

古典籍とJIS漢字についての再考察 － 何が変わったか、変わらないでいるか －

The re-considering about the Classic text and the JIS-KANJI

當山日出夫
TOUYAMA Hideo

立命館大学グローバルCOE (DH) 客員研究員
Ritsumeikan - University Global-COE (Digital Humanities)

京都市北区等持院北町 56-1
56-1, Toujiin- Kitamchi, Kita-ku, Kyouto
E-Mail : htoym@kcn.ne.jp

あらまし：本稿は、13年前の、第1回の「人文科学とデータベース」シンポジウムで発表した論文を、再デジタル化し掲載する。それに対して、現在の視点からの考察をこころみる。古典籍に限らず、資料のデジタル化（文字テキスト）では、そこに、研究者（作成者）の解釈がある。解釈なしに、エンコードはあり得ない。そして、テキストの解釈こそが、研究者にとって最も重要な課題である。

Summary : The author released a paper, the " classic text and JIS kanji ", 13 years before. The author released this paper by the symposium of the first. The author rereads the previous paper from the present viewpoint and reevaluates it. " We don't solve the problem of the classic text and JIS kanji " even if the character increases. The basic investigation about the text and the character is important

キーワード：古典テキスト JIS漢字 0208 0213 ユニコード

Keywords : Classical-TEXT JIS-KANJI 0208 0213 Unicode

1. 本稿の意図について

筆者は、この本稿が掲載になる『人文科学とデータベース』シンポジウムの第一回（大阪電気通信大学、1995年）で、『古典籍とJIS漢字－テキストの本文校訂との関係において－』と題する発表を行った。今からふりかえれば、この1995年という年は、Windows95が発売になった年であり、インターネットが急速に一般化した時期でもある。そして、この論考は、さらに8年前に書いた論考『コンピュータではあつかえない漢字－「和漢朗詠集」の場合－』（『汲古』、汲古書院）を、ふまえて書い

ている。昭和62（1987）年。つまり、合計すれば、21年前の論考になる。

ここでは、まず、このとき論考を再掲載してみたい（個人的には、再デジタル化してみたいということであるが）。その意図について、述べれば、

第一に、すでに引用されることもない人目につかない資料になってしまった論考ではある。それを、現在の時点では、どのように読めるか、提起した問題は解決されたのか、考えてみたい。

第二に、これは、コンピュータと文字についての論考

である。その原稿を書いたコンピュータと文字の環境は、この間に大きく変化している。日本のコンピュータ(Windows95)では、JIS漢字(0208いわゆる83JIS)だけの時代である。それが、Windows Vistaになり、「JIS X 0213:2004」や「Unicode」が使える環境になって、どう書けるのか、自ら考えながら書き直してみたい、のである。

まず、原則として全文を再掲載する(注はのぞく)。

【1】はじめに—JIS漢字の論点—

JIS漢字をめぐるのは、これまでに各所で様々な議論されてきている。筆者もまたいくつかの問題提起をおこなってきた。その論点を整理すると次のようになる。

- (1) .字体・字形・書体・フォントなどJIS漢字を論じるための諸概念の整理
- (2) .収録字数・字種の範囲の問題
- (3) .78年版と83年版以降の改訂をめぐる新旧JIS漢字の問題
- (4) .そもそもJIS漢字とは何を規定したものなのか
- (5) .これからのJIS漢字はどうあるべきか
- (6) .JIS漢字の有効利用と漢字検索の問題

このような、JIS漢字についての議論は、古典テキストをあつかう人文学研究においてコンピュータ利用がはじまった時から既に、最大の問題点であった。研究者が自分の研究目的でコンピュータを使うようになった時、まず直面したのは、「必要とする文字がない」「近い字はあるが字体が不満である」という、JIS漢字の字種・字体についての問題であった。それが、近年になってようやく上記のように整理されてきた段階、と言えるであろう。だが、依然として、JISに無い字をどうするかということは、未解決のまま残された問題点である。

【2】JISに無い字

JISに無い字への対処策としては、現時点では、次のようなものがある。

- (1) .外字作成。最近では、アウトラインフォントを簡単に自作できるようになってきた。
- (2) .代用。アルファベットや仮名文字、あるいは、大漢和番号など、JISコード内の文字で記述可能な方法

で代用する。

- (3) .あきらめる。JISに無い字は、「■」(活版印刷でいうところのゲタ)とでもしておき、紙に書いた一覧表を付属させる。
- (4) .他の文字コード系を使う。現に、中国文学研究者などは、日本のJISコードで日本語文と、中国の簡体字を使う中国語文を、一つの文書に同居させることを行っている。

どの方法がよいのか、それぞれの研究目的・利用目的によって異なる。

だが、ここで改めて考えてみるべきことは、そもそもJISに無い字とは何であるのか、ということではないだろうか。一見、分かり切ったことのように思えるテーマであるが、考えてみると、広範囲にわたる歴史的検証と厳密な論理構成が要求される、きわめてやっかいな問題である。以下、『和漢朗詠集』を材料として、JISに無い字とはいったい何であるのか、あらためて考えてみたい。

【3】『和漢朗詠集漢字索引』の方針

『和漢朗詠集』は、平安時代、藤原公任の撰になる秀句集で、およそ800程の和歌と漢詩句を、テーマごとに分類して編纂したものである。王朝貴族文化の精髓であるとともに、その後の文学作品に多大の影響を与えた作品として、文学史上に名をとどめている。多数の写本・版本のテキストが伝えられているが、最も標準的に用いられるのは、伝藤原行成筆御物本である。現代の校訂注釈本の多くは、このテキストを底本に採用している。

この写本(複製)に基づき、2種類の校注本(岩波日本古典文学大系・新潮日本古典集成)を参看して、本文の表記にもちいられている全ての漢字を検索することを目的として、漢字索引を作成した。

凡例から、字体の認定にかかわる箇所を引用する。

[1] 本文

- 3 本文は、次の方針による。
 - a ①の影印本文を本行とする。
 - b なるべく正字体をもちいるが、原本(影印本)での書写字体にも配慮する。
 - c 異体字については、索引としての検索のため、原則的に正字体に統一する。

- d ②③の校訂本文とともに（ ）で示す。ところにより、校注者の判断の違いがあるが、それらを区別することはしない。また、必ずしもすべての異同を示すこともしない。わずかな字体の違いなどで、索引としての検索に支障のないものについては、採用しなかったものもある。

[6] 字体

- 1 本索引は、その作成はパーソナルコンピュータによったものである。印刷については、レーザプリンタで印字したものを、そのままの形でオフセット印刷してある。したがって、使用字体および字種は、すべて原則的にJIS規格（JIS C 6226 78）によらざるをえなかった。
- 2 JIS規格に無い字については、編者が外字として作成した。
- 3 そのため、すべての漢字を正字体に統一することは不可能であり、結果として部分的に新字体・正字体の混乱が生じている。しかし、索引として、「和漢朗詠集」の漢詩句の漢字を検索するのに、実用上の不都合は生じていないはずである。
- 5 編者が外字として作成したのは、「和漢朗詠集」本文の漢字では次の63字である。

最終的に、上記のような方針にもとづく漢字索引を作成したのであるが、古典テキストのコンピュータ処理という観点からは、次の問題点を指摘できる。

(1) 字体の統一処理

多くの場合、古典テキストの校訂にあたっては、字体を正字体に改めて統一的に処理することが多い。古典を、いわゆる「正字」（まさしく「正しい文字」）で表記するのは当然のことのように思われるかもしれない。しかし、実際の写本等では、必ずしも「正字」ですべて書かれているということはない。むしろ、当時の通行の書写字体で書かれるのが普通であり、現在の『康熙字典』を規範とする正字意識を杓子定規に適用すると、かえって、テキストの本来の姿を理解できないことになってしまう。

(2) 索引としての字体処理

古典テキストのコンピュータ処理が、電子化テキストとして、文字・語彙の検索に利用することを目的とす

る場合、あるいは、索引の作成を目的とするような場合、検索の便宜のための字体処理が必要になる。具体的には、

A. テキスト全体を統一的に処理する。

B. 検索システムに、異体字シソーラスのようなものを用意する。

この何れかが必要になる。

(3) 本文校訂

凡例3のdとした、校注本の字まで検索の対象とするか否かである。『和漢朗詠集』を一般の研究者が読む場合、実際には通行の校注本によっている。校注者により、本文校訂の判断が異なる場合があり、現実には、校注本に独自の本文が生じることになる。これらをどこまで採用するかが問題である。また、底本（写本）の誤写とおぼしき箇所であっても、簡単に廃棄してしまつてよいのかも疑問である。

以上の3点について総合的に考えたうえで、また、索引としての実用的性格にも配慮して、実際の漢字索引作成となる。

[4] 『コンピュータであつかえない漢字』

『和漢朗詠集漢字索引』の刊行にさきだつて、筆者は、『コンピュータであつかえない漢字—「和漢朗詠集」の場合—』という論文を発表した。これは、索引作成にあたって、準備的に考えたこと、特にJIS漢字関係の問題点を、まとめたものである。この論文を書いた時、筆者の使ったパソコンに搭載されていたJIS漢字の規格は、「JIS C 6226 78」であった。現在では、「JIS X 0208 90」になっており、この間に、78年版から83年版への変化、いわゆる新旧JIS漢字の改訂がおこなわれたことは周知の事実である。

しかし、古典テキストの本文校訂とJIS漢字についての、もっとも基礎的な問題点については、現在にいたるまで十分に議論がつくざれたとは言い難い状況であり、今から、8年前に書いたものではあるが、この論文についてふりかえってみたい。

結論として、『和漢朗詠集』における非JIS漢字として、64字を認定した。

一般に、ある文献におけるJIS漢字・非JIS漢字を問題にする時、まず問題になるのは、JIS漢字でどの程度入力

可能か、非JIS漢字はどれくらい出てくるのか、ということであろう。それについてみれば、『和漢朗詠集』の総字数は約10,000字、異なり字数は約1,800字である。したがって、総字数に対する非JIS漢字の割合は約0.7%であり、異なり字数に対しては約3.5%ということになる。

だが、筆者は、このような数字にそれほど意味があるとは思わない。まったく無意味とするわけではないが、重要なのは、何%かという数字ではなく、その中身である。JIS漢字・非JIS漢字と判定した、具体的な字の種類であり、さらには、その判定基準である。

表には全部で64種類の漢字が並んでいる。おそらく古典テキストの本文校訂の現場には疎い、強いていえば理系のコンピュータ研究者の目からは、どれも同じようにJISに無い字と見えるかもしれない。

たしかに、誰がどう見ても、JISに無い字（非JIS漢字）であると判断する字はある。そして、それが一覧表に表に示した非JIS漢字のかかなりの部分をしめることは事実である。だが、中のいくつかの字については、そう単純に非JIS漢字であると判定してしまうには躊躇される、微妙な学問的判断が要求されている。テキストの本文校訂とかかわる場合である。

[5] 所属部首とJIS漢字

まず、筆者の作成した『和漢朗詠集』の電子化テキストは、最終的には部首画数順配列の漢字索引を作成することを目的とした。そのため、異体字等については、次のような原則でのぞむことになった。

- (1) .なるべく正字体をもちいることにするが、原本（平安時代の写本）で使用される字体にも配慮する。
- (2) .その字の所属部首が変わってしまうなどの特にいちじるしい支障がないかぎりにはJIS漢字をもちいる。

ここで注目してもらいたいのは（2）の方針である。もし、部首画数順ではなく、他の配列方式、例えば、四角號碼配列や音訓配列による索引を作成するのであったならば、また異なった判断を下すことになったからである。具体的には、次の諸例である。

A.所属部首が同じであるので、統一処理した例。

- [1] .崑 所属部首は「山」。「山」を上を書く字（JIS）と、左に書く字（非JIS）とがある。所属部首が同じであるので、「崑」に統一的に処理することにした。

[2] .楫 所属部首は「木」。原本では、この字の右旁が「戈」になった字体をもちいている。しかし、所属部首は「木」であるので、「楫」に変えた。

[3] .苑 所属部首は「艸」。原本では、全8例のうち2例が、「苑」。残りの6例は、「艸冠」の下に「ウ」ないし「ワ」を書く字。特に区別して用いた形跡も認められないので、「苑」に統一。なおこの字については、現代の校注本の書き下し文は、「苑」を使用している。

B.反対に、所属部首が同じであっても、統一せずに別の字とした例がある。

[5] .寧 原本では普通に「寧」をもちいる。しかし、1例だけ下部を「用」に作る字が使用されている。それは、固有名詞（人名）で使用されている。こと固有名詞の用字については、原本どおりの字体（非JIS）を使用することになる。ただし、この箇所、所属部首という点では、同じ「ウ冠」に属する。

C.似通った字ではあっても、所属部首が異なるために別の字として処理することになる例。

[6] .「徘徊」の熟語で使われるが、「徘徊」と書けばJIS漢字で表記可能。この「徘徊」の「彳」を「人偏」に作る「俳」はあるが、「徊」を「人偏」に作る字はない（非JIS）。一般に、古写本の書写字体において、「彳」と「人偏」とは、さほど厳格に区別されているとは言いがたい場合が多い。特に行書や草書で書かれた場合には、同じように書かれてしまう。

しかし、明朝体で印刷する場合、区別される字となるし、また、所属部首が異なってしまう。なお、この箇所、新潮古典集成では「徘徊」としている。

ただし、最終的な漢字索引においては、「徘徊」とした。それは、その当時使用したパソコン（NEC 9801M2）の能力では、作成できる外字数に厳しい制限があったからである。やむをえず、「徘徊」の例に限って、原則を曲げて妥協することとした。

以上は、所属部首とJIS漢字の判定がからんでいる例である。原則は、原本で使用される異体字（非JIS漢字）が、部首を同じくするならばJIS漢字で処理し、異なるならば外字を作る、ということである。これはあくまでも部首画数順配列の漢字索引作成という目的に即して考えた場合のことである。四角號碼配列索引であるならば、[1]

の「崑」や[2]の「楫」は、別の位置に配列されてしまふことになる。しかし、[6]のような「俳・俳」は同じ位置になる。

【6】本文校訂との関連

本文校訂との点で問題になる例がある。

[7] .僻 「僻」の「人偏」を「土偏」に作る字が原本では使われている。これは誤写である。単純な誤写と判定してしまい「僻」に本文を改めるならば、「土偏」の字（非JIS）は不要になる。原本の表記を尊重するが故に、非JIS漢字が必要となる箇所。なお、「土偏」の字は普通の漢和辞典には見いだせない。

[8] .護 正しい本文は、偏が「言」ではなく「水（サンズイ）」。「上」の例とは逆に、原本どおりの表記で済ませるならば、JIS漢字の「護」だけで足りる。

これらの例のうち、「7」の「僻」については、一見すると無意味な処置と思えるかもしれない。単純誤写について、おそらく架空と思われる字を、外字として作る必要があるのかと、疑問に思うのが普通かもしれない。しかし、そうしなかったのは、次の例があるからである。

[9] .瑩 現代の校注本（岩波古典大系・新潮古典集成）ともに、この字である。しかし、原本の字は、「玉」ではなく「火」に作る字である。『和漢朗詠集』で「瑩」はこの箇所の他に5例見いだせるが、それらを観察すると、あきらかに「玉」と「火」は区別しなければならない。

ところで、この字を含む漢詩句は、『白氏文集』からの引用である。平安時代撰の『和漢朗詠集』は『白氏文集』から多くの漢詩句を採用しているが、平安時代に実際に日本で読まれたテキストが、幸いなことに現存している。その古いテキストを見ると、該当箇所の原文は「火」に作っている。これは、現代の校注本が誤ったさかしらな校訂を加えてしまった箇所である。

つまり、[9]の例について、校注本をそのまま信用して、「火」につくるのは誤写と判定してしまい、「瑩」を本文とするならば、外字作成は不要になる。だが、本文校訂としては誤った処理になってしまう。一見すると誤写のように見ながちな箇所であっても、調べてみると、意外とそれが本来の正しい本文を伝えているという例は、古典テキストの研究において希ではない。したがっ

て、あえて誤写の本文であっても、作字の必要があるのである。

『和漢朗詠集』の漢字索引を作成するといっても、ただ機械的に漢字をデータ入力して、索引作成のプログラムを走らせればよいというものではない。あくまでも、『和漢朗詠集』研究に資するものを、目指さねばならない。そのためには、既存の校注本を盲信するのではなく、学問的批判の姿勢が不可欠である。

【7】字が増えれば解決するか？

以上、指摘してきた『和漢朗詠集』のJIS漢字をめぐる問題は、コンピュータで使える字が増えさえすれば解決する……と見なされがちかもしれない。だが、そう簡単な問題であろうか。

『和漢朗詠集』の非JIS漢字64字のうち、上述の9字を除く55字については、その字がコンピュータで使える字として存在すれば、問題なく解決する。コンピュータで使える字は、多ければ多いほどよい。

しかし、上記に指摘した9字については、単に使える字が増えれば解決するという質の問題ではない。仮に非JIS漢字とした字が使えたとしても、索引作成を目標とした本文校訂という視点からは、問題が無くなるわけではない。

ここにあげた字は、たまたま必要とすべき字が非JIS漢字であったために、出てきてしまったものにすぎない。JIS漢字内部で処理可能であるが故に、JISに無い字として、議論の対象にはならないが、しかし、本文校訂の問題としては重要な問題をはらんだ文字が、これ意外にも多数存在するのである。

上記の「崑」に類似した例として、「峯」と「峰」の場合がある。「峯・峰」は、ともにJIS漢字であるために、非JIS漢字の問題点にはひっかかってこないが、テキストの本文校訂と字体処理については「崑」とまったく同じように考えるべきものである。

このような本文校訂にかかわる問題は、どんなに使える字が増えても決して解決しないし、また、自由な外字作成が可能になったとしても解決しない。

コンピュータで古典テキストをあつかう時、JIS漢字の制限のことは必ず問題になる。それに対して、どれだけの字数があれば十分かという方向での議論、つまり、何

字あれば何%をカバーできるかというたぐいの発想では、本文校訂の質を論ずることはできない。そして、古典テキスト研究者にもとめられるのは、より良質の本文校訂にもとづく電子化テキストの作成である。この研究の原点を常に認識しておかねばならない。

【8】文字とコンピュータ

最近のコンピュータと文字をめぐる議論は、「ISO 10646」およびその日本規格である「JIS X 0221」に話題が集中しているように見受けられる（本稿執筆の時点では、まだ規格が決まった段階で、それを実装したコンピュータは登場していない。）

ISO 10646について考える前に、そもそもコンピュータと漢字をめぐる議論では、次の各レベルを区別して論じる必要があることを見ておきたい。

(1) .文字集合

例えば、「JIS漢字・教育漢字・常用漢字」などのように人為的に定めた文字集合（キャラクターセット）である。

ここでは、どのような文字を、どのような字体で、どれほどの数、集めるのが妥当かどうか、が議論される。異体字については、教育漢字・常用漢字などは、意図的に、その文字集合に限定する限り、排除する方針である。JIS漢字の場合は、（その方針が無定見であるとはいえ）ある程度の異体字を取り込んでいる。

(2) .コード系

上記の(1)で定められた文字集合を、具体的にどのような形でコンピュータで使うかである。今日一般に使われている「シフトJIS」などがこれにあたる。

(3) .文字属性

漢字というものが「形音義」でなりたつものである以上、(1)の設定の段階で、既にある判定が下されている。そうでなければ、文字の選定は出来ない。だが、一旦決まった文字集合に対して、ユーザの側がどのような認識を持って対応するかは、また別の問題である。

JIS漢字についても、すぐに全体を問題にするから議論が混乱するのであって、実際的な利用法として、中にふくまれる教育漢字だけを使う、常用漢字だけを使う、という選択肢は可能である。

常用漢字中心の使い方をした場合、例えば「余・餘」

は新字体・旧字体の関係として認定されるが、正字体中心の使い方でのぞむならば、「余・餘」は別の字として使い分けることになる（「余」は「自分」の意味、「餘」は「あまり」の意味）。また、「芸」と「藝」の新字体として使うか、「ウン」と読んで植物の名称として使うかは、利用者の解釈に依存する。

あるいは、コード表の制定は、それがISOであれJISであれ、個々の利用者の漢字に対する属性認識まで拘束するものではない、と考えるべきであろう。言い換えれば、自然言語の表記として、人間は自由に文字を使うのである。

(4) .文字検索

文字集合（コード表）にある字であっても、利用者が実際に探せなければ、存在しないに等しい。JISにある漢字であるにもかかわらず、ワープロの仮名漢字変換で出せなかったためか、その字の箇所だけ手書きになっている文書を、目にするのは希ではない。

文字集合が現実にも有効に活用されるためには、そこにふくまれるすべての漢字を、もれなく速やかに検索可能なシステムを必要とする。場合によっては、それに無い字については、無いことの確認が求められる。

特に、古典テキスト研究の場合、JISに無い字が頻出するので、JISに無いことを確認する作業が、重要な意味を持つてくる。

(5) .フォント

コンピュータのディスプレイ上で、または、プリントアウトで、実際に目にする字のかたちである。上述の(1)の文字集合、(3)文字属性、(4)文字検索、これらは、「字体」レベルで、ある程度抽象的な概念として文字をとらえることになる。だが、現実には、文字は個々の具体的な文字デザインとして存在する。

具体的な文字デザインと、抽象的な字体概念とは、すくなくとも概念的には区別して議論しなければならない。

以上、各レベルの諸問題は、コンピュータで人為的な文字集合を使用する場合、必然的に発生することである。これは、JIS漢字だけに固有の問題ではない。将来、ISO 10646に的確に対応するためにも、上記の各概念についての認識は不可欠である。

[9] ISO10646 (JIS X 0221)

ISO10646の実現によって、古典テキスト研究は、どのように変わるであろうか。はたして、多大の恩恵をこうむるであろうか。

確かに、ISO10646が使えるようになれば、使用可能な字は増える。非JIS漢字を外字作成でしのぐような例は、かなり減るに違いない。しかし、だからといって、100%大丈夫かといえばそんなことはないであろう。古典テキストで、これまで「非JIS漢字」が問題になったのと同じように、これからは「非ISO10646漢字」という問題が発生することは必至である。

いや、実は、問題の本質は、字が足りるとか足りないとかではない。いわば、自然言語としての文字表記に対して、人為的な固定的文字集合 (JISやISO) で対応する場合の、文字論・表記論にかかわる原理的問題が未解決なままなのである。論点を絞っていえば、

(1) 数の不確定

人為的な固定的文字集合では、絶対に数が不足する。なぜなら、漢字は、部品 (部首) の組み合わせによって、新たに作ることが可能な文字である。

(2) 属性のゆれ。

漢字の属性 (形音義) は、一義的に決定不可能である。異なる文字集合間で、漢字属性の整合性をどのように考えるか。

このようなことがらについて、文字論的考察の視点が定められていないのが現状である。

これは、上記の (1) 文字集合および (3) 文字属性にかかわる問題であるが、その他にも、(4) 文字検索が、ISO10646でどうなるのかも、不安材料である。

JIS漢字でも、実際に研究者が、古典テキストの研究に使うためには、オリジナルのコード表「JIS X 0208」そのもののだけでは不十分で、市販のJIS漢字コード辞典の類や、JISコード付の漢字辞典を、座右におかなければならない。

JIS漢字は、一応、日本の漢字を対象としているので、漢字の属性認定については、そう大きな問題は生じない。しかし、ISO10646になれば、中国・台湾・韓国といった、漢字文化圏とはいっても、日本とは異なる言語文化における漢字までをも対象にしなければならない。日本の既存の漢字辞典の中に、どうやって外国の漢字を

とりこめばよいのであろうか。

[10] おわりに一ある提言として一

JIS漢字 (およびISO10646) については、近年、特に東洋の古典テキストを扱う学問領域においては、重要な議論のテーマとなっている。

ISO10646は、研究者に福音をもたらすであろうか、それとも、より混沌とした混乱状態を招来するに終わるであろうか。期待もある反面、なかなかその実像がつかめないでいる。

今、ここで我々がなすべきことは、過度に期待することでもなければ、問題点の指摘に終始することでもない。人為的文字集合 (キャラクターセット) とはいったい何であるのか、異なる文字集合間における漢字属性の整合性とはいったい何であるのか、という問題について、冷静に文字論・表記論的に考察をすすめることである。

人文学の研究者の世界である。研究者の数だけ考え方があろう。文字についても、必要とする文字はそれぞれに異なっている。研究者としての文字の共有化は困難かもしれない。しかし、文字についての、基礎概念の共有化であれば、可能かもしれない。我々に求められているのは、そのための資料収集であり、考えることである。

今まさに、ISO10646 (JIS X 0221) が登場し、同時にJIS漢字 (X 0208) の改訂も進行している。今後、コンピュータと文字は、どうなっていくか予断を許さない。だが、どのような状況を迎えようとも……すくなくとも「この字が無いのは問題だ」と言うことだけにとどまるのは、もう止めにしようではないか。

2. コンピュータと漢字を考える視点

以上が、今から13年前 (1995) 年の『人文科学とデータベース』(第1号) に書いた文章の、脚注を除いたほぼ全文である。用語について確認すると、文中で、「ISO10646」と表現しているものは、現在いうところの「Unicode」である。

あらためて確認しておきたいのは、何故、今になって、古い原稿を取り出してきたのか (そして、それをここに再掲載したのか) ということである。

第一には、「コンピュータであつかえない漢字」といっ

ても、それは、一義的に決まるものではない。あくまでも、本文校訂とともにある。そして、その本文校訂は、そのテキストをどのようにあつかいたいのか、研究者の判断によってきまる。原本で使用する文字を、異体字など区別して可能な限り忠実に再現したい場合もあるし、また、逆に、異体字などは統合して検索の便宜を優先する場合もある。どのような方針でのぞむかによって、校訂の結果は異なり、必要とする文字も異なる。したがって、「無い字」も異なるものとなる。前稿および本稿でしめたのは、あくまでも、筆者の判断によって必要とした文字である。このことを踏まえないで、一律に、文字の有無を論ずることには意味がない。

第二には、コンピュータと文字、というのは、人文情報学（デジタル・ヒューマニティーズ）の歴史そのものであると言ってもよいからである。すくなくとも、多くの人文学系研究者が、コンピュータ（パソコン）を使うようになったのは、PC-9801 などにおいて、第二水準漢字まで実装可能になったのが、大きな契機となっている。日本語、あるいは、漢字表記文献の、テキスト処理のはじまりである。

それまでは、コンピュータで漢字をあつかうことは不可能ではなかった。しかし、それは、大型計算機で利用するしかなかった。それが、個人レベルで使用する、パーソナルコンピュータで可能になったことの意義は大きい。そして、現在は、この流れの中にある。

また、初期のパソコンの処理能力では、画像データの処理などは、不可能であった。デジタルカメラなどもまだ登場していない。その後、今日、画像処理やGISなど、多様な処理が可能になっていることは、周知のごとくである。

しかし、画像データなどであっても、そのアーカイブのメタデータの記述としては、文字から離れることはできない。

全体として、このような経緯を俯瞰するとき、コンピュータで文字をあつかうことについて、どのように考えられてきたかを、「歴史」としてふりかえてみることは、決して無意味であるとは思えない。いや、過去の「歴史」をふまえることによってしか、未来への展望はないであろう。

第三に、たしかに、今となつては、すでに古びてしま

った内容がいくぶんふくまれている。だが、筆者自身が、かつての自分の文章を読み直して、まだ、この問題は解決されていない、と感じる箇所も無いではない。特に、現在、コンピュータと文字は、新たな局面をむかえている。その論点としては、

- ・ Windows Vista によって実装されている「JIS X 0213:2004」

- ・ Unicode

- ・ (新) 常用漢字表（仮称）

などがある。

現在、審議中の「(新) 常用漢字表」のゆくえによって、現在の文字コード (0213:04) や、Unicode にまで、その影響がおよびかねない。

3.JIS X 0213:04 と『和漢朗詠集』

本稿の執筆の時点では、『和漢朗詠集』全体（約1万字）の再検討は、はたしていない（今後、これが、筆者にとっての当面の研究課題の一つである。）ここでは、「コンピュータであつかえない」とした、63 (64) の漢字について、簡略にはあるが、再確認した結果を示しておきたい。0213:04 にある漢字については (3) (4) で、水準を表示した。「U+」とあるのは、Unicode。

01 芎 (4)

02 蕙 (3)

03 莢 (3)

04 遄 (4)

05 𦵏 (4)

06 庾 (3)

07 縹 U+7E3F

08 媿 (4)

09 擎 (3)

10 𦵏 (3)

11 翎 (3)

12 熒 (3)

13 捷 (4)

14 𪔐 (3)

15 𪔐 (4) ※「并」「𪔐」を包摂。

16 飄 U+98BD

17 𪔐 U+5675

18 飡 (4)
19 鄺 (3)
20 嶠 (4)
21 嚶 U+5593
22 𪔐 (3)
23 醅 (4)
24 醕 U+9191
25 屨 (4)
26 𦏧 (3)
27 𪔑 (3)
28 渚 (4)
29 𪔒 (3)
30 漪 (3)
31 𪔓 (3)
32 𪔔 (3)
33 𪔕 (4)
34 𪔖 (4)
35 𪔗 (3)
36 𪔘 U+58C0
37 𪔙 (4)
38 𪔚 (3)
39 𪔛 (4)
40 𪔜 (3)
41 𪔝 (3)
42 𪔞 (3)
43 𪔟 (3)
44 𪔠 (4)
45 𪔡 (4)
46 𪔢 (4)
47 𪔣 (4)
48 𪔤 (4)
49 𪔥 (4)
50 𪔦 (3)
51 𪔧 (3)
52 𪔨 (3)
53 𪔩 (4)
54 𪔪 U+5EB3
55 𪔫 U+8F5D
56 𪔬 (3)
57 涇 (3)

58 𪔭 (4)
59 𪔮 (4)
60 𪔯 (3)
61 𪔰 (3)
62 𪔱 (4)
63 𪔲 「𪔳」 + 「𪔴」
64 𪔵 U+76A4

結果は、以上のごとくであり、『和漢朗詠集』については、JIS X 0213:04 をもちいれば、かなりの文字が記載可能である。また、いくぶんは、この範囲外の文字もある。しかし、それも、Unicode Ext.A の範囲までもちいれば、いくぶんはおぎなえる。

だが、それでも、書けない文字が残る。それは、63番とした、「𪔳+𪔴」の字である。これは、「𪔵」(4)の異体字である。ここも「𪔵」を使用すれば、すべて書けることになる。しかし、本文での使用を見ると、801番の漢詩句で「𪔶嶺」として、固有名詞(地名)で使われている。固有名詞で使用する文字は、軽々に改めるべきではないと判断するならば、この字だけが残ることになる。この字の大漢和番号は31165。

さらにメタレベルでは、次のことが指摘できる。筆者が前稿を執筆したのは、0208規格によっている。今回、同じ内容を0213:04規格で書くとすれば、次の点が変わってくる。0213では、「部首」が単独の文字として入っている。したがて、前稿で「サンズイ」「艸冠」と表記していた箇所が、現在では、「𪔷」「𪔸」などが、使える。無い字はあるとしても、その字がどのような字であるのかの表現がたやすくなっている。

しかし、本文校訂全体として見るならば、文字が増えることは、異体字処理の問題を増やすことでもある。例えば、0208では、「崑」一種類しか文字がなく、やむを得ず統合することになって、その制限があるがゆえに、異体字処理ができた。だが、0213に拡大することによって、「𪔹」(4)が使用でき、「峰」「峯」に類する問題を生ずることになる。

この観点からは、『和漢朗詠集』という古典テキストが、現在の文字コード(0213:04)で、どのようにエンコード可能であるか、ゼロから、本文校訂を再検討しなければならない。

筆者自身が、ここでおこなったごとく、0208 で無かった字が、0213:04 で使えるようになったかどうかだけの調査では不十分である。といよりも、この視点からだけの文字についての調査は、学問的に批判に耐えるものではない。文字コード系が変わるということは、異体字処理をふくめて、すべての文字表記にかんする環境が変わることである。今後、筆者自身の課題として、このことを確認しておきたい。

前稿で指摘した問題点、「【7】字が増えれば解決するか?」は、依然として残された問題なのである。また、最後に記した、「どのような状況を迎えようとも……すくなくとも「この字が無いのは問題だ」と言うことだけにとどまるのは、もう止めにしようではないか」は、今でも同じ思いでいる。特に、現在の文字論の課題である、「新常用漢字」について考えるとき、この視点にたちかえるべきと考える。

4.おわりに

本稿は、蒼惶としてしたためたものであり、前稿の再掲載ということで、大部分の紙数をついやしてしまっている。しかしながら、かつて、0208 の環境で書いた自らの論文を、新しく、0213:04 の環境で、書き直してみることには、やはり、筆者なりには意義があったと感じている。それは、文字について論ずるとき、その論ずる論文を書く環境によって、論の方向、すくなくとも、執筆の有り様が、大きく影響を受けるということを実感できたからである。

このことは、いうまでもなく、古典テキストのみならず、あらゆる文字データ資料について、考えなければならない問題である。写本・版本をデータ入力する場合はもちろん、近代以降に活字として出版されたテキストを、入力する場合にも、同様の問題を考えなければならない。

これは将来において、現在の文献資料（ワープロ作成の文書）を、再デジタル化（文字コードで表現）するとき、オリジナルの資料が、どのような文字環境で書かれたものか、重要な意味を持ってくるはずだからである。

今の時代の「資料」を、「史料」として、我々は、次の世代にアーカイブズとして残していかななければならない。このとき、原資料の執筆環境（文字環境・使用のコンピュータや実装フォント）についても、また、残すべきデ

ータとして、考慮しなければならない。

また、筆者自身として、本稿の執筆を契機として、かつて若かりし時の仕事である『和漢朗詠集』（0208 による）および『神田本白氏文集』（これは、完全に手書き原稿による）などの、古典テキストについて、どのような文字環境でエンコードすれば、どのような本文校訂上の問題が生じるのか、再考察・再調査する必要を、強く感じる次第である。

この新たな課題の第一歩として、本稿をしたためた次第である。

5.付記・謝辞

本稿のテーマ、文字環境の違いによる本文校訂の問題、および、再調査、とういことは、以前より筆者の念頭あったことである。今、あらたに「新常用漢字表」の議論が起こっている。文字の規格を決めるとはどういう行為であるのか、そこでしめされる字の「かたち」とはいったいなんであるのか、この根本的なところから議論を深めていく必要がある。

拙稿につき、発表の機会を与えてくださった、同志社大学文化情報学部の関係者のかたがた、特に、津村宏臣先生には、深く謝意を表するものである。

参考文献

- 當山日出夫（1988）、『和漢朗詠集漢字索引』。勉誠社
當山日出夫（1995）、「古典籍と JIS 漢字一テキストの本文校訂との関係において」、『公開シンポジウム 人文科学とデータベースー「データ」を読む・観る・解くー』
『ユニコード漢字情報辞典』。ユニコード漢字情報辞典編集委員会（編）。三省堂。2000