また，文献情報の生成という面から見ると，これまで，図書館と国語年鑑編集の2つのセクションが相互に関係を持ちながらも，比較的独立した形で仕事を進めてきている．ここでは，これを一貫した形でシステム的に統合したものとするべく作業を進めている．業務の中でのもの(本，雑誌等)や情報の流れの管理についても問題となる点が出てきているが，その管理もDBMS上で行なうよう検討している．

作業の現段階では，2つのセクションがデータを共有することから生ずる問題について，問題の洗い出しを進めている．内容に関する問題としては，共有することになるデータについて，各々が別々のやりかたでデータを取ったり，表現したりしているところにある．書誌的事項として，書名や雑誌名の認定の相違，目録記述をする際の転写の原則の相違などがある．また，すでに内部で機械可読形式として蓄えてあるデータとの間でも，同様の問題が生ずる．

蓄積されるデータの公開については，その内容や方法は今後の作業における検討課題となっているが，なんらかの形で一般に利用可能なものにすることを考えている．

日本古代・中世の漢字表記文献の機械可読化とその利用における諸問題
——「白氏文集」「和漢朗詠集」「古事記」等を例として——

當山 日出夫

日本の古代・中世の文献・史料等については，そのかなりの部分が漢字を使って，いわゆる漢文で表記されたものである．これらの文献は，文学・歴史・言語等の研究の諸分野に共通して利用されるものが多い．また同時に，これらの文献・史料は，東アジア漢字文化圏における漢字表記文献という立場からも考察される必要がある．現在，各所で，これらの文献の機械可読化（コンピュータによるデータベース化）が進行しているが，将来における学際的なデータの共同利用という展望が必要な時期にさしかかっていると思われる．

1．パーソナル・コンピュータをもちいて「和漢朗詠集」「千載佳句」「新撰朗詠集」（日本・平安時代に作られた，漢詩の秀句集）の漢字一字索引をすでに刊行した（いずれも勉誠社刊）．個人で利用するパーソナル・コンピュータで処理し，パソコン用レーザープリンタで印字したものを版下に利用して刊行した．これは，漢文の文献の漢字索引を作成することは，現実的に十分可能であることを立証したものである．

2．「白氏文集」（中国・唐の白楽天の作品集）の漢字一字索引も計画を進行中であるが，これもパーソナル・コンピュータによって処理することは可能である．日本に残る古い写本の系統のテキストと，現在流布している版本の系統のテキストと，複数の本文による検索を可能とする予定．

3．漢文表記の文献の機械可読化は，データ入力よりも，校正や校異の付加が重要な問題である．単純に書物の本文を入力すればよいというものではない．漢字の字体の統一的処理や本文の区切りなど，個別の文献ごとの事情をふまえて，厳密に対処しなければならない．特に字体の処理と校異はきわめて高度な判断が要求される．

4．機械可読化されたテキストは，その文献の専門家よりも，むしろ，周辺の文献を対象としている研究者にとって便利な性格をもっている．

5．特に漢文で表記される日本古代の諸文献「古事記」「日本書紀」「続日本紀」「万葉集」「風土記」などは，文学・歴史・言語等の研究の諸分野で共通に利用するものであり，これらの文献の機械可読テキストは，学際的に供給され，利用されることが望ましい．現在，そのほとん

どが入力されているにもかかわらず，共通の基盤の上でデータベースとして提供され得る状況にはない．

6．日本における文学や語学の研究の実態・研究方法から言って，学術的データのオンライン・データベース（ホスト・コンピュータでの集中管理）はうまく機能しない可能性がある．

7．CD-ROMやフロッピー・ディスク等により供給されるデータは，できるだけ単純な形式のテキスト・データであるべきである．そして，それを個々の研究者の研究目的に即して再加工し，自由に自分のものとして使いこなすことができる状態が望ましい．

歴史学研究に対するパーソナル・コンピュータの応用

京都大学 大型計算機センター　星野　聰

筆者は日本古代史を中心とする研究の支援に適当なシステムを設計している．このシステムは，文字情報と画像情報に関するものに分けられる．本研究の目標は，安価なシステムによって，これらの分野の研究を十分支援できるツールを作り，実際に専門的な資料を準備し，この分野の研究に役立つことを実証することである．

過去数年間に於けるパーソナル・コンピュータ関連の機器の性能の向上が著しい．そこで，研究に適合したツールを開発できれば，従来の歴史研究方法に大きい影響を与えることは確実である．

そこで，古典テキストの編集に適したエディタを作成した．このエディタでは大きいテキストを編集できるようにした．また，文字種が多いので，最大63文字の外字パタンファイルを必要に応じて複数個作成し，これらの外字をすべて表示に用いられるようにした．但し，外字に対するコードはコードの未定義領域（シフトJISコードで，第1バイトがef以上）を割り当てる．行末の連続する空白は物理的にファイルには格納していない．また，ファンクション・キーの指示により，送り仮名をサプレスできる．外字の文字パタンはレーザプリンタにダウンロードしている．なお，続日本紀・日本後紀・続日本後紀・文徳天皇実録・三代実録（巻1～10）の範囲では，401個の外字を作成した．

文字列検索には，テキスト内の同一文字間をリンクするファイル（以下ではLFと呼ぶ）を用いている．続日本紀の全テキストに対してLF作成に約4分45秒を要する．続日本紀から「岡田臣」を検索すると，LFをRAMディスクに格納した場合に約1.5秒，格納しない場合には約4秒を要した．該当件数は2件．いずれも最終の第40巻にある．また，一種のブロウズ機能を有している．

テキストには，ヨミ，句読点，返り点，割り注など種々のものが付加される．このようなデータを編集するため，一行に属性，テキストのうちの一字，この文字に始まる文字列に関する注記を，別々のカラムの範囲に記述するようにした．

このエディタで作成した，注釈を含むテキストデータ・ファイルを入力として，頭注を有する冊子体形式のテキストに変換するようにした．これをそのまま研究成果とすることができよう．

次に，日本の歴史と関連が深い画像データに，明治・大正・昭和にわたる地形図の利用がある．そこで，地形図を光磁気ディスクに格納し，必要に応じてスクロールして近接地域を表示できるようにした．地形図の画像データ量を縮小するために1ピクセルを1ドット（白黒）で格納している．