

月刊

SINICA

しにか

特集●古典とコンピュータ

2

1992

VOL.3/NO.2

平成2年6月16日
第3種郵便物認可
1992年2月1日発行
(毎月1日発行)

第3巻 第2号 通巻23号

大修館書店

月刊

しにか

SINICA

「JISに無い字」をめぐって

豊島正之

【文中〈文字観念〉等は、JIS規格票・付属解説からの引用です】

「JISに無い」わけ

ようやく文献学的研究にも計算機が文房具並みに用いられる時代になり、研究対象テキストの計算機可読ファイル化もあちこちで行われているので、「……の字がJISに無い」というグチを聞く機会も増えた。JIS 78からJIS 83への改訂に伴う異体字入れ替え、規格票の印刷字形の変更(野村雅昭、一九八四)の影響で、「鶯」のつもりがプリントしたら「鶯」になった、「鷗」が「鷗」に化けた、等という経験をお持ちの方も多いだろう。

「JISに無い」というのにもいろいろあって、
1 JISが同一視する字を区別したい。

青と青、葛と葛、祇と祇

この類は、JISは同一字体(〈文字観念〉)の字形(図

形としての文字実現)上の変異と見なし、〈字形の違いがわずかな〈同値〉の文字として、同じコードを与えている。

2 JIS内に異体字はあるが、別の異体字が使いたい。

幫 に対する 幫 (三十六字母の一つ)

減 に対する 減 (図書寮本名義抄引干禄字書に俗字)

嘩 に対する 哱

3 異体字すらJIS内に無い。

哈 (切韻の韻目の一つ)、疇 (人名)

JISの立場から言えば、32区36の文字実現(印刷字形)は「青」でも「青」でもよいので、そのような字形上の小差を包摂した「抽象的な文字観念」に対して32区36というコード(コードポイント)を与えているだけであ

る。「青」は「JISに無い」というより、JIS内で「青」と「区別されない」だけである。

「JISに無い」字に関しての議論の混乱の源は、その字がJISで他の字体と同一視されているだけなのか、それとも(区別はしているのだが)表内には登録されていないのか、規格自体からは判然としない点にある。上記の2の異体字関係が典型で、表内に無い異体字は、JISが「区別していない」のか「登録していない」のか、直接的には知るすが無い(1の〈同値〉の挙例にしても規格票「解説」の例で、規格本文には〈同値〉への言及は無い)。

もともと、文字表(コード表)とは本来そういうものなので、アルファベットの古い国際規格ISO 646でも、アクセントの有無(a対ă)、テイルの有無(n対ñ)等で異なる文字は、規格が「区別しない」のか、単に「登録しない」のか不明であった。「区別」の有無は、ISO 6937(二重重ね合わせとして)、ISO 8859(一字として)のように、これらを網羅的に登録する規格が登場するまで判明しない。JISも「補助漢字」の登場によって、上記「幫」(28区29)や「減」(19区02)は、JIS 90には「登録されていない」ことが判明したが、「補助漢字」にも無い「哱」は依然不明である。

また、上記のように、JISは〈同値〉の文字の具体

的な字形には関知しない。というより〈同値〉に限らず特定コードポイントの字形にはまったく関知しないのがJISの基本的立場だが、これも世界のコード表に共通することで、特にJISがサボっているわけではない。

文字の差と文字観念

「学」と「學」、「孝」のようにコード表内に異体字が複数登録してあれば、それらが表内で区別されていることが分る。コード表の規定する〈文字観念〉とは、いわばコード表という「構造」の中でコードポイントとして弁別的(distinct)かどうかによって定まる構造的な価値(valeur)であるわけで、表内に「無い」字が「区別しない」のか「登録しない」のか不明なのは、このためである。

仮に、現実の言語の文字観念もこのような構造的なvaleurとして定まるものだと考えてみると、コード表の文字観念は、出来るだけ現実言語のそれに近付けるべきだという考え方があり得る。

しかし、それはかなり不都合なコード表になりそうである。何しろ同じラテンアルファベットでも、言語、政府等々によって文字数(レパートリー)が異なるので、ポルトガル語のa(ă, âもある)とフランス語のa(âはあがăが無い)とはvaleurが異なるから別コードだ、とい

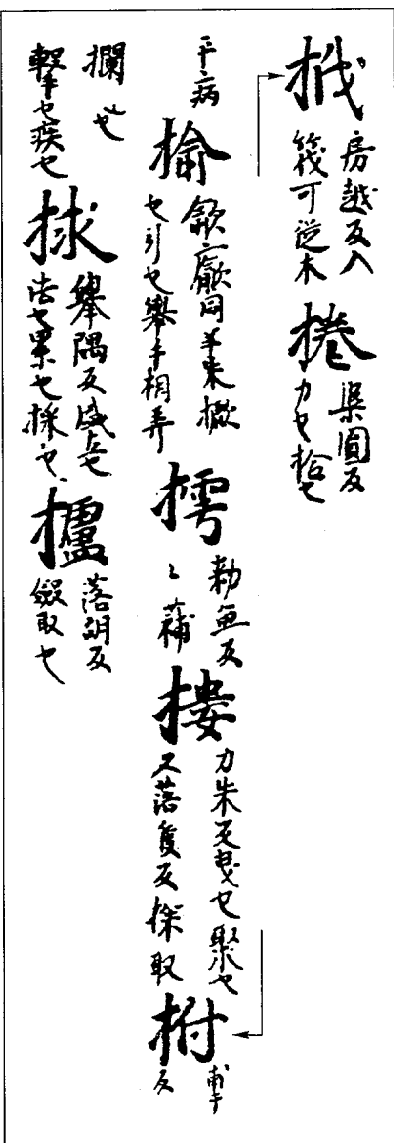


図1 天治本新撰字鏡
10-16 ウ「房越反入、筏、可逆木」とある。「棧」字は、切韻入声月韻の「棧」であろうし、10-9オ「拊」字は切韻平声虞韻甫无反の「拊」字であろう(貞刈伊徳、1961、p.11)

実の言語の「文字観念」にはさらに厄介な問題が多い。
江戸期以前の平仮名を主体とする文献では、「ハ」が語頭にあってもワ音を表す仮名として用いられることがある(『安田章、一九七三』)。他の「ハ」の「変体仮名」には見られないこうした「ハ」の用法は、それを他の仮名とは別字として区別した「文字観念」を要請するであろう。しかし他方、同時代でも「ハ」が語頭ではハ(またはバ)音しか示さない文献もあるわけで、言語・時代だけでなく、文献ごとに「構造的な文字観念」が異なることになる。
近世以前の写本では本篇と手篇はともに三画で書かれることがあつて紛らわしく、院政期に法隆寺の僧がかな

りの短時日で書写した「新撰字鏡」天治本や「玄応一切経音義」大治本には本篇なのか手篇なのか見ただけでは分らない見出し字が珍しく無い。前者には本篇の字が手篇部に収められている例もあつて(図1)、字書・音義類においてすら両者の截然とした区別を疑わせる。
金属活字印刷の漢字字書「落葉集」にも、本篇と手篇を見分けるのが困難な字がある(図2)。「落葉集」と同一活字の「ぎやどぺかどる」字集は、熟語を部首分類するユニークな字書だが、上巻で本篇に分類した「柎械枷鎖」を下巻では手篇にしている(図3)。同一文献内ですら「文字観念」が揺れているわけである。
歴史的に「文字観念」が一貫しないのはもちろんで、

しかし、これではコード表の文字観念と言語のそれとの対応を保証するものが何も無くなってしまうので、こうしたコード表規格の本文には必ず「legend」(スクリプト)という文字観念説明書が含まれており、「a: small n with ide」等と一字一字に注釈を加えることになっている。つまり、言葉による文字の注釈を含んだものがコード表

実際の問題としては、コード表の「文字観念」を現実の言語の「文字観念」と完全に一致させるのは、「構造」自体違うのだから明らかに不可能である。上記の通り、アルファベットの国際規格はこれをはじめから諦めている。実際には、現実の言語の「文字観念」が構造的な value として定まるか否か不明で、特に日本の文字の場合、現

文字観念の揺れとコード表

である。
JISの漢字には、この注釈がまったく無い(「解説」にも無い)。「祇」が「祇」でないといって通産省に祇園から抗議が来るといった騒ぎの遠因は、一つにはJISが「文字観念」の説明を「字形」を掲げるだけで済ませようとした「注釈無きコード表」である点にある。
確かに漢字の「注釈」は困難な作業だが、漢字字書(漢和辞典)はその具体例の一つである。もし、JISに部首、部首内画数、総画数、対応する説文解字・宋本玉篇・康熙字典・諸橋大漢和等の字・番号等を明記した漢字字書が附属していたら、利用者の混乱・当惑はかなり少なかったのではないかと思う。なにしろ、今のところJISの字は規格が字形に閏知しないため(画数すら不明なのである(ただし、「補助漢字」には多少この種の注釈がある))。

うことになりかねないし、そもそも別言語は別コードにということになりそうだからである。これでは、国際ネットワークでの電子メールのやりとり等でも困惑する。アメリカの大学に在籍しているポルトガル人にスペインからフランス経由でメールを出す場合、中継地全部で読まれる宛名には、何語のコードを使うべきだろうか？
それゆえ、アルファベットのコード表国際規格ISO 6937、ISO 8859はありとあらゆるアルファベット(キリル、ヘブライ等も含む)を網羅的に集めて、言語を超越した規格として制定されている。このコード表内にはaもaもあるからといって、aの無いフランス語にあってはaのvalueが異なる、等という文句は誰も言わない。それは、コード表の問題というより、むしろ規格の運用の問題である。コード表の文字観念を現実言語の文字観念と一致させよという主張は、実際に利用に供するコード表としてはあまりに狭量に過ぎる憾みがあると言えよう。

というのは、音声記号による音声・音韻表記の慣習とまったく同じである。

音声記号は、(しばしば誤解されているような)言語から独立に音価を表現するものでもなければ、一方特定言語の特定共時態専用の記号であるわけでもないというヌエ的存在である点で、コード表とまったく同じ便宜的存在である。異なる言語の表記に同じ音声記号が用いられても、それらが同じ音価である保証は無いし、一方同じ言語の同じ音の表記でも、研究者によって記号表現は異なり得る。しかし、適切な説明と diacritics、さらに見本のテープがあれば、大きな誤解は生じない。

文献学的に厳密なテキスト表現に、工業規格のコード表を利用するのは無理だという意見もあるが、上記のような注意を払えば決して困難ではない。音声表記に精密表記／簡略表記の別があるように、テキスト翻刻にも研究の意図に従って精疎があり得る。JIS規格票の印刷字形を見ただけで文献学的利用に耐えないとするのは、ジョーンズ発音の基本母音(cardinal vowels)を聞いただけで、日本語の母音と異なるから使えない、と言うのと大差無い早合点である。

なお、以上の議論はJIS90と「補助漢字」で足りるか否か、という問題とは独立である。「JISに無い字」はどんなにJISを太らせても残る。ことはコードポイ

ントの数という量ではなく、コードポイントの解釈という質の問題なのである。

引用文献

- 歌舞伎評判記研究会 一九七二〜七七 翻刻覚書(「歌舞伎評判記集成」月報に連載、岩波書店)
貞荊伊徳 一九六一 新撰字鏡の解剖(要旨)付表(下)(訓点語と訓点資料 14, 1-28ページ)
野村雅昭 一九八四 JIS C 6226 情報交換用漢字符号系の改正(標準化ジャーナル 14-3, 4-9ページ)
安田章 一九七三 吉利支丹仮名遣(国語国文 42-9, 1-20ページ)
ISO 646 7-bit coded character set for information interchange
ISO 6937 Coded character sets for text communication
ISO 8859 8-bit single-byte coded graphic character sets
JIS78 JIS C6226-1978 情報交換用漢字符号系
JIS83 JIS C6226-1983 情報交換用漢字符号系
JIS90 JIS X0208-1990 情報交換用漢字符号
補助漢字 JIS X0212-1990 情報交換用漢字符号—補助漢字

(注)

このようなJIS解釈漢字字書の計算機可読版は、各大型計算機メーカーが作成したものもあるし、金水敏氏、古田啓氏と稿者が共同で作成したKdicもパソコン通信pc-van (ORIENT OLS) pos (pool pdd) 等で無償で公開している。

(とくしま ままゆき・北海道大学)