

日本語・日本文学研究におけるパソコンの汎用性

金 田 文 雄

Broad Applicability of Personal Computers in Japanese Language and Literature Research

Fumio KANEDA

Abstract

In recent years personal computers have become a virtual necessity for research in the sciences. Of course in the humanities as well they are already being put to work in many areas, but their broad applicability still offers many possibilities. This article, whose scope is limited to Japanese language and literature, examines how computers have been used in their research, and also explores the possibilities for their future use.

In research concerning the linguistic study of Japanese .. which is actually a very broad field of study .. we find not only textual data bases, but also data bases of audio, visual (including maps), and other kinds of information, which indeed makes full use of computers' multimedia capabilities. In view of this situation, one might even say that future research which does not use computers will not have currency.

Meanwhile, in the area of Japanese literature research there are grand conceptions such as one for a "Japanese classical literature data base," but at this time there are only data bases of the texts themselves, which are used only by individuals or, at most, small groups. Nevertheless, the groundwork is quickly being laid by commercial developments such as the publication of the CD-ROM edition of the *Kojien* dictionary. There are still many challenges such as networking by electronic mail, but in the future personal computers are expected to make significant contributions to the consolidation and systematization of research.

パソコンをめぐる環境

日本におけるパソコンの普及は、アメリカなどに比べると大きく立ち遅れていると言われているが、それでも研究の領域では、理系は言うにおよばず、文系においても随分活用されつつあるのが現状である。コンピューターが「電子計算機」などと訳されていた時代、(それとてもしさほど遠い過去のことでもないが) そうしたものは、いわゆる文系-社会科学はともかく、

人文科学，なにかんずく日本語・日本文学の研究者にとっては無縁のものと思われていたことであろう。ところが，コンピューターのハード面での技術展開は，近年めざましいものであり，同時に各社による開発競争と普及にともなって価格も下がり，これまでよりもいっそうコストパフォーマンスの高いものになってきた。また，これと並行して，有用で使いやすいソフトも次々に開発されている。

かつて，パソコンを扱おうとする場合，プログラミング言語の習得は必須の課題であった。そして，このことが多くの人をしてパソコンから遠ざけていた最大の要因であった。ところが，今やそうしたことを全く知らなくても，——やや極端に言えば導入したその日からさえ——活用することができるようになったのである。もっとも，今もプログラミングの必要性を強く説く人もいるのは事実である。もちろんできるにこしたことはないのだろうが，それは後からもし必要が生じたら，その時でよいのではないだろうか。

昨今では，日本語・日本文学の研究者の間でもワープロは随分普及しているようだ。ところが，パソコンの導入までにはまだ一步の階梯があり，この段差が意外に大きいようだ。しかし，ワープロもその用途が限定されているとはいえ，パソコンの一種であると考えることができるし，まさしくその通りなのである。ただし，せっかくパソコンを手にしながらかワープロとしてしか使わないとすれば，これは大いにもったいないと言わねばならないだろう。なにしろ今時のパソコンはマルチメディア性をこそ売り物にしているのであるから。もっとも，ワープロ機能としてのみ比較しても，パソコンのもつ記憶容量の大きさはけた違いであり，十分にその価値を認識することはできる。

さて，マルチメディアとしての特質であるが，ここでこそパソコンはその本来の機能を発揮すると言ってよい。すなわち，文字処理（ここではワープロ，及び表計算等のいわゆる演算機能を含むものとする）以外に音声や画像（動画も可能である）をも記憶させたり，また処理することができるからである。後ほど詳しく述べるが，こうした音声や画像は日本語・日本文学の研究，あるいは授業にも大いに寄与するものである。そして，この面でも CD-ROM を中心に，すでにいくつものソフトが市販されており，これらを活用することもできるのである。ちなみに，CD-ROM の容量は標準が540メガ・バイトであり，文字なら2億7千万文字，すなわち数十巻の百科事典でもそっくり1枚に入るだけの容量がある。この CD-ROM の長所は，これほどの大容量を持つわりには安く，しかも持ち運びが簡単だという点にある。ただし，CD-ROM の弱点は，「リード・オンリー・メモリー」というその名の通り，読み込み専用で書き込みができないことである。したがって，この意味では昔ながらの「単方向メディア」であるといわれてもしかたがない。つまり「インタラクティブ（双方向）性」がないのである。¹⁾

また、パソコン自体のもうひとつの売り物である、このインタラクティブ性についてであるが、残念ながら現状では先の CD-ROM がそうであるように、まだまだ改良と、そして同時に使用者の側にも工夫の余地を残しているようである。

周辺機器と市販ソフト

まず、パソコンの本体であるが、NEC の PC98 シリーズをはじめ、Apple の Macintosh、富士通の FM、IBM の DOS-V 機など各社とも実に多様な製品をラインアップさせている。しかも、競争が激しい業界だけに新製品が次々と発売され、どれを選んでいいのか目移りしそうなほどである。そのうえ、定価や価格もあってないかのごとき感さえ呈している始末だ。現にこうした情勢を反映して、Macintosh の一部の製品などのようにオープン価格として、初めから定価のないものまでが登場している。しかし、それぞれの製品の機能や演算能力といったものは、概ね実勢価格に比例するもののようだ。つまり、同じくらいの値段のものであれば、それほど性能的には大きな差はないといっているのではないだろうか。ただし、パソコンの場合はハード本体とともに、その機種によって使用できるソフト面での違いが大きい。そして、残念ながら共用できるものはゲーム・ソフトのハイブリッド版など、ごく一部に限られているのが現状である。もっとも、CD-ROM は、今後こうしたハイブリッドが主流となりそうな情勢であり、またユーザーのほとんどもそうなることを願っているが。しかし当面はまだまだソフトによる制約が大きく、このことがハードの特性以上にパソコンの個性となっているのではないだろうか。よく指摘されていることだが、一般に日本語の文字機能なら NEC の PC98 シリーズが、また画像処理をふくめたマルチメディア機能ならば Macintosh が、それぞれ現在のところ一日の長があるとされているようだ。

さて、そのソフトだが、最初に基本的なものだけをあげると、まずワープロソフトとしては現在「一太郎」がもっとも良く普及しており、それだけの評価の高さと安定度とを示している。もちろんこれ以外にも多数のワープロソフト、エディタソフトがあり、文書作成を便利なものに行っていることは言うまでもない。次いで、大量の数値データを扱う場合には（データの絶対量が多ければ多いほどパソコンはその本来の能力を発揮するはずだ）表計算ソフトが有効であるし、また文字・音声・映像などのデータ処理にはデータベース・ソフトが欠かせないであろう。なお、このデータベースについては、後ほどあらためて詳しく述べるが、ここではデータベース・ソフトの一例として、すでに広く活用されている＜桐＞をとりあげたい。

たとえば、逆引き辞典や索引などの作製には、従来であれば手作業によるカード化の方法が採られてきた。一語一語をカードに書き込んで、しかる後に語末から再配列し直すといったかなり面倒な作業が要求されることはいうまでもない。ところが、これをデータとして＜桐＞に

取り込み（もちろんデータ入力という作業は残る）、これを文字列反転させれば、ほとんど瞬時にして逆順の文字列が得られるのである。そして、これを五十音順に並べ変えさえすればたちどころに逆引き辞典は完成してしまうのである。もちろん逆引きに限らず、任意の検索条件を指定して特定の語群を選び出すこともでき、使い方次第で実に多様な研究上のニーズに耐えるであろう。²⁾

今のことにも関連するのだが、CD-ROM ソフトの一例として、『広辞苑』（CD-ROM 版・岩波書店）を取りあげたい。筆者などは、パソコンを導入する以前は、これがどのようなメモリを持っているのか全く想像もつかず、キーボードに打ち込むよりも辞書を引くほうが早いのにとさえ思っていた。「見出し語の語句の意味を知る」というごく初歩的な辞書使用法しか思い浮かばなかったからにはほかならない。ところがこの『広辞苑』CD-ROM 版は従来の書籍としてのそれとは、いわば全く別物とさえ言ってよいものである。たとえば、そこでは見出し語検索以外につぎのようなことができる。

- (1) 表記形検索—これは、漢字またはアルファベットで入力すると、それを表記形とする項目を検索するものであり、同じ漢字表記で読み方が異なる語などが一覧できる。日本語学習者などには便利であるかもしれない。
- (2) 前方・後方一致検索—語句の語頭あるいは語尾の部分だけを入力し、それを構成要素とする語を一覧できるもの。逆引きの応用と考えてよさそうだが、一語の途中に含まれる要素によっても検索できる点がミソか。
- (3) 条件検索—5つまでの AND 条件と、それぞれの AND 条件について3つまでの OR 条件を指定し検索できるもの。これなどは、いかにもデータベースらしい特性を生かしたものだといえそう。
- (4) 色見本一覧—234の色名がカラーでディスプレイされる。ただし、これはカラー版の辞典なら書籍でも可能だったのであり、それほどすごみが感じられないかもしれない。
- (5) 鳥の鳴き声一覧—60種の鳥の鳴き声を収録している。御愛敬程度という気もしないではないが、これなどは一応、パソコンの持つマルチメディア性を生かしたものだとは言えるだろう。もっとも今どきの学生たちには案外役にたつものかもしれない。少なくともパソコンに親しむ初期段階にはおもしろがるというぐらいの効果は期待できるだろう。

これら以外には、自分専用の編集版を手軽に作成できるなどの特徴があるようだ。ただ、何といっても『広辞苑』の電子辞書版なので、語彙数にも限界はあるが、やがてこのシステムにさらに改良を加えつつ、『日本国語大辞典』なみの大部のものも生まれてくる可能性はあるだ

ろう。いずれにしても、先駆的な試みとして評価したい。³⁾

さらにまた、こうしたものの以外にも多様なソフトが市販されている。DTP ソフト（文書等への画像の取り込み）、画像処理ソフト、音声処理ソフトなど、それぞれ目的に応じて用いることができる。もっとも、必ずしもこれらを即時に導入し活用するようなことはきわめて稀であろうし、通常はおおよそ今列記したような順で展開し、応用範囲を拡げていくことになるだろう。なお、先ほど少し触れたデータベース・ソフトとも関連するのだが、データベース作成をサポートするものとして、OCR（文字読み取り）ソフトがある。これはスキャナーで読み取った活字を自分のパソコンのハード・ディスクに取り込んで、データベース化等の処理を可能にするという実に画期的なものである。もっとも、現状の能力では100%の取り込みは不可能であり、往々にして誤った文字として認識してしまうために、取り込んだ後でキー・ボードによる修正・校正が必要である。⁴⁾

一方、周辺機器にもいろいろなものがあるが、さしあたってはスキャナーが有効そうである。これは要するに、プリントされた画像や図などをパソコンのハード・ディスクに取り込むものであり、さらにはその後のある程度自由な処理をも可能にするというものである。ちなみに、明浄短期大学の伊藤鉄也氏は、こうしたスキャナーと画像処理ソフトとを用いて、「探幽筆三拾六歌仙」の小野小町像を、ものの見事に、しかも色も鮮やかに再現して見せた。⁵⁾

日本語・日本文学研究におけるパソコン活用の現状

日本語学と日本文学の研究は、その性質の上から重なるところの多いことは当然であるが、ここではあくまでもパソコンの活用を問題とするので、一応それぞれ別個のものとして考察を進めていきたい。

(1) 日本語学

まず日本語学であるが、日本文学に比べると、こちらのほうがおそらくはパソコンへの取り組みも早く、またすでにその汎用性への模索を含めて成果も着々と上がっているといえるだろう。これはひとつには、日本語学研究の領域の中で、例えば方言研究や語彙の研究などにおいて、これまでも相当量のデータを蓄積し、これを統計的な意味でも処理するという方法がとられてきたからであろう。そして、このようなことを行うにはまさしくパソコンがその本領を発揮するからにほかならない。また、方言あるいは音声学の研究では、音声を記録し解析できることのメリットも実に大きいであろう。もっとも、単に音声を記録するというだけならば、従来から用いられてきたテープレコーダーで十分なのであるが、パソコンは、これを時間軸に沿って記録するにとどまらず、データとして自由な解析処理を可能にしたのである。さらには

方言地図の作成や、果てはインフォーマントの画像といったヴィジュアルなものまでも取り込むことができる。そこでは、まさしくパソコンの持つマルチメディア性をいかんなく発揮できるということにはかならないのである。

ちなみに「日本語学」(明治書院)は、1991年8月号で「新しいデータ・新しい研究」という特集を組み、ほぼ全ページをデータベース関連の論考にあてている。幸と言うべきか、あるいは不幸にしか、筆者は日本語学を専攻する者ではないが、これなどを読んでみると、いまや日本語学を研究するにはパソコンを扱えないようではどうにもならないのではないかと、といった気がしてくるほどである。また、同誌ではその翌月から「わたしのパソコン言語学」の連載をスタートさせ、さまざまな方法的試みを紹介することで、これを継続化させていた。そもそもこうした日本語学(国語学)と、コンピューターの結び付きは存外に古く、早くも1966年には国立国語研究所に、文科系の研究所としては初めて、コンピューターが導入されていたのである。⁶⁾ また、データベースを研究に用いていくことについては、1973年に中野洋氏が「言語研究用データにおけるデータベースの考え」等を発表している。⁷⁾ あるいはデータベースそのものも、1988年には『日本語研究文献目録・雑誌編』⁸⁾ が出されるなど、着々と成果が蓄積されていったのである。

さて、ここからは日本語学のいくつかの領域での成果を見ていきたい。まず、マルチメディア性をもっとも生かせそうな方言学からとりあげることにする。「日本語学」1991年10月号に、宮島達夫氏が『茨城方言・民俗語辞典』について書いているが、これは期せずして従来の方言資料集と、方言資料の新しい方向性を示すものとして興味深いものとなっている。氏によれば、もともとこれは活字で組むことを予想していたものが、出版社との交渉の過程でパソコンを利用することになったものであるらしい。経費を安くあげることが目的であったらしいが、結果としては茨城方言のデータベースが得られることになったのである。宮島氏は、試みにこれを用いていくつかの統計や検索を試みている。⁹⁾ そして、彼は将来の方言資料のあり方を展望して、次のように語る。

長期的には、パソコンは方言集出版の可能性を保証するよりも、その必要性をなくする方向にすすむのではないか。つまり、データベースとして各県の中央図書館なり大学なりにおいてあれば、それに自由にアクセスして情報をえ、さらに拡充することができる。それは、さらに全国的なネットワークにつながる。

たしかに、こうした方言の分野では研究そのものが、通時的にも、また共時的にもデータベースの蓄積によって大いに発展するはずである。

ところで、方言研究においては歴史的に音声を記号化して記述する方法が確立されてきた。もっとも、テープレコーダーが登場してからは音声そのものを記録できるようになったわけだ

が、これをデータベース化することによって一層便利に、しかも様々な分析が可能となった。たとえば、ある地方の方言において、70代の女性だけであるとか、あるいは年齢を問わずにすべての世代の女性だけとかのデータがたちどころに得られるのである。個々人の収集するデータだけを見ればたいしたことがないように思えるかもしれないが、インフォーマントの数や、それにともなうデータの総量が膨大なものになった時、あるいはそうなればなるほどに、パソコンはその威力を発揮するはずである。

方言研究のデータベースによる音声的側面の研究では、1989年度から1992年度にかけて行われた文部省重点領域研究「日本語音声の韻律的特徴の実態とその教育に関する総合的研究」があり、これらの全部はデータベース化されているわけではないものの、これもまた今後のひとつの方向性をさし示すものであろう。

また、方言学の一領域として言語地図の作成があげられる。そして、この方面でもパソコンは有為な働きをしてきた。しかし、その一方で「日本語学」1992年5月号の「コンピュータが描いた『糸魚川言語地図』をめぐる」によると、それは「内容が特殊であり、本の形も大きく、ばか重く、その上、値も張るので、少数の人の目にしか触れていないと思う」¹⁰⁾とされており、既述の宮島氏の説をみごとに裏づける形となっている。すなわち、もはや言語地図の作成にはパソコン(コンピュータ)はほとんど必須のものであり、同時に享受者の側はデータベースとしてこれを利用するといった形態がそこに浮かび上がってくる。この報告の最後のほうで柴田武氏はこう語っている。

最後までコンピュータでは無理らしく思われることが一つある。話者に面接して、その答を記録することである。方言調査における面接は、十分に時間をかけて該当することばを思い出してもらい、語形についての内省情報も求め、さらに、もっと別の言い方はないかと尋ねる「会話」である。この「会話」の記録こそ最後の人間の世界ではないかと思う。

ここでの語りには、ほとんど哀惜の念さえもが漂っているように感じられる。すなわち、コンピュータの登場以降、人間にできる最後の砦はもはやここぐらいしかないのではないかというような。しかし、如何にコンピュータのハードやソフトが進んだとはいえ、そしてその処理能力が人間の力をはるかに凌駕するものであったとしても、所詮コンピュータは与えられた命令を実行するだけであり、そのコンピュータに何をさせるかを考え出すのはまさしく人間にほかならないのである。そして、これほどにコンピュータの開発が進展した今、それをどのように使うかを考えることが各分野で求められているのではないだろうか。この時、日本語・日本文学がこれまでに行ってきた学問的蓄積、方法意識はかならず生かされるはずである。もっとも、同時にその科学性があらためて問いなおされるであろうが。

また、同じ「日本語学」の同年8月号に高橋頭志氏による「『四国言語地図』を作る」があ

るが、ここでは言語地図作成の手順とともに、それを高知女子大学の国文科での授業に生かしていく試みが紹介されている。大学での授業におけるパソコンの導入については後述したいが、これはそうした意味でも参考になる論考である。¹¹⁾

あるいは、地図作成という点では同じだが、学問的には幾分か位相を異にしたものとして、独自のプログラムを用いた「方言認知地図」の仕組みとその成果、及び将来の方向性が「パソコン国語国文学」に紹介されている。¹²⁾

そして、こうした研究成果が次々と生み出されていく一方で、これをさらに補強・推進するような形で、「言語地図作成支援プログラム」といったものも発表されている。¹³⁾ このプログラムの作成者の肩書きは国立国語研究所員（1993年当時）となっているが、従来のいわゆる国語・国文学の研究者という枠組みからは随分遠いのではないだろうか。まさしくパソコンは学際的であり、また必然的にそのことを要求しているのである。

次には方言研究とも関連の深い音声にかかわるものの例としてとして、上野善道氏の「アクセントデータベースの作成」¹⁴⁾ をとりあげたい。氏は「データベースの作成は訂正作業の連続だ」と語っているが、そうした訂正作業を繰り返すことで、逆に日本語の姿が浮かび上がってくるといった側面もあるのではないかと思う。また、このデータベースの汎用性についても「アクセント研究の様々な分野で利用できることは言うまでもない。それ以外にも、音韻、語彙、意味、あるいは音韻と語彙にまたがる分野で、これを利用してやってみたいこと、やってほしいと思っていることは多い」と述べ、その中から「モーラ音素の数量的分布」を例示している。

このような音声データベースはまた、日本語教育にも有効である。堀口純子氏は、上級日本語学習者を対象としたカード型データベースの作成例を紹介しているし、¹⁵⁾ また才田いずみ氏も「外国人日本語学習者のための音声学習支援システム」¹⁶⁾ にとりくみ、論稿の最後につぎのような問題点と提言を掲げている。

今後 CAI 教材開発のニーズも高まり開発件数も増えていくと予想されるが、マルチメディア化が進んでいくことにより、コストも開発労力も増大していくに違いない。こうした時に、複数の教育機関で類似のコンセプトを持った CAI 教材開発が同時に行われているとしたら、結局は費用とエネルギーを浪費していることになりはしないだろうか。日本語学習者の多様化が言われている昨今、教材也多様である必要がある。それぞれの CAI 教材が、使用現場での実情に合わせてコースウェアや使用データを手直しできる形で作成され、安価に、あるいは無料で提供されるようになれば、CAI 開発を行う教師は、互いに競合しない教材開発を考えていくことができ、日本語教育全体としては、かなりの節約になるはずだ。

こうした日本語教育の分野でもパソコンの持つマルチメディア性が最大限に発揮できそうで

あり、今後もさまざまな CAI 教材の開発が期待できそうである。また、そうした際に、日本語教育に限らず、才田氏の提案するようにネットワーク化を進めて行く必要があるはずだ。なにせネットワークもまた、パソコンはひじょうに得意としているのであるから。

一方、音声や画像を用いないのでマルチメディア性を生かしているとは言い難いものの、文字機能を駆使したデータベースだけでも十分に有効である。そうした研究にはたとえば、宮島達夫・中野洋・鈴木泰・石井久雄編『フロッピー版古典対照語い表』（笠間書院1989年9月刊）を活用したものとして、古田啓「文法研究のためのデータベース利用」¹⁷⁾や、金水敏「AWKによるテキスト型データベース活用術」¹⁸⁾などがあり、こうした論考の中でも応用例などが紹介されている。

(2) 日本文学

次に日本文学の研究領域におけるパソコンの活用状況を概観したい。日本語学では、先に掲げた以外にも相当に幅広く、また多様にパソコンが駆使されている。ところが、これに比べると日本文学は、まだあまりパソコンと関わりを持っていないと言わねばならない。たとえば、日本語学の方は、「日本語学」が特集号を組み、また「計量国語学」といった、ほとんどこうした領域を専門の範疇として取り扱う学会誌¹⁹⁾までが存在する一方で、日本文学の分野では「文学」「国文学」「国文学解釈と鑑賞」「日本文学」などの代表的な全国誌では、こうした特集はおろか、紹介記事もほとんど見られないのが現状である。

では、どうしてこれほどの違いが生じたのだろうか。ひとつには、日本語学の研究領域においては従来から統計的な方法が用いられてきたことがあり、この限りではパソコンがきわめて有効なこと、また先にもふれたマルチメディア性を生かせることが大きかったであろう。その一方で、日本文学研究ではその対象とするテキストの長さや、その絶対量の多さなど²⁰⁾がパソコンの導入や活用に不利に働いてきたのであろう。しかし、最近ではようやく本文のデータベース化も試みられるようになり、これらを研究者が相互に利用できるようなれば、状況も大いに変わってくるに違いない。²¹⁾ また、商業ベースでも岩波書店が「八代集」の CD-ROM を出し始めるなど、この分野でも急速に展開していく時期にさしかかっているようだ。²²⁾

まず、そもそも本文の校合そのものにデータベースが用いられるようになった。ただし、このデータ入力作業は相当に物理的な時間と、さらには細心の注意力とが必要とされる。もっとも、それとても従来は手作業で行っていたことをパソコン入力に代替するに過ぎないとも言える。先掲の『パソコン国語国文学』に、源氏物語「若紫」の巻の冒頭部を「CSV」形式を用いた諸本の例が掲げられている。²³⁾ そして、ここで新たに浮かび上がってくるのが文字の問題である。すなわち、古典文学のテキスト本文は、本来の写本では旧漢字などが用いられてお

り、JIS 漢字コード12546字では明らかに不足するということである。この問題については統一的な見解はまだないようだが、中村一夫氏は先掲書で、データベースとしての汎用性からこうしたケースでは、任意の記号をあて、置き換え対照表を作成する方法を勧めている。

他にもまだまだデータの互換性や、著作権の問題など、これから克服してゆかなければならないことがらは多いが、一方で電子化されたテキストならではのメリットも多い。その最大のものは、何よりも電子テキストの持つ柔軟で動的な性格であろう。すなわち、これまでの手作業による校訂本や語彙索引は、多大な労力の末に一旦完成してしまえば、その改訂は非常に困難であろうが、電子テキストでは誰でもが新たな項目の追加や、修正などを行うことができるのである。すなわち研究が続く限り、常により理想的なものを指向し続けられるということなのである。また、コンピュータ・ネットワークを用いれば研究の共同化もより有利であるし、それだけ開かれたものにもなるであろう。

次に、こうして作業の結果生まれたものや、あるいは校訂本文によって作成されたデータベースは、電子テキストとして、さまざまな研究に活用することが可能である。一例を挙げると、任意に選ばれたキーワードを本文全体から摘出してみせることなどがたちどころにできることである。源氏物語などは、すでに語彙索引があるのだが、データベースの場合は、使い方次第でより広範に、しかも多彩な利用が可能であると思われる。たとえば、和歌や俳諧などの研究において（もちろん物語や散文にも有効であるが）、特定の歌語の用例検索は当然のこととして、2語、あるいは3語を含む歌の同時検索なども、ほとんど瞬時にして可能である。あるいは品詞ごとに、またあるいはジャンルごとの語彙群（源氏物語にあらわれる植物など）を次々に探し出すといったこともできるのである。それでそれが何になるのか、と言ってしまうとそれまでだが、少なくともこうしていろいろ検索しているうちに、何か面白い発見が期待できるかもしれない。あるいは学生たちにやらせてみると、思いがけず独創的な発見がそこに生まれてくるかも知れない。彼等がパソコンに親しみ、そして同時に作品の持つさまざまな世界に触れるだけでもうけものだと思うのだが。

これに類することで、面白い報告があったので以下に引用してみたい。

出版物のリストを検討することから、文学史の常識を反省するきっかけを作った論文がある。岩波の『文学』1988年9月号に載った高木元氏の論文である。「江戸読本の形成一貸本屋の出版をめぐる一」と題するこの論文は、江戸の文学における出版者がどのように出版物に関与したかを巡って新鮮な視野を持たせてくれる。この論文で高木氏はまったくコンピュータ利用に触れていない。しかし、氏から直接聞いたところによれば、氏は、入力した江戸時代の出版目録を様々にソートして、まさしく遊んでいるときに、その並び順から一つの発見をしてこの論に至ったと言う。

この事は二つの意味をもつ。まず、第一に、近世文学のように二次資料が豊富にある場合には、その資料もまた電子媒体に落しておくことによって、次のステップに進めるということと、第二に、川喜田二郎氏の『発想法』にあるような、“発想”をコンピュータも我々に与えてくれるのだということである。²⁴⁾

ところで、英米文学の領域においては、研究者の絶対的な多さや、コンピューターのより広範な普及をも反映して、電子テキストもある程度蓄積されているようだ。また、それにとまって、こうしたデータベースを用いた研究の試みも数々なされている。『英語英文学研究とコンピュータ』によると、たとえば、特定の語彙のその作品における登場頻度を調査したり、その作品に固有のイメージの成立過程を考察するといった方法が報告されている。²⁵⁾ また、同書には Steinbeck の“Cup of Gold”を例にとりて、イメージ造形の方法の解析や、作家の表現上の特質や、個性といった文体論的な考察も試みられており、今後のいっそうの発展の可能性を感じさせる。²⁶⁾

また、毎年膨大な量の研究書や論文が発表されているが、これらの情報源としてもデータベースは実に便利である。ちなみに、伊井春樹氏の「古典文学総合時点データベース構築の試み」²⁷⁾によれば、『国文学年鑑』昭和61年度版には『源氏物語』に関するものだけで、研究書25冊、論文総数は実に277点にもものぼるそうであり、たしかにこれだけでも随分な量になるし、ましてや過去5年間や10年間ということになれば、膨大なものになるだろう。年度ごとの『国文学年鑑』を調べれば、そのことは把握できるわけではあるが、もしこれがデータベース化されていれば、著者別、掲載誌別などの多様な検索方法が可能になり、いわばデータそのものが立体化してくるであろう。

こうした研究支援システムの一貫として、国文学研究資料館のオンラインデータベースも、『マイクロ資料目録』、『和古書目録』にはじまり、今後もいっそう充実したものとなっていくはずである。

お わ り に

これまで人文科学の研究においては、かならずしもじゅうぶんにパソコンが活用されてきたとはいえないであろう。しかし、一方ではこれまでに概観してきたように、こと日本語学、日本文学の研究領域に限ってさえ、ずいぶん様々な試みがなされていることもまた確かである。しかも、これまで繰り返し述べてきたように、パソコンによるデータベース化は、蓄積が進めば進むほど累加級数的にその本来の機能を発揮するはずである。また、単なる検索だけではなく、それは我々の発想そのものにも革新をもたらすかも知れない。あるいは、これまでの研究のあり方や、形態にも変化をもたらさそうである。いずれもまだまだその端緒についたばかり

だといってもいい。なにしろ、すでに伝説化された有名な物語、すなわちスティーヴ・ジョブズが彼の家のガレージに Apple 1 号機を持ち出したのは、1976年のことなのであるから。パソコンにとって、今後の20年が、その日からの20年にもまして激変するであろうことは疑いない。そして、これを利用する研究環境もまた、相互にこれを促していくはずである。

- 1) CD-ROM の特性等については、西垣 通『マルチメディア』(岩波新書)を参照した。
- 2) <桐>については『パソコン国語国文学』(西日本データベース研究会編著・啓文社1995年1月刊)の第1章(西端幸雄氏担当)を参照した。なお、同書で氏は年間約4万首の和歌の語彙データベースを作成し、これが実に50万データに達すること、これをカードにして積み上げれば50mになるが、光磁気ディスクならばわずか1枚に収まることなどを述べている。実にわかりやすい例解だと思う。ましてその後の処理作業を想像した時、こうした研究にとっていかにパソコンが有用であるか、もしくははやそれなくしては不可能ではないかとさえ思われるのである。
- 3) 『広辞苑』CD-ROM 版については、増井 元『『広辞苑』CD-ROM 版』(『日本語学』1991年8月号所収)を参照した。また、早川聞多氏の「自在な検索システムについて」(『人文科学データベース研究』第3号所収)にもこの機能等についての説明がある。
- 4) 特に和文の中にアルファベットの混じる文書や、注などで文字サイズの異なる場合などにこうした誤読が生じやすいようである。なお、日本語 OCR の特性と問題点等については、酒井恵美子「OCR を使って KWIC を作る」(『日本語学』1993年3月号所収)に詳しい。
- 5) 先掲の『パソコン国語国文学』の表紙にこの画像が採用されている。ちなみに、小野小町が書見台にパソコンを置いているというアイデアはなかなか秀逸だと思う。
- 6) 熊谷康雄「現代日本語のテキストデータベース」(『日本語学』1991年8月号所収)にこの指摘がある。
- 7) 中野洋「言語研究用データにおけるデータベースの考え」(国立国語研究所 LPD 月報別冊12 1973年所収)
- 8) 国語学会・国立国語研究所『日本語研究文献目録・雑誌編』<フロッピー版>(秀英出版1988年)。またこのデータベースについては中野洋「日本語データベース」(『日本語学』1991年8月号所収)に詳しい解説がある。
- 9) 宮島達夫『『茨城方言・民俗語辞典』の印刷と検索』(『日本語学』1991年10月号所収)
- 10) 柴田武・井口三重「コンピュータが描いた『糸魚川言語地図』をめぐって」(『日本語学』1992年5月号所収)
- 11) 高橋顕志『『四国言語地図』を作る』(『日本語学』1992年8月号所収)
- 12) 先掲『パソコン国語国文学』第5章2(ダニエル・ロング)
- 13) 前川喜久雄「言語地図作成支援プログラム」(『日本語学』1993年4月号所収)
- 14) 上野善道「アクセントデータベースの作成」(『日本語学』1992年2月号所収)
- 15) 堀口純子「話しことば研究のためのカード型データベース」(『日本語学』1992年10月号所収)
- 16) 才田いずみ「マッキントッシュによる音声教育」(『日本語学』1992年12月号所収)
- 17) 古田啓「文法研究のためのデータベース利用」(『日本語学』1991年12月号所収)
- 18) 金水 敏「AWK によるテキスト型データベース活用術」(『日本語学』1992年6月号所収)
- 19) 計量国語学会の学会誌。
- 20) もっとも、日本語学の研究対象も原則的には無限量といえ、この限りでは違いがないともいえるのであるが。

- 21) 先掲の『パソコン国語国文学』に、本文データベースについてわかりやすく紹介されている。
- 22) 大阪大学の伊井春樹氏の提唱する『古典文学総合辞典』のようなものが出来上がれば文学研究だけではなく、授業や学生の事前学習などにも便利であろう。
- 23) 『パソコン国語国文学』第2章2（執筆担当中村一夫）。また、このあたりの本文データベースについても同書を大いに参照した。
- 24) 内田保廣「近世文学の『世界』から」（『人文科学データベース研究』第2号所収）
- 25) 西村道信「特定作家の文体研究」（齊藤俊雄・編『英語英文学研究とコンピュータ』英潮社1992年所収）には Robert F. Allen の次のような公式を紹介している。
- X=Absolute or observed frequency
ANF=Adjusted norm frequency
Z=Central limit theorem
- $$Z = \frac{(X - ANF)}{\sqrt{ANF}}$$
- 26) スコット・ピュー「スタインベックの小説—コンピュータによる文体分析の可能性—」。前掲『英語英文学研究とコンピュータ』所収。
- 27) 伊井春樹「古典文学総合時点データベース構築の試み」（『人文科学データベース研究』創刊号所収）