化III類を行うために必要なデータを作成するプログラムを開発した．プログラムはC言語で書かれており，PC9800上で使用可能である．主な機能として

○誤字など回答文を修正する機能
○例えば「お金」,「金」というような同じ意味，同じカテゴリーに属する単語や文章を統合する機能
○回答をアイウエオ順に並べ変え，回答の頻度表を作成する機能
○数量化III類を行うために必要な0-1型データを作成する機能

などを挙げることが出来る．このプログラムをパソコン用の数量化III類のプログラムにつなぐことによって，パソコン上で対話形式で自由回答データの数量化III類の分析が可能となる．

国語学研究文献情報データベース化の方策

国立国語研究所　熊谷　康雄・江川　　清

国立国語研究所では1989年度より，新たに情報資料研究部を発足させ，国語学研究を推進させるための各種データベースの構築とその活用の研究に着手しており，文献情報データベースはその中に位置付けられるものである．

国立国語研究所の図書館では日本語研究関係の文献を中心に図書の収集を行なってきている．また，1953年以来，毎年，1年間に発行された刊行図書，雑誌掲載論文などの文献情報を集めた「国語年鑑」の編集・刊行を行なってきており，国語学関係の研究情報の情報源となっている．国語年鑑は国立国語研究所の図書館で収集した図書雑誌，および，各種の目録から集めた情報を整理・編集し，分野別に刊行図書および雑誌論文を収めている．一方，1989年には，国語学会と国立国語研究所の共同で,「国語年鑑」所載の雑誌論文をデータベース化したフロッピー版の「日本語研究文献目録 雑誌編」（約8万4千件）が作成されている．また，刊行図書についても同じ年に「国語学研究文献総合目録」（約2万3千件）が発行され，データは機械可読になっている．いずれも，「国語年鑑」をベースにして作成されたものである．

すでに機械可読化されている過去のデータ（前記の各目録のファイルや電算写植化されてからの国語年鑑の電算写植のデータ等）については，これをデータベース化して，システムに取込むための作業を進めている．

これらの状況も踏まえつつ，現在，手作業が主体である図書館の業務と文献情報の収集と国語年鑑編集・出版の業務の機械化を進め，各業務の計算機による補助，データの共有，機械可読データの蓄積による情報の多面的な活用を図るため，システム作りを開始した．これによって，国語学研究文献情報の収集，そのデータベース化と利用までの一連の流れをシステム化する．研究文献情報としては，流通ルートには載らない所謂gray literatureも含め，研究情報として広く収集する方策を取る．

この作業の前提となることには以下のようなことがある．図書館，国語年鑑の編集のいずれの担当者もその担当分野の専門家ではあるが計算機については非専門家である．現在，それぞれの部門が行なっている業務は中断せずに，これと平行して順次機械化を進めていく．図書館と国語年鑑の編集のセクションは相互に関連しながらも独立して仕事をする．当面，それぞれの担当者が扱うハードウェアはスタンドアロンのパソコンである．これらの前提となる制約から解決すべきいくつかの問題が出てきている．

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