

「赤嶺コード」について

赤嶺 守：琉球大学法文学部

最近、コンピュータの開発が進むにつれ、大量の情報をデータベース化する作業が歴史の研究分野でも行なわれるようになり、国立奈良文化財研究所では、木簡のデータベース化、東京大学史料編纂所の島津家文書目録データベース・国立歴史民俗博物館の東大寺文書目録・荘園関係文献目録など、各研究機関での史料のデータベース化が積極的に進められている。歴史研究で、コンピュータを利用する目的は、膨大な量のデータを効率よく処理し、より実証的な歴史研究を推し進めることにあり、いまやそうした研究分野において史料の電子情報化は不可欠なものとなっている。

平成6年度に発足した科学研究費補助金・重点領域研究「沖縄の歴史情報研究」(領域代表 筑波大学 岩崎宏之)は近年急速な発展を遂げている情報科学の成果を援用して、琉球・沖縄の政治・社会・文化並びに環東シナ海世界の地域間交流に関する実証的な歴史研究を行うことを目的とし、歴史情報の収集と集積においては、琉球・沖縄史及び環東シナ海世界の地域間交流史に関する研究文献情報、史料情報を集積して統合的な把握を可能にし、また主要文献史料のマイクロフィルム化・画像データベース化、テキスト・データベース化を行ない、そしてその検索システムの開発に研究の主眼がおかれている(「沖縄の歴史情報研究」ニューズレターNo3、柴山守「東洋学研究における大量マルチメディア情報の提供方式の研究と実用化」)。「琉球・沖縄の対外関係史」研究班(代表琉球大学 金城正篤)では、『歴代宝案』の各文書に文書番号を付し、年代、文書形式、発信者、受信者、写本・影印本の現存状況などの歴史情報を提供するデータベースを作成する作業を進める一方、その他に関連史料を統一した形で情報化する基礎的作業として中国第一歴史档案館所蔵の『清代中琉関係档案続編』『清代中琉関係档案選編』や『中山世譜』の全文テキストおよびデータベースの作成作業を進めた。琉球・沖縄の対外関係史研究を進める上での基本的な一次史料となる『歴代宝案』『清代中琉関係档案続編』『清代中琉関係档案選編』『中山世譜』といった漢文の史料群を統一した形で情報化する際に、特に問題になったのが外字をいかに処理し、全体の中でどう標準化するかということであった。

コンピュータの漢字文字については1978年に「情報交換用漢字符号系」が制定され、1字ごとに「文字コード」が割り振られている。これを通称「JIS 漢字コード」といっているが、現在普及しているワープロ漢字辞典では<JIS>の第1水準が2,965字、第2水準が3,388の合計6,353字で、第1水準や第2水準にある漢字はコンピュータ入力および出力が可能である。それに含まれていない漢字は「外字」となり、作字をして入力すること

は可能であるが、勝手に作字してしまうと文字の情報交換がうまくいかずオンライン化をはかる際の障害になる。なお、JIS 漢字には印刷現場などで用いられることの多い第1水準、第2水準以外の漢字5,801字をコード化した「情報交換用漢字符号系 補助漢字」(JIS X 0212)という新しい規格が90年に追加されており、国立歴史民俗博物館のデータベースの外字処理は、外字の部分を≡で表示し、JISの第1水準、第2水準にない外字を、この「情報交換用漢字符号系 補助漢字」を利用し、その部分の漢字については対照表で確認できるようにしている。しかし、「情報交換用漢字符号系 補助漢字」は印刷業界などからの要望にこたえたもので、現段階では補助漢字を搭載した製品は少なく、補助漢字の本格的な普及は、さらなる半導体メモリーのコストダウンと表示装置や印刷装置の改良、および基本ソフトの世代交代が行われてからであろうといわれている(『ワープロ漢字辞典』旺文社)。

「沖縄の歴史情報研究」プロジェクトでは、この問題を解決するために、(1)外字の1字ごとに字形のイメージを本文テキストに張り付ける。花園大学国際禅学研究所のウルスアップ氏、クリスティアン・ウイッテアン氏によるIRIZ漢字ベースの方式で、MS WORD上では外字イメージを本文に張り付け、コード &C3-5043 のようなコードを付加する。通常このコードを非表示にしておく。(2)外字の1字ごとにその外字を形成する部分品を { や } で囲み、本文の中に埋め込む。例えば 金+平 のように表し、本文テキストに埋め込む(柴山守「<歴代宝案>情報化の問題点」)といった幾つかの手法が試みられているが、「琉球・沖縄の対外関係史」研究班では外字の1字ごとに<1A><Ab>といったアルファベットと数字の組み合わせによって2バイトで表現する「識別子(赤嶺コード)」を付加し、本文テキストには識別子を用いて入力する方法をとった。全文テキスト・データベースの作成にあたって、「識別子(赤嶺コード)」を設定して外字処理をする方法を採用したのアルファベットと数字の組み合わせによって2バイトで表現する「識別子」を付加することによって、『清代中琉関係档案続編』『清代中琉関係档案選編』『中山世譜』といった抬頭のある文書特有の書式(例えば『清代中琉関係档案続編』の場合、各紙幅6行、各行20字、抬頭2字、平18字)を変えずに、文字化けせずにオンライン化による情報交換を可能にすることを優先したからである。また何度となくおこなわれる入力の校正の際にも、原文書と同じ書式の機械可読形式版文書であるので、校正をかけやすいという利点も考慮にいられた。記号置換した文字の確認は「外字・記号対照一覧表」でチェックする方式を取り、また「外字・記号対照一覧表」のみでは外字の記号入力する際のコードの確認が煩雑で時間を要するため、他に「外字・記号対照部首索引一覧表」を作成し、コードが外字の部首でもひけるようにしておいた。

档案文書には一つの字種に対して、異体字が多く用いられている。異体字が多いとコンピュータは検索洩れを起こしてしまい、必要な情報が得られなくなってしまうことがある。異体字を一つの字体に統一することは、検索洩れを防止する意味でも重要である。異体字の中にも <JIS>の第1水準、第2水準に含まれない外字が少なからずある。そうした同

一字種とされる外字の異体字をどう統一するかは、色々と見解のちがいもあるが、本研究班では沖縄県の歴代宝案室作成の「異体字統一表」を参考にしながら、以下の原則で外字の異体字を統一処理した。

1. <JIS コード> にない異体字（俗字・別体字等）は原則的に旧字の正字に統一し、記号コードを付加する。

2. 外字の正字が<JIS コード> 含まれず、常用漢字が<JIS コード> にある場合、常用漢字で文字統一をする。これは外字をできる限り少なくするための措置である。

1998年1月31日現在、「外字・記号対照一覧表」および「外字・記号対照部首索引一覧表」には約1,400字の外字が収録されている。Windows版の最大外字登録可能数は1,800字なので、外字フォント（京都大学人文研のe漢字<電子漢字>ホームページのe漢字資料集参照、<http://www.zinbun.kyoto-u.ac.jp/~ekanji/data.html>）を作成し、それをWindows95の外字エディタを使ってロードすれば、機械可読形式版の外字はWindows95版のアプリケーション上に表示される。こうした機械可読形式版文書は市販のテキストデータの文字列検索や加工を行う「GREP」「AWK」といったソフトでの検索や並べ替えにも十分に対応できる。

現在、「外字・記号対照一覧表」および「外字・記号対照部首索引一覧表」には『歴代宝案』『清代中琉関係档案続編』『清代中琉関係档案選編』『中山世譜』簞崇業『使琉球録』陳Sf『使琉球録』夏子楊『使琉球録』徐Sg光『中山伝信録』『古文書等緊急調査報告書』『日中・日朝関係研究文献目録』『沖縄県郷土資料総合目録1』『沖縄県郷土資料総合目録2』『沖縄県郷土資料総合目録3』『沖縄県立沖縄図書館郷土史料目録』『琉大郷土資料目録』『沖縄歴史関係主要論文目録』『南島文献資料目録』『九学会連合沖縄研究文献目録』等の外字が収録されている。