

記録類の全文入力を試み

——『石山本願寺日記』を素材に——

水 藤 真

-
- | | |
|------------------|-----------------------|
| はじめに | 3. 『石山本願寺日記』データベースの内容 |
| 1. 全文入力に何を期待するか? | 4. コンピューターの効用のほど |
| 2. 入力作業の実例 | 5. 今後の課題 |
-

論文要旨

この「記録類の全文入力を試み」というのは、日記や編さん物の全文をコンピューターに入力して、その検索・統計などの用途に益ならしめようとするものである。例えば、「祭」という言葉が何という日記の何年何月何日条に出て来るかを、コンピューターを使って、瞬時に知ることは出来ないか、などという期待を実現出来ないかという試みである。

当然、過去に記された日記には色々な特徴があり、コンピューター化に適したものとそうでないものがある。例えば、日記に記された文章は、それぞれの日にちで長さが異なる。また、文章の途中に書き直したものや、後で追記されたものが挿入されるとか、また現在では使われていない文字が使われているとか、様々なことがあり、単純にただ文章をコンピューターに打ち込めば良いというものではない。この為の諸条件の整備が最初の難関であった。

次には、コンピューターの操作が難しいという問題があった。

こうして、いくつかの関門を突破しつつ、次の努力目標を掲げた。

誰でもが使えるような易しいシステムであること。あらかじめキーワードを準備するのではなく、任意の文字列で検索出来ること。この2点である。こうしてシステムは兎も角動くようになった。しかし、一旦稼働することが分かると、次には新たな問題が持ち上がった。例えば検索の為の時間である。当初、時間はいくらかかっても、任意の文字列で検索出来ることを至上目標とした。当然の結果とはいえ、時間は40万字の検索で12、3分を要することとなった。それでも、初めて検索結果の出来た時は感動であった。しかし、2度、3度と異なる言葉を検索しようとした時、これではかなり障害となる。そこで今度は時間短縮を計ることが重要課題となった。そして、今、1件当たり約1、2分で検索出来る。本稿はこの悪戦苦闘の記録である。

はじめに

国立歴史民俗博物館歴史研究部では、昭和63年度を皮切りに、改めてコンピューターの活用をより深く考えることとし、同部の西沢奈津子・福田豊彦・湯浅隆・水藤真の4名が中心となって、この作業にあたることとした。そして数ある問題の中から、ひとつは古記録の全文入力と試験的に取り組むこととし、昭和63年度は古代の記録（担当西沢）・平成元年度は中世の日記（担当水藤）・平成2年度は近世の日記（担当湯浅）を順次入力してみることにした。

何分初めてのことで全てが暗中模索・試行錯誤の繰り返しであり、未だ実験段階の域を出るものではない。いずれは公開というか、多くの人に利用して頂きたいとの夢もあるがまだまだ先のことと思う。それよりもまず当面は、果たして、歴史研究にコンピューターが本当に利用出来るのか？ また古代から近世までの史料、その中で日記類に限った場合でも、同じ書式でコンピューターに入力し稼働させることが出来るものかどうか？ これらを確認することが先決であった。次には実際の入力をどういう風にするのが最も良いかなどが検討課題となる。例えば割り注をどうするか、傍注をどうするか、第2水準までにない漢字をどう表現するかなどなど、頭の痛いことが山積していた。こうした状態からのスタートである、この試みは未だ優れて実験・検討段階であることを最初にお断りしておきたい。

この試みは上記の4人の言わば共同研究の成果でもある。たまたま私が連絡役をしたということもあるので、各位の御了解を得て、取り敢えず自身が担当した『石山本願寺日記』の入力の経験を発表させて頂くこととする。多少ともこの種の作業の参考になればと思うからである。なお本稿で果たし得ない不備な点についてはいずれ、それぞれを分担した方々から、より充実したものとして御発表頂けるのではないかと期待したい。

1. 全文入力に何を期待するか？

日記本文の全文入力をするに当たって、一体何を期待するか？ わざわざコンピューターまで使って、どういう便利さを求めるのか？ 日本史の研究者が、コンピューターに期待することは何か？ その最低限の要求・最大限の要求は何か？

当面、日記の全文がコンピューターに入力されたと仮定して、ある文字・単語・熟語で検索して、その結果が画面に出力されることが大事である。次にその検索の場合あらかじめ検索の為の単語が指定されている、つまりキーワードを設定するのではなく、任意の文字列で検索出来ることが望まれる。すなわち、好きな時、好きな単語・言葉で検索出来ること。これを最低限の要求項目とした。

こうしてまとめた第1次案はおよそ以下の様なものであった。

参考資料1：PC全文入力システム開発

1. 記録類（日記・編さん物）の全文を原則として1日単位で入力する。

従ってそのまま全体をプリントアウトした時は年月日を追った日記（＝本）そのものとなる。

2. 検索は任意の文字列で出来る。

その結果は、

- (1) その文字列を含む日にちの、年月日一覧表として出力出来る。
- (2) その文字列を含む日にちの、本文を1日単位で見ることが出来る。

3. 一度検索した結果を2次テーブル（例えば、地名テーブル・人名テーブル）に登録出来る。

従って、一度検索したもの（文字列）は、2回目以降は2次テーブルでより早く検索出来、ここからも本文を見る事が出来る。当然、ダブリの登録はしない。

4. データの構成

1日単位のデータは以下のものからなる。

- (1) 年月日（和暦）
- (2) 年月日（西暦コード・数字及び段落の12桁）
- (3) 出典
- (4) 本文 本文は750字を基本とし、それを越える場合はさらに750字、さらにそれを越える場合はさらに500字まで都合、最大限2,000字まで入力出来るものとする。

5. 2次テーブルの項目

- (1) 固有名詞テーブル（地名テーブル）
- (2) 固有名詞テーブル（人名・官職名テーブル）
- (3) 事項テーブル
- (4) 備考テーブル
- (5) その他 今後予想されるいくつかのもの（例えば西暦和暦対応テーブル・十干十二支対応テーブルなど）

6. その他の希望

- (1) AND・ORの検索は出来るだけ多く出来る事が望ましい。
- (2) 50万字の検索を1分以内にする。
- (3) ヒットした文字列の数及び日数（1日単位のヒット数）を表示出来る。
- (4) 本文検索の場合ヒットした文字列を強調表示出来る。

データの項目としては極めて単純である。すなわち、(1)和暦、(2)西暦コード、(3)出典、(4)本文、のたった4項目しかないのである。

さて、次には入力すべき本文の問題がある。

これについては、『石山本願寺日記』の入力の場合には次の様にまとめてみた。

参考資料2：石山本願寺日記全文入力 入力要綱

1. 漢字は原則として常用漢字を用いる。
常用漢字に無いものは、JIS第1水準・第2水準の漢字を使用する。
同音・同意の漢字が2種以上ある時は、易しい方の漢字を使う。
第2水準までにない漢字は、当面「■」記号を付し、その読みを単語単位で《 》で括る。
2. 入力は、1日単位で行う。
一日条が長い場合は、各750字以内の段落に強制的に分割し、年月日を示す数字列の後にA、B、C……を付す(次項参照)。段落の無い場合(750文字以内の時)は、0を入れる。
3. 各段落ごとに年月日を記す。
年月日は、年が4桁・判断領域が2桁・月が3桁・日が2桁、及び段落のA、B、C等、計12桁とする(このやり方は、東京大学史料編さん所に教えて頂いたものを基本とし、それに段落分1桁を加えた)。
4. 和暦の年月日を各段落本文のはじめに記す。
この際、前項3と連動し、自動的に相互の読み替えが出来る様にする。
5. 文節の区切りは、底本の区切りとする。
6. 割り注は、()に入れる。
7. 傍注・编者注・当て字などは備考欄で処理するか、当該の箇所、[]で入れる。
(注) すなわち()は原文にあるもの、[]は後の注ということになる。
8. 普通、漢字で表記される言葉のかな表記・漢字かな混じりの表記は、単語・熟語単位にその直後に、[]で囲み、当該の漢字を入れる。
9. 推定の判読文字は、単語・熟語単位に、その直後に、[]内に単語・熟語単位で漢字を入れる。
10. 引用文は、「 」に入れるか、2段落とする。
11. 虫喰い欄は、□□□を推定字数分入れる。
字数不明の場合は、□……□とする。
12. 見せ消ちは、消された文字の先頭に：を入れ、最後にも：を入れ、書き直しの文字は()に入れる。
13. 抹消は、：(抹消)：とし、書き直しの文字は、()に入れる。

14. 底本の括弧は、〈 〉に置き換える。

… (以下略) …

あまりに煩瑣になってはと大分省略したが、おおよそ以上のような約束ごとを決め、基本的にはこれにのっとって入力を開始した。

その前に今少しいくつかの了解事項を持った。今それを簡条書きに列記すると以下の様なものである。

- (1) コンピューターはあくまで器械であって、我々の生活・研究を手助けするはずのものである。従って、有用でない場合また却って研究の阻害になる場合は、躊躇なく廃棄すること。
- (2) 我々はコンピューターの専門家ではない。コンピューターの勉強は一切必要ないこと。全くのユーザーとして、人文系研究者として、日頃扱う史料・データの整理・研究に役立つものであるか否か、この点のみを確認すること。
- (3) 上記(2)に関連して、その操作は全くの素人でも行えること。少なくともワープロが出来る人間であれば誰でも操作出来ること。
- (4) とはいえ、やはり新たな試みに挑むのであるから、多少の負担は生ぜざるを得ないだろう。従って入力するデータ（具体的には日記）は各担当者が直ぐに自分の研究に生かせるものを選ばせて欲しいこと。つまりこの試みに挑む人々の研究のプラスになるものであること。
- (5) 上記(4)と関連して、その代わり当面コンピューターに関心のない人・コンピューターを毛嫌いする人にこの仕事を押し付けない、またコンピューターに無関心であると非難しない。
- (6) またコンピューターに取り組む人々においても、その取り組む時間は、1日2時間、週2日、つまり1週間当たり計4時間（＝半日）を越えてはならない。なぜなら他にやるべきことが沢山ある。コンピューターは、他にやるべき仕事、すなわち本務を手助けする為のものはずである。もしこれ以上の時間を要するなら、研究支援とは言えない。様々な仕事のコンピューター化が我々の仕事ではない。

以上である。では次に実際の入力作業の手順を紹介したい。

2. 入力作業の実例

データベース作成の最も大変なことは2つある。ひとつはデータをどういう形で入力し、何をさせるか？ つまり入力の書式（フォーマット）とソフトをどうするかという問題である。書式の設定はコンピューターに何をさせるか？ ということ、最も大切なことである。これは十分検討する必要がある。当面我々のものは先に示した通りである。ソフトはソフト開発会社・あるいはコンピューターの専門家に依頼する。あるいはマニアの人にお問い合わせするという方法もある。いずれにせよ我々は素人である。

見通しはたったと思う。従ってこの先はこれをしばらく稼働してみて種々の改良を加え、改めてより良いデータベース作成に取り組む必要があるのではないかと思う。

では、兎も角も当面作成したデータベースの中身を見ることにしよう。

3. 『石山本願寺日記』データベースの内容

まずパソコンのスイッチを入れ、このデータベースを呼び出すと、次の画面が出る。

画面 1

```

記録類 全文入力データベース
(1) データの追加入力
(2) データの修正・削除
(3) データの検索
(4) データの読み込み
(5) システム管理
(6) R B A S E   P R O
(7) 終      了

```

画面1の様な6種のことが出る。検索をしたい場合は、ここで、「(3)のデータの検索」にカーソルを移動し、リターンキーを押すと、次の画面が出る。

画面 2

```

記録類 全文入力データベース
(1) データの追加入力
(2) データの修正・削除
(3) データの検索
(4) データの読み込み
(5) システム管理
(6) R B A S E   P R O
(7) 終      了

【ESC】-終了  検索処理選択メニュー
(1) 条件検索
(2) 西暦・和暦一覧表 (文字列含む)
(3) 西暦・和暦一覧表
(4) 2次テーブル検索 - 本文データ参照

```

ここでは、画面2の様に4種の検索が出る。今は「(1)条件検索」と「(2)西暦・和暦一覧表(文字列を含む)」を紹介しよう。

そこで、「(1)条件検索」にカーソルを移動し、リターンキーを押すと、次の画面が出る。

画面 3

【ESC】-終了 テーブル選択
(1) 本文テーブル
(2) 官職名テーブル
(3) 地名テーブル
(4) 事項テーブル
(5) 備考テーブル

ここでも画面3の様に5種の検索が出来る。今は「(1)本文テーブル」を選んでみよう。「(1)本文テーブル」にカーソルを移動し、リターンキーを押す。すると、次の画面が出る。

(注) 画面3の(2)以下は、2次テーブルで1回検索した結果をそれぞれのテーブルに登録出来るようにしてある。従って二回目以降の検索はこの2次テーブルを使うと早くなることを予想して作成したが……、結果は必ずしも思わしくない。目下検討中。

画面 4

本文テーブル検索画面

西暦年月日段落	和暦年月日
[本文]	出典
1	
2	
3	

ここでも、それぞれの欄で検索出来る。つまり西暦年月日段落・和暦年月日・及び本文である。そして本文欄では最大3つまでの単語(文字列)を含む本文を検索できる(例えば、「越前」と「朝倉」と「路次」を含む本文があるかどうか)。では例えば「祭」を含む本文を検索してみよう。まず画面4の本文欄でカーソルを「本文」1に移動する(カーソルはリターンキーまたは→キーで移動出来る)。ここで「祭」と入力する(画面5)。

画面 5

本文テーブル検索画面

西暦年月日段落	和暦年月日
[本文]	出典
1 祭	
2	
3	

入力が終わったら、リターンキーを押して行く。すると画面6が出る。

画面6

E/再編集 D/取消		処理を選択して下さい	[ESC]-検索実行
西暦年月日段落		和暦年月日	
[本文]		出典	
1 祭			
2			
3			

ここでは画面上部の指示のごとく、エスケープキーを押す。すると画面7が出る。

画面7

実行 中止

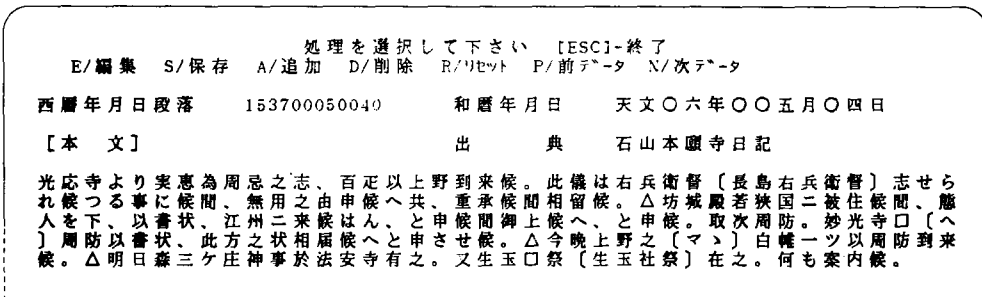
既に実行が光っている（カーソルがここにある）ので、そのままリターンキーを押す。すると画面8の様に「しばらくお待ち下さい……」と出る。

画面8

実行 中止
しばらくお待ち下さい.....

実際しばらくすると、ピッと音がなり、画面9が表示される。

画面10-2



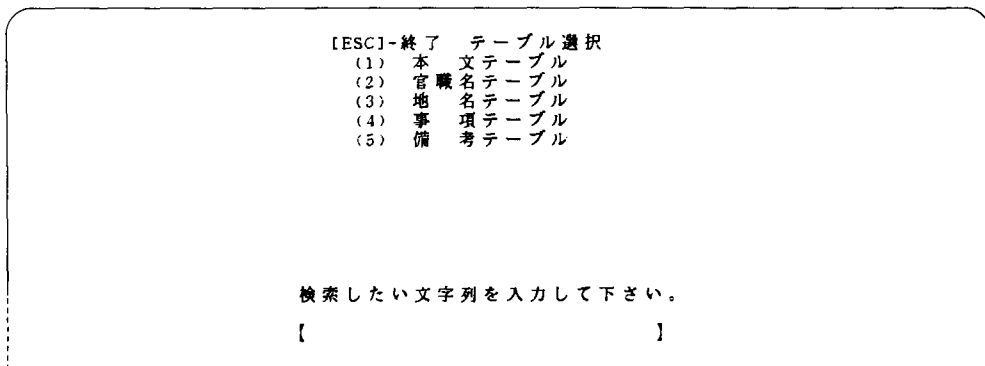
これでひとつの検索は終わった。すると画面は最初の画面（画面1）に戻っている。ここで止めたければ、「(7)終了」を選びリターンキーを押せば良い。また検索を続けたければ、今したのと同じことを繰り返せば良いのである。まず1回終わろう。「(7)終了」を選びリターンキーを押す。するとこのデータベースを呼び出した時の画面に戻る。

(注) この最初と最後の操作は、最初に設定したもので異なる為、今は省略しておく。

なお、このシステムはRBPROを基に若干の改良を加えたもので、PC98シリーズ及び日立2020・2050のどちらでも稼働する。

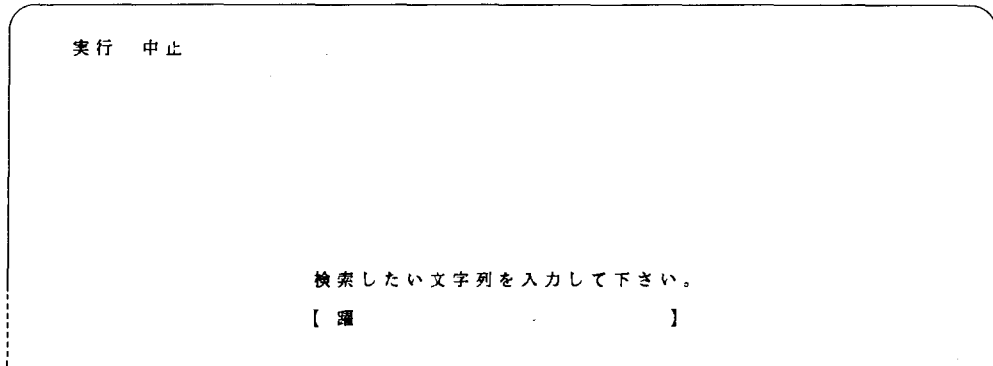
では次にもう一度、今度は「(2)西暦・和暦一覧表（文字列を含む）」という検索を試みよう。先の画面2で、「(2)西暦・和暦一覧表（文字列を含む）」を選びリターンキーを押す。すると画面11が出る。

画面11



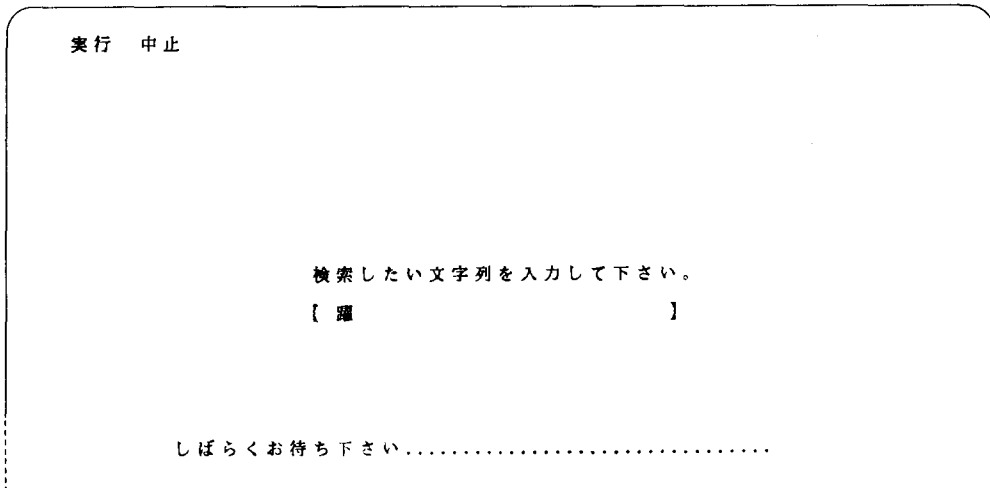
ここでまず「(1)本文テーブル」にカーソルを合わせ、リターンキーを押した後、好きな文字列を入力する。例えば「躍」と入れてみよう（画面12）。

画面12



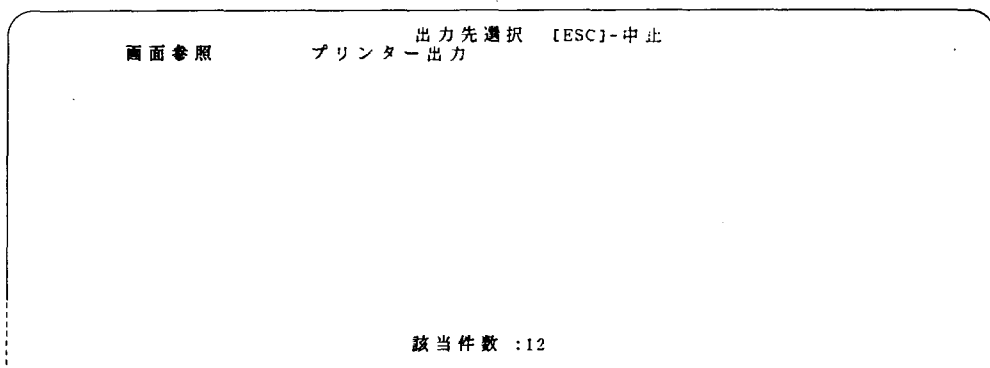
そして実行すると画面13となる。

画面13



しばらくすると、ピッと音がして画面14が出る。「躍」を含む本文は12件あったのである。

画面14



ここでも「画面」と「プリンター出力」の両方が選択できる。そのプリント結果は画面15のようになる。

画面15

[ESC]-終了		西暦	和暦
△今祝	朝之寺	153700070110	天文六年〇〇七月十一日
有於	小兒	154300070150	天文十二年〇〇七月十五日
寺之	所了	154300070160	天文十二年〇〇七月十六日
返、	祐等	154300070190	天文十二年〇〇七月十九日
之番	人等	154300070200	天文十二年〇〇七月廿日
院延	夕相	134400080130	天文十三年〇〇八月十六日
自日	新殿	154600070160	天文十五年〇〇七月十七日
自北	興有	154600070170	天文十五年〇〇七月十八日
代、	去七	154600070180	天文十五年〇〇七月廿二日
	日北	154600070220	天文十五年〇〇七月廿七日
	去七	155100070170	天文廿年〇〇七月廿一日
	日北	155100070210	天文廿年〇〇七月廿一日

これが終ると画面16が表示される。

画面16

2次テーブルに登録しますか [Y/N]:

この場合検索項目が「躍」なので、もし2次テーブルに登録したければ、「Y」と押し「事項テーブル」に登録する、その必要がなければ「N」と押す。今「N」と押す。すると画面は最初の画面1に戻る。

以上が当面二通りの検索である。

ところで、「躍」の検索結果（画面15）を見て載きたい。全12件中、11件は7月、今1件は8月である。すなわち『石山本願寺日記』で「躍り」と言った場合、それは盆踊りのことなのである。意外なことが分かる可能性を秘めているのである。

当面開拓した日記類の全文入力の概要は以上である。勿論というか、残念ながらといおうか、最初に期待した参考資料1：「PC全文入力システム開発」、参考資料2：「石山本願寺日記全文入力 入力要綱」の目論見通りにはっていない。この点は未だ今後しばらく課題として残る。

一方、「躍り」の検索結果に見る様に多少は有益でありそうでもある。

次には、こうしたパソコンの利用から何が分かるか？ どう利用したら有用であるか？ しばらくの間、検索に熱中してみたく思う。その上で今後の問題点をまとめてみたい。

4. コンピューターの効用のほど

さて『石山本願寺日記』の書名は本願寺が大坂石山・紀伊鷲森・和泉貝塚にあった時の日記・書札案などを昭和5年(1930)大阪府立図書館長今井貫一氏の在職25年を記念して、上松寅三氏が編さん・校訂したものを刊行したものである。この上巻に本願寺第10代法主証如上人が、天文5年(1536)正月から天文23年(1554)8月まで記した日記が収められている。それはまた「天文日記」・「本願寺日記」・「証如上人日記」とも呼ばれている。

すなわち、天文5年から天文23年までの日記であるが、もともと記されない日があったり、少し散逸しているために、現在4,041日分が残っている。ただし、今回の入力はおよそ1日単位にしたものの、1日条をA・Bに分割したものもあり、また目次も加えたので、全部で4,171レコードが入力されている。そしてその全文字数はおよそ60万字程度である。

(注) 1日条を複数に分割したのは、全4,041日中、僅かに81件である。全体を考える上で大きな数値ではないので、本稿の考察ではダブリを無視してヒットした数で示していく。

天文期(戦国期)の本願寺と言えば、直ぐに一向一揆が連想される。そして一向一揆と言えば、またすぐ加賀一向一揆が思い出される。ではこの加賀と本願寺とのつながりはどうであったろうか? まず加賀で検索してみる。すると39件がヒットする。しかし、良く見ると平井加賀守・小田加賀入道・藤木加賀法眼など人名がほとんどで、地名の加賀はたったの4件に過ぎない。ところが加州で検索すると362件ヒットし、全て地名である。正確にはこの内12日分は段落分割した際の重複分なので、350日条に加州関連の記事が記されていることになる。すなわち全4,041日中350日は加賀のことが記されている。ほぼ10日に一度は加賀のことが記されていることになる。では加賀に隣接した越前・越中・能登・美濃はどうか?

加賀	39(4)	加州	362(362)
越前	73(42)		
越中	95(92)		
		越州	4(4)
能登	72(42)	能州	21(11カ)
美濃	44(16)	濃州	55(55)

(ヒットした件数とその内、地名のヒット件数を括弧内に示した。)

ということになる。やはり加賀・加州は、この時期周辺諸国を抜いて本願寺と深く関わっていたのである。言ってみれば当たり前のことではあるが、それが数字で裏付けられた訳で、こうしたコンピューターの使い方は有効ではないかと思う次第である。

さて戦国時代の人々の間、特に武士やその周辺の人々の贈答には種々のものがあったが、中でも馬と太刀は多いものである。太刀は450件・馬は445件ヒットする。但し「馬」を含むものは、例えば「対馬」とか「馬場」とかが含まれるため、動物の馬はこの内323件である。その中で御太刀・御馬と敬語を付す例は、御太刀7件・御馬5件である。御太刀7件は、近衛殿の関わるもの2件、上意の関わるもの2件、若君の関わるもの1件、御内書と関わるもの2件の計7件である。すなわち証如は全く稀にしか「御」の敬語を使わないということになる。逆に言えば証如が御の字を使う時、それは幕府将軍か、公家最上層に関わる記事であるとの推定が出来よう。これが本当に当てはまるものかどうかは、今後精緻に検討していかねばならないが、検討素材を提示したものと言うことは出来よう。

(注)「御」で検索すると、922件ヒットする。実に様々なもの・行為に「御」の字は使われている。これらの多くが「御太刀」と同じ理由で使用されたか否かは、別に詳細に検討したい。

馬の記事を見ていると、「馬代」という語が目につく。例えば、

尾張国織田藤左衛門方より、同名差上、書状(并)太刀(金)馬代千疋到来候。

(天文五年三月十一日条)

山崎喜山返事同披露状也。為返、一腰・馬代三百疋来候。

(天文六年十月五日条)

などである。この様に馬代とは馬の代金のことで、多くはここに引いた例の様に三百疋(三貫文)から、千疋(十貫文)までである。中には五千疋という大変高額の場合もあった。馬代は200件ヒットする。すなわち、馬323件中/馬代200件、つまり3分の2は実際の馬ではなく、その代価が贈答されていたのである。一方太刀とは言えば、「太刀代」とあるものは太刀450件中僅かに18件がヒットするに過ぎない。ここからは、太刀は現物が送られたが、馬はその輸送の困難さのためにお金に替えて送られたのではないかとの推測が導かれる。もっともこの場合も他に、太刀1本を指して単に「一腰」という表現もあるので、これら関連の言葉使いを総合的に見て判断しないと、誤りを犯すことになる。

次に葬・葬送・葬儀・葬式を検索してみよう。すると「葬」は8件、「葬送」は4件のみヒットする。そして、「葬儀」・「葬式」に至っては「該当のデータはありません」と出る。これが少ないのは、証如が本願寺法主としてこれらに関わらなかったことを示していよう。本願寺は直接には葬送に関わらなかったといえる。ここにたまたま記された葬・葬送は証如の身内のそれに関わるものであったものと思われる。本願寺が直接かつ一般的に葬送に携わっていれば、もっとこの言葉は多用されたはずである。

では次に「屋」の付くものにどんなものがあるかを探ってみよう。238件がヒットする。その中で商家の屋号ではないかと思われるものを拾うと以下の46件が認められる。

- | | |
|---------------------------|--|
| (1) 越前国いぼ屋 | (2) 檜物屋又四郎 |
| (3) 高師屋(堺之者) | (4) 加州四ツ屋 |
| (5) 坂本四屋左近次郎 | (6) 堺朱屋 |
| (7) 杭瀬屋彦左衛門 | (8) 寺内新屋敷の絹屋後家 |
| (9) (或号磯屋十郎トモ) | (10) 古手屋 |
| (11) 太刀屋 | (12) 新屋敷木津屋跡職之儀
(木津屋与三衛門)
(木津屋善四郎) |
| (13) 寺内田辺屋勘解由 | (14) [庄屋のカ] 十郎左衛門 |
| (15) 葺師屋 | (16) 大屋新三郎 |
| (17) 与津屋 | (18) 鳥屋 |
| (19) 堺円教下塩屋次郎兵衛 | (20) 墨屋浄宗 |
| (21) 南町屋御厨屋次郎左衛門 | (22) 堺舳松米屋 |
| (23) 昆布屋三郎次郎
(昆布屋甚三郎) | (24) 菊屋之宗左衛門 |
| (25) 堺越後屋 | (26) 大坂円屋 |
| (27) 桶屋浄芸 | (28) 竹屋 |
| (29) 堺万屋 | (30) 銭替者針屋(祐西) |
| (31) 腹巻屋 | (32) 寺内山口屋 |
| (33) 荒川屋四郎衛門
(荒川屋源左衛門) | (34) 塗屋道専 |
| (35) 平野屋 | (36) 湊ニ屋[そうげんや] |
| (37) 寺内坂東屋宗二郎 | (38) 賀島屋 |
| (39) 葉屋藤左衛門 | (40) 油屋髭新左衛門
(油屋浄円) |
| (41) 八文字屋 | (42) 堺木屋次郎左衛門 |
| (43) 堺大和屋□兵衛 | (44) 越後蒲原之秤屋 |
| (45) 金屋正乗
(金屋乗珍) | (46) 木村屋藤衛門 |

中には、苗字であって屋号ではないものがあるかも知れないが、一応以上の46例が認められるのである。通常の索引は、各単語・用語の読みを頭から、多くは、あいうえお順・イロハ順・あるいはアルファベット順に並べてあるので、こういう語尾や中間の読みや文字での検索は行い得ない。こうした検索はさすがにコンピューターならでのものと思う。

さてこの屋号を見ると、(1)国名・地名を付したもの、(2)職種を示すもの、(3)苗字を付したものの、(4)その他である。こうした屋号の付け方はもう近世のものとは変わらないのではないか。またその所在を見ると寺内・大坂と堺が多い。この時本願寺が大坂に所在していたからでもあるが、さすがである。しかしこうした近郊のみに止まらない。越前・加州・坂本・越後蒲原など広い地域にわたっているのである。

ついでと云っては何であるが、せつかくその広がりに触れたので、先にも少し見た国別の出現頻度を見てみよう。つまりこの日記が日本全国の中で特にどの地域と深く関わりを持ったかを見てみたい。それは別の言い方をすれば、門徒はどこに多く分布したかという問題とも言えよう。

北国 9 坂東 6 西国 1 東国 1 中国 0 北陸 1 関東 2 九州 13 四国 18
 若狭 28 越前 42 加賀 366 越中 92 越後 4 越州 4 能登 52 佐渡 0 山城 3
 近江 64 美濃 72 飛騨 23 信濃 10 上野 1 下野 0 野州 1 伊賀 1 伊勢 55 尾張 47
 三河 20 遠江 3 駿河 3 甲斐 17 相模 1 安房 0 上総 0 下総 3 常陸 4 出羽 5
 陸奥 0 ? 大和 20 河内 48 摂津 15 和泉 30 播磨 52 備前 4 備中 1 備後 15
 安芸 2 周防 0 長門 0 紀伊 29 淡路 2 讃岐 5 伊予 1 阿波 0 ? 土佐 9
 丹波 12 丹後 0 但馬 9 美作 2 因幡 2 出雲 6 隠岐 1 筑前 0 筑後 2 肥前 0
 肥後 2 豊前 2 豊後 6 日向 3 薩摩 2 大隅 0 沓岐 0 対馬 0
 京都 38 大坂 42

以上である。ここでの例えば若狭は若州など別称を含む数値である。また、人名と覚しきものは除いた。どちらとも取れるもの・判別のむずかしいものは適宜判断した。

改めて、加賀が断然出現頻度の高いことがうなづかれる。次には越中の92、美濃72、近江64、伊勢55、播磨52、能登52件が続く。以下、河内48、尾張47、越前42、和泉30、紀伊29、若狭28、飛騨23、大和20、三河20が続く。いずれも門徒の多い地域と重なっている。それはまた一向一揆が起きた地域でもある。しかし、これだけで判断するのは早計である。と言うのは、山城は3件であるが、京都は38件ある。また摂津は15件であるが、大坂は42件ある様に国より狭い地域の名で記されることも多いからである。それにしてもヒットした数だけは確実に記されている訳で、この時期本願寺の影響が及んだ地域・本願寺と交流を持った地域と言うことが出来る。それは優れて畿内・近国であったのである。中でも越前・加賀・能登・越中の北陸と美濃・尾張・伊勢・三河の東海がその主たる基盤であったと言えよう。

こんな調子で検索を重ねて行けばもはやキリはない。まあ、せつかくだからもう少しだけ続行してみよう。

先程、言葉の末尾に「屋」の付くものを検索して職種の多様さを探ってみた。今度は「銭」で検索して銭がどんな風に使われていたか、どんな種類の銭があったかを考えてみよう。

銭は118件ヒットする。その種類は以下の如くである。

出銭・頭銭(斎または風呂の頭銭として出現することが多い)・一銭勸進・礼銭・借錢・米銭・飯米銭・器銭・路銭・放銭・風呂銭・末寺銭・連銭葦毛・斎銭・灯明銭・年貢銭・地子銭・折紙銭・矢銭・聖供銭・段銭・替銭・官銭?

ここでは多くは本願寺がその配下に課した銭または役を銭で収めさせていたことが知られる。それにしても様々な名目で賦課していたものと改めて感心してしまう。

もうひとつ今度は食事に関して「餅」を検索してみよう。238件がヒットし、これも結構な種類がある。

菱餅(ひし餅・ひしの餅)・樽餅・いたたきの餅(小頂餅) [戴餅]・温餅(温キ餅・あたたけ)・赤餅・小豆餅(小豆並餅)・祝之餅・柿餅・栗粉餅(栗粉之餅)・餅鏡・捻餅・夕後餅・小麦餅・杏形餅・棚餅[駢餅]・葛餅・砂糖餅・惣者餅

その素材から付いた名称・その形から付いた名称・その目的から付いたもの・その色から付いたもの等がある。

ところで色と言え、と言ってすぐに「白」・「赤」で検索してみる。すると即座に「白強飯」・「赤強飯」の言葉を見つけることが出来る。その色はその時・その席に応じて使い分けられたものと思われるのである。どうやらコンピューターは、その使い方によっては、色々な問題点を探ることが出来そうである。その可能性を確かめた今、しばらく検索を休み、今後の課題を次に考えてみることにしたい。

5. 今後の課題

今後の課題としては、大きく分けて次の3つの問題があるような気がする。

- (1) データ構造(フォーマット)とソフトの問題。
- (2) その利用の仕方の問題。
- (3) 著作権など諸権利との関係。

(1)のデータ構造やソフトの問題は、多分この先いくらかでも便利になる様な気がする。問題は(2)と(3)である。

(3)は当面良く分からないので、省略したい気持ちである。ただ交通信号の様に赤は止まれ(ダメ)、緑は進め(良い)と、誰にでも分かる様なルールを示して欲しい。もうひとつは、本当に守られるべき権利は何かを考えて欲しい。と言うのは、我々が扱う史料は多く既に何百年も経過している。それらは基本的には既に国民共有の財産と言うべきではないのか。その編さん・出版と言って、それに要する労力・知恵・資金はやはり大きなものがあると思う。しかし国民共有の財産とどちらが守られるべき権利かというような疑問である。要は分かり易いルー

ルを示して欲しいということである。一方、多少このコンピューターと付き合ってみると、全く恐ろしいとの印象も持つ。と言うのは先にも触れたが、ここまで来るのにほぼ3カ年を要した。しかし、このデータのコピーはその気になれば、今こうしてワープロを打っている、同じ行為でほんの数文字分をキーボードで叩けば、全く僅かな時間で出来てしまう。多分30分かからない。こうした現状を見ると、果たしてこの先本当に知的所有権は守ることが出来るのか、どうすれば良いのか、全く途方に暮れる現状ではある。

さて、問題は(2)の、こうしたデータベースの利用の仕方である。先にいくつか試みに検索してみた。確かに有効でありそうである。しかし、そうした有効性はこの試みで既に分かったことでもある。この先、いくらこうした検索を繰り返しても、それは検索する文字・言葉が違うというだけのことである。確かにひとつひとつの言葉の使われ方をより正確に把握することは出来るようになった。様々な史料から特定の言葉を選び、その性格・意味を探り、ひいてはその時代背景にまで迫っていくなどの、従来の歴史研究のひとつの手法は、もはや、それら史料（原典）がコンピューターに入力されていさえすれば、誰でもが、何時でも検索結果の検討から導き出すことが出来る。少なくとも、我々の様な全くのコンピューターの素人さえがここまで出来る時代なのである。すなわち、これらはコンピューターに入力された時点で、既に共通の常識になってしまう。それもさしたる思考もなく、僅かな指先の手作業で、何時でもその常識を取り出すことが出来るのである。正直、初めて検索しその結果が出た時は感動の一瞬であった。しかし、今は次々と打ち出されるデータ（結果）を前にして、いささかうんざりもし、途方にも暮れている。

より良いフォーマット・ソフト造りと並行して、こうした新たな悩みを抱え込むことになった様な気がする。増える情報・データをどう処理するのかばかりが問題であった時期もあろうが、情報・データを減らすことも一方の課題かも知れない。

手作業では出来ない、あるいは困難なことをコンピューターならやってくれるのではないかと期待は大変大きかった。その可能性は確かに見えて来た。このデータベースを使えば既に表示した様な検討は出来る。ただ、出来ることが分かってしまうと、もはや空しささえ漂う。

多分、この先は、現在既にある漢字OCR（文字自動読み取り装置）が改良されて、活字はもちろん、毛筆の文字もかなりのものは読み取ってしまうことと思う。そうした時代はもう目前である。コンピューターをどう利用するのか？ これは今後の大きな課題である。もはやコンピューターは何が出来るとは出来ない。コンピューターはおそらく我々がこれに期待することには何でも応えてくれよう。要はどう利用するかである。

（注） およそ10年程以前、国立歴史民俗博物館にコンピューターが導入された時、館員の多くは薔薇色の夢を持った。しかしその夢は無惨にも打ち砕かれ、今度は絶望の淵をさ迷うこととなった。当初、我々がコンピューターに期待したことは、例えば、時間を

かければ出来る、ソフトを開発すれば出来る、と様々な仮定の条件が付いていた。全く詐欺に
あった様な感じであった。その印象は別として、客観的に必ずしも当館のコンピューターが上
手く作動しなかった理由としては、以下の様な諸条件があった。

- (1) 館が展示オープンに忙しく、コンピューターどころではなかった。
- (2) 未だ、余り普及しておらず、その操作がむずかしく思えた。
- (3) コンピューターの実体が分からず、何を期待するか? どう利用出来るのか見通しを持
てなかった。
- (4) スタッフが不備であった。

こうした条件の全てが解決した訳ではさらさらない。しかし、この内のいくつかの条件は改
善されている。

また本稿が紹介した様に、パソコンではあるが、多少の利用価値は確認出来た。今後は、や
はりどう利用するかが最も重要な課題であると思う。逆に言えば具体的利用目的もないままに
ただ入力しても余り意味はないということになろう。機能的な問題は、程々既にある機能の活
用で充足されるだろう。もし、ない場合でもその開発は時間の問題である。要はどう利用す
るか? が今後の最大の課題と思う。もはや、最初の約束ごとのひとつにあった、コンピュー
ターに無関心な人に……は通用しない。もう、コンピューターは実用の時代に入ったと言える。

当初、この「今後の課題」では、もっと具体的な入力の仕方・フォーマットの在り方等をこ
まごまと書こうと思っていた。しかし、もうそれは本当に小さいこと・常識に思えて来た。多
分今後ますますコンピューターは好むと好まずに関わりなく我々の生活の中に入り込んで来る。
もう拒否出来ない。利用しうること・必ず入って来ることが分かった今、その取り組みは新た
な段階に入ったと思う。

(国立歴史民俗博物館 歴史研究部)

Attempt at Total Input of Records

SUIRŌ Makoto

This "attempt at a total input of records" aims to utilize the full text of diaries and compilations for retrieval and statistics by inputting them into a computer. This is an attempt to be able to know immediately, for example, in what article of what date the word "festival" appears in a diary, by using a computer.

It is natural that diaries so far written have a variety of characteristics. Some are suitable for input, but others are not. For example, the length of sentences written in diaries differs day by day. Some were rewritten or inserted after the original was written, and some include Chinese characters which are not used today. Therefore, it is not simply a matter of inputting sentences. The first barrier was the arrangement of conditions for input.

The second one was the difficulty of computer operation.

Anyway, the following two goals were set, and some barriers overcome.

The system must be easy to operate so that anybody can use it.

Data must be retrievable by any optional character string, without preparing keywords in advance. At last, the system started operating : But once it started, new problems occurred, for example, the time required for retrieval. At first, the ultimate goal was to be able to retrieve data by any optional character string for any length of time. As a result, the system was found to take 12 or 13 minutes for a retrieval from 400,000 characters. I was excited when I succeeded in retrieving data for the first time, but this length of time was a considerable barrier to the repeated retrieval of data by more than one word. Then, the curtailment of retrieval time became an important task. Now retrieval time has been reduced to only one or two minutes. This paper is the record of this hard struggle.