人文科学とコンピュータ　11－1
（1991．10．14）

# 読書とコンピュータ

## テキストデータで本が読めるのか

當山　日出夫

円満寺

テキストデータの利用方法について、読書とコンピュータ、あるいは、人文科学とコンピュータという観点から概観すると、次の問題点が指摘できる。

1．テキストデータは、書物と併用してこそ有意義に利用できる。
2．テキストデータと比較することによって、従来の書物の持っている文化的・歴史的な意味が明らかになる。
3．テキストデータの出現によって、書物を主な研究対象とする人文科学の研究は、根本的に変革していくであろう。

また、ＴＥＩについて考える時、これらの点は、重要な意味を持ってくるものである。

## READING AND COMPUTER

Hideo TOYAMA
ENMAN-JI

631 NARA-SHI NIMYO-CHO 1502, JAPAN

About Usage of Text Data
The points of this essay is the following three points.

1. Text data can be used meaningfully when it is used together with books.
2. The conventional cultural and historic meaning of the books become clear by comparing the books with the text data.
3. The study in liberal art science with books as its main studying objects will be changed fundamentally by the appearance of the text data.

## 1．はじめに

「読書とコンピュータ」などと、いささかおおげさなタイトルをかかげてしまったが、本発表は、何か結論を出すというものではなく、テキストデータ（注1）について、こんなことも考えてみる必要があるのではないか、という提言をしたい、その程度のものであると御理解ねがいたい。

## 2．2つの課題

人文科学におけるコンピュータ利用について、私見としては、次の点が問題あるいは課題であると思っている。

（1）まだ使っていない人に対して
　　→今後どのような啓蒙活動が必要か。

随分と利用されるようにはなってきたけれども、それでも、全体から見れば、人文科学研究におけるコンピュータ利用は、まだまだ少数派である。大多数の研究者は、まったく無関心か、あるいは、関心はあるのだがどう対処してよいかわからない、という状態であると言ってよかろう。これらの人々に対して、コンピュータ利用をどのように啓蒙していく必要があるのか。その便利さを訴えるだけでは、もはや効をそうさない時代に入ったのかもしれない。

ただ、研究室にコンピュータやパソコンがあるから、仲間の研究者が使うから、という理由で使わざるを得ない、という利用形態は、本来のあるべき姿とは言えないかもしれない。

ただ便利かどうかという観点から見るならば、道具としての効率のみが問題とされる結果になる。そうではなくて、人文科学研究の分野にコンピュータ利用が入りこむことによって、何か質的な変化が生じるかもしれない。その意味で、コンピュータを使う使わないの次元ではなく、たとえ使わないにしても、コンピュータ利用によって生じる質的な変化を視野にいれることの必要を考えねばならない。

（2）すでに使っている人に対して
　　→その使い方に問題はないか。

では、すでに使っている人に対しては、もう使っているのだからという理由で、何も働きかける必要はないのか。ひょっとすると、コンピュータの使い方やそのもつ意味を、根本的に反省してみる必要があるのではなかろうか。

コンピュータの登場によって、人文科学研究そのものが大きく本質的に変革をせまられているのかもしれない。コンピュータは、既存の研究方法を支援してくれる道具にとどまるものではなく、人文科学研究の存立の基盤そのものを動かしていくかもしれない。

この点を、研究者自身としてどのように考えるべきであろうか。

## 3．人文科学研究とテキスト

ところで、今回の発表のタイトルを、「読書とコンピュータ」としたのは、一般的に言って人文科学研究における、最大の研究資料となるのは、テキスト・書物、である。活字印刷されたものもあれば、木版による板本のものもある、また、写本で伝わっているものもある。どのような形態であるにせよ、本を読み理解し自分の解釈を見出していく営みが、その研究活動の根本に位置する。

コンピュータはこれからもどんどん発達していくだろうし、ソフトウエアもさらに良いものが出来てくるに違いない。が、そのようなコンピュータやソフトの発達にともなう研究環境の変化と、コンピュータ利用の有無による研究環境の変化とは、まったく次元の違うものであると見なすべきである。

文章を書くためのワープロソフト（エディタ、日本語ＦＥＰ等）の発達は、例えれば、筆で和紙に字を書いていた時代から、万年筆や鉛筆でノートに字を書くようになった変化に類すると言ってよい。それに対して、コンピュータの登場は、文字の無かった時代から、文字と紙を持つ時代へと変っていった時の変革に類すると言ってよい。テキストを電子化することによる、記録・通信・伝達・保存・検索等の処理についての無限の可能性は、文字の無かった時代に文字が登場し、紙の無かった時代に紙が登場した、その革命的なできごとに匹敵する。

ここで、文字で書かれたテキストを読み解釈することに基盤をもつ、人文科学研究の多くの分野が、その本質において、まったく影響を受けないということは、有り得ないであろう。そして、このことに研究者自身が無関心ではいられない・・・はず、である。が、現実にはどうであろうか？

## ４．テキストの歴史

ここで簡単に本と読書の歴史を振り返ってみたい。（一応、日本の場合を例にとる。また、話しを簡略にするため、きわめて大雑把な言い方になるが。）

常識的な知識として・・・昔は、本は無かった。いや、本がない以前に、文字も紙もなかった。

文字と紙の発明によって、はじめて、人類は本格的な書物というものを手にすることができた。文字も紙も無い時代、人々が、なにかまとまりのある言葉を伝えようとすれば、それは、口承による他はなかった。日本の最古の文学作品と言われる「古事記」「万葉集」など、その最初の成立は、文字で紙に書かれた書物としてではなく、口承によったものである。

日本における書物の歴史は、本格的には奈良時代から始まる。中国から多くの書物が伝えられてきた。それらは日本で読まれ、また、日本においてそれを写した写本も作られる。そして、前時代から口承によって伝えられてきた文学作品（「古事記」「万葉集」など）が、紙の上に文字で記録され固定した書物として成立する。つまり、口承された作品としての「古事記」の成立と、書物としての「古事記」の成立とは、別次元のことがらに属する。

そのうち、日本人自身の著作も書かれるようになる。平安時代に書かれた「源氏物語」など、日本人による創作の物語であるが、これは、写本として成立した。

長い間、書物とは写本であった。筆写によってしか、書物は作り伝えることができなかった。これは、室町から江戸時代はじめ頃までつづく。

写本の次には、木版印刷がおこる。仏教の経典などは、比較的早くから木版印刷が行われたが、一般の読者を対象にした商業的な印刷業が発生するのは、江戸時代以降である。江戸時代の作品として著名なものの多く、例えば、井原西鶴の作品（「好色一代男」など）は、板本として刊行され流布し読まれたものである。

次に、活版印刷術の時代をむかえる。明治の文明開化とともに、西洋から活版印刷の技術が導入され、それが主流となり、現在にまで及んでいる。（実際には、現在では、活版にかわって写植が中心になってはいるが。）近代の文学作品、例えば、夏目漱石・芥川龍之介・三島由紀夫などの作品は、活字印刷による出版によって、刊行され読まれたものである。

以上の概略を整理してみると、本発表の趣旨からは、次の２点が指摘できる。

（１）メディアの変化（口承→写本→印刷）

が、作品の成立に大きな影響を与えてきた。このことを逆に言えば、新しいメディアが生まれることは、それまでとは質的に違った内容の作品を生み出していくことにつながる。新しいメディアの出現は、より新しいより広い読者層の開拓をともなう。

（２）新しいメディアが出現したとしても、それまでのメディアがすたれて無くなってしまうのではなく、共存する。印刷が始まったからと言って、世の中から、写本がすべて姿を消したわけではない。むしろ、その役割分担が発生して、写本は写本としての役割をにないつづけている。

## ５．テキストデータの出現

現在、コンピュータの発達・普及にともなって、テキストデータが登場した。

では、テキストデータの出現が、書物を書く（あるいは読む）というこれまでの歴史の流れの中において、どのように位置するのであろうか？　この問いに対して答えを出せる段階にはない。最初に述べたごとく、本発表は、むしろ、このような問いかけが必要なのではないか、という提言をしたいのである。

が、それだけでは、あまりにも無責任と思われるので、現在の時点で考えられる点を以下にいくつかあげておこう。

## ６．テキストデータと書物

すでに作成され流布し始めているテキストデータに「源氏物語」がある（注２）。これを例にする。「源氏物語」テキストデータは、いかなる利用が可能であろうか。どのように読まれるであろうか。

もし、私自身が「源氏物語」を読もうと思った時、どうするであろうか。書架から、日本古典文学大系（岩波書店）「源氏物語」を取り出してきて、開くであろう。現時点の研究の水準からして、岩波古典大系の「源氏物語」は、最良のテキストであるとは言い難い。しかし、作品として「源氏物語」を読もうと思った時、まよわずこの本に手がのびる。

何故か・・・最初、大学で国文学・国語学など勉強しはじめた時、読んだテキストがこれだからである。客観的に見れば、滑稽かもしれない。学問的には、より進んだ良質のテキストに依拠しなければならないのは当然である。だが、最終的に論文を書いて、用例として引用する箇所等について他のテキストの異文を適宜参照するということはあっても、「源氏物語」を作品として読んでみたい時には、やはり他の本では馴染めないのである。

これは、非常に心情的・情緒的な思いこみのようなものかもしれないが、文学作品を読むということには、どうしてもつきまとうことなのかもしれない。

しかし、語彙の研究において、なにか特定の言葉を、「源氏物語」から探し出そうという場合、古典大系「源氏物語」を、最初から読んでいって目で探そうとは、思わない。

以前であれば、「源氏物語大成」の索引を使った。

今であれば、「源氏物語」テキストデータを使うであろう。「源氏物語」テキストデータは、単純に本文をほとんどそのまま入力しただけのものであるので、同一語の異表記という問題が検索作業において生じる。が、これは、さほど大きな障害にはならない。むしろ、問題は、これは本文が小学館の日本古典文学全集「源氏物語」であり、馴染みのある岩波古典大系のテキストではない点である。

ところで、さかのぼって、「源氏物語大成」の語彙索引が世に出る以前はどうであったか。「源氏物語」の語彙の研究をしようと思った研究者は、ひたすら、「源氏物語」（全部で５４巻ある、現在の普通の活字本にしても、少なくとも５～６冊以上の分量になる）を最

初から読んで求める言葉を探していかねばならなかった。

以上のような状況から言えることは、少なくとも、研究者の立場からして、書物としての「源氏物語」が持っている意味・役割は、テキストデータの出現によって、大きく変ったということである。簡単に言えば、テキストデータの出現以前は、書物「源氏物語」が、「源氏物語」のすべてであった。が、テキストデータは、その役割のいくぶんか・・・語彙の検索のために読む・・・を、分担するようになった。

これは、研究にとってプラスであろうかマイナスであろうか。語彙検索の効率という点からは、あきらかにプラスである。が、その結果として、研究者が、本を読まなくなる、読まなくても仕事ができるという状況が生じつつあることも、一方の事実である。これはあきらかにマイナスと言えよう。昔であれば、語彙検索のためには、やむをえず本を読まねばならなかった。労力は大変だし、見落しの可能性もある。しかし、ひたすら本を読むことによって多大の副産物を得ることができたことは、確かである。

それでは、「源氏物語」テキストデータは、ったい何のために研究者に提供されるのであろうか。もちろん、提供者の意図も重要だが、実際にそれを使う側の意識はどうであろうか。

さらに一般に、「源氏物語」以外の諸々のテキストデータについては、どうであろうか。ＣＤ－ＲＯＭ等を使った電子出版という語がさかんに使われるようになってきている。これも一種のテキストデータとみなしてさしつかえない。辞書などは別にして、はたして、普通の書物が電子出版として、読書の対象、つまり、本を読むために利用されるようになるであろうか。あるいは、電子出版と、既存の書物とは、並立・共存し、相互に役割を補完するものであろうか。

また、人文科学とコンピュータという視点からは、どうであろうか。書物そのものを研究対象とする人文科学において、書物の持つ意味・役割の変革にどう対処すべきであろうか。あるいは、テキストデータの出現によって、既存の書物とはいったい何であったのか、その持つ意味を根本的に考え直してみる契機になりはしないであろうか。これまでは、ある文献・作品がすでに書物として存在することを、あまりにも当然のごとくに考えてきたのではなかろうか。

書物（文字による記録）とは何であるかを考える契機は、これまでは、口承（無文字）の段階から、文字による記録が始まった段階への移行の考察が主であった。あるいは、文字を有してからも、文字化されないもの（今でも、口伝として残っている）の考察であった。つまり、書物以前の段階から書物へ、ということである。

これに対して、今、書物から再び書物以外のもの（テキストデータ）へ、という段階にさしかかってきている。ここにきて、あらためて書物のもっている意味を問いかけなおす必要が生じている。

テキストデータは、たしかに便利である。が、テキストデータから欠落してしまっているものを、見落してはいけない。

簡単に思いつく例をあげれば、書物の装丁デザインである。テキストデータと比較することによって、装丁デザインをふくめて書物なのだ、ということに、あらためて気づく。

筆跡も重要。以前、私自身の仕事として、「和漢朗詠集」の漢字索引をパソコンで作成したことがある。もちろん、このもとのデータとして本文を入力してある。「和漢朗詠集」テキストデータからは、もとの写本の筆跡は欠落してしまっている。これは当然のことだが、重要である。「和漢朗詠集」は文学作品としての受容はもちろんだが、書道史においても貴重な作品である。というよりも、文学と書道とが融合して「和漢朗詠集」は伝えら

れ受容されてきたとも言えるのである。やや極端に言えば、テキストデータとなった瞬間に、総合的に「和漢朗詠集」をとらえることが不可能になってしまうとさえも言える。

活字本についても同様の指摘はできるが、印象としては、テキストデータの方が、原本からは遠いという感じがしてならない。

装丁デザインとか筆跡のことは、画像データによってある程度おぎなえるのかもしれない。が、テキストデータと画像データとの関係については、まだまだ、未開拓であると思われるので、ここでは、ふみこまないでおきたい（注３）。

要するに、総合的に書物がもっていた意味を、ここで考え直してみる必要がある、ということを指摘しておきたいのである。

## ７．著者と読者の新しい関係

テキストデータの出現というよりも、むしろ、パソコン通信等の一般への普及にともなう現象とみるべきことかもしれないが、伝統的な著者・読者の関係の崩壊ということが指摘できよう。

これまでの書物について言えば、本を書いて出版できるというのは、ごく一部の限られた人間の特権であった。一部の著者と大多数の読者は、切り離された存在であった。情報の流れは、一方通行であることが原則であった。

これからは、この関係は、部分的にであるかもしれないし、徐々にであるかもしれないが、しかし、確実に崩壊していくであろう。これまでの著者・読者の関係は、近代的な印刷・出版が膨大な設備と資本を必要としたことに起因する。が、コンピュータやパソコン通信の普及にともない、電子メディアによる発言が可能になる。これは、強いて言えば、出版以前の段階、つまり、写本の段階にもどったとも言えるのである。

写本の時代、識字層は限られたものであったかもしれないが、その狭い人口のなかに限定して見ると、文字が書けさえすれば、誰でもが書物の著者になりうる可能性があった。これは、活字による近代的な出版の時代において、ある市民が本を書いて商業的に出版できる蓋然性よりもはるかに高いものである。

平安時代の昔、貴族たちにとって和歌を読み詠むことは日常の行為であった。この時代、作者と読者とは分離していない。上手な作者がいて、その作品がよろこんで享受されたということはあったとしても。誰でもが、和歌を詠むと同時に他人の作品を読みもしたのである。これは、やはり、写本というメディアの特性を無視しては、考えられない。

現在、ある人がワープロで文章を書いて、そのフロッピーディスクをコピーして多数に配布したとして、著作権は著者のものであるし、これはこれで、十分に一種の電子出版として成立する。また、パソコン通信のボードに書いたメッセージも著作物であり、最近では、単なる情報交換・おしゃべりの域を越えて創作活動の場としても機能しはじめている。

例えば、「新・青べか物語」（注４）という書物は、ＡＳＡＨＩパソコンネット（注５）というパソコン通信において、「御符の新青べか物語」として連載されたものをもとにして出版された。これは、パソコン通信から出版へという道をたどった例であるが、もともとのパソコン通信の文章の方が、ネットの一般参加者からのコメントが随時よせられていて、読物としては、むしろこちらの方が面白いかもしれない。が、パソコン通信の文章は、活字でもないし書物でもない。原則的に、電子的な媒体（ネットのホストコンピュータの記憶装置、あるいは、各会員がダウンロードしたもの）の中にのみ存在する。

また、本稿執筆時点ではまだ具体化していないが、「朝日新聞」の連載小説を、この秋から作家・筒井康隆氏が担当することになり、

それへの読者からのコメントが、ＡＳＡＨＩパソコンネットを通じて連載と同時に進行する予定である。もちろん小説自身はいずれ単行本として刊行されようが、ほぼリアルタイムで読者からの反応がよせられ、それに答える形で小説家が作品を書いていく、このようなことは、電子的なメディアの出現によって始めて可能になったとも言える。そして、読者からのコメントをふくめて小説全体を読みたい場合、書物として刊行された小説ではなく、パソコン通信にアクセスして、ディスプレイ上の文字として読まねばならない。（ダウンロードして、プリントアウトしてということも可能ではあるが。）

島戸一臣氏の「御符の新青べか物語」や筒井康隆氏の新聞連載小説へのコメントは、ネット参加者であれば、誰でもが自由に書き込める。そして、それは、誰でもが自由に読める。つまり、ここでは、画一的な著者・読者の関係は、もはや存在しない。

前項では、テキストデータは、読まれるものではない、読書の対象とは成り得ないだろうと言った。が、ここでは逆に、電子化されたテキスト（それのみで存在する、書物が共存しない）は、読書の対象として成立することになる。そして、注意しておくべきことは、これは、これまでの著者・読者の関係ではなく、誰でもが著者であり読者である、という関係において成立する、ということである。

では、このことが人文科学とコンピュータという視点からどのような意味を持つであろうか。正直なところ、現在、このような現象が人文科学全般にどのような意味を持っているか、予見することはきわめて難しい。

人文科学の研究対象は書物であると言ったが、究極的には、それを書いたり読んだりする人間と言うべきである。その人間が、コンピュータの出現によって、どのように変っていくのか、そして、それにともなって人間を究極的には研究対象とする人文科学も、どのように変っていかざるをえないのか、という問題設定によって考えることになろう。

少なくとも、今の時点で言えることは、既存の研究分野・研究対象では補足できない事態が生じつつあることは、確かである。

問題を言語に限定してみても、誰でもが、不特定多数の人間に対して発言者でもあり読者でもある、という状況に対応して、人間の言語行動はどう変っていくであろうか。おそらく、既存の言語学・国語学の研究領域では捉えることが難しいだろう。場合によると、新たに情報言語学とでもいうべき新分野の開拓が必要になってくるかもしれない。

## ８．ＴＥＩについて

最近、人文科学とコンピュータ関連で話題になりつつあるものに、ＴＥＩ（注６）があげられる。本年６月のＴＥＩワークショップならびに人文科学コンピューティングシンポジウムでも、発言したことであるが、ここで一部その内容を再確認しておきたい。

ＴＥＩは、テキストデータの利用を前提にした企画である。

では、そのＴＥＩ化されたテキストデータは、いった誰のために、何のために、提供されるべきものなのであろうか。はたして、テキストを読むために使うのであろうか（そうであるならば、タグだらけのテキストでは、読みにくいので通読のためにそれを排除して表示するツールが必要になる。）あるいは、読むためではなく、学術的な語彙検索や統計的な処理のためだけに利用されるのであろうか。すでに書物としてある作品・文献については、書物との関係をどのように考えるべきなのか。その如何によって、ＴＥＩとして付加されるべき諸情報やテキストデータの形式は、大きく変ってくるにちがいない。

ＴＥＩは、我々にとって未知の企画である。具体的な日本語対応はかなり先のことになろ

う。

が、ＴＥＩについて考える時、タグ付加やその利用のためのソフトウエア開発といった技術的側面にのみとらわれるのではなく、いったい誰のための何のためのテキストデータであり書物であるのか、という視点から、絶えず問いかける姿勢が必要である、と思う次第である。

1991/09/05

注１　「MACHINE READABLE TEXT」のことであるが、「機械可読テキスト」「テキスト・データベース」「電子化テキスト」等、種々の呼び方が現在行われている。本稿においては、「テキストデータ」で統一した。

注２　長瀬眞理氏によって作成されたもの。底本は、日本古典文学全集「源氏物語」（小学館）である。
長瀬眞理「日本語－英語対象「源氏物語」のテキスト・データベースの作成に関する基礎的研究」『情報知識学会誌』１巻１号（１９９０）

注３　かりに画像データとして「和漢朗詠集」が存在したとしても、それが、読書の対象となり得るかどうかは、疑問であると言わざるをえない。物理的に、紙に書かれ（印刷され）製本されたものとしての「書物」と、その視覚的イメージだけを抽出した画像データとは、やはり、異質であると見なすべきである。

注４　「新・青べか物語」島戸一臣著　朝日新聞社　平成２年１０月刊

注５　朝日新聞社系列のいわゆる商用パソコン通信ネット。ネット事務局は、〒１０４東京都中央区銀座８－１０－４　和孝銀座８丁目ビル　（株）アトソン内
ＴＥＬ０３－３２８９－７０６１
ＦＡＸ０３－３２８９－７０６６

注６　「TEXT ENCODING INITIATIVE」これをめぐって、１９９１年6月２８日に、ＴＥＩワークショップ（於東大大型計算機センター）、翌２９日に、人文科学コンピューティングシンポジウム（於機械振興会館）が行われた。なお、ＴＥＩについての私見は、「ＴＥＩと日本文学研究」と題して、ＴＥＩワークショップにおいて発表した。