

「絶対漢字」と漢字の「字体」

— デジタルファースト時代の漢字コード問題¹⁾ —

朱 一 星

〈提要〉

本論文は近年筆者が漢字符号理論中の基本概念——漢字の理論値（“漢字単位”）の思考を更に分析討論し、試みとして“絶対漢字”という新術語を以て効果的に記述符号学上の理論漢字概念を説明する。 “絶対”という語、ここでは“概括的”と“抽象的”、且“超越”特定の字体字形を指す。 本文は改めて漢字の“字体（字形）”“符号化”等々をしばしばコンピュータの端末ユーザが漢字を利用するに際しての語彙を、漢字の科学的原理と符号の本質を、同時に統一コードの符号化原則を説明し、その符号化対象は“文字（Characters）”而非“字形（Glyphs）”であることを述べた。 本文はかつて漢字符号化第一線の日本の漢字学者が漢字の世俗観念を保持し、独自の有益な探索をしたこと、主張としてグローバルな漢字の障害なく、優先課題は現代漢字の理論値、即ち“絶対漢字”のグローバルな規範を、国家を越えての“漢字の同期”、而非は純粋な文脈で現れることのみを“視覚方言”の漢字の筆劃の形を訴求する。

〈キーワード〉

ユニコード (Unicode) 符号化原則 絶対漢字 漢字字体 漢字包摂 異体字セレクト

2016年1月より、日本で個人番号(マイナンバー)制度の本格的な実施と共に、各自治体では人名漢字登録に関してある措置が行われた。それは、住民基本台帳ネットシステム統一漢字と戸籍統一漢字など約6万字ある文字情報基盤の漢字を過去の関連告示等をもとに整理し、最終的には約1万字のJIS X 0213 (JIS第1水準～第4水準)文字表への「縮退」である。本論文では、この「縮退」の実質的意義から「一意の漢字符号」とは何かを考え、かねてから問題視してきた「漢字コードのあり方」について見解を述べたい。筆者の漢字符号化に関する見解は現在主流の見方とは必ずしも一致しないが、漢字符号化の問題については、内外の研究者や政策立案者と問題意識を共有できることを切に願うものである。漢字コードに関心のある方が本論文を単なる「思考実験」として読んでいただいても結構である。

一、三種類の漢字「包摂」：ユニコード符号化原則は字体規範意識に踏み込むものではない

経済産業省傘下の、IT国家戦略実現のための独立行政法人である情報処理推進機構が公開した「縮退マップ利用ガイド」²⁾では、「縮退」の理由を次のように説明している。

インターネットを使った情報公開や電子申請など、不特定多数の一般利用者を対象とするサービスでは、利用者の使う多様なデバイスで表示、入力ができるよう、使う文字の範囲を JIS X 0213 の範囲(約1万文字)とするなど、必要最小限に絞ることが求められます。³⁾

また、同「利用ガイド」には、「図形的類似・異体字・誤字・正字・本字等の文字の関係を整理」する際、万が一適切な候補がない場合、「漢字一文字への縮退変換は行わず、読み仮名へ置き換える、あるいは、複数の漢字からなる熟語等へ置き換える」などのように、強引な解決策を容認する印象さえ与える内容であった。

人によって多様な解釈が可能であろうが筆者の考えでは、一連の措置は実質的に、自治体窓口担当者と共に住民が自らの手で行う大規模な人名用漢字の「包摂」にはかならない。

かつて国際漢字符号集の漢字包摂が日本で大きな議論を巻き起こしたと記憶しているだけに、今回の措置も賑やかな議論が湧き起こるかと思いきや、そのような気配はなかった。

漢字の「包摂」(ユニコードでは「統合」とも言う)について、巷では多くの人が漢字の「統一」と混同しているようである。しかしユニコードの漢字「包摂」は、原則的には漢字を使用する国と地域の漢字字体規範(標準漢字)の文字政策に介入するものではない。事実、ユニコードが本来主張している符号化原則はその逆であって、文字の字体に対する地域ごとの標準字体に影響を与えない仕組みである。漢字に対する字体・字形の図形上の値(Glyph)を規定せず、抽象的な符号(Character)のみを対象とするのが、「包摂」という行為の実質的な意義である。

漢字に対する「包摂」は、次の三種類に大別できると思う。

一種類目は、本論文冒頭に触れたマイナンバー制度の実施と共に行った情報処理推進機構の「縮退マップ」による措置である。厳密に言うと、これはユニコードの漢字包摂と一線を画すべきであるが、ユニコードの漢字包摂とまったく同じ方向の発想である。このことには異論を挟む余地がない。

二種類目の「包摂」は、「CJK統合漢字」の形成過程に行った措置である。それは、中国大陸・台湾・日本・韓国の各地域の同一漢字について、公式の規範では字形がわずかに異なる場合、同じ符合位置にそれら僅差異形漢字を置き、互いに「同等」と認めた漢字群のことである。このことを明確に示すため、初期のユニコードには中国語大陸漢字／中国台湾漢字／日本漢字／朝鮮語漢字の四種の例示字体を横一列に並べた。例えば「骨／骨／骨／骨」のようである(のちにベトナム漢字と中国香港漢字も追加)。この方法でユニコードは漢字について特定の図形的規範意識に踏み込むものではないことを表明している。ユニコードの漢字レパートリーは国際通信用に設計されたもので、特定の字体規範を基準にしてしまうとローカルの規格にとどまってしまう。同一漢字に対して字体規範を特定しないことは、ローカル漢字規範の境界線を超えて通信が可能となる条件でもある。そのため、CJK統合漢字の当初の目標は、漢字符号集の「統合」であり、漢字字体の「統合」ではないと筆者は認識している。

三種類目の「包摂」は、いわゆる「互換漢字」と称される領域のことである。これは既存の(ローカルの)文字コードで本来包摂すべき漢字が何らかの事情で重複符号化した場合、それらを収めるために設けた領域を指す。例えば、韓国の国内規格 KS X 1001 では漢字を音配列として認識するため、同一漢

字でも異読のある場合は重複して符号化した(同字異音異コード)。特定の語構成にある漢字を正しいハングル文字に置き換えるため、韓国語ではこれらの異読漢字には何かしらの区別が必要である。しかし、このローカルニーズによる漢字感覚は他の漢字使用国からは理解を得ることは難しい。このズレは「互換漢字」によって吸収され、ユニコードと KS X 1001 との往復変換を保証しながら重複符号化による不都合を防いだ。

これらのことから分かるように、互換漢字が名称通りの「互換」等価でなく、より厳密な完全に同一文字と見なす「正規」等価に指定するユニコード側の認識は正しい。

繰り返すが、漢字「包摂」の究極の意図は、ユニコードの符号化漢字は一意であることを保証するためのものであって、多様な書体、文字規範、字形のデザイン仕様を否定するものではない。これはデジタル情報通信では、データの一貫性が命であるため、同一文字符号が理論上重複しては困るという文字符合化技術の要求と表裏一体である。上記三種類の「包摂」の形態こそ違うが、本質的な目的は同じである。

では、なぜ符号化文字の同一性が必要で、しかも実現可能になるのか。それを理解するためには、われわれの置かれているデジタル情報化という時代背景を理解しなければならない。

二、デジタル文字の本質：「筆跡」という文字の形状値の喪失

日本政府の閣議決議である「世界最先端IT国家創造宣言」を踏まえて政策立案を行う経済産業省商務情報政策局が2017年3月に公表した「情報共有・連携に向けた政府での取り組み状況」⁴⁾の中に、次のような内容が記されている。

情報システム間でのデータ連携等、行政機関内外における行政データの円滑な流通を可能とするよう、データレイアウト、語彙、コード、文字等の標準化・共通化など、行政情報システムに係る規格の整備や相互運用性の確保に向けた取組を行う。(第4 デジタル・ガバメント推進方針)

また、首相官邸の「高度情報通信ネットワーク社会推進戦略本部(IT総合戦略本部)」第22回新戦略推進専門調査会電子行政分科会が公表した「新たな電子行政の方針についての考え方」⁵⁾では、「デジタル技術を徹底活用した利用者中心の行政サービス改革」や「サービスデザイン思考に基づく業務改革(BPR)の推進」などを謳い、「紙から電子へ」、「デジタルファースト」などをキーワードとする方針を公表している。ここから、日増しに増えるデジタルデータを、更なるデジタル技術で活用するデジタル・ガバメントの実現が中軸となっていることが分かる。

メディアもこの動きを見逃していない。日経BP社のサイトに「政府が電子行政の推進方針を大転換する」と題する次のような内容のコラム記事があった。

新方針に基づく行政サービス改革では、マーケティング技術を活用した利用者ニーズの把握、ユーザーインターフェース部分のプロトタイピングやベータ版のリリースなども駆使しな

がら、デジタルファースト(デジタル処理前提)、ワンスオンリー(一度提出された情報は再提出を求めない)、ワンストップの実現を目指す。⁶⁾

このような一連の政策の背景には、デジタル情報化というキーワードがある。そこでまず、デジタル情報化時代の文字の本質について少し考えてみる。

人類が「文字」という文明社会の発明品を手に入れてからつい最近まで、文字情報の伝達においてほとんど変化していないことは、発信者の手(またその延長である印刷機など)によって適切な材質(石、粘土から動植物の皮、金属、紙など)に「痕跡(筆跡)」を残す行為だった。書き手が残した「痕跡(筆跡)」は、基本的に書き手の意図した図形として、半永久的に残る。「字体・字形」として具現する文字の「痕跡」は「非揮発的」であるため、読み手の目に触れる文字は、書き手(また印刷物)の「痕跡(筆跡)」そのものであった。

ところが、デジタル情報化時代においては、それらのプロセスが全く異なる。デジタル文字は最初から目には見えない電磁電子信号である。それを確認するためには端末機器のディスプレイや印刷機に頼らなければならないが、しかもそこに現れる文字図形でさえ「仮の姿」に過ぎない。そもそも書写行為ではないため、当然「筆跡」など存在しない。デジタル機器を使って文字を「書く」ことは、正確には「痕跡を残す」ことではなく、「文字フォント」と呼ばれる電子的文字図形を「選ぶ」ことを意味する。

その文字フォントデータは、印刷して「痕跡」を作らない限り、永遠に「無形」の信号に過ぎない。留意して欲しいのは、ディスプレイに映し出された「文字」は、プリンタで印刷された「文字」と必ずしも同じ形であることを保証しないということである。パソコンの(あるいはプリンタの)性能が向上した今日、その差を意識する人はほとんどいなくなっているとしても、この事実は本質的には変わらない。

デジタル情報の場合、文字の「痕跡」が移動・転送するのではなく、形のないデジタル信号がネットを経由して瞬時に伝達するだけである。そして、世界のあちらこちらのディスプレイ、またはプリンタなどの端末機によって「形」が表示される。電子信号の「文字」に形を与えるのは「文字フォント」であるが、発信者の使用した文字フォントが遠方の端末機器にも備えられているとは限らない。極端な場合、適切な文字フォントがなければ電子信号は何ら「形」をも映し出せない。本大学でも毎年多く見られる学生名簿の人名漢字における表示・印刷障害には、そういった仕組みが背景にある。

この基本的な事実は、われわれの「文字」概念を根底から揺るがすものである。

一言でこれを表わせば、デジタル情報化までの伝統文字には文字の書き手の「筆跡」あるいは印刷文字の「痕跡」が伴うが、デジタル情報の「文字」にはそもそも「痕跡」など存在しないということである。

三、二種類の漢字「字体」：字源学の「字体」とデザイン「字体」

漢字は伝統的に「形・音・義」の三要素を持つと言われてきた。ここでしばらく、その中の「形」について掘り下げて議論したい。伝統文字学(漢字学・字源学)の世界で言う狭義の「字体」の代表例は「異

体」「別体」「繁体」「簡体」などである⁷⁾。この「字体」の定義は、明治書院の『漢字百科大事典』に次のように説明されている。

文字の形態素的な考え方により個別の文字ごとに他と区別される示差的特徴によって規定される⁸⁾

この説明の通りであれば、「字体」は漢字ごとに論ずべき概念で、例えば「體／体」、「學／学／孛」のようなものであるが、実際複数漢字で偏旁が共通の場合（「會／会；繪／絵」）もある。この「字体」概念を本論文では「字体(1)」とする。

長い漢字書写歴史の中で、ある漢字は意味用法によって違う漢字に分離するケースもあるが、逆に何かのきっかけで違う漢字同士が共通意味を持つようになることも起こりうる。多くは篆書から隸書・草書・行書・楷書に変わる過程で、書き手によって多様な字体解釈がなされたことに起因するが、歴史の中で、ほぼ同一概念の漢字に多様な字体が存在するようになった。いわゆる異体字問題として取り上げられるケースのほとんどは、この字体(1)のことである。

一方、同一文字概念のいわゆる「正・通・俗」諸字体の中に、たいてい難しい正字字体と同一読み同一意味の、平易な「俗字体」が同時に存在、流布した。その「俗字体」が、今日の「簡体字」や「簡略字」として標準字体となっている場合もある（例えば「竜／龍」など）。

この字体(1)の特徴は次のように要約できる。

- 1a. 字体(1)は個別漢字の字源学的事象で、同一漢字(偏旁)以外に適用不可である。【排他性】
- 1b. 字体(1)の形状的特徴は、異なる書体にわたって共通する。【書体間連動】

次に、上記字体(1)に対して、ほとんど別の概念と考えられる字体がある。

毛筆で漢字を書く時代に美しいとされた筆の撥ね具合を上手く運用した筆画の「流麗」さは、活字を彫る際には、彫刻刀を制御することが難しいため、活字制作の効率に支障を来す。そこで、印刷業で活字を量産できるようにするために刻字職人が考案したのは、効率的に刻字作業進めることができるそれまででない「字体」であった。それは直線をベースにし、曲がり角に「うろこ」を付けるなどして独特のデザイン要素を取り入れたものであった。漢字は「書く」ものから「彫る」ものになり、制作方法が大きく変化したと言える。

書写文字時代では「篆」「隸」「草」「行」「楷」の何れも「書体」と言うが、印刷文字の「明朝体」「宋体」などのように、単なる「～体」と呼ばれるようになった。ここでいう漢字の「字体」は、各々個別漢字のことでなく、集合名詞的な、体系全体を意識するものである。それは、あらゆる漢字に関わるし、統一感が求められ、デザインの性格が強い概念として理解されるべきである。言い換えると、ここで言う漢字字体は、同一文章内では基本的にはほかの漢字と一致するべきものである。ちょうど特定の書き手の抄写文字

が、文章内で同じ個性を持つ字体になるのと同様である。この種の字体概念は多分に文化的、芸術的感覚と通底するものである。

写真植字時代を経て、今日のデジタルフォントの設計は、伝統的な書風以外にも、中抜きや陰付きなどのアート系字体、子ども風やギャル風の個性化字体など、多様な字面(じづら)が容易に実現可能になった。それらのグリフセットも大抵「字体」と言う。更に、アルファベット文字デザインの影響で「斜体」や「ゴシック体」などの「字体」も加わるが、こちらも同じ概念である。

本論文では、このような字体(1)と異なる概念の字体を以下では字体(2)とする。

全体として、字体(2)はデザイン(書風)の問題であるため、文書の外観的要素が常に優先されることになるが、同一漢字の異なる書体間では読みと意味を共有していることに異論はない。これは異体字がデザイン要素と無関係である字体(1)と決定的に違う。字体(2)は理論上、全文字リストを対象に統一感が求められるのは、重要視すべきがデザイン性であって個別漢字の事象ではないからである。

例えば「曾/曾」に見る「ソ/ハ」、「紙/紙」の糸偏の足、「父/父」の筆押さえなどの字形差異は大抵この字体(2)に属するものである。「辻/辻」問題を取り上げるならば、その他の「厶/厶」との整合性も度外視できない。要するに字体(2)の多くが書体・書風に付随する点を見逃すことはできない。

要約すれば、字体(1)に対し、字体(2)の特質は次の通りになる。

- 2a. 字体(2)はデザイン要素であり、他の漢字と風合いが共通する傾向が強い。【デザイン調和性】
- 2b. 字体(2)の形状的特徴は、他書体と必ずしも共通しない。【書体間独立】

字体(1)と字体(2)の特質を観察すると気付くが、本来字体(2)は2bの性質にある通り、必ずしもほかの書体にまで広がる発想がないのに対して、字体(1)概念との混同で、あるいは文字デザイナーの方針によって、1bのように他書体に波及することもある。一旦そのようなことが起こると、今度は2aのデザイン統一意識によって瞬間に別漢字にまで波及し、その別漢字において更にこの一連の現象を繰り返す……。このように字体問題はまるでハウリングのように自己増殖し、さまざまな「字体」を副次的に産出することがあり得る。

字体(1)のバリエーション(異体字)は、通常「正字・俗字」、現代社会では標準漢字や規格外漢字などの話になる。一方、字体(2)には、この「標準字体」の考え方はない。「字体問題」を指す時も、基本的に字体(1)を念頭に置くべきことが多い。

例えば、明朝体が標準で隷書が俗体だ、などと考えるのはナンセンスである。漢字の「正誤」意識は本来字体(1)の個別漢字ごとの問題であるにもかかわらず、字体(2)との混同などで、単純なデザイン要素の差異に過剰に反応することで、歪んだ正誤判断が増幅し、蔓延することになる。

明朝体漢字だけを対象に、上記の二種類の「字体」を区別せずに議論する人が少なくないが、字体(1)と字体(2)を本質的に異なるものと捉えていない可能性が高い。

ここで重要なことを明確に提示する必要がある。それは、字体(1)も字体(2)も「同一漢字」という前提に基づいているということである。字体(2)は単なるデザイン差であるのはもちろんのこと、字体(1)も同一抽

象概念のバリエーション(異体字)である。単純な話だが、これらのさまざまな「視覚的」字体差は、音声言語に置き換えると「訛り」や「方言」に相当するものと言える。この観点から見れば、現在の地域ごとの「標準字体」も、考え方によっては漢字概念の「視覚方言」と言える。

更に、この問題を複雑化する別の混同もある。それは「字体」概念を、例えば「学/字/宇」や「刀/刃」のような文字間の差異にまで適用する考え方である⁹⁾、さすがにこの発想に第三の「字体」を設けるには躊躇する。いわゆる「字体問題」は、読みと意味が同じである漢字の異形態問題で、「学/字/宇」も「刀/刃」はすでに読みと意味の異なる別「字」であり、「体」は蛇足である。このように「字体」概念と「字」の概念を重ねるのは、後述のように「漢字概念」同士の差異弁別に外形上の要素を引き込むという理論上の不具合を生じさせることになる。

それでは、記号理論から見る文字コードや漢字符号化はどういったものであろう。

四、漢字体系の記号論的最小単位は「絶対漢字」である

デジタル文字の概念が文字歴史の革命であると認識した上で、漢字そのものを改めて考えてみよう。

前述の通り、漢字をより正しく理解するためには、漢字の文字体系の側面から深く理解しなければならない。文字学(伝統漢字字源学)の研究歴史は長いが、近代言語学の記号理論からアプローチする研究は未だに学問分野として成り立っていない。研究者のなかでも漢字の正しい科学的な姿に対する誤認は根強い。筆者はデジタル符号集における漢字のあり方について、近年一貫して次の立場に立脚している。すなわち、あらゆる文字がそうであるように、漢字も自律的に存在する記号体系である。

言語記号の場合、個人の(具体的な)パロールに対して、同じ言語を話す社会が共通項としてのラングという考え方がある。同様に、個人の筆跡を伴う「漢字」は具体的な「形」のある文字であるが、社会共通項としての「漢字概念」は、個人の字体字形意識に左右されている文字が本来持つコミュニケーション機能を果たすことができなくなる。ここで注意していただきたいのは、抽象的な「漢字概念」は、「形がない」とか「形が自由」などのように理解されては困るということである。デジタル漢字(符号化漢字)は、一種の記号と見なす場合、唯一で絶対的な「形状値」が存在しないということは、さまざまな字体バリエーションが存在することを意味する。記号は理論的概念で、具体的、あるいは明示的に提示できない。一方、記号としての「漢字概念」に対して、日項目にする書籍や日常生活の文字コミュニケーション現場の(筆跡やインク痕跡などの)漢字は、個々の書き手による特定の、視覚的に捉えることができる「漢字表現」である。「漢字表現」はなにかしらの「字体」といわれる「形状値」を持つことになる。

同じ漢字概念の様々な「現象」としての漢字表現が実質的に同じ記号であることを明確にする努力は、漢字記号同一性問題の研究である。

未だに多く流布している誤解は、「漢字概念」を大まかな「輪郭」や「骨組み」などの「字体」に求める考えである。しかしながら、「骨組み」が酷似している別漢字もあれば、「骨組み」が大きくかけ離れている同一漢字もある。漢字の同定は決して統計学的相似性の問題ではない。単純に「形」に焦点を合わせると、中国本土の簡体字である「芸(藝)」、「沪(滬)」などと日本の簡略字体である「芸(藝)」、「沪(滬)」などを安易に同一視する誤謬につながる。これは例えば数学記号の「1」とアルファベット文字の「l」と

同一視したり、長音記号「ー」とダッシュの「-」と漢数字の「一」と同一視するのと同じように、文字の外形に捕らわれるがゆえに起こる同定エラーである。

漢字をコミュニケーションツールとする漢字圏においては、ことは単純な異体字問題に片付けられない。日本漢字、中国大陸側の簡体字、台湾・香港で使う繁体字¹⁰⁾を常に接する筆者のようなものにとつて、漢字を異なる字体の背後にある抽象概念として判断する感覚が要求される。

そして、漢字の同一性問題は、単なる漢字知識の問題でなく、漢字を記号体系として捉え、デジタル符号化する際の文字コードの理論問題でもある。

記号論では、「重複のない」記号を「差異体系」と言うが、この場合の「差異」も見た目の違いという意味合いではない。丸山圭三郎がその原理を次のように的確に説明している。

私たちは、この関係的同一性と関係の差異に操作されて行動し、時にはこれを実体的同一性、差異と混同してしまう。(中略)「何からできているかを知るよりは何において異なるかを知る」ことの重要性の指摘は、「関係の世界において意味をもつものは差異だけである」というソシュールの根本的認識から出ているのです。

したがって、関係的存在というものは「ではない」という否定的要素によってしか定義できず、「である」という実定的な要素によって規定できるものではありません¹¹⁾

記号論的最小単位は体系内では「唯一」であり、「ほか」のなにもものでもない。このことは、記号論の基本原則であることが一般に知られている通りである。そして、「記号」はいかなる物質的、素材的な要素に左右されないうえ、「記号」の外観(「実体」)とも無関係である。

ラングというのは信号体系である。(中略)信号の素材そのものは非関与的なものと見なすことができよう。

(中略)

そもそも同一性というものは二つの視点から決定されるのであって、一つは物理的・生理的な次元における実体的な同一性。もう一つは関係的・構造的同一性です。私たちが物理的に生理的に同じだと思っているものも、ある一つの体系、関係の中ではまったく異なることもあれば、その逆の場合もあります。¹²⁾

記号とは精神的なものであり、概念的なものである。それはわれわれの視覚・聴覚で感じ取れるような具象的なものそのものではない。このことは、われわれが「漢字概念」の抽象性を論じる時もさほど変わらない。さまざまなシーンでわれわれが目にし、他人に「これだ」と示せるものは、永遠に例示的であり、刹那的な「仮の姿」に過ぎない。理論漢字に関するこの根源的な認識はまだほとんど根付いていないため、たとえ符号漢字を「抽象概念」として認識する一部の学者でも、結局この「抽象概念」を大まかな「骨組み」という「形」から解き放していないのが現状である。

デジタル情報通信における漢字符号を考える際、筆者は繁体字・簡体字・異体字・俗字などのすべてを含む概念として「漢字単位」という表現を提唱したが¹³⁾、本論文ではこの概念を文字論の、あるいは記号論の立場から再度提起し、「絶対漢字」と表現してみることにする。記号体系としての漢字の「抽象概念」も、現代漢字理論のイーミックの「漢字単位」も、今後、この「絶対漢字」という用語で説明していくつもりである。

筆者は、更に漢字を記号論的立場から検証し、「漢字概念」、すなわち「絶対漢字」は、「絶対音価」と「絶対形状値」を持たない記号であり、それは視覚的に確認できる全ての異体異形を包摂する共時的規定的抽象的文字単位だと考えるようにしている。「絶対漢字」は、図形的には明確に提示できないながらも、「視覚的」文字単位である。

これは一見「非常識」な解釈であるようだが、しかしちょうど音韻論という音素が音声学的に絶対的でないながらも「聴覚的」言語記号であるのと同じく、記号論的事実である。

デジタル情報化時代にこの事実を確認する意義は極めて大きい。なぜなら、「絶対漢字」に基づく漢字の解釈は、文字コードにおいて符号化漢字の考え方に直接関わってくるからである。すなわち、ユニコードにおいて、あるいはまさに後述のような符号化原則に適用すれば、符号化の対象は漢字の「字体」ではなく「絶対漢字」そのものだからである。

五、デジタル情報化時代の主役「絶対漢字」は、「符号」であり「字形」ではない

従来の文字生活では、抽象概念としての「絶対漢字」は実感の伴う「形」で提示できないため理解されにくいものであった。ところが、デジタル情報化時代における文字(漢字)符号化の対象はほかでなく、抽象概念としての文字(漢字)でなければならない。

電磁電気信号としての文字は、人間ではなくコンピュータが扱う以上、人間のように融通は利かない。デジタル文字符号の重複符号化は情報通信にとって、もっとも避けなければならない忌諱とされる。重複符号化の文字は、データ通信において情報の精度を損なう。

ユニコード規格書第二章「General Structure」では設計原則についての説明のなか、次のような原則がある。

Characters, not glyphs : The Unicode Standard encodes characters, not glyphs.¹⁴⁾(ユニコードの符号化する対象はキャラクターであり、グリフではない)

この基本原則がなぜかあまり広く語られていないような気がするのは筆者だけであろうか。

ケン・ランデー(Ken Lunde)は『CJKV日中韓越情報処理』の中で、規格書に対する世間の誤解について次のように語る。

CJKVの文字集合規格は、それらの規格書に記載されている文字の字形を定義しているわけではない。残念なことに(あるいは場合によっては幸運なことに)、多くの人がこれらの規格書に出ている字形を正式なものと誤解している。¹⁵⁾

ケン・ランデーは文字コードについて概ね正しい知識を持っているが、日本の「常用漢字表」に対する説明には疑問を感じざるを得ない部分もある。

しかしながら、公式の常用漢字表が、すくなくともこれに含まれる1,945文字について字形を定めていることにも注意してほしい。¹⁶⁾

ここで、重要なポイントを確認する必要がある。まず「常用漢字表」の漢字はあくまで「例示字体」であって、決して「字形を定めている」ものではないという点である。次に、漢字のローカル規格の「字形」(例示字体)をもって、ユニコードの符号化原則を変えられるかどうかという点である。これはちょうど中国の漢字標準字体をもってユニコードの符号化原則を変えられないのと同じことである。

検討すべき点は、ほかにもある。

(Unidodeの漢字は)厳密且つ精確な規則に従ってこれらの文字を統廃合した結果、重複のない20902文字となった。¹⁷⁾

この判断は事実ではない。後述のようにユニコードの漢字統合規則は厳密でも正確でもなく、「符号(Characters)」としての「漢字」、すなわち本論文で言う抽象概念の漢字(「絶対漢字」)のかなりの部分が重複符号化されてしまったことは明らかな事実である。このことについてケン・ランデー自身の著書の中で「統合できそうな漢字」として「劍・劍・劔・劔・劔・劔」の例¹⁸⁾を挙げている。そして次の記述によっても伝えられている。

残念ながら、初期の文字集合規格では字形差の符号化モデルに矛盾があり、結果として、…異体字を別々に符号化してしまった。¹⁹⁾

という具合である。

本論文の立場からすれば、「異体字を別々に符号化」することは「字体」を「符号(Characters)」と同列に扱うことになり、ユニコードの原則から離脱していることになる。これは抽象概念の漢字概念、すなわち「絶対漢字」に「形」を持ち込むのと同じことで、「字形差の符号化モデルに矛盾があり」、欠陥を抱えたものになる。

この問題は漢字側からも解明できる。絶対漢字の本質がその記号論的抽象性にあることが理解できれば、字体(1)であれ字体(2)であれ、漢字の異形同字であると捉えるべきで、それを理論上の漢字概念と同列に扱うことはできない。その字体差異をコードポイントに押し込む時点で、ユニコードの信頼性が損なわれていることは理解されるべきである。

ここまでの話を踏まえた上で改めてCJK統合漢字の統合規則を見ると、その統合規則の混乱ぶりが容易に見えてくる。実は、ユニコードの符号化原則に合致する漢字統合規則は、以下に挙げただけ一つの規則だけである。

R2 Noncognate Rule. In general, if two ideographs are unrelated in historical derivation (noncognate characters), then they are not unified. (非同一視ルール。一般的に、歴史的由来から異なるとみなされる二つの漢字(非同系文字)は、統合しない。)²⁰⁾

このR2以外の統合規則については、規則として網羅的かどうかという疑問を乗り越えて、抽象的な「符号」に対して外形上の要素で統合(非統合)すること自体が符号の抽象性ゆえ不適切であるということになるが、このことは残念ながら広く認識されていない。この辺の詳細な解説は他の識者か、今後の機会に別稿で行うこととしたい。

ここで特筆すべきは、現在 R1 とされる原規格分離規則 (Source Separation Rule) がほかの重要統合規則と矛盾することは当初から専門家の間で知られていたという事実である²¹⁾。その規則の存在だけで漢字統合規則を事実上「無規則」にしてしまっている。なぜなら、ローカル規格(原規格)で重複符号化された漢字(R1の対象漢字、例えば「戸/戸/戸」など二百組近く)は、本来、前述の「第三の包摂」(互換文字)の対象文字群に入れるべきだと考えるのが順当だからである。

詳細な解説には更に時間と労力をかけるべきであろうが、ここまでの説明だけでも、複数の深刻な問題が重なった結果、ユニコードの漢文字符号集合は、字形上の差異で重複符号化漢字をプレーンテキストに押し込み、デジタル漢字の情報通信や情報交換に不確定要素を与えたことが事実である。

ユニコードの符号化原理の存在意義について、トニー・グラハムが次のように述べている。

Unicode標準の文字レパートリを“軸(pivot)”として使い、ある符号化文字集合方式をUnicodeに変換し、次にUnicodeから目的の符号化文字集合に変換するという方法で2つの符号化方式間の変換を行うことができます。Unicodeを軸として使うのは、多数の符号化文字集合の間で変換が必要な場合に有効です。²²⁾

「キャラクターであり、グリフではない」という原則は、文字理論に置き換えて言えば、「軸」となる国際漢文字符号集合の符号化対象が漢字記号の離散的最小単位でなければならないことを意味する。

情報化時代の漢文字符号化集合は、科学的合理的に設計すれば、漢字の記号論的働きを正しく表現できる強力な道具となるものである。言い換えれば、文字のデジタル化は、多数の言語圏や地域で使用できる漢字の合理的要素を引き出す理想的なプラットフォームを提供できたはずである。しかし、現行ユニコードの漢文字符号化の設計モデルは、そのような理想の実現にはほど遠い。

次に、文字コードの関係者の中でこれまでに出了意見を拾ってみよう。

六、漢字の「符号」本質に気付いた人たち

日本のメーカー社員としてユニコードの会議に参加し、ユニコードの発足からの流れを見ていた小林龍生が1998年に一度書き、十三年後の『ユニコード戦記』に再度取り上げた論述は、貴重な意見として注目に値する。それはデジタル漢字の抽象性についての次のくだりである。

この翻刻を頼りに原文を眺めていて、おもしろいことに気がついた。芭蕉が書いた文と翻刻に用いられた文字の字形が、必ずしも一致しているわけではない。例えば、冒頭の「海浜にさすらえて」という個所。この「浜」の文字を芭蕉は「濱」に近い字形で(毛筆とこの書籍で用いられる明朝体の字形の相違はあろうが)書いており、翻刻は「浜」の字を用いている。思い立って、手元にあった角川書店版の松尾芭蕉全集(昭和三七年初版、筆者の手元にあるのは、昭和四三年の再版)を繕いてみると、案の定「濱」の文字が用いられている。一方、国の字は、芭蕉の自筆は「国」、角川本は「國」、自筆本の翻刻は「国」の字を用いている。²³⁾

小林は漢字の字体について、実際に古文書を翻刻印刷するプロの編集者たちの間、異体字同士の同一性はきちんと成立しているということを感じて理解していた。つまり漢字の記号論的性質に気付いていたのである。しかし、小林には同時に、漢字に対する日本の社会的習慣をも重んじていた一面もあり、別の場面では次のような矛盾を感じさせる内容を述べている。

このような〈辺〉の異体字の使い分け需要は日本にしかない。他の漢字圏諸国・諸地域にとって、不要な統合分離は実務的にも迷惑千万なことだ。一方で、『高麗大蔵経』の版本に含まれる大量の異体字が韓国から提案されてきたことがあった。仏典や印刷の歴史の研究者たちにとっては、たとえば使われている異体字の相違は、時代や版本の系統を調べるうえで重要な要素となるだろう。実際の研究は原本や原本の写真による複製によって行うにしても、論文執筆の際には電子テキスト上で異体字の使い分けができるに越したことはない。しかし、これもまた日本で戸籍業務を司る役所の担当者にとっては迷惑千万なことだ。このように、要求される異体字の粒度は地域や目的によって千差万別で、国際的な合意が得られるわけがない。²⁴⁾

漢字記号が、歴史上すべての時代のすべての記号表現を、同じシステムで実現することは理論上ありえない。これは記号システムの共時性原理である。歴史上の漢字はすべて同じシステムで処理できると考えるのは、漢字について、長い歴史の中でほとんど字源学の研究しか展開されず、記号論的に体系化されていないために生まれた楽観的な希望に過ぎない。

漢字字体問題が複雑なため、漢字同一性問題を「粒度」として解釈する研究者がいる。この「粒度」説は、分割不可能な「絶対漢字」を用途に合わせて細分化できる「単位」と見なしている。この方法は漢字を離散型符号としてではなく、符号化の対象となる文字の境界線が極めて曖昧な連続型データに仕立ててしまう。これは明らかに現代漢字の本質を認識していなかった時代の、矛盾を孕んだ妥協的な表現である。ところが、多くの研究者や文字コード技術者がこの認識に迷い込んだ結果、漢字の記号論的同一性の追及という根源的問題に踏み込めずに今日に至っている。

そんな迷いの中でも、漢字の抽象性を無視し、無闇に異体字を符号化しようとする考えを「字体至上主義」だと言い切る小林龍生の議論は、漢字の「符号」としての本質を捉えようとした努力が伝わって

る。まさにこの努力の結果として小林龍生と樋浦秀樹が提案したIVS(異体字セレクト)技術が生まれたことは一般に知られている話である。²⁵⁾

かつてJIS委員会の重要な存在だった野村雅昭も、ある意味で、難問に挑戦した一人であった。彼は漢字コード問題について、『漢字の未来』の中で次のように書いている。

ユーザーのなかには、自分の必要とする字が、この規格のなかにはないという不満がではじめた。たとえば、自分の姓は「辻」であるが、ワープロでは「辻」と印字されてしまうというようなものである。実は、これは一部のメーカーやユーザーによる誤解であった。もともと、この規格は、文字概念とコードとの対応関係を定義したものであり、字体や字形をさだめてはいない。「祇園」の「祇」を「祇」とデザインしても、その逆であっても、いっこうにさしつかえない。ただし、情報交換では、両者を同じコードとしてあつかうというものであった。²⁶⁾

あるメーカーでは、「辺」の旧字体の「邊」の異体字を百種類以上も用意しているという。それらをすべて満足させなければならないとしたら、何万字あっても、たりないはずである。²⁷⁾

野村雅昭は80年代のJIS規格に深く関わった人物だが、「世間」は彼が任期中に行ったJIS変更を否定的にしか評価していないようである。いわゆる83JISの非互換の変更問題は、結果的に多くの矛盾を抱え、使用者の困惑を招くものとなった。しかし、字体の変更は単なる「例示字体」の変更であり、本来理論的な発想では「符号」としての漢字に影響するものではないはずである。印刷ニーズとしての字体要求は本来文字コード(プレーンテキスト)の問題ではなく、マークアップ言語やリッチテキスト形式のような上位層プロトコル、またはフォント次元の実現課題である。すなわち野村の異体字に対する認識には科学的、かつ合理的な要素があったと言える。私見では、あの時の野村雅昭の失敗は、彼が正しい認識を大勢に受け入れてもらえなかったために中途半端な結果になったことぐらいである。

1990年代当時はおろか、今日の漢字圏においても、文字研究界並びに漢字符号化に関わる人びとが果たして漢字問題の本質を正確に捉えているかどうかは甚だ疑問である。更に言うならば、国際符号化漢字集合について考えるにあたり、現存する複数の漢字規格同士の漢字理論値の非対応(1=n問題)も議論すべきであろう²⁸⁾。すなわち、「絶対漢字」が符号化となる場合、漢字圏共通の「絶対漢字」表の建設がその前提となる。

このような展望については、今後も議論が行われてゆくであろう。漢字の多言語共有を前提とする情報通信の環境にも耐えうる符号化の設計には、「絶対漢字」の同期が不可欠であり、この漢字圏共同事業は今後注目されるであろうと筆者は考える。

七、むすび

本論文では、デジタルファーストの情報化時代における文字そのものの革命に伴い、漢字の文字概念はデジタル情報化の世界でどのように振る舞うかを説明し、不十分ながら、デジタル情報化時代の漢字符号の根源的な問題提起をした。

漢字の文字概念としての抽象性は、一部の学者が理解するような「骨組み」ではなく、幾何学的形状が関与しない記号論的「絶対漢字」として捉えなければならない。この基本的理解がなければ、漢字コード問題の混乱から抜け出すことは困難であろう。

難局を乗り越えるためには、もはやローカルな漢字規範意識だけで対応できない時代にわれわれが入ったのだという認識を社会と共有しなければならない。本論文で繰り返し説明しようとする真実は、漢字の符号化にあたり、ユニコードの原理から言えば、符号化の対象である符号(Characters)は抽象概念の「絶対漢字」であり、「字体」ではないということである。目標実現に先んじなければならないのは、漢字理論値の「同期」(1=n問題の解消)である。その上で初めて、IVS(異体字セレクタ)技術の真価が発揮されるのだということに関係者に理解していただきたい。漢字は文化を担いながらも、これからの時代では、更にテクノロジーとしての側面がますます重要視されよう。漢字記号論の先に、漢字工学という分野の誕生も決して現実離れた話ではないはずである。

漢字という四角い東洋文字は、特定の言語意味にアクセスしながら、言語圏ごとにそれぞれの字体と字音を付与することのできる技術である。近年の発掘調査と先端技術による分析調査を精力的に取材し、漢字誕生の深層を伝えたNHKスペシャル番組では、阿辻哲次が漢字のこの特質を捉えて次のように述べた。

ある一つの漢字を殷の人がどう読もうが周の人がどう読もうが、その漢字が表している意味はそれぞれの話言葉、方言にあるので、音声で話し合っ通じなくても、文字を見たら意味をお互いに理解できる。話し言葉が違っていても、漢字の意味を理解できるという特徴があつて、殷から周は、その一番早い例ではなかったかと思います。²⁹⁾

これが「絶対漢字」の古代からの真の姿と働きである。古代漢字の仕様を今日の言語事情に置き換えると、中国語、日本語、韓国語のいずれの言語で使用されても、ほとんどの漢字が表している意味はそれぞれの話し言葉、言語にあるということである。漢字という文字技術の未来は漢字民族の手に委ねられている。目指すべきは漢字字体の多様さを認めながらの、「絶対漢字」概念の確立とその完全同期であると信じる。

先人の知恵はわれわれにそう教えてくれている。

注

- 1) 本論文は筆者が2011年から漢字記号や漢字コード問題について書いてきた一連の考察のワンステップである。その間、小林龍生、安岡孝一ほかより貴重なご教示をいただき、この場で謝意を表わしたい。本論文に関するご意見は、京都外国語大学国際言語平和研究所までお寄せ下さい。
- 2) 独立行政法人情報処理推進機構(IPA)公開サイト(<http://mojikiban.ipa.go.jp/4141.html>)を参照。
- 3) 「縮退マップ利用ガイド」(上記IPA 公開サイト、2015年3月31日) p.1。

- 4) 情報共有基盤推進委員会2016年度第5回情報共有基盤推進委員会公開資料, 独立行政法人情報処理推進機構公開サイト(<http://opendata.ipa.go.jp/node/20170302/index.html>) を参照。
- 5) 「第22回新戦略推進専門調査会電子行政分科会」配付資料。首相官邸ホームページ(http://www.kantei.go.jp/jp/singi/it2/senmon_bunka/densi/dai22/gijisidai.html)
- 6) 井出一仁(2017)。
- 7) 明治書院『漢字百科大事典』, p.21, p.27, p.28 を参照。
- 8) 明治書院『漢字百科大事典』, p.13。
- 9) 文化庁『常用漢字表の字体・字形に関する指針 文化審議会国語分科会報告(2016年2月29日)』 p.7を参照。
- 10) 「繁体字」は本来中国本土の簡略字体の旧字形を指すが、ここでは台湾香港などの標準漢字を指す一般的な習慣に従った。
- 11) 丸山圭三郎(2013), p.55。
- 12) 丸山圭三郎(2013), pp.53-54。
- 13) 朱一星(2011) を参照。
- 14) 2.2 Unicode Design Principles, *The Unicode® Standard Version 9.0.0*, p.14。
- 15) ケン・ランデー(2002), p.17。
- 16) ケン・ランデー(2002), pp.17-18。
- 17) ケン・ランデー(2002), p.121。
- 18) ケン・ランデー(2002), p.123。
- 19) ケン・ランデー(2002), p.122。
- 20) 18.1 Han より。(*The Unicode® Standard Version 9.0.0*, p.680)。
- 21) 小池建夫(1993), 宮澤彰(1992), 松岡栄志(2010) を参照。
- 22) トニー・グラハム(2001), p.106。
- 23) 小林龍生(1998), p.218, 小林龍生(2011), pp.94-95。
- 24) 小林龍生(2011), p.224。
- 25) IVS(異体字セレクタ)については, 小形克宏(2009) に参照されたい。より正確な情報はUnicode協会 HP(Unicode Ideographic Variation Database <http://www.unicode.org/reports/index.html>)にある。
- 26) 野村雅昭(1988), p.209。
- 27) 野村雅昭(1988), p.210。
- 28) 1=n問題とは, CJK漢字の中, 伝統的に別漢字(非異体字関係)の記号論的対立を解消(中和)する規範措置を指す。広義的にはCJK漢字の中で異体字関係認定の異同も含む。日本常用漢字の弁[=辯/辨/瓣/(辦)], 台[=台/臺]; 中国本土規範漢字の机[=机/機], 后[=后/後]など, 数百組あると考えられる。
- 29) NHK番組「漢字誕生 王朝交代の秘密」(NHKスペシャル「中国文明の謎」第2集 2012年11月11日 初回放送)。

参考文献

海保博之 『漢字を科学する』(有斐閣選書(892)), 有斐閣 1984年11月

橋本萬太郎・鈴木孝夫・山田尚勇編著 『漢字民族の決断-漢字の未来に向けて』, 大修館書店 1987年6月

宮澤彰 東アジア地域の漢字コードの統一をめざして, (『しにか』, 3(2), 大修館書店 1992年2月)

- 小池建夫 統合化 CJK 漢字集合の実際, (『しにか』, 4 (2), 大修館書店 1993年2月)
- 佐藤喜代治 ほか 『漢字百科大事典』, 明治書院 1996年1月
- 野村雅昭 『漢字の未来』, 筑摩書房 1988年7月
- 野村雅昭 『漢字の未来 新版』, 三元社 2008年4月
- 小林龍生 要求する側の責任ということについて, (『電腦文化と漢字のゆくえ』, 平凡社 1998年1月)
- 小林龍生 『ユニコード戦記 —文字符号の国際標準化バトル』, 東京電機大学出版局 2011年6月
- 豊島正之 JIS漢字批判の基礎知識, (「日本語の文字と組版を考える会」第8回講演資料 1998年3月8日)
- 豊島正之 「原文に忠実な翻刻」をめぐる, (日本語の文字と組版を考える会第8回講演資料 1998年3月8日)
- 小池和夫・府川充男・直井靖・永瀬唯 『漢字問題と文字コード』, 太田出版 1999年9月
- トニー・グラハム (乾和志、海老塚徹訳、関口正裕監修) 『Unicode標準入門』, SHOEISHA 2001年5月
- ケン・ランデー (小松章、逆井克己訳) 『CJKV日中韓越情報処理』, O'REILLY 2002年12月
- 小形克宏 印刷文字から符号化文字へ, (張秀民ほか著『活字印刷の文化史』, 勉誠出版 2009年5月)
- 松岡栄志 『漢字・七つの物語 中国の文字改革一〇〇年』, 三省堂 2010年9月
- 朱一星 漢字の国際コード規格をどう考えるべきか —漢字コード規格の根拠となる「国際漢字単位」の提言,
(京都外国語大学『研究論叢』, No.76, 2011年1月)
- 朱一星 如何界定漢字的理論單位, (京都外国語大学『研究論叢』 No.81, 2013年7月)
- 朱一星 論證“和諧體”的理論優勢 科學規劃數碼漢字, (香港中國語文學會『語文建設通訊』, 第110期,
2015年12月)
- 朱一星 關於統一碼 (Unicode) 漢字編碼框架的理論問題, (京都外国語大学『研究論叢』, No.86, 2016年1月)
- NHK「中国文明の謎」取材班 『中夏文明の誕生』, 講談社 2012年12月
- 丸山圭三郎 ソシユールを読む (『丸山圭三郎著作集 第II巻』, 岩波書店, 2013年12月)
- 榎並利博 電子行政における使用漢字の問題, (『情報管理』57 (5), 2014年8月)
- 文化庁 『常用漢字表の字体・字形に関する指針 文化審議会国語分科会報告 (2016年2月29日)』, 三省堂
2017年1月
- The Unicode Consortium *The Unicode® Standard Version 9.0.0 – Core Specification*, 2016年7月
- 井出一仁 政府が電子行政の推進方針を大転換, (『ITpro』, 日経BP社 2017年2月23日)