

〔学界動向〕

「歴史研究と電算機利用の可能性——歴史系（日本史・経済史・社会史） 大学院院生を主たる対象とするワークショップ——」の開催

小 口 雅 史

近年のいわゆるパソコンの普及によって、人文・社会科学系の研究分野においても、その利用による新しい研究の可能性が拡大されつつある。とくに英米文学や国文学、経済学や経営学の分野では、電算機利用が活発になされている。しかし歴史研究の世界では、最近の、中野栄夫氏による一連の事例紹介（「コンピュータ歴史学のすすめ」歴史手帖一九—四）や、パソコンを利用して戦国大名の分析をした、山室恭子氏による『中世のなかに生まれた近世』（吉川弘文館など）はあるものの、なお電算機アレギーとも言うべき状況が顕著であろう。Association for History and Computing が結成され、充実した報告書・学会誌が刊行されている欧米に比して、はるかに立ち後れていると言わなければならない。

そうした中で、去る九月二十一日・二十二日の両日に互って、中央大学駿河台記念館（東京都千代田区神田駿河台）において、表記のワークショップが開催されたことは注目すべきであろう。開催の母体は、情報知識学会人文・社会科学系部会（部会長・安澤秀一（明海大学教授・国立史料館名誉教授）で、実務は永村眞日本女子大学助教授によるところが大きかったようである。このワークショップ立案に関係したグループ

は、既に二年ほど前、「文字記録史料と電算機応用に関する課題と解決」研究集会を開催し、大きな反響を呼び起こしている。歴史研究の集会でありながら、既存の歴史学会の主催としなかったのは、この問題に関心のある歴史研究者であるならば、日本・東洋・西洋の別を問わず、また所属学会をも問わず自由に参加できるようにとの安澤氏の配慮によるものである。

歴史研究における電算機利用についての研究集会としては、これまでに既に六回ほど開催されている、星野聰京都大学大型計算機センター教授・勝村哲也京都大学人文科学研究所付属東洋学文献センター助教授らを中心とする「東洋学支援データベースの研究」シンポジウムがあり、また対象は幅広い分野にまたがっているが、テキスト・データベース研究会が毎年二回ほど開催されている。しかし歴史系大学院における電算機設備の状況は、一般に、院生たちがそうした成果を享受するにはあまりに不十分な状況にあり、カリキュラム上も配慮されていない。そこで本ワークショップでは、柔軟な思考をもち、次代を担うべき院生たちを対象の主体に据え、これまでの成果の公開と普及をはかることを大きな目的としていたようである。

参加者は延七〇人を越えたようであるが、主として首都圏の大学院生あるいは自治体史編纂関係者であった。開催者の意図通り参加者の専攻分野は時代・フィールドとも多岐に亘り、盛会であった。弘前大学からも、大学院人文科学研究科修士一年の安達健次君（日本古代史専攻）が参加する予定であったが、諸般の事情によりそれは果たせなかったものの、筆者の参加は予定通りなされ、歴史研究の世界における新しい動きについて、多くのことを知ることができた。ここではそうした有意義な体験を紹介してみようと思う。本ワークショップに対する関心を高め、歴史研究における電算機の積極的利用の刺激の一端となることができれば幸いである。

本ワークショップのプログラムは以下の通りである。

- ① 基調報告（安澤秀一）
- ② パソコンを利用した資料整理の一事例（内田金生）
- ③ 大名家臣の系図分析におけるパソコンの利用（村越一哲）
- ④ 歴史研究の作業環境の整備―統合型ソフトの利用による資料の整理と
 作表・グラフ化―（五島敏芳）
- ⑤ 漢字OCR装置デモ（TAU技研）
- ⑥ 多摩市史の目録作成とパソコン（桜井昭男）
- ⑦ テキストファイルを利用したデータの応用について（荒巻喜光）
- ⑧ 自治体史編纂とパソコン―『茅ヶ崎市史 現代 資料編 新聞集成1』
 （仮称）の索引編成にあたって―（佐藤正広・柴田貴行）
- ⑨ パソコンによる歴史資料目録作成の一事例（保坂裕興）

このうち、①～④までは、筆者が当日、別な研究会とかけもちであったため、残念ながら拝聴することができなかった。もともと、本ワークショップ開催に深く関係した内田氏自身による②の報告は、本人の病气によって中止となったようである。

当日配布されたレジュメによってその内容を推測すると、②は、合資岡谷製糸会社『電信簿』四冊のデータを、代表的な表計算ソフトであるLotus1-2-3によって資料整理し、明治期における国内向生糸の流通状況を明らかにしようとしたもので、データ入力項目は、「分類コード・取引年月日・取引地・取引相手・取引銘柄・取引価格・取引数量・電報の内容・返信の有無・返信の内容・意味不明の記号」である。それらをデータ処理し、「出荷先別にみた取引数量・月別にみた出荷先の変化・出荷先と主要取引銘柄の対応関係・出荷先別の取引生糸価格の変動状況・どのような場合に返信が行われているのか・意味不明の暗号の推定」などを出力しようとしたもの。ただこうした並べ替え・検索はLotus1-2-3のような表計算ソフトでは限界があり、入力データが増えてくれば、むしろデータベースソフトの方が威力を発揮するであろう。

③は、『蜂須賀家・家臣成立書併系図』に書かれた御目見以上の家から無作為に知行取一〇〇家を抽出し、その歴代当主の子供の記載を表計算ソフトを用いて入力して、徳島藩知行取の家族規模（子供数）推計のためのデータ作成を行うもの。

④は、病人の旅人・行倒人などに対する処理の規定・事例を集めてデータベース化し、通年的にみて法令がいつ出されているか、幕府と藩とはどう発令状況が違うか、その法令はどう変化しているか、事例では

どんなケースが多いか、何を目的にして動いている人が保護される事態におちいることが多いか、どのように処理されることが多いか、移動する人の出身地への連絡はどの程度かなどを表・比較するもの。そこから得られた新しい「材料」を基に、行路病死人の道中方での把握縮小・処理の簡略化、一方での実際の事件の増加といった結果・仮説を導き、実事件に対応しきれなくなっている幕府の状況を問題化しようとしている。

⑤以下は、筆者自身も参加したもの。⑤は、自然語検索（文字列表現を直接に検索）を可能にするための大前提である、史料全文テキスト・データベース構築のための漢字OCR（光学的読取システム）「PC-Star」（TAU技研）のデモ。価格はパーソナル・レヴェルの類似品のそれをはるかに超え、システム全体で四〇〇万円以上らしいが、それだけの読取能力は確認できた。普通の活字の史料集なら、相当の入力時間の省力を可能にしよう。もちろんイメージ・データも管理できる。

⑥は、パソコンによる『多摩市史関係所在文書目録』作成の事例報告。Let'sアイリスを用いて、表題・年代・差出（作成）・受取・リール・コマ・形態・綴方・数量・備考・番号・年代コード・分類を入力している。最後の二項目は目録には出力されないが、カード管理を能率的にするために必要なものである。「分類」については、その具体的内容、特徴、キーワードの要などについて議論があった。カード並び替えのために不可欠な「年代コード」については、データベース構築担当者それぞれの手法があるであろうが、データ交換のためにも、何らかの統一的方法が今後求められるであろう。なお桜井氏は引き続き、黒川能で勇名な山形

県柳引村における、文書所蔵者宅で行なう、パソコン持参による史料調査についても報告した。現場で史料の封筒入れや表題取りを行ない、さらにカードを取り、後にパソコンに入力するという従来の方法に比して、数倍の早さで仕事ができたとのことである。筆者も、関係する自治体史編集委員会において、この方法を提言したことがあったので、興味深く聞かせていただいた。スピードばかりでなく、転写の際の誤写を防げる点からも、この方法がすぐれていると思うのであるが、編集委員会では、文書所蔵者に対する配慮を理由に、実現しなかった。しかし柳引村の例では、パソコン持参に対し所蔵者は快く応じてくれたとのことであり、類似の例も多いようであるから、今後この方法が普及していくのではなかろうか。

⑦は、歴史研究に限らず、個人としてパソコンをどのように活用しているかということについての事例報告。個人蔵書のデータベースをつくり、その卒論文章への引用法、史学文献目録（史学雑誌）「西洋史」のデータベース化などについて紹介した。ソフトはdBASE III PLUS。前者では、「書名・書名ヨミ・副書・副書ヨミ・編著者1ヨミ・編著者2ヨミ・編著者2ヨミ・訳者1ヨミ・訳者2ヨミ・訳者2ヨミ・発行所・叢書名・出版地・出版年1・出版年2・ページ・サイズ・定価・購入価・受入日・分類番号・登録番号・注記」を、後者では、「著者・著者ヨミ・訳者・訳者ヨミ・論文名・サブタイトル・雑誌名・ページ・発行年・分類1・分類2・目録・注記」を入力する。ここでも「分類」の内容、記号コード化の問題があろう。

⑧は、膨大な新聞資料の中から茅ヶ崎市に関係するものを抜き出し、

データベース化する試みの事例報告。ソフトはNinja3Pro'Zero'Zero'Zero'など。年月日・曜日・見出し・地域・分類・新聞名・特記事項・項目(索引用語)などを入力する。ここでは自然語検索が採用されているが、新聞などのように同一の事項が様々な表記でなされる場合には、何らかの統一を図ることや、末尾からの検索による統一の必要があることも問題点として述べられた。

⑨は、学習院大学史料館における内膳司浜島家文書(追加分)受入のための仮目録作成に関する事例報告。ソフトはThe CARD3+。項目仕様は「文書番号(枝番号)・表題・処理番号(資料劣化、利用制限など)・和暦年月日・西暦年月日(数値9桁)・作成者・宛先・形態・件数・点数・寸法縦・寸法横・状態・部門・受入形態・内容・備考」。最後に所蔵機関のデータベース群として、各データベース間を連結可能にする改造・統一、相互利用の拡大の必要性などについても述べられた。

全発表終了後、参加者全員による感想・意見交換があったが、大きな論点の一つは、検索時における、シソーラス(thesaurus)体系の構築か、自然語検索かということであった。本ワークショップ主催者の安澤氏は、自然語検索で十分であるということを繰り返し力説しておられた。これはかつてデータベース用ソフトが発達していなかった時代に、どうしてもシソーラス体系を構築する必要がある、その当時の発想がいまだに尾を引いていることへの警鐘である。秀れたシソーラス体系を構築することには大きな困難がともない、それがデータベース化の大きな障害になっていたことは事実である。しかし近年のソフトの飛躍的進歩は、自然語検索にも十分対応できるほどになっている。安澤氏の主張は当然

のことであろう。ただ、永村氏も指摘しておられたように、多様な歴史的表现を自然語検索だけでカヴァーするには問題があり、例えば人名などについて、官途と実名を事前に処理するなどの作業も不可欠であろう。

また⑨報告で触れられた、データベースの相互利用も重要な課題である。ソフトに様々なものが使用されるのは当然であるが、データそのものはテキストファイル形式でやりとりするといった方式でのデータの共有化が最も簡単である。自治体史編纂でパソコンが頻繁に利用されるようになれば、オンラインによる資料の相互利用が容易になる。自治体史編纂作業の大幅な省力化も夢ではない。ただその場合でも、年代コードの統一化程度はどこかの機関でしておかないと、データのやりとりの際して不都合が生じるので、各方式を比較して早急に検討する必要がある。

筆者の専門とする日本古代史の史料全文データベース化においては、漢字の問題がある。現在、JISは第二水準までしか統一しておらず、第三水準以降は各メーカーでまちまちであるが、これでは不便極まりない。もつとも仮に諸橋の大漢和辞典にある漢字全てを登録したとしても、なお完全でないことも明らかであるから、この問題の解決は容易ではないが、何らかの方策を検討する必要がある。現在、日本古代史の分野では、大型コンピュータを利用した一字索引がぞくぞく刊行されつつあり、先に触れたデータの共有のためにもこうした検討が急務である。

なお筆者としては、今回の諸報告の中で、データベース・ソフトとして、筆者の研究室で論文目録データベース作成などに日常利用されている

る桐V3が登場しなかったのは意外であった。Lotus-2-3やdBASE III PLUSなどのアメリカ製のソフトは、日本語データを扱う上でいろいろ不十分な点が目立つ。また前者は表計算ソフトとしては秀れているがデータベースとして処理できるのはせいぜい一〇〇〇件程度で、長期的には必ず容量不足をきたすであろう。桐V3の標準日本語FEPである松茸に若干問題があることは否定できないが（とくに桐V3上で動くときに松茸が受ける制限の大きさがあげられる）、柔軟性・スピード・自動振仮名など、それを乗り越えるだけのメリットは十分備えている。ワークショップ終了後の自由時間を利用して、筆者が構築中の日本古代史文献目録の実例を紹介できたのは幸いであった。なお⑧で問題とされた末尾からの検索についても、桐V3では簡単にできる。日本語データの検索では、Ninja3Proが、また表引きではThe CARD 3+なども秀れており、当然のこととはいえ、日本語データの処理には日本製品を利用した方が良いでしょう。

ちなみに文科系学部での、学生・院生が自由に触れることのできるコンピュータの整備の遅れが指摘されていたが、当弘前大学人文学部には幸いコンピュータ室があり、まがりなりにも新鋭器五台を中心に複数の機器が準備されている。まだ利用は筆者のゼミをはじめとして一部の学生・院生に限られているようであるが、今後の活発な利用を指導していきたい。

このワークショップは、引き続き開催されていくようであるが、ますますの充実拡大を祈念して擲筆することとする。

（弘前大学人文学部助教授）