

国文学文献資料のデータベース

宮澤 彰 (国文学研究資料館)

1. はじめに

文献データベースと呼ばれるものは数多くあり、そのほとんどは一定分野の学問の研究論文を対象とするものである。これらに對し「和古書」のデータベースは、文献データベースであるとしてもやや異った性格も持っている。それは「和古書」がそれ自体で文献学の対象としての直接資料であることによるものであろう。研究の二次資料である論文に較べ、直接対象であるために、正確さを要求される、構造についても表現する必要が有る、研究成果を反映できるようなものであることが要求される、などのむづかしさを持つているわけである。この点は化学での化合物データベースなどに似ているかもしれない。しかしながら、ものとしては文献であり、タイトル、著者など項目としては一般の文献データベースに似たデータである。

ここでは国文学の文献資料である「和古書」について説明し、その構造をデータベース化について紹介する。

2. 和古書について

2.1 範囲

「和古書」というのは、江戸以前の日本の本をさす。といへば話は簡単だが、文献学的にはその定義だけで問題になるようである。まず第1の問題は「江戸以前」という年代的なもので、明治になっても本の出版形態が一時に変化したわけではなから、江戸期とほとんど同じ形の本が出版されている。これを厳密に区別することはむづかしいから実際的には明治時代出版の和古書も認めざるを得ないだろう。しかしこれも大正、昭和まで拡張することには問題を感じるのが普通である。

第2の問題は「日本の」という地理的な問題である。一般に「漢籍」——中国の本——は和古書に加えて有り、これも当然のようだがむづかしい問題になる。江戸以前に中国から輸入された本は数多いし、日本の本もまた中国のものをまねて作られた。素人には和古書か漢籍か区別のつかないことがしばしばある。まして本が破損して表紙などがなくなれば専門家でも見分けるのはむづかしくなる。これに日本で表紙をつければこれは漢籍か和古書か、なげと存せし法律論議かクイズの問題としか思えないが、データベース入力の實際に、そのような本を前にして迷わないための原則はやはり必要である。

最後の問題は(一語々々難くせをかけたようだが)「本」とはなににかにある。例えば江戸時代から続く店の蔵から古い大福帳がでてきたとして、これは「本」だろうか。多くの人はノーと答えるだろう。では、五代前の御先祖の日記がでてきた、これは本だろうか。勿分あまり本らしくはなっているが、本のよるかも知れないところではないだろうか。しかし、冷泉家の土蔵から出てきた藤原定家の日記と存せ、これはもう立派に本として適用しえうである。いずれかの段階で版本となって出版されたら、もう文句のつけようなく「本」と存する。出版されていない本——「写本」——と「本」とはついている文書(もんじょ)類との境界

は、このようにあまりは、きりやはしていない。ただ、歴史学の史料としか存在しないものが文書(もんじょ)で、文学的研究対象になったものは「本」ということはいえるだろう。

これらの「和古書」周辺のものは、それらを対象とする分野が異なため、当然整理のされかたも異なっており、「和古書」と同じデータベースに統合することは不可能であるし、意味もない。しかしながらそれらを完全に排除してつくることも現実的には無理なため、一部でたとえば文書(もんじょ)を和古書扱いして整理することになる。この場合には文書としての検索は有効にできなくなるが、やむをえないことだろう。

さて、上記のような和古書は、日本の各地の図書館、寺社、個人、そして古本屋に所蔵されている。その数は推計で2~3百万点をもそれ以上とも言われる。2~3百万点の数字は比較的数を掴みやすい図書館蔵のものからの推計であるから、個人蔵をいれたり、文書(もんじょ)との境界をゆるくすれば1千万点程度にはなるだろう。国文学研究資料館では国文学の文献資料であるこれらの和古書について調査し、マイクロフィルムとして収集している。

2. 2 形態など

和古書の多くはいわゆる「和装本」であった。他には卷子、一枚ものなど多くの形態もあるが、数は少ない。典型的な版本では1点が数冊(5冊、10冊など)からなり、1冊あたり数十丁のふくらみである。タイトル、著者、出版事項などの主な書誌事項は、表紙、扉(表紙の次の丁)、目録(目次のこと)、序文、本文の第1ページ、最終巻の本文最終ページ、跋文、最終巻最終丁にある刊記などから得られる。刊記は現在の新刊書での奥付にあたるもので、刊年と本屋などが書いてある。とくに本屋が発行者という意味の版元だけでなく、売り捌いている本屋までふくめて多人数書いてあるのが、現在の奥付との大きな相違かもしれない。もちろん地名の存在、本屋名の存在、刊年の存在もいろいろにみられる。

写本の場合、版本ほどの一定の様式はないが、大体版本に準じた造本のものが多く、扉、表紙などの用語は版本に準じて使われる。もちろん刊記はないが、ある場合には奥書と呼ばれるこの本の由来、書写者などを記した文が、本文の後に書かれている場合がある。書写事項は多くこの奥書からとられるが、次に写す人が前の奥書も含めて写し、自分のことは書かない場合などもあるため、奥書がそのまま信用できるわけではない。

2. 3 記載題と統一書名

2. 2 に述べたように、タイトルは本の表紙などから得られる。主なものをあげると

外題：表紙上にある題 — 表紙の上に直書してあることもあるが、典型的な版本では別紙に題名だけ刷って貼付けてある。(これを題簽という)

内題：広義には表紙以外のところにある題のことだが、狭い意味では本文先頭にある題名。

扉題：扉に書いてある題——扉は表紙と本文との間に入れた丁で、ここには題名の他に著者、本屋名なども書いてある場合がある。

目録題：目録（目次）の先頭には必ず「XXX目録」というように書いてある。ここに記された題である。

柱題：版心（ふくすまじのためめに折、た紙の折り目の部分）にある題名、ここにある題は略称のような場合が多い。

これらのタイトルがなぜ細かく問題になるかと言うと、内題と外題が一致しないというのがよくふつうにあるからである。それらは《春色梅の辻占》と《春色梅辻占》というようにごくわずかの表記の違いであり、《愚秋抄鶴本》と《愚秋抄鶴》という一部の省略形であり、たりすることもあるが、《女大学》と《女今川雲井鶴》というように全く異なる場合もある。さらにいわゆる角書（つのがき）をもつ場合もある——《新版繪入／忠盛祇園櫻》。これらの違いが本の系統を調べるなどの文献学的研究に必要なデータとなっている。またある題名がどこに記載されているかということは、次に述べる統一書名を決定する手がかりになるのである。

上のような本に書かれていたタイトルは記載題という。記載題がいろいろある時にはふつう、それらのうち1つを統一書名とする。統一書名の必要理由は、次のような場合を考へるべし。

ある図書館で《女大学》を整理して《女大学》という一本に収めたい。後でまた同じ本を整理した時には《女今川雲井鶴》という一本に収めたい。こうすると実は同じ《女大学》の2セットが、別のものとなってしまふ。

場合によっては記載題が一切ないという本を整理することもある。この場合は何かの方法で統一書名をみつけなければならぬ。ある場合には全く書名が書いてなくても、内容を読むと実は《伊勢物語》だとわかるかもしれない。またある場合には、記載題は別でも、統一書名は《伊勢物語》にしなければならぬこともある。《左五中將物語》が《伊勢物語》だというのはよく知られた例であろう。

このように見ると、統一書名という概念は個別の図書ごとの記載題とははなれて、作品全体につけられた書名を考へた方がわかり易くなる。この考へは文書（もんじょ）に近いものではややぼけるかもしれないが、典型的な文学作品の場合には問題ないだろう。

2. 4 著者項目

著者名の場合にも、タイトルである、たのと同様のことが見られる。第1に、著者名は一般に本文冒頭（内題の下）、扉、刊記などの場所に書かれるが、それらの間で別々の書き方をされる。例えば《八犬伝》第一輯の本文冒頭では、曲亭主人とあり、刊記には曲亭馬琴とある。さらに序では装笠陳人鮮などとある。これらはすべて一人の馬琴の名であるが、それを見分けするのはある程度の知識を要する。

第2に同じ作品である、ても、本が違えば別の名前が書いてある場合がある。ある本には著者名がなく、ある本には著者名があるという場合があるだろう。それ

5を総合してある作品の著者を決定するのは(ある場合には)さらにむづかしい仕事となる。

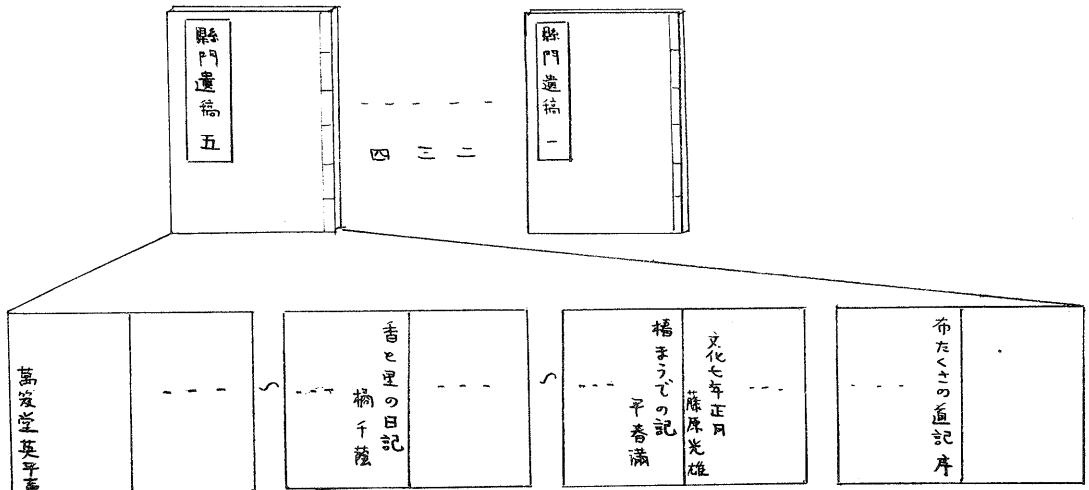
第3にこのようにして著者名が集ま、たとえて、四方赤良が蜀山人で大団南敵と同一人物であることを識子のはまた別の話である。江戸以前の日本人は1人でいくつも名を持っており、何回でも改名することが多いため、かなりの実際の問題である。しかしながら著者名からの検索を有効にするためには避けられなかったらう。

2.5 叢書・合集など

これまで述べてきたのは和古書の中でも単行書——図書1点に作品1つが入ったもの——についてである。これに対して新刊書では逐次刊行物というのがあるが、和古書では叢書とか合集などの概念がある。これは新刊書にもあることで「銀河鉄道の夜他2篇」という本は1冊の中に3つの作品をおさめてある。我々は和古書でのこのような例を合刻(版本の場合)とか合写(写本の場合)とか合集などと呼んでいる。合写・合刻はかなり多く見られ、点数でいって全体の10%をこえた本がこのような非単行書ではないかと推計している。とくに写本では合写はかなり多いし、版本でも例えば「伊勢物語」と「古今和歌集」とを上下に合わせたという本などもある。

合写・合刻によく似たものに叢書(広い意味の)や我々が仮に「セットもの」と呼んでいる形態がある。この典型的なものに「八代集」がある。「八代集」は「古今」「後撰」「拾遺」「後拾遺」「定葉」「詞花」「千載」「新古今」の8つの執撰和歌集を、後世総称してそう呼ぶようにな、たものであるが、実際の本にも記載題として「八代集」をもちた8冊本のこの8つの歌集があらわれている。このような例はかなり多い。これがさらに大まな現様で見られるのが叢書である。例えば「群書類従」は膨大な量の作品がさらさら一大叢書であるが、もう少し少ない作品数(数十点)のものもかなりある。これが組みあわせて2レベルのネストをもち例を図-1で紹介しよう。

図-1 レベル2の叢書例



この例では《奥門遺稿》という叢書があり、その中に《ふたくせ日記》(記載題では《ふたくせの道記》)というセットものが含まれ、それが実際の作品として《櫛まうでの記》と《香セリの日記》とからなっているという構成である。2段階の叢書は1%に満たないが、無視できない数あることは確かである。

3. データベースについて

3.1 データ項目

このデータベースに関するデータ項目を、作品のレベル、個々の図書レベル、著者のレベルにわけあげてみよう。各項目もさらに細かい項目からなっているが、大きな単位でまよめておく。またしコード管理用の情報(整理した年月日など)等の細かい項目は略す。

表-1 作品レベルのデータ項目

項目	内容	備考
統一書名	・伊勢物語, 古今和歌集 など	
著者表示	・八文字白笑一世, 八文字其笑一世著, 昌琢, 昌俣, 具野著 など	くり返し
成立年代	・安永2年, 慶長年間, 平安後期, 文化11~天保13年 など	
分類等	・黄表紙, 俳諧, 物語 など	くり返し
注記	・一般注記, 典拠注記 など	

統一書名は1つの作品に1つであるが、残念ながら同名異書と呼ばれる作品群があり、プライマリキーはなさない。

著者表示というのは著者名と役割表示(著, 編, 画など)をあわせてもので複数のことになりありし、また十返舎一九著画といふものもある。多い場合は脚で略す。もう一つ、著者表示としては必ずしも本名に統一するというわけではない。伝統的に狂歌の著者表示では大田南畝といふ方はいないようである。

表-2 図書のレベルのデータ項目

項目	内容	備考
記載題	角書(冠称)や尾称などを含むこともある。記載部所(外題, 内題など)と組み合わせる。	くり返し
記載著者名	・記載部所(扉, 巻首など), 役割表示と組み合わせる	くり返し
形態事項	刊/写の別, 5冊, 10丁, 1軸 など	
書写/出版事項	・写年と書写者名, 刊年と地名, 書肆(本屋)名	
注記	・系統注記, 伝来, 書き入れ など	
叢書・合巻関係	・xxxと合, xxxの内 など	

記載題では内題、外題などの別を組み合わせたものが、同じ題が外題と扉にせうこせは多し。また文冊以上の本の場合、各冊によつて内題が異なるといふこともある。

記載著者名では役割表示の他に、その名がどこに書いてある、不かも必要の場合がある。

書写/出版事項は一般には刊(写)年1つと、1人の書写者または、複数の地名、書肆(本屋)名の組からなるが、まれに刊年が何年かにわたつて出版されている場合もある。

叢書、合集関係については後述する。

表-3 著者のレベルのデータ項目

項目	内容	備考
名前	姓、名、氏、世系、号などからなる	くり返し
生没年	文化2年~慶応3年、宝永3年没、 鎌倉時代中期 など	
職業・身分	歌人、国学者 など	くり返し
注記	典故など	

名前は1人の人物がいくつも持つ。曲亭馬琴と瀧澤解など。そしてその名前の各部分が姓だ、たり、号だ、たりする。従つて色々な組み合わせがあり得るが、1人の人が実際に使う場合、組み合わせは限られているようである。

3.2 可変長とくり返し

上記のように整理したうち、書名や人名などの「名前」項目は、日本語の特性として必ず表記(漢字)と読み(カタ)からなる。これら「名前」項目は短いものから長いものまでさまざまあり(これは一般的に文献データの特徴かもしれない)、和古書の場合殆どくり返し項目に存在している。

参考として既存のデータ(国文学研究資料館蔵マイクログラフ資料目録)から著者名、書名の長さでくり返し数についての統計を示そう。著者名の長さには姓名などの区切りに1文字分を数えている。

表-4 名前の長さでくり返し数に関する統計

	最大	最小	平均
著者名の長さ(表記)	13字	1字	4.53字
(ヨミ)	23字	1字	7.95字
作品あたりの著者数	17	0	0.85
書名の長さ(表記)	49字	1字	5.64字
(ヨミ)	96字	1字	10.36字
記載題の個数(図書あたり)	21	0	3.06

3.3 データベースの構成

表-1~表-3まではほぼ、作品・本・著者に対応してそれをあつちすレコードでし得るのだが、モード一つの問題がある。この問題には叢書・合集の問題がかさんでいる。

たとえば1冊の版本が《伊勢物語》と《古今集》の合刻で出版された時、表-2のデータ項目のうち、形態や出版事項はこの合刻された1冊に関する属性である。しかし記載題や記載著者名は分離してとらなければならぬ。これに討して同じ《伊勢物語》と《古今集》を1人の人が1冊に写した。しかし奥書をつけて《伊勢物語》は何年何月、《古今集》はいついつとわかっているはず。この場合には形態だけがこの合字された1冊に関する項目で、他は分離してとらえなければならない。叢書やセットものに対する場合も似たことがおこる。ある場合は同時に出版されてセット全体に対する出版事項があるだろう。別の場合には形態もかわって、記載題だけが共通の項目かもしれない。

実は単行書と思われるものでも、出版年が何年かにわたって同様のことが見られた場合もあるのである。これは形態という物理的な本の属性と、タイトルなどの内容的な属性とをずれたら生じる問題ではあるが、これを正直に反映させることは（不可能では無いが）あまりに煩雑で実際的で無い。我々は出版/書写事項も形態も学に分離してとり、合冊や叢書の全体をまとめて記述は別におくことにした。

この考え方で、表-1~3の構造をみると、次のようになる。

図-2 データ構造

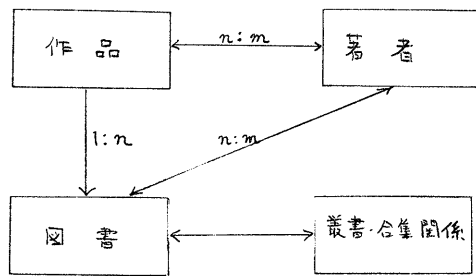


図-2 厳密なバックマン
ダイアグラムではない。

ここでは各レベルの項目に表-1~3のようなくり返し項目を許して、そのため $n:m$ の関係がでてしまう。しかしながらこの見方は、図書-作品-著者という伝統的な見方に近く理解はしやすい。

残念ながらこの先のインプリメントをどのように行なうかは、計画の進行中のため、未定である。この構造をなるべく素直に表現できて、しかも叢書・合集関係の表わし易いような方法が望まれている。

4. おわりに

かなり細かく複雑な泥臭い話になつてしまつた。このような紹介がデータベースの研究に益するこゝとができれば幸いです。