書院。

[小野] 小野芳彦 「文科系の計算機利用Ⅱ－データ入力のユーザーインターフェイス（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8集 平成5年3月。

[川口1] 川口洋、中山和彦 「「宗門改帳」データベース（DANJURO）の開発」 「1989年情報学シンポジウム講演論文集」 pp.125-134. 1989年。

[川口2] 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「宗門改入別帳家別帳」のデータベース化を実例として－」 「歴史地理学」 151 pp.19-33. 1990年。

[川口3] 川口洋 「「宗門改帳」データベース・システム（DANJUROの改良）」 「情報処理学会研究会報告 人文科学とコンピュータ」 pp.12-14 1992年。

[浜野、黒須、森本] 浜野潔、黒須里美、森本修馬 「江戸時代は皆婚社会か？」