

字形と字体の狭間 (情報標準の考古学試論)

小林龍生^{†1}

JIS X 0208 の 1983 年改正は、一部の例示字形の変更と符号位置の入れ替えによって、その後四半世紀以上続く混乱を市場に引き起こした。本稿では、現在の JIS および ISO/IEC の語彙の定義に立脚して、“字形”“字体”“字種”“図形文字”など、符号化文字集合に係わる用語とその関係を再定義し、その観点から、JIS 漢字コードの規格本文及び解説を読み直し、改めていわゆる新旧 JIS 問題の本質を探る。

Between Glyph and Glyph Image (An Essay on Archeology of Information Standardization)

TATSUO KOBAYASHI^{†1}

Some changes of glyph images and code positions in JIS X 0208:1983 made ensuing long confusion in Japanese market. This paper discusses the definitions of glyph, glyph image, graphic character and character. With the definitions, the text of standards and attached commentaries are critically re-read to seek the true reason of the confusion.

1. はじめに

1978 年に発行された JIS X 0208[a]は、7 ビットと 8 ビットの 2 バイトを用いる情報交換用符号化文字集合としては、世界でも先駆けとなるもので、その後の特に東アジアの漢字文化圏の情報標準に大きな影響を与えたものとして、高く評価されている。この規格は 1983 年の改正の際、複数個所の例示字形の変更と符号位置の入れ替えが行われた。この変更は後に新旧 JIS 問題と通称される大きな混乱を市場に引き起こした。

本稿は、符号化文字集合で用いられる“字形”“字体”“字種”“図形文字”などの用語を、最新の規格群[b]と矛盾しない形で再定義し、その定義に立脚して、当時の規格本文と解説を読み直すことにより、改めて、新旧 JIS 問題の本質を解き明かし、併せて、長く続いた混乱を収束させる道筋を示す。

2. “字形”“字体”“字種”“図形文字”

最初に、用語の定義を行う。本稿では、“字形”“字体”“字種”“図形文字”とそれらの関係について規定する。“文字”一般についても触れるが、情報規格において扱われる“文字”の定義と一般社会や言語学を中心とする研究者集団での用いられ方に大きな隔りがあることもあり、厳密な議論は行わない。

^{†1} 有限会社スコレックス
Scholex Co., Ltd.

[a] 本稿では、都合 3 度改正されている JIS X 0208 の具体的な版の違いに言及する場合以外、JIS X 0212 や JIS X 0213 も含め、JIS 漢字コードとして言及することがある。

[b] JIS 漢字コードの他に ISO/IEC 10646(UCS)を想定している。

(1) 文字

下記は、Wikipedia の文字の項目の記述である。

“文字（もじ）は、言葉（言語）を伝達し記録するために線や点を使って形作られた記号である。” [c]

また、下記は、小松英雄の漢字の定義である。

“漢字は、〈形=字形〉〈音=発音/語形〉〈義=意味〉の三つの要素を総合的に表す表語文字(logogram)である。”

[d]

本稿の議論は漢字のみを対象とするが、用語としての“文字”と“漢字”を区別なく扱う。本稿の議論の一部が、“文字”一般に適用できる可能性は否定しないが、本稿では一切議論しない。

下記は、同じく小松英雄の論述である。

“文字の機能は、一定の順序に組み合わされることによって発揮される。制約的社会慣習（以下、煩瑣を避けて「社会慣習」と呼ぶ）とは、それに従って行動することが不文律として要求されているところの、当該社会に定着した慣習である。書記について言うなら、特定の順序で組み合わされた表音文字の連鎖や表語文字の特定の字形が、どういう事物や概念を表すかは、社会通念として定まっており、その慣習に従わなければ

[c] <http://ja.wikipedia.org/wiki/%E6%96%87%E5%AD%97#.E6.96.87.E5.AD.97>

[d] 小松 pp.4-5.

ば意図どおりの情報伝達が保証されない。” [e]

筆者は、この小松の主張に全幅の賛意を表す。漢字の特定の字形（字体）が表象する事物や概念は、社会通念として共有されている。その上で、社会通念そのものが、時代や地域により、さまざまな変化と多様性を持っているということを、確認しておく。

(2) 字形

字形とは、漢字の形状の具体的視覚表現である。

ちなみに、JIS X 0208:1997 では、下記のように定義している。

“字形(JIKEI): 字体を、手書き、印字、画面表示などによって実際に図形として表現したもの。” [f]

字形には、敢えて英語を付加せず、ローマ字表記をしているが、Shape もしくは Glyph Image に対応すると考えられる。

(3) 字体

字体とは、漢字の形状の抽象的概念である。

下記は、JIS X 0208:1997 における定義である。

“字体(ZITAI): 図形文字の図形表現としての形状についての抽象的概念。” [g]

字体については、Glyph に対応すると考えて差し支えない。

(4) 字形と字体の関係

ある字形は、対応する何らかの字体を持つ。対応する字体を持たない視覚表現は、漢字ではない。

しばしば、“字形差”“字体差”といった議論がなされる場合があるが、本稿では、“字形が等しい”とは、二つの図形が幾何学的な意味で相似関係にあることのみをいうこととする。また、2つの異なる字形があるとき、それらが共通の1つの字体に対応する可能性と異なる2つの字体に対応する可能性があるが、その対応関係も、社会通念に依存する。

下記の図は、国立国語研究所の高田友和らによるものである。

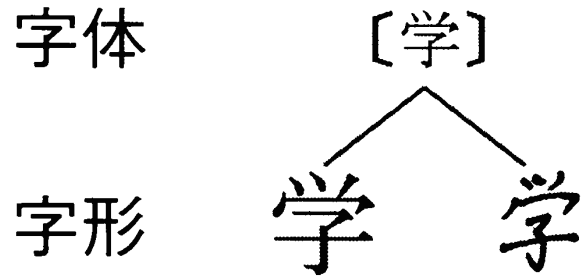


図 1 字体と字形の関係[h]

Figure 1 Relation between Glyph and Glyph Image

(5) 字体弁別粒度

字体弁別粒度とは、ある二つの字形を単一の字体に対応させるか別個の字体に対応させるかの判断の拠り所となる、視覚的差異の度合いをいう。

常用漢字表における明朝体のデザインに関する記述は、この字体弁別粒度に係わっている。常用漢字表では、相対的に字体弁別粒度を粗くとり、微細な字形差をデザイン差として字体を区別しない立場を取っている。

(6) 字種

字種とは、形状の異同に係わらず、意味的、音声的、概念的に定まる、漢字の異同の単位である。

字体より上位の階層に属する概念と考えられる。本稿では、高田らに倣い、常用漢字表での用法に準じる。 [i]

後述の“文字概念”は、この字種に近い意味であると考えられる。

(7) 字種と字体の関係

ある字体は、対応する何らかの字種を表象する。

ある字体が表象する字種は、文脈又は対象となる社会慣習により異なる場合がある。また、一つの字体が、同時に二つ以上の字種を表象する場合がある。

同形異字がこの議論に属する。例えば、日本で藝の常用漢字体として用いられる芸は、中国語の文脈では、全く異なる別の意味を持つ。

また、JIS X 0208 や JIS X 0213 で単一の符号位置を付与されている柿の字は、「かき」を意味する場合と「こけら」を意味する場合がある。元来「かき」と「こけら」は字体も異なる別の字種であるが、歴史の中で字体が混同されて使われ、符号位置としても同一視される結果となっている。上記の字形、字体、字種の間を、同じく高田らによる図によって示す。

[e] 小松 P.6.

[f] JIS X 0208:1997,p.3.

[g] Ibid.

[h] 高田他、行政用文字の研究—汎用電子情報交換環境整備プログラム—p.101

[i] Ibid, p.100-101.

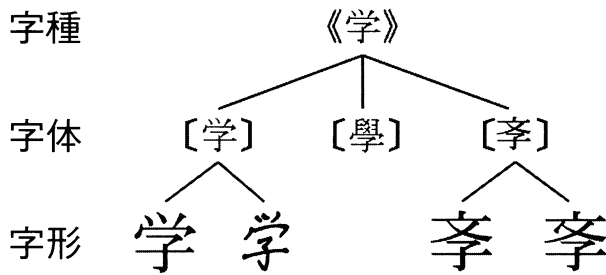


図 2 字種・字体・字形の関係[j]

Figure 2 Relation among Character, Glyph and Glyph Image

(8) 文字の表

有限個の互いに異なる字形で構成される表を文字の表という。

この用語は、本稿独自のものである。例えば、常用漢字表や戸籍法施行規則別表第二（いわゆる人名漢字表）などは文字の表である。JIS や UCS などの符号化文字集合の規格票も文字の表の要件を満たしている。広義には、康熙字典などの漢字字典類も文字の表の要件を満たしている。

(9) 字体集合

字体集合とは、互いに異なる有限個の字体を要素とする集合である。

(10) 文字の表と字体集合

文字の表は、それぞれの字形に対応する字体を要素として持つ字体集合を表象すると考えることが出来る。

文字の表と字体集合との関係も、本稿の議論の眼目の1つである。字体は抽象的な概念であるため、視覚的に認識できる形で示すことが出来ない。一方、字形が文字として認識されることにより、(社会や文脈によって異なるとはいえ)ある字体に結びつけられる。このような個々の字形と字体の関係を、文字の表と一対一に対応する字体集合との関係に拡張して考えよう、というのが本稿における著者の主張である。

一つの文字の表を構成する互いに異なる字形が、単一の字体に対応するときには、そのことを明言しなければならない。一つの漢字の表の中で互いに異なる字形は、原則的には、それぞれ互いに異なる字体を表象する。

先に挙げた常用漢字表の明朝体のデザイン差に関する記述は、この典型的な例である。

(11) 図形文字

字体に対応する固有の名前を図形文字という。図形文字は、一つ又は複数の字体に対応する。

下記は、JIS X 0208:1997 における図形文字の定義である。

“図形文字(graphic character): 制御機能以外で、通常、手書き・印字・表示の可視的表現をもち、一つ以上のビット組合せからなる符号化表現をもつ文字。” [k]

ちなみに、この JIS の定義で用いられている文字は、一般社会での文字の用法とは著しく異なっている。

“文字(character): データの構成、制御又は表現に用いる構成単位の集合の要素。” [l]

本稿における著者の提案の1つは、この図形文字の定義にある。JIS の定義における“手書き・印字・表示の可視的表現をもち”という部分は、ある字形がある字体に対応することに相当する。また、“一つ以上のビット組合せからなる符号化表現をもつ”という部分は、文字そのものの定義の繰り返しに過ぎない。

この JIS の定義を読み替えてみよう。

“字形として視覚化可能な符号化文字集合の要素の固有名”

さらに、“字形として視覚化可能な要素”とは、すなわち、字体集合の要素としての字体そのものであると考える。

このような手順を踏むと、“図形文字とは字体に付けられた固有名である”という定義が自然に導き出せる。ただし、字体弁別粒度が社会通念による多様性を持つため、広い範囲での情報交換を目途とする汎用的な符号化図形文字集合では、ある文字集合においては異なる字体とされる要素に対して単一の固有名を与える場合がある、ということである。

3. 文字概念の桎梏

上記の概念定義を前提として、いわゆる新旧 JIS 問題が何であったかについて、再考する。

JIS X 0208 の 1983 年改正の経緯については、1997 年版の解説に詳細な記述がある。本稿でも、基本的にはこの 97 年版解説を出発点に、再考を試みる。

3.1 事実

JIS X 0208 は 1983 年改正に際して、下記の符号位置に例示されている字形（以下例示字形）が変更された。

表 1 JIS X 0208:1983 で変更された例示字形

Table 1 Glyph Image changed in JIS X 0208:1983

新 JIS 字 形	UCS 符号 位置	JIS X 0213 面 区点	旧 JIS 字 形	JIS X 0213 面 区点	UCS 符号 位置

[k] JIS X 0280:1997, p.3.

[l] Ibid, p.4.

[j] 高田他,op.cit.,p.101

唾	U+5516	1-16-02	啞	1-15-08	U+555E
焰	U+7114	1-17-75	焰	1-87-49	U+7130
鳴	U+9D0E	1-18-10	鷗	1-94-69	U+9DD7
嚙	U+565B	1-19-90	嚙	1-15-26	U+5699
俠	U+4FA0	1-22-02	俠	1-14-26	U+4FE0
軀	U+8EAF	1-22-77	軀	1-92-42	U+8EC0
嶮	U+9E78	1-24-20	嶮	1-94-74	U+9E7C
麩	U+9EB9	1-25-77	麩	1-94-79	U+9EB4
屢	U+5C61	1-28-40	屢	1-47-64	U+5C62
繡	U+7E4D	1-29-11	繡	1-90-22	U+7E61
蔣	U+848B	1-30-53	蔣	1-91-22	U+8523
醬	U+91A4	1-30-63	醬	1-92-89	U+91AC
蟬	U+8749	1-32-70	蟬	1-91-66	U+87EC
搔	U+63BB	1-33-63	搔	1-84-86	U+6414
驛	U+9A28	1-34-45	驛	1-94-20	U+9A52
箆	U+7BAA	1-35-29	箆	1-89-73	U+7C1E
摑	U+63B4	1-36-47	摑	1-84-89	U+6451
填	U+586B	1-37-22	填	1-15-56	U+5861
顛	U+985B	1-37-31	顛	1-94-03	U+985A
禱	U+7977	1-37-88	禱	1-89-35	U+79B1
澆	U+6D9C	1-38-34	澆	1-87-29	U+7006
囊	U+56A2	1-39-25	囊	1-15-32	U+56CA
澆	U+6E8C	1-40-14	澆	1-87-09	U+6F51
醜	U+9197	1-40-16	醜	1-92-90	U+91B1
頰	U+982C	1-43-43	頰	1-93-90	U+9830
麵	U+9EBA	1-44-45	麵	1-94-80	U+9EB5
萊	U+83B1	1-45-73	萊	1-91-06	U+840A
蠟	U+874B	1-47-25	蠟	1-91-71	U+881F
攢	U+6505	1-58-25	攢	1-85-06	U+6522

※ 便宜的に JIS X 0213:2004 の符号位置と MS 明朝体を用いている。

また、以下の符号位置に示されていた例示字形が入れ替えられた。

表 2 JIS X 0208:1983 で交換された符号位置
Table 2 Code positions changed in JIS X 2028:1983

UCS 符号位置	JIS X 0213 面区点位置			JIS X 0213 面区点位置	UCS 符号位置
U+9BF5	1-16-19	鯀	鯀	1-82-45	U+9C3A
U+9D2C	1-18-09	鶯	鶯	1-82-84	U+9DAF
U+86CE	1-19-34	蛎	蠣	1-73-58	U+8823

[U+64B9	1-19-41	攪	攪	1-57-88	U+652A
U+7AC3	1-19-86	竈	竈	1-67-62	U+7AC8
U+6F45	1-20-35	灌	灌	1-62-85	U+704C
U+8ACC	1-20-50	諫	諫	1-75-61	U+8AEB
U+981A	1-23-59	頸	頸	1-80-84	U+9838
U+783F	1-25-60	礪	礪	1-66-72	U+7926
U+854A	1-28-41	蕊	蕊	1-73-02	U+8602
U+8ACC	1-31-57	韌	韌	1-80-55	U+8ACC
U+8CCE	1-33-08	賤	賤	1-76-45	U+8CE4
U+58F7	1-36-59	壺	壺	1-52-68	U+58FA
U+783A	1-37-55	礪	礪	1-66-74	U+792A
U+68BC	1-37-78	禱	禱	1-59-77	U+6AAE
U+6D9B	1-37-83	澆	澆	1-62-25	U+6FE4
U+8FE9	1-38-86	迤	迤	1-77-78	U+9087
U+877F	1-39-72	蠅	蠅	1-74-04	U+8805
U+69D9	1-43-74	楨	楨	1-84-02	U+69C7
U+4FAD	1-43-89	俚	俚	1-48-54	U+5118
U+85AE	1-44-89	藪	藪	1-73-14	U+85EA
U+7BED	1-47-22	篋	篋	1-68-38	U+7C60

※ 便宜的に JIS X 0213:2004 の符号位置と MS 明朝体を用いている

事実としての変更は、以上である。しかし、この変更は、衆知のように、爾後四半世紀あまりにわたる大きな混乱を市場に引き起こした。

3.2 83 年版解説の記述

この変更について、83 年版解説では、下記のような説明がなされている。

“これに関連して注意が必要なのは、今回の改正では、主として第 1 水準内の常用漢字等以外の漢字の中に、表中の字形を改めたものがあることである。これは、第 1 水準内の常用漢字等の字形との整合を図ることを目的としたもので、別の文字概念を採用したわけではない。また、二つの水準に振り分けたもののうち、水準間で字形を入れ替えたものがあるが、これも同様の趣旨からの措置である。” [m]

3.3 97 年版解説の解釈

一方、JIS X 0208 の最新版（1997 年版）の解説には、下記のような記述がある。

“しかし、1978 年の第 1 次規格で、規定外である‘解

[m] JIS X 0208:1983, p.71.

説’の2ページ目に1箇所、用語を間違えて用いた箇所がある。《この規格は、文字概念とその符号を定めることを本旨とし、その他字形設計などのことは範囲としない。》‘解説’が規格本文の規定に反することを述べてはならないことを考慮すると、この一文は明らかに間違いである。しかし、この一文が、そのまま理解され、その後の改正を間違った方向に誘導し、一般にも広く受け入れられてしまったと考えられる。”[n]

3.4 字体と文字概念の狭間

97JIS 解説の立場は、下記のように要約することが出来る。78JIS の解説では、《この規格は、文字概念とその符号を定めることを本旨とし、その他字形設計などのことは範囲としない。》との誤った記述によって、図形文字の一属性しかすぎない文字概念が一人歩きすることを許してしまった。そのため、83JIS において、“文字概念に変更はないのだから例示字形の変更や符号位置の入れ替えは許される”という誤認を生んだ。しかし、筆者は、新旧 JIS 問題の本質は、もう少し異なるところにあったと考える。先に提案した定義を用いて、83 改正の際に起こったことを再考する。

まず、78JIS や 83JIS で用いられている文字概念という言葉を検査する。

78JIS 本文では、用語の意味の中で、備考の図形文字の特徴の1つとして、名称（又は読み）、用途（又は用法）、字形、符号と並べて登場する[o]。言い換えれば、78JIS は、図形文字をこのような属性をもつ何か、と規定している。この時点で、78JIS の開発者たちは、すでに“字形”という用語を用いている。しかし、注意を要するのは、この時点では、“字体”という用語が用いられていない点にある。一方、78JIS の解説には、“3.3 異体字の扱い”として、同一文字概念に属する字形の相異についての議論が記述されている[p]。この議論は、本稿の用語を用いれば、字形の相異を字体の相異として扱うか否か、すなわち、字体弁別粒度を奈辺に置か、という議論であることは明白である。

以上の考察を踏まえると、83JIS の解説には、2つの異なる層での問題が内包されていることが明らかとなる。

第一の問題点。“字種”と“字体”との区別についての認識が未成熟であったこと。83JIS の開発者たちは、筆者の定義に従うと字種と字体として区別して扱うべき概念を文字概念という曖昧な概念として混同していたために、不要な混乱を招いた。

第二の問題点。JIS 漢字コードを受け入れた社会は、83JIS の開発者たちが想定していた文字概念よりも細かい粒度での字体の弁別を必要としていた。83JIS が開発された当時は、情報機器の一般への普及は未だ黎明期にあり、JIS 漢

字コードで想定されていた用途も一部の工業分野に限定されていた。しかし、その後の急速な情報機器の普及発展により、JIS 漢字コードは単なる工業標準に枠を超えて、文芸等の文化的創造活動や、行政を含むさまざまな実務での人名の表記等に使われるようになった。そして、そこで求められる区別の粒度は、JIS 漢字コードの開発者たちが想定していたものとは、異なるものであった。

97JIS の解説では、78JIS 解説の記述を“明らかな間違いである”と断言している。しかし、筆者には、この 78JIS 解説の記述を明らかな間違いであると断言することにはためらいがある。むしろ、単に、符号化文字集合によって符号化される対象が何であるか、についての認識が未成熟だったというのが実際のところではなかろうか。規格は、市場で用いられて初めて価値を生む。そして、市場の批判を受け入れて修正を加えていくことで、安定したものとなっていく。

筆者が本稿で提案する“字形”“字体”“字種”“図形文字”の定義とそれぞれの関係は、30年におよぶ混乱を経て初めて明確になってきたものと言えるかも知れない。

以上の考察で、83JIS における例示字形の変更については、その意味するところがおおむね明らかになったと考える。しかし、符号位置に交換については、依然として、規格論的に大きな問題がある。

83JIS が 78JIS の解説にあるように字体概念を符号化したものであったとして、たとえ例示字形の変更は許されても、符号位置の入れ替えは許されなかった。ある集合（それが字体概念の集合であれ、字体集合であれ）を符号化することは、その集合のそれぞれの要素に対して排他的な一意の符号を付与しなければならない。それは情報技術分野においては、符号化するという行為に原始的な要請である。

83JIS は、一部の符号位置を入れ替えることによって、78JIS との互換性を放棄する結果となった。もしくは、暗黙の内に同一の文字概念もしくは字体に対して、複数の符号位置を付与するという結果を招来することとなった。

4. 混乱の收拾

2013年6月日本政府は世界最先端 IT 国家創造宣言[q]を公表した。この中に、下記のような記述がある。

“特に文字の標準化・共通化に関しては、今後整備する情報システムにおいては、国際標準に適合した文字情報基盤を活用することを原則とする。”[r]

ここで、言及されている文字情報基盤とは、独立行政法

[n] JIS X 0208:1997, p.解 12.

[o] JIS C 6226-1978, p.1.

[p] Ibid., p.70.

[q] <http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou1.pdf>

[r] Ibid., p.18.

人情報処理推進機構が内閣官房、経済産業省と共に進めている事業である[s]。国際標準整合性を担保しつつ、戸籍・住民基本台帳を中核とする行政実務に用いられる人名用の漢字の利用環境を整備することを目途としている。

この事業では、UCS では統合される字体を弁別するための機能として UCS で規定されている Variation Selector(VS)を用い、IVD(Ideographic Variation Database)[t]に積極的にコレクション登録し、UCS では符号化されていない文字については、情報処理学会情報規格調査会 SC2 専門委員会を通して、積極的国際提案を行っている。

IVD は、UCS では統合規則によって1つの符号位置が付与される(ある利用者集団にとっては)異なる“字体”を Variation Selector (VS, 字形識別子)を用いることで区別し、字体識別粒度が利用者集団や利用目的によって異なるという問題点を、基底文字と VS のシーケンスの集合をコレクションとして、利用者集団や利用目的ごとに、個別にデータベースに登録することで解決するものである。

5. おわりに

新旧 JIS 問題の本質は、情報規格が提供する字体弁別粒度と社会が求める字体弁別粒度との乖離にあった。その背景には、字体弁別粒度が、利用者集団や利用目的による多様性を持つことがあった。本稿冒頭で引用した小松英雄の言を俟つまでもなく、情報標準は、このような文化や社会慣習の多様性を毀損することへの恐れを忘れてはならない。

謝辞 本稿の“字種”“字体”“字形”“図形文字”の定義は、筆者が私的に作成した“論理哲学論考風”と呼ぶメモが元となっている。このメモの作成に当たって、貴重なコメントを頂戴するとともに、議論に応じてくださった下記の方々に謹んで感謝の意を表す。

高田智和氏 (国立国語研究所)

田代秀一氏 (独立行政法人情報処理推進機構)

坂倉基氏 (角川書店)

神崎正英氏 (Xenon Limited Partners)

参考文献

- 1) 小松英雄：日本語書記原論 [補訂版] 新装版，2006 年，笠間書院
- 2) 高田智和，井手順子，虎岩千賀子：行政用文字の研究—汎用電子情報交換環境整備プログラム—，『日本語科学』23 (2008 年 4 月) .pp.95-11.
- 3) 情報交換用漢字符号系 JIS C 6226-1978，日本規格協会
- 4) 情報交換用漢字符号系 JIS X 0208-1983，日本規格協会
- 5) 7ビット及び8ビットの2バイト情報交換用符号化漢字集合 JIS X 0208:1997，日本規格協会

[s] <http://mojikiban.ipa.go.jp/>

[t] <http://www.unicode.org/ivd/>