

現代における漢字の新要素

A New Element of the Chinese Characters at the Present Time

高田時雄

京都大學人文科學研究所

1 外字の問題

コンピュータが普及して以来、漢字をとりまく環境が一変した。コンピュータでは漢字一字一字は共通のコードによって規定されているために、演算によって高速な検索や置換が可能である。以前は、一枚ずつカードをとって、気の遠くなるような時間をかけていたコンコードス（一字索引）のようなものも、ほぼ一瞬にして出来上がる。これは明らかな利点である。

しかし一方、コード化されていない漢字については、利用が非常に制限されるという、きわめて深刻な問題も起こってきた。いわゆる外字の問題である。漢字が手書きされていた時代には、どのような文字であっても時と場所をえらばず自由に書き記すことができた。活字印刷の時代になってからは、印刷所に手持ちの字がない場合に若干の不便があったけれども、必要な活字を鑄造すればそれで済んだ。その工程を支えるシステムも存在していた。つまりコンピュータ以前には、いわゆる外字の問題はほとんど存在しなかったのである。

しかし今は違う。人文系の研究現場において七〇〇〇字に満たないJISコードの漢字だけで仕事ができるような分野はどれほどあるだろうか。少なくとも中国や日本の古い文獻を扱うような研究領域では、はるかに多数の漢字が必要である。コード化された漢字数の不足という問題はきわめて切實であり、緊急にその解決が迫られている。

さて、コンピュータで外字を扱おうとするとき、初期に用いられていた方法は、システムの外字領域にユーザが個別に定義した漢字を登録して使用することであった。これだと自分のシステム内部では外字の同一性を保持し得るけれども、他人のシステムに持ち込むことができない。最近のようにインターネットの利用が一般化してくると、この種のユーザ定義外字では對處不可能であり、誤解の危険もある。したがって現在この方式は急速に廢れつつある。

それに代わって、インターネット上でわりによく見られる方法は、外字を畫像ファイルとして貼り付けるやり方である。これだと發信側と受信側で文

字化けの起こる心配がない。しかしこれでは検索が不可能で、コンピュータの利点がまったく発揮されない。問題は多いのである。

では、どのようにすれば外字問題を理想的に解決できるのでしょうか。

2 大きなコード

簡単に思いつくのは漢字コードの数を増やすことである。現在、三四一六八字からなるいわゆるユニコード（多くの矛盾点を内包していることが指摘されているが、今はそれに觸れない）が J I S 規格として公表されており（J I S X 0221-1995）、それが実用化すれば、かなりの需用が満たされることも事実であろう。しかし、それでもやはり文字数が充分でないことは簡単に想像できる。ある版本ないし寫本を底本にしてテキストデータベースを作ること考えてもらえばよい。そこに出現する漢字の形態は、俗字、誤字を含めてさまざまである。それをそのままの形で忠實にデータベース化しようとするれば、このような漢字数ではまかなえないのは見易い。その場合、必要な字数が最終的にどれほどになるかは見當もつかないほどである。それにもかかわらず、誤字を含めてすべての漢字がコンピュータで扱えるようになってほしいというのは漢字を扱う研究者のすべての願いであろう。異體字はおろか、誤字のたぐいは標準字體に書き換えればよいではないかという議論も當然あり得る。しかしそれではコンピュータを使う限り、そもそもの誤字や異體字に言及することすら出来ないということになる。例えば「的」の誤字として「𠄎」があったとして（敦煌寫本には実際にある）、この文字をそのままに表示できないとすれば、「𠄎の白の部分が自になった字」と説明する以外にないだろう。

ただし現実問題としては、無限数の漢字にコードを割り振るということは不可能であり、學術用でしか必要のない大きな漢字コードを、J I S のように現代社會一般に必要とされる漢字を念頭においたセットと同一延長線上に構想することはあまり効率的とは思えない。もちろん一般のパーソナル・コンピュータで使用できる漢字数が増加することは入出力や検索の手間を考えると、大いに歓迎されることであるが、おそらく學術用の大きな漢字コードの體系は J I S や B i g 5 や G B など各國別のコード體系（ユニコードはこれらを無理矢理に合併したものだから便宜上ここに含める）とは獨立したところで構想されるのが望ましい。それはまた國際的な學術用統一コードとなるべきである。

現在、いわゆる外字問題の解決のために、五萬から八萬字程度のフォントセットがいろいろなかたちで提供されている。日本国内では、京大人文研の e K a n j i や文字鏡研究會の「今昔文字鏡」などがあり、東大明朝というものも公開されつつある。今のところ普通には、これらのフォントは、上に觸れたように畫像として貼り付けるのに用いられている。このようなフォントセットを基礎として、非營利のセンターが外字のサービスを行えるように

なればたいへん都合がよい。サービスは当然インターネット上で行われ、すでに存在する漢字については、世界の研究者はオンライン上でそれをもらってくる事が出来るし、また研究者の要望に基づいて漢字の追加も随時行われる。普及度が高くなれば、そのコード体系が學術用漢字コードとして使用される可能性もある。これは一種、夢のような話でもあるが、決して非現實的でもなく、緊急の課題でもある。

3 漢字の検索と漢字データベース

大きなコード体系が現實のものとなり、それを縦横無盡に取り扱う必要が出来てくると、また別の問題が生じる。つまりある一つの漢字を全體から選び出そうとする場合にどのようにすればよいかということである。また当該の漢字があることは分かっているにもかかわらず入力すればよいか迷うという場合もある。漢字の大海のなかからある一つの漢字を選び出すという点では、なるほど「検索」と「入力」とは同じと言ってよい。この問題は漢字の数が多くなるに従って一層深刻になることが想像される。もしコード番號だけしか頼れるものがないとすれば、求める漢字を探して五萬なり十萬なりの漢字を頭から見ていく以外にない。前のほうにあれば幸いで、運の悪いことに後ろのほうに列んでいたらいくら時間があっても足りない。実際には、部首や畫數、發音といった漢字の屬性が利用できるのも、比較的容易に目的の漢字を探せるようにはなっている。しかしどの部首に屬するかを言い當てるのが難しい漢字はいくらかあって、たいていの漢字字書には畫數で配列した「難檢字表」といったものがついている。今日われわれは「かな漢字變換」という便利なツールで漢字の入力・検索を行っている。これは音・訓という漢字の読み方を利用しているのだが、読み方が分からない漢字はどうすればよieldろうか。異體字や誤字となると専門家でも読み方が分からない漢字は無數にあるに違いない。要するにいくら多くの漢字があっても、他と區別されたものとしてそれを指定できなければ實際に使えることにはならないのである。テキスト・データベースでも異體字を含めて検索するのか、あるいは個々別々に検索するのかは研究目的によって違って来るが、それが自由に出来なければならぬ。

異體字相互の關係だけを解決するためならシソーラスがあればよい。シソーラスとは本來分類語彙表のことで、漢字についていう場合には異體字關係を指示したリストである。これを利用すれば劍・劍・劒・劒などの異體字を時には一まとめにしたり時には區別したりできる。しかし上で述べた検索・入力の困難はシソーラスでは解決できない。そういった場合に不可缺となるのが漢字データベースである。漢字データベースは分類に用いることが可能なすべての漢字屬性を集めたものとなるはずである。既存の各種コードはもちろん、部首や畫數、ピンインによる現代中國音、かな表記の漢音・吳音、そ

その他さまざまな音表記、四角號碼等々、さらに古代の反切や、個々の漢字がどこでどのように用いられているかなどの情報が登録されていればさらにより。異體字関係を明らかにするシソーラスの要素も当然取り込む必要がある。それも時代と地域などによって肌理の細かい指定をしておく必要がある。その漢字データベースを参照することによって、比較的容易に個々の漢字を検索したり、入力したりすることができるようになる。

4 漢字の第四要素

中國では漢字の分類法として昔からさまざまなものが考案されてきた。ほとんどの場合それは漢字の三要素である形・音・義を利用したものであった。そのため歴史上の漢字辭書のスタイルはだいたいこの三つの要素による分類によって決定されている。『爾雅』などの訓詁を集めた書物は義による分類、『切韻』以来の韻書は音による分類、『説文』『玉篇』から『康熙字典』に至る「字書」の系列は形による分類と、大まかには見なしてよいと思われる。これら辭書には見出しとして出された文字に対して幾つかの異體字が掲げられていることが多い。その限りにおいて人は辭書によって正字體と異體字の関係を知ることが出来る。しかしあくまで正字體を基準として配列してあるために異體字の側から検索するのは困難をともなう。一方で『龍龕手鑑』のような部首配列の異體字辭書も存在はするが、佛典という一定の範囲を越えるものではなく普遍性に乏しい。もし上に述べたような漢字データベースに蓄えられた情報を全面的に利用し得るなら、漢字のありようはまったく変わったものになり得ると思われる。

漢字は、その長い歴史の中で、さまざまな関係を作り上げながら発展してきた経緯がある。いくつか簡単な例を挙げてみよう。

「体」は「體」に對應する現代日本通行の常用字體であり、また中國の簡體字體でもある。しかし字書を繙けばわかるように、本来この文字は『廣韻』に「羸、又劣也」とあり、「そまつで見劣りする」意の字であった。發音は反切が「蒲本切」であるのにしたがえば、現代中國音 bèn、漢音ホンとなろう。つまり今日「笨」と書かれる文字にあたる。一方、「体」が「體」の簡略字（あるいは俗字）として用いられるのは、「躰」とならんでとりわけ中國近世（明清時期）では一般的である。このように時代に應じて異體字関係が異なっていることは注意されねばならない。同様に「傘」は日本では「傘」の俗字だが、現代中國では「食」の俗字として（簡體字ではない）用いられている。以前に中國のある都市で「付傘店」という看板を見たことがある。これも前例と同じく異體字関係が必ずしも場合によって一対一でない例である。

さんずいへんの「得」は字書には普通「水也、一曰水貌」（大廣益會玉篇本は「水」とのみいう。ここは『康熙字典』の引用する所に據る。『大漢和』も「1. 水貌、2. 水名」として、「得」に通ずる場合を指摘しない。）とある

が、この文字は敦煌寫本などでは一般に「得」の異體字としてきわめて普通に用いられる。このことを言うのは『正字通』で「今俗以得爲得」と指摘する。もちろん「ぎょうにんべん」と「さんずいへん」が寫本ではほとんど一般的通用関係にあったことが、その背景にある。したがって「復」もまた「復」と書かれることがある。この場合の異體字関係が「得 得」の一方向的な関係で、反対に「得」と書いて「得」を意味することがないのは、「得」の出現頻度がきわめて低いからに過ぎない。同じように敦煌寫本で、「節」を「節」、「第」を「第」、「等」を「等」と書いたりするのは、「たけかんむり」と「くさかんむり」が一般的に混乱しているからである。これも「たけかんむり」で書くべきものを「くさかんむり」にする傾向が強いが、「苦」を「苦」のようにした場合もある。このような寫本時代における一般的通用関係も漢字データベースは包括しなければならない。

このように漢字は同一字形であっても異なった語詞に用いられ、反対に同一文字であっても異なった形に書かれたりすることがある。その関係は歴史的にも地域的にも、あるいはまた文獻の種類によってもさまざまに異なっているのが實際である。この複雑な関係はこれまで學問の世界の中である種の閉ざされた知識として傳承されてきたか、あるいは充分には研究されず、したがってよく知られないままできた。しかし、この漢字相互の關連性が明らかにならないと、現代のコンピュータ社會の中で漢字が有効に機能し得ない。ことばを換えて言えば、漢字という書記體系の中に占める各漢字の位置を特定する必要が、コンピュータ利用によってますます必要となってきた。

傳統的に漢字は形・音・義の三要素から成り立っていると説かれてきた。現代では、それに加えて新たな要素「縁」があらわになってきた、と言ってもよい。「縁」とは異體字関係を含めた形・音・義さまざまの漢字の諸関係のことである。オンライン上で漢字を使用する場合に、入力・検索などの作業はこの「縁」という要素を通じて展開する。「縁」の要素はオフラインでもあり得るかもしれないが、ネットワークにより漢字データベースに繋がった状態（おそらくそれが近い未來に一般化するにちがいないのだが）ではじめて顯在化してくることになる。網の目のように、あるいは蜘蛛の巣のように縦横に張りめぐらされた「縁」の見えない糸によって、漢字は全體の中に然るべき位置を見出さざるを得なくなっているのである。

この駄文は、平成10年度科學研究費基盤研究C（企畫調査）「インターネット上で漢字を利用するための漢字管理機構の制定」（代表者：勝村哲也）の研究分擔者としてその會合に出席し、代表者をはじめ安岡孝一、桶谷猪久雄、クリスティアン・ウィッテアン他の諸氏との討論の過程で啓發されたことに基づいている。