システム/制御/情報，Vol. 45, No. 9，pp. 528–535，2001

解　説

日本における最新文字コード 事情 (前編)

※本原稿は著者によるゲラ刷りであり，最終稿とは異なっている．
本原稿を引用する場合は，必ず印刷された最終稿を確認すること．

安岡 孝一*

1. 文字コード 前史

1946 年 11 月，当用漢字表 [1,2] が告示された．この表は，前文に「現代國語を書きあらわすために，日常使用する漢字の範囲を，次の表のように定める」とあるとおり，漢字制限を主眼としたものである．すなわち，日本で用いる漢字は，この表の 1850 字に制限するというのが，当用漢字表の最大の目的である．文字コードの立場から考えた場合，使用する漢字の数が制限されているというのは，有限のビット数で漢字を表すことができるという点で，非常にうれしい．この漢字制限が徹底されていたならば，日本の文字コードはもっと簡単なものになっていただろう．しかし，当用漢字表の漢字制限は，いくつかの点で不徹底であった．一つは字体に関してであり，当用漢字表の字体は，簡易字体といわゆる旧字体とが混然となったものであった．もう一つは，まえがきに「固有名詞については，法規上その他に関係するところが大きいので，別に考えることにした」とあるとおり，人名や地名にはこの漢字制限を及ぼさなかったという点である．

当用漢字字体表 [4] は，当用漢字表の簡易字体をさらに徹底させるために，1949 年 4 月に告示されたもので，当用漢字表の各漢字に一対一に対応する 1850 字の表である．これにより，例えば，当用漢字表での「國」は「国」に，「漢」は「漢」に，「鬪」は「闘」に，それぞれ字体が改められた．あるいは，当用漢字表で既に簡易字体になっていた「礼」(原字は「禮」) は，当用漢字字体表では「礼」に改められた．当用漢字字体表によって，当用漢字表に掲げられた字体は全て改められ，日本で用いることのできる漢字は 1850 字に制限されたはずだった．

ところが，人名はそうはいかなかった．司法省 (現，法務省) は 1948 年の時点では，子の名に用いることのできる漢字として，当用漢字表の 1850 字のみを認めていた [3]．だが当用漢字字体表が告示されるに至って，当用漢字表と当用漢字字体表の両方を，子の名に用いることのできる漢字とみなしたのである [5,6]．この結果，字体が改められたはずの「國」も「漢」も「鬪」も「礼」も，「国」や「漢」や「闘」や「礼」と共に，子の名として用いることができるということになったのである．その上，1951 年 5 月に人名用漢字別表 [7] が告示された結果，子の名に用いることのできる漢字は，さらに 92 字が追加されることになった．

加えて 1954 年 3 月に，国語審議会から当用漢字補正案 [8] が報告されるに至って，日本の漢字制限はグラグラと揺らぎ始めるのである．

2. 文字コードの黎明期

1950 年 11 月に中川機械 (現，松下冷機) と毎日新聞によって，日本初のリモートコントロール邦文モノタイプ「SC-R」が開発されて以後，日本では数多くの漢字テレタイプや電算写植機が製作されていた．これらにおいては，各社が各様の文字種や漢字コードを用いていたため，会社が違えば互いにデータ交換できないというシロモノであった．1960 年代までは，それでも特に問題はなかったのである．当時のテレタイプやネットワークは，基本的に各社がそれぞれ社内でだけ用いるものだったからである．会社をまたいでデータ交換をおこないたい場合は，例えば 1959 年 8 月に制作された 6 社協定新聞社用コード表 (CO.-59)[9] のように，交換用の文字コードを決めておいて，それをオペレータが打ち直せば，特に大きな問題はなかったのである．

実際，1969 年 6 月に制定された JIS 初の文字コード[1]である JIS C 6220 (現，JIS X 0201) には，漢字は全く含まれていなかった．JIS C 6220 は，基本的には 8 ビットの文字コード[2]であり，02/1 ～ 07/14 (16 進数で書くと 21 ～ 7E) にはローマ文字[3]，10/1 ～ 13/15 (16 進数で書くと A1 ～ DF) には片仮名，00/0 ～ 02/0 および 07/15 (16 進数で書くと 00 ～ 20 および 7F) には制御文字が収録されていた．ただし，JIS C 6220 に漢字が含まれていなかったのには，別に大きな理由がある．JIS は文字コードの開発に際して，国際規格 (ISO) との整合性を第

* 京都大学人文科学研究所附属漢字情報研究センター

[1] これに先行して 1961 年 11 月に，JIS C 0803 印刷電信機のケン盤配列および符号 (1987 年 3 月 JIS X 6001 に改称) が制定されているが，当時の国際的文字コードだった International Telegraph Alphabet No.2 とも互換性がなく，結局 1994 年 6 月に廃止されている．

[2] 7 ビット版も規定されているが，ここでは割愛する．

[3] 03/0 ～ 03/9 が数字，04/1 ～ 05/10 が英大文字，06/1 ～ 07/10 が英小文字で，残りは特殊文字である．05/12 の「¥」，07/14 の「‾」を除いて，ASCII と同じである．

一に考えていた．文字コードは国内のみのデータ交換にとどまるものではない，という認識からである．JIS C 6220にしても，10/1 ～ 13/15の部分に，日本で最低限必要な片仮名を追加しているが，00/0 ～ 07/15の部分でISO 646と整合性を保っている．しかしながらJISは，国際規格と整合性を持った漢字コードの開発を，あきらめていたわけではない．日本が必要とする漢字コードに，国際規格の方を合わせるべく，虎視眈々と狙っていたのである．

その頃ISOでは，ISO 646の21 ～ 7Eの部分に，別の文字コードを呼び出すための仕掛けを，国際規格として開発中であった．この94字の部分を切り替えることで，複数の文字コードを扱えるようにする仕掛けである．しかし，漢字を含む文字コードを考えると，94字ではあまりに少ない．1971年10月に情報処理学会漢字コード委員会が報告した標準コード用漢字表(試案)ですら，6086字を含んでいた[1]ので，単純に計算しても65種類の文字コードに漢字を分散し，それらを次々に切り替えながら使うしかない，ということになってしまうのである．そこでJISはISOに対し，この94字の部分に，複数バイトの文字コードを呼び出せるよう提案した．すなわち94 × 94=8836字，94 × 94 × 94=830584字，あるいはそれ以上の字数を収録可能な文字コードを使えるようにしよう，というアイデアであり，日本はこれで漢字コードを開発しようと考えたのである．結果としてこの提案は承認され，国際規格ISO 2022に取り入れられることとなり，さらにその翻訳規格[2]であるJIS C 6228 (現，JIS X 0202)が，1975年3月に制定された．これで，94字を超える字数を収録可能な文字コードが，国際規格に整合する形で開発可能となったのである．

3. 世界初のISO準拠複数バイトコード

1974年4月，日本情報処理開発センター(現，日本情報処理開発協会)でおこなわれた漢字符号標準化調査研究委員会に，行政管理庁行政管理局(現，総務省行政管理局)から，一つの分厚い手書き資料が提出された．行政情報処理用標準漢字選定のための漢字使用頻度および対応分析結果，という長い名前のその資料は，情報処理学会漢字コード委員会標準コード用漢字表(試案) 6086字，6社協定新聞社用コード表(CO.-59) 1985字[3]，1973年8月現在の内閣調査室収容漢字表3181字，1973年8月現在の日本科学技術情報センター収容漢字表1864字[4]，1973年8月現在の大蔵省主計局収容漢字表4276字，1973年8月現在の国立国会図書館収容漢字表3956字[5]，1973年8月現在の日本生命収容人名漢字3044字，1972年度の国土行政区画総覧使用漢字3251字，の8つの表の対応を作り，官報における漢字の使用頻度(1972年)を加えて，行政管理庁が作成したものであった．この資料に基づき行政管理庁は，行政情報処理用標準漢字(案)[6]2817字を選定していた[7]が，それはこれらの表に現れる頻度が高い漢字を集めたものであった．例えば「楓」や「樫」や「槻」は，この資料によれば，国土行政区画総覧使用漢字に「頻度1」で現れるのみであったことから，行政情報処理用標準漢字(案)には含まれていなかったのである．

ところが，この資料に対する漢字符号標準化調査研究委員会の視点は，行政管理庁とは異なっていた．委員会は,「楓」や「樫」や「槻」が日本の地名に使用されている[8]ならば，これらを日本の漢字コード規格に収録すべきだと考えたのである．そこで委員会は，この資料から，もっと大きな漢字集合を制作することにした．すなわち，行政情報処理用標準漢字(案)と情報処理学会漢字コード委員会標準コード用漢字表(試案)を合わせた上に，地名として国土行政区画総覧使用漢字を，人名として日本生命収容人名漢字を，全て加えた漢字集合を制作したのである．

さらに，この漢字集合を第1水準と第2水準に分けるために，委員会は全部で37の漢字表を集めた．それら37の漢字表から2934字を抽出した後，当用漢字字体表[4]，当用漢字補正案[8]，人名用漢字別表[7]，都道府県名，当時の市区町村名および郡名，を加えて[9]第1水準とした．そして，先の4漢字表で制作した漢字集合から，第1水準の漢字を除いて第2水準としたのである．この上で，第1水準の漢字は五十音順，第2水準の漢字は康熙字典順とすることが決められ，その順序に並べるという作業がおこなわれた．

これら第1水準，第2水準の漢字に，数字，ローマ字，平仮名，片仮名，ギリシア文字，ロシア文字，そして108個の特殊文字を加え，JIS C 6228に準拠するように94

[1] [12]には，なぜか，漢字6100字とある．

[2] ISOの規格は，基本的に英語かフランス語で書かれている．これに対し日本では，対応するJISを制定することになっている．

[3] [9]によれば，漢字は1987字(漢数字の「〇」を含む)である．

[4] [10,24]によれば，この時点での漢字数は1861字のはずである．

[5] [11]によれば，漢字は3965字収容されていたはずである．

[6] 翌年，行政情報処理用基本漢字という名称になった．

[7] 実際の資料作成および漢字選定は，受託先である谷村株式会社新興製作所(現，新興製作所)がおこなった．

[8] 「楓」に関しては，奈良県生駒郡平群町に楓原(しではら)という地名がある.「樫」に関しては，和歌山県日高郡印南町に樫川(ほくそがわ)という地名がある．しかし,「槻」を含む日本の地名は発見されておらず，岡山市の槻(さい)を誤った可能性が考えられる．

[9] 作業漏れがあったらしく，例えば，千葉県印旛郡や熊本県菊池郡泗水町は，第1水準だけでは書けなかった．

×94の形に並べた文字表が，漢字符号標準化調査研究委員会の報告書(1976年3月)にまとめられた．この後，委員会はJISの委員会として作業を継続し，1976年6月に告示された人名用漢字追加表28字[13]への対応や，JIS C 6228で用いるエスケープシーケンスの取得[10]などをおこない，1978年1月にJIS C 6226 (現，JIS X 0208)を制定した.「通常の国語の文章の表記に用いる図形文字の集合とその符号」としての，世界初のISO準拠複数バイトコードの誕生である．

JIS C 6226では，94×94の文字表の各符号位置は，1区1点から94区94点までの区点番号で示されており，7ビットあるいは8ビットコードの21～7Eの部分に呼び出して用いることが想定されている．区は1バイト目，点は2バイト目に対応し，21～7Eの部分に呼び出された時，例えば16区1点の「亜」は30 21という2バイトで表される．なお，特殊文字は1区1点～2区14点，数字は3区16～25点，英大文字は3区33～58点，英小文字は3区65～90点，平仮名は4区1～83点，片仮名は5区1～86点，ギリシア大文字は6区1～24点，ギリシア小文字は6区33～56点，ロシア大文字は7区1～33点，ロシア小文字は7区49～81点，第1水準の漢字は16区1点～47区51点，第2水準の漢字は48区1点～83区94点に，それぞれ収録されており，残りの符号位置は空きである．

このように，多大な労力を傾注して開発されたJIS C 6226だったが，開発当初の時点で既に重大な問題をいくつか抱えていた．一つは，JIS C 6226が，JIS C 6220の上位互換になっていなかったことである．そもそも当初の計画では，JIS C 6226の1区にはローマ文字，2区には片仮名，3区には平仮名，そして4区以降に漢字を収録する予定だった．つまり，JIS C 6226の1～2区は，JIS C 6220のローマ文字および片仮名の部分と上位互換にする計画だった．ところが，実際のJIS C 6226では，英数字だけはJIS C 6220と上位互換となるように3区にまとめて収録された[1]ものの,「!」や「?」のような特殊文字は1区にバラバラに収録され[2]，片仮名は5区にJIS C 6220とは異なる順序で収録された．この結果，JIS C 6226はJIS C 6220の上位互換ではなくなってしまい，JIS C 6220のデータは，JIS C 6226に変換できなくなってしまったのである．

もう一つの問題は，JIS C 6226が地名に関して，二次資料である行政情報処理用標準漢字選定のための漢字使用頻度および対応分析結果を典拠としたため，日本国内の実際の地名とは乖離してしまったことである．例えば，鹿児島県の吐噶喇(とから)列島の「噶」は，JIS C 6226に含まれていなかった．沼津市の鷭町(くまたかちょう)，岡山市の穐(さい)なども，JIS C 6226では書けなかった．あるいは,「楾」がJIS C 6226の60区51点に収録されていても，この字がどういう読みでどういう用例があるのかは，ほとんどの人には全く謎だったのである．

さらなる問題は，JIS C 6226における人名に関する扱いが，当時の法務省の扱いとは必ずしも合致していなかったことである．一例を挙げよう.「富」と「冨」は「人名で使い分けられている」ことから，41区57点と41区58点という別々の符号位置が与えられている．これに対し,「巣」と「巢」は「字形の違いがわずかであると認める」ことから，33区67点という符号位置に合併されている．つまり，JIS C 6226では,「富」と「冨」は書き分けることができるが,「巣」と「巢」は書き分けられないのである．ところが，当時の法務省の基準では,「冨」は子の名に用いることができなかったが,「富」や「巣」や「巢」はいずれも子の名に用いることができた[3]．つまり,「巣」と「巢」は，子の名において使い分けることができるにもかかわらず，JIS C 6226では使い分けることができなかったのである．

このような問題を，開発当初から内包していたため，シフトJISと呼ばれる規格違反の符号化や，外字による文字化けが，JIS C 6226にはつきまとうことになるのである．

4. 常用漢字表と人名用漢字許容字体表

当用漢字表，当用漢字字体表に代わるものとして，1981年10月に告示されたのが，常用漢字表[16]である．この表は，前文に「一般の社会生活において現代の国語を書き表すための漢字使用の目安」とあるとおり，制限ではなく，あくまで目安である．すなわち，1946年以来おこなわれてきた1850字の漢字制限はこれで終わりを告げ，新たに1945字の目安が登場したのである．なお，当用漢字字体表の1850字中1849字に関しては，常用漢字表ではそのままの字体が踏襲されており，現実的には「燈」から「灯」への字体変更と，95字の追加となっている．また，常用漢字表のうち355字には,「いわゆる康熙字典体」が括弧書きで添えられている.「これは明治以来行われてきた活字の字体とのつながりを示すために添えたものであるが，著しい差異のないものは省いた」と説明されている．

これに合わせて法務省も，子の名に用いることのできる漢字に関して，扱いを変更することになった[15]．概

[10]エスケープシーケンスは,「ESC 2/4 4/0」(16進数で書くと1B 24 40)だった．

[1]21～7Eに呼び出された時,「0」～「9」は23 30～23 39に,「A」～「Z」は23 41～23 5Aに,「a」～「z」は23 61～23 7Aになる．

[2]特に，JIS C 6220では1つの符号位置02/2に合併されている「“ 」と「 ”」が，JIS C 6226では1区40点と1区41点に分離されてしまった[14]．

[3]当用漢字表に「富」と「巢」が，当用漢字字体表に「富」と「巣」が含まれていたが,「冨」はこれらの2表，人名用漢字別表，人名用漢字追加表のいずれにも含まれていなかった．

要は以下のとおりである．

(1) 当用漢字字体表が廃止され，常用漢字表がそれに代わるものとなったことから，常用漢字表1945字を全て子の名に用いることのできる漢字とする．
(2) 従前の人名用漢字別表92字と人名用漢字追加表28字から，常用漢字表に採用された8字を除き，新たに54字を追加して，166字の人名用漢字別表とする．
(3) 当用漢字表1850字のうち，常用漢字表に括弧書きで添えられた195字に，従来，子の名に用いることができるとされていた10字を加え，205字を人名用漢字許容字体表とする．

すなわち，従来の当用漢字字体表，人名用漢字別表，人名用漢字追加表に含まれていた漢字は，基本的にそのまま子の名に用いることができるが，当用漢字表に含まれていた漢字に関しては，扱いが変わったのである．この点に関して，多少，細かい解説を加えよう．

常用漢字表が登場する以前は，当用漢字字体表も当用漢字表も，いずれも子の名に用いることができるということになっていた．「漢」や「国」や「巣」や「闘」や「類」や「礼」と共に，「漢」も「國」も「巣」も「鬪」も「類」も「禮」も子の名に用いることができたのは，先に述べたとおりである．ところが当用漢字表が廃止されたことによって，これら「漢」「國」「巣」「鬪」「類」「禮」などに対する扱いをどうするかが，問題となったのである．この問題に対する一つの解決法は，このような漢字を，子の名に用いることを一切認めない，とするものである．しかしこの解決法は，従前に用いることができた字体を一気に否定することになり，あまりに非現実的である．これとは逆のもう一つの解決法は，このような漢字を全て，引き続き子の名に用いることができる，とするものである．しかしこの解決法は，既に廃止された当用漢字表を，引き続き根拠として認めることになり，非常にまずい．

結局のところ，民事行政審議会がこの問題に対して取った解決法は，上記の二つの解決法のいずれとも異なっていた．審議会の結論は，このような漢字のうち，常用漢字表に括弧書きで添えられているものに限り，子の名に用いることを認める，というものである．すなわち，この時点で用いることができた字体をある程度残しながら，根拠としては常用漢字表の括弧書きの字体，すなわち常用漢字表が著しい差異があるとした字体を用いる，というものであり，これに基づいて作成されたのが，人名用漢字許容字体表である[1]．先の「漢」「國」「巣」「鬪」「類」「禮」を例に取ると，これらに関する常用漢字はそれぞれ「漢(漢)」「国(國)」「巣(巣)」「闘(鬪)」「類(類)」「礼(禮)」という形の括弧書きを持っていたことから，「漢」「國」「巣」「類」は人名用漢字許容字体表に収録され，「鬪」「禮」(および「鬮」「禮」)は収録されなかった．

以上，述べてきたとおり，常用漢字表の告示は，こと人名においては，当用漢字表と当用漢字字体表の差異を浮き彫りにすることになった．この解決策として登場したのが，人名用漢字許容字体表だったのである．

5. シフト JIS の誕生

1982年4月，三菱電機から日本初の16ビットパソコン「MULTI 16」が発売された．暫定的な独自拡張を施したCP/M-86をOSとするパソコンだったが，OSレベルで漢字をサポートしている[2]という点で画期的だった．この独自拡張のために開発された文字コード [17,18] が，シフト JIS の原型だったのである．

シフト JIS は，JIS C 6220 の隙間に JIS C 6226 の文字を埋め込んだ8ビット1～2バイトコードである．すなわち，JIS C 6220 の 00～7F と A1～DF はそのままにして，94×94 の JIS C 6226 の文字表を (31+16)×(63+125) の形に変形し，1バイト目には 81～9F と E0～EF を，2バイト目には 40～7E と 80～FC を用いるのである．JIS C 6226 の区点番号との対応は，81 40 が1区1点[3]，81 41 が1区2点，…，81 7E が1区63点，81 80 が1区64点，…，81 9E が1区94点，81 9F が2区1点，…，81 FC が2区94点，82 40 が3区1点，…，9F FC が62区94点，E0 40 が63区1点，…，EF FC が94区94点となっている．

開発当時のシフト JIS は，単に JIS C 6226 の規格違反の実装の一つに過ぎなかった．しかし，1982年12月に「MULTI 16」の日本語CP/M-86が完成[4]し，同じ月に東京芝浦電気(現，東芝)の「パソピア16」に搭載された MS-DOS ver2.0 漢字版が，やはりシフト JIS を採用したことで，日本の16ビットパソコン市場は，急速にシフト JIS に傾いていったのである．

6. JIS C 6226 の 1983 年改正

工業標準化法第15条により，日本工業規格 (JIS) は，5年毎に見直しをおこなうことが定められている．その際に，改正・廃止の必要があれば改正・廃止し，なければ確認をおこなう，ということになっている．1978年1月に制定された JIS C 6226 に対しても，1983年の見直しの際に，常用漢字表や人名用漢字別表の影響が問題となり，1983年9月に改正がおこなわれた．この改正は，1978年版の JIS C 6226 と「できるだけ互換性を保つ

[1] 民事行政審議会は [2] を参照しておらず，その結果「敍」([2] で「敘」に訂正された)が，人名用漢字許容字体表に含まれてしまっている．

[2] CP/M-86 上で動作する M-BASIC plus では漢字が使えたが，CP/M-86 無しに動作するスタンドアロン版 M-BASIC では漢字は使えなかった．

[3] [17,18] によれば，1区1点の間隔だけは特別に 20 20 を使うこと，となっていた．

[4] かな漢字変換が，OS レベルでサポートされた．

ように努力した」はずだったが，結論から言うと，文字コードの改正にしては非互換な変更があまりに多く，その後の日本の文字コードの混乱を引き起こす最大の原因となった．

非互換な変更の中で，最も大きな混乱を引き起こしたのは，既存の文字の符号位置に対する変更である．例えば「堯」は，1978 年版の JIS C 6226 では 22 区 38 点に収録されていたが，1983 年版ではこれを 84 区 1 点に移動してしまった．さらに 1983 年版では，22 区 38 点に「尭」を収録したのである．このような変更をおこなった理由は,「常用漢字等[1]では，新字体を第 1 水準に，旧字体を第 2 水準におく」という原則を，新たに人名用漢字別表に追加された「尭」に対しても適用しようとしたため [19] である．しかしこれは，既存の文字の符号位置を変更してはならない，という文字コードの大原則を，無視したものである．このような符号位置の変更をおこなえば，それ以前に蓄積されたデータに対する互換性が完全に失われ，文字コード規格そのものに対する信頼性が，破壊されることになるのである．

しかも「常用漢字等では，新字体を第 1 水準に，旧字体を第 2 水準におく」という原則を守ろうとした，という説明自体も非常にアヤシイものである．例えば,「瑤」は 64 区 86 点から 84 区 4 点に移動させられ，64 区 86 点には代わりに人名用漢字の「瑶」が収録された．しかし，そもそも 64 区 86 点も 84 区 4 点も第 2 水準であり，第 2 水準内で文字を移動するということを，平気でおこなっているのである．あるいは，41 区 16 点の「檜」と 59 区 56 点の「桧」は，1983 年版では入れ替えられてしまった.「檜」が 59 区 56 点に,「桧」が 41 区 16 点に，それぞれ移動させられてしまったのである．しかし,「桧」も「檜」も人名用漢字ではない．もちろん，常用漢字でもない．しかも，この入れ替えによって，東京都西多摩郡檜原村が，第 1 水準だけでは書けなくなってしまった，というオマケ付きである．

その上，非常に不思議なことに，JIS C 6226 の 1983 年改正は，1981 年 10 月に施行された人名用漢字許容字体表に対しては，一瞥もくれていない．JIS C 6226 の「常用漢字等では，新字体を第 1 水準に，旧字体を第 2 水準におく」という原則とやらは，人名用漢字別表と人名用漢字許容字体表との間には，全く適用されなかったのである.「巣」と「巢」は相変わらず 33 区 67 点に合併されていたし,「漢」と「漢」は 20 区 33 点に合併されたままだった.「国」と「國」はそれぞれ 25 区 81 点と 52 区 2 点という別の符号位置が与えられていたが，これは 1978 年の制定時からのもので，別に 1983 年の改正時に変更したわけではない．すなわち，例えば「巣」と「巢」は，常用漢字表が字体の差異を認めているにも関わらず，JIS C 6226 では「字形の違いがわずかであると認めるもの」のままで放置してしまったのである．この時点で，常用漢字表の括弧書きの字体に対する JIS の見解は，一般の社会と齟齬をきたすことになった，と言っても過言ではない．

JIS C 6226 の 1983 年改正における，もう一つの大きな非互換変更は，常用漢字や人名用漢字以外の漢字に対して，字形を大きく変えてしまったものがある，ということである．これは「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」と解説にはあるが，元の字形が JIS C 6226 から消滅してしまったという点で，先の例よりさらにタチが悪い．例えば，18 区 10 点の「鷗」は「鴎」に，32 区 70 点の「蟬」は「蝉」に字形が変更された．あるいは，39 区 25 点の「嚢」は「嚢」に変更されたが，この結果，兵庫県美嚢郡が JIS C 6226 では書けなくなるという，致命的な事態が起こっている．1978 年の制定時に，都道府県名，市区町村名および郡名を第 1 水準に補充した，という事実を，解説に記しておきながら，1983 年の改正では，当時の郡名や市区町村名が書けなくなる可能性すら，全く考慮に入れない変更をおこなってしまったのである．

以上述べてきたように，JIS C 6226 の 1983 年改正は，文字コードの改正としてはあまりに非互換なものであり，かつ一般の社会の文字通念からも乖離したものであった．JIS C 6226 の当初の設計意図も，JIS C 6228 が規定する文字集合の概念も，ましてや常用漢字表と人名用漢字別表および人名用漢字許容字体表の関係も，全く理解しないまま改正をおこなったのではないか，という疑念すら持たれる．この結果，JIS C 6226 の 1983 年版は，1978 年版とは全く違う文字集合であるとみなされ，1978 年版とは異なるエスケープシーケンス (ESC 2/4 4/2) が付与されるという結果になった．その上，日本電気など一部のメーカーが 1983 年改正に従わず，1978 年版の JIS C 6226 の実装を続けるという，異常な事態に発展していくのである．

7. JUNET コードと EUC

1984 年 10 月，日本のインターネットのさきがけとなる JUNET が，慶応大学と東京大学と東京工業大学を結ぶことで，スタートした．JUNET において，JIS C 6220，JIS C 6226，JIS C 6228 の，いずれも 1982 年以前の版を元に考案されたのが，当時の JUNET コードである．当時の JUNET コードは,「1B 24 40」を漢字イン,「1B 28 48」を漢字アウトとして，21 〜 7E の部分を切り替える 7 ビットコードであった．7 ビットコードを採用したのは，ARPA インターネットでのテキストメッセージ (すなわち電子メール) が，RFC 822 (現，RFC 2822) によれば，7 ビットの ASCII だったからであり，これと整合性のある文字コードとしたかったからである．

[1] 当時の JIS C 6226 の解説には,「常用漢字表および人名用漢字別表を合わせて常用漢字等と呼ぶ」とある．

同じ頃，日本語UNIXシステム諮問委員会は，UNIX上で漢字を扱う方法について，議論を重ねていた．1984年11月にJIS C 6228が改正されて，複数バイトコードがA1～FEにも呼び出せるようになった[2]こともあり，それに準拠した形での漢字コードを模索していたのである．1985年5月，委員会は，報告書をAT&Tに提出し，これがEUCの原型となった[20,23]．

EUCは，JIS C 6228準拠(すなわちISO 2022準拠)の文字集合を，最大4つまで同時に使える8ビットコードである．文字集合0は，21～7Eに割り当てられ，通常はASCIIが用いられる．文字集合1は，A1～FE(あるいはA0～FF)に割り当てられる．文字集合2も，A1～FE(あるいはA0～FF)に割り当てられるが，各文字の前に8Eを必要とする．文字集合3も，A1～FE(あるいはA0～FF)に割り当てられるが，各文字の前に8Fを必要とする．日本語版EUCにおいては，文字集合0にASCII，文字集合1にJIS C 6226，文字集合2にJIS C 6220の片仮名を，それぞれ用いることになっていた[1]．すなわち日本語版EUCでは，JIS C 6226の1区1点～94区94点は，それぞれA1 A1～FE FEという2バイトで表されるのである．あるいは，JIS C 6220の片仮名10/1～13/15は，それぞれ8E A1～8E DFという2バイトで表されると考えてよい．

一方，JUNETコードで使用されているエスケープシーケンスが，実は間違っているということが，1986年3月，東京大学の和田英一によって指摘された[21]．すなわち，漢字アウトに「1B 28 48」が使用されているが，これはNATSスウェーデン名前用コードを呼び出すためのものであり，JIS C 6220のローマ文字には「1B 28 4A」を用いるべきだ，という点である．ただし，この間違いは，JIS C 6228初版の解説に「エスケープシーケンスの登録の現状」の項があり，JIS C 6220のローマ文字に対しては「ESC 2/8 4/8」が登録されている，と書かれていた[2]のが原因である．さらに和田は，1978年版のJIS C 6226に対する「1B 24 40」と，JIS C 6220のローマ文字に対する「1B 28 4A」に加えて，1983年版のJIS C 6226に対する「1B 24 42」と，ASCIIに対する「1B 28 42」も使用できるようにすべきだ，と提案している．また，漢字イン，漢字アウトという考え方は正しくなく，エスケープシーケンスは常に「GOTO文」だということも強調している．

1988年2月に作成されたJUNET利用の手引(第1版)では，「JUNETにおける漢字利用の約束」として，和田の提案に基づき，これら4つのエスケープシーケンスのいずれを使ってもよい[3]が，「1B 28 48」は使わないことが協約された．これが新しいJUNETコードである．なお，JIS X 0201 (JIS C 6220が1987年3月に改称)の片仮名は，JUNETコードでは使わないことが協約されている．したがって，シフトJISのA1～DF部分あるいは日本語版EUCの8E A1～8E DF部分の片仮名は，それぞれJIS X 0208 (JIS C 6226が1987年3月に改称)中の対応する文字に変換しなければ，電子メールで送ることはできなかった．

8. 補助漢字と平成明朝体

JIS C 6226への文字追加を要求する声に押され，1985年4月，漢字符号系調査研究委員会が日本規格協会の下で発足した．委員会は，追加希望漢字を調査し，全部で30の漢字表(単純合計で17742字)を収集した．しかしながら委員会は，これらの漢字表から漢字を選定する作業はおこなわず，日本印刷産業連合会に選定リストの作成を依頼する，という方法を取ることにした．というのも，同じ頃，日本印刷産業連合会の文字コード委員会においても，JIS C 6226に含まれていない漢字の調査をおこなっており，9つの漢字表(単純合計で41772字)を収集していたからである．

日本印刷産業連合会では，これら39の漢字表から，大漢和辞典[22]に収載されている漢字の抽出をおこない，異なり字数で12924字を得た．さらに，この12924字から，「字形の違いがわずかである」漢字がJIS C 6226に含まれているものや，39の漢字表のうち1つの漢字表にしか現れないものを削除し，5790字とした．また，大漢和辞典に収載されていなかった漢字のうち，3つ以上の漢字表にあるものを選びだし，53字を得た[25]．

情報交換用漢字符号系改正原案作成委員会(1987年4月に漢字符号系調査研究委員会が改組)は，日本印刷産業連合会が選定したこれらの漢字5843字に，非漢字266字を加えた文字集合案を，1987年12月に完成した．この段階で委員会は，JIS X 0208 (JIS C 6226が1987年3月に改称)への文字追加はおこなわないことを決定した．すなわち，JIS X 0208はそのまま[4]にして，これら6109字を，JIS X 0208とは別に規格化することにしたのである．ただし，規格化にあたって，規格票を印刷するためのフォントが必要である，と委員会は考えた．こ

[2]例えば，JIS C 6226の16区1点の「亜」をA1～FEに呼び出すと，B0 A1となる．

[1]日本語版EUCの文字集合3に，JIS X 0212を追加することが合意されたのは，1991年12月のことであり，これが現在のEUC-JPである．

[2]1982年6月に正誤表が出され，「ESC 2/8 4/10」に訂正された．

[3]JUNET利用の手引(第1版)の「メールの文字コード」の項に，「JUNETではメールには“新JISコード”を使うことになっています」とあることから，「1B 24 40」は使えないと解釈する向きもある．しかしながら，少なくとも1988年2月の時点では，エスケープシーケンスを3つに絞るという合意はなかった．

[4]ただし，1984年11月に改正されたJIS C 6228 (1987年3月にJIS X 0202に改称)と整合させるための改正はおこなうこととされた．

れが，文字フォント開発・普及センターの設立 (1988 年 11 月) の契機となったのである．

文字フォント開発・普及センターの最初の仕事は，JIS X 0208 の 6930 字[1] および文字集合案 6109 字の明朝体と，JIS X 0208 の 6930 字の角ゴシック体を開発することであった．センターは，1989 年 1 月に明朝体の公募をおこない，リョービイマジクスのデザインが入賞した．後に平成明朝体 W3 と名づけられるフォントの，委託開発の開始である．また，1989 年 3 月におこなった角ゴシック体の公募では，日本タイプライターのデザインが入賞した．後に平成角ゴシック W5 と名づけられるフォントの，委託開発の開始である．開発完了は 1990 年 5 月という，はっきり言ってムチャなスケジュールであった．

ところが，フォント開発完了も間近の 1990 年 3 月，法務省がおこなった人名用漢字別表に対する 118 字の追加 [26] により，JIS X 0208 の見直しをおこなう必要が生じた．情報交換用漢字符号系改正原案作成委員会としては，JIS X 0208 には文字の追加をおこなわない方針だったが，118 字をチェックした結果，少なくとも「凜」は追加する必要があったのである．残る 117 字については，規格票に用いる平成明朝体の字形を，人名用漢字に合わせてそれぞれ手直しすればよいだろう，と思われた．例えば，64 区 2 点の「燿」は，字形を「燿」に変更すればよい，と考えたのである．しかし，63 区 70 点の「熙」の字形を「熙」に変更するのは，問題があった．JIS X 0208 規格票本文の「用語の定義」の中に,「主としてこうき (康熙) 字典の部首による」という一文があった[2]のである．委員会は，JIS X 0208 規格票本文も，平成明朝体で印刷することを決めていたので，もし平成明朝体の「熙」を「熙」に変更するならば，規格票本文のこの部分をも変更するのかどうかが問題になる．結論として委員会は,「熙」と「熙」は別字であると確認し,「凜」を JIS X 0208 の 84 区 5 点に,「熙」を 84 区 6 点に追加して，規格票本文の「主としてこうき (康熙) 字典の部首による」は変更しなかった．これが 1990 年 9 月改正の JIS X 0208 であり，JIS X 0202 でのエスケープシーケンスも，従来の JIS X 0208 の更新 (ESC 2/6 4/0 ESC 2/4 4/2)[3]となったのである．

同時に委員会は，文字集合案の各漢字に対する調査を続行し，JIS X 0208 への移動が確定した「凜」，大漢和辞典で義未詳となっている漢字，他の漢字と構成要素が同じと見られる大漢和辞典非収載漢字など 42 字を，文字集合案から削除した．委員会はこの 6067 字を，JIS X 0202 に準拠するように 94 × 94 の形に並べ，JIS X 0202 で用いるためのエスケープシーケンスを取得[4]し，平成明朝体で印刷して，JIS X 0212 として制定 (1990 年 10 月) した.「JIS X 0208 で規定している“ 通常の国語の文章の表記に用いる図形文字の集合 ”に含まれていない図形文字を必要とする情報交換のために，JIS X 0208 の補助として用いる図形文字の符号」としての,「補助漢字」の誕生である．

JIS X 0212 では，94 × 94 の文字表の各符号位置は，JIS X 0208 と同じく区点番号で示されている．呼び出し先は，21 ～ 7E のみならず，A1 ～ FE も想定されている．特殊文字は 2 区 15 ～ 25・34 ～ 36・75 ～ 81 点，ギリシアアルファベットは 6 区 65 ～ 69・71・73・74・76・81 ～ 92 点，キリル系アルファベットは 7 区 34 ～ 46・82 ～ 94 点，ラテン系アルファベットは 9 区 1・2・4・6・8・9・11 ～ 13・15・16・33 ～ 48 点に加えて 10 区 1 ～ 24・26 ～ 87 点と 11 区 1 ～ 27・29 ～ 35・37 ～ 87 点に収録されており，これらに関しては，JIS X 0208 の収録文字とは区点番号が衝突しないように工夫されている．漢字は 16 区 1 点～ 77 区 67 点に収録されており，JIS X 0208 の漢字と区点番号が重なっている．なお，常用漢字表の括弧書きの字体に対する委員会の見解は，基本的に 1983 年時点のものと変更はなく，したがって「巣」や「漢」などの常用漢字表の括弧書きの字体は，JIS X 0212 にも収録されていない[5]．

JIS X 0212 に収録されている文字は全て，JIS X 0208 には収録されていないはずだが，当時の規格票を丹念に読むと，誤ってこれらの両方に収録されてしまっている文字があることがわかる．例えば，当時の JIS X 0208 の解説によれば，JIS X 0208 では「σ」と「ς」[6]を 6 区 50 点に合併し，1 つの符号位置で両方を表していることになっている．ところがこれに反して，JIS X 0212 は,「ς」を 6 区 88 点に収録しているのである．すなわち「ς」は，この時点で，JIS X 0208 と JIS X 0212 の両方に収録されてしまった，ということになる．また，JIS C 6226 の 1983 年改正で消滅してしまった漢字に対して，この時点の JIS X 0208 の解説も「別の文字概念を採用したわけではない」と主張している．それにもかかわらず，消滅してしまった漢字のうち 28 字が，JIS X 0212 に再収録されている [27] のである．例えば,「鷗」は，JIS X 0212 の 76 区 31 点に収録されており，JIS X 0208 の

[1] JIS X 9052 ドットプリンタ用 24 ドット字形 (1983 年 9 月に JIS C 6234 として制定， 1987 年 3 月に改称) が規定する字数で，当時の JIS X 0208 の 6877 字に加え，縦書き字形 53 字が含まれている．

[2] ちなみに常用漢字表は，1981 年当時の大蔵省印刷局活字を用いていたため,「康熙字典」となっていた．人名用漢字の「熙」は，これに倣ったものと推測される．

[3] 1984 年 11 月の JIS C 6228 (現，JIS X 0202) 改正により，複数バイトコードに与えられるエスケープシーケンスは，指示先に応じて 4 種類となったが，残りの 3 つはここでは割愛する．

[4] 代表的なエスケープシーケンスは,「ESC 2/4 2/8 4/4」である．

[5] 74 区 18 点に「鬪」が収録されているが，常用漢字表の「闘 (鬪)」の括弧書きとは字体が異なっている．

[6] ギリシア語では，語末の σ は ς と記す．

18区10点の「鴎」の「文字概念」との間で，齟齬をきたす結果になっている．これらの齟齬は，JIS X 0208の本文や解説を書き変えることを，委員会が極度に嫌ったために，引き起こされてしまったものである．すなわち，1983年版のままの本文や解説を，1990年版にも忠実に反映してしまったことが，「熙」の問題も含めて，JIS X 0208の1990年改正の最大の問題点であるといえる．

後編目次

参考文献

[1] 吉田茂: 内閣告示第32号，当用漢字表; 官報，昭和21年11月16日号外，pp.1-2，1946年11月16日．
[2] 昭和21年11月16日官報号外内閣告示第32号中正誤; 官報，第6118号，p.63，1947年6月9日．
[3] 鈴木義男: 司法省令第94号，戸籍法施行規則; 官報，昭和22年12月29日号外(4)，pp.1-25，1947年12月29日．
[4] 吉田茂: 内閣告示第1号，当用漢字字体表; 官報，昭和24年4月28日号外(1)，pp.1-3，1949年4月28日．
[5] 昭和24年6月29日民事甲第1499號民事局長回答; 戸籍，第1號，p.7，1949年9月．
[6] 昭和24年6月29日民事甲第1501號民事局長回答; 戸籍研究，第24號，p.25，1949年9月．
[7] 吉田茂: 内閣告示第1号，人名用漢字別表; 官報，第7310号，p.402，1951年5月25日．
[8] 国語審議会(漢字部会): 当用漢字表審議報告; 国語審議会報告書—付 議事要録—，昭和27年4月～29年4月，文部省/秀英出版，p.6，1954年9月．
[9] 岡本榮一: 漢字テレタイプ機種別の解説; 新聞印刷技術，第13号，pp.2-11，1960年3月．
[10] 高橋達郎，森田朗，広田広三郎: 漢字入出力機器について; 情報管理，Vol.12，No.6，pp.309-319，1969年9月．
[11] 小田泰正: 国立国会図書館の機械化準備; 国立国会図書館月報，No.117，pp.2-9，1970年12月．
[12] 規格委員会: 1971年における規格委員会の活動; 情報処理，Vol.13，No.7，pp.495-503，1972年7月．
[13] 三木武夫: 内閣告示第1号，人名用漢字追加表; 官報，第14869号，p.5，1976年7月30日．
[14] 西村恕彦: 漢字のJIS; 標準化ジャーナル，No.171，pp.3-8，1978年5月．
[15] 奥野誠亮: 法務省令第51号，戸籍法施行規則の一部を改正する省令; 官報，昭和56年号外第88号，p.1，1981年10月1日．
[16] 鈴木善幸: 内閣告示第1号，常用漢字表; 官報，昭和56年号外第88号，pp.2-77，1981年10月1日．
[17] 長谷川均，岡崎健: 漢字CP/M®のコード体系; 情報処理学会マイクロコンピュータ研究会資料，26-2，1983年3月．
[18] 山下省治郎，安藤澄夫，成岡祥匡: 三菱パーソナルコンピュータ《MULTI 16》日本語CP/M-86の特長と機能; 三菱電機技報，Vol.57，No.11，pp.52-56，1983年11月．
[19] 野村雅昭: JIS C 6226情報交換用漢字符号系の改正; 標準化ジャーナル，Vol.14，No.3，pp.4-9，1984年3月．
[20] NEレポート: 日本語版Unixの標準案を提案; 日経エレクトロニクス，No.370，pp.111-113，1985年6月．
[21] 小川貴英: JIS no kaisetu to junet heno teian (In Japanese); fj.kanji，Message-ID: <134@tsuda.UUCP>，1986年3月．
[22] 諸橋轍次: 大漢和辞典(修訂版); 大修館書店，1986年4月．
[23] 小野芳彦: UNIXの日本語化の実現方法; 情報処理，Vol.27，No.12，pp.1393-1400，1986年12月．
[24] 日本科学技術情報センター三十年史編集委員会: 日本科学技術情報センター三十年史; pp.256-257，1988年3月．
[25] 田嶋一夫: JIS漢字補助集合案の設定と今後の課題; 情報処理学会研究報告，Vol.89，No.13 (情報学基礎研究会報告No.12)，89-FI-12-1，1989年2月．
[26] 長谷川信: 法務省令第5号，戸籍法施行規則の一部を改正する省令; 官報，平成2年号外第21号，pp.1-11，1990年3月1日．
[27] 内田富雄: JIS X 0212 (情報交換用漢字符号—補助漢字)の制定; 標準化ジャーナル，Vol.20，No.11，pp.6-11，1990年11月．
[28] 安岡孝一，安岡素子: 文字コードの世界; 東京電機大学出版局，1999年9月．
[29] 小林龍生，安岡孝一，戸村哲，三上喜貴[編]: インターネット時代の文字コード; bit別冊，共立出版，2001年4月．

著者略歴

安岡 孝一（やすおか こういち）

1965年2月18日生．1988年3月京都大学工学部情報工学科卒業．1990年3月京都大学大学院工学研究科情報工学専攻修士課程修了．同年4月京都大学大型計算機センター助手，1997年8月同助教授．2000年4月京都大学人文科学研究所附属漢字情報研究センター助教授，現在に至る．京都大学博士(工学)．1996年10月から2000年3月まで符号化文字集合調査研究委員会WG2委員．2000年9月いきなり二女の父となり，2000年12月から情報処理学会文字コード標準体系専門委員会委員．文字コードに関する著書に[28,29]がある．電子情報通信学会会員．yasuoka@kanji.zinbun.kyoto-u.ac.jp

解　説

日本における最新文字コード 事情 (後編)

※本原稿は著者によるゲラ刷りであり、最終稿とは異なっている。
本原稿を引用する場合は、必ず印刷された最終稿を確認すること。

安岡 孝一*

前 編 目 次

9. 国際符号化文字集合への道

日本で JIS X 0212 の開発が進んでいた (前編 8 章参照) 頃，ISO では，漢字を含む 32 ビットの文字コードを国際規格として制定する動きが，具体化していた．これが，ISO 10646 である．1989 年 1 月に作成された ISO 10646 のドラフト第 1 版では，32 ビット中の各 8 ビットで 20 〜 7E と A0 〜 FF のみを使用し，合計で $(95+96)^4$=1330863361 字を収容可能な文字コードとしていた．各 8 ビットには，上から順に Group，Plane，Row，Cell という名称がつけられており，Group 20 の Plane 20 にあたる (95+96) × (95+96) の部分は，BMP (Basic Multilingual Plane) と呼ばれていた．さらに日本の提案では，BMP の 20203021 〜 20207E7E の部分に，JIS X 0208 の 16 区 1 点〜94 区 94 点をそのまま埋め込むことで，JIS X 0208 の漢字部分と互換にすることが期待されていた．また，BMP の 2020B021 〜 2020FE7E に中国の GB 2312 [1] の 16 区以降を，2020B0A1 〜 2020FEFE に韓国の KS C 5601 (現，KS X 1001)[2] の 16 行以降を，それぞれそのまま埋め込むことを提案していた．これによって，日中韓の主要な 94 × 94 の文字コードと，ある程度互換性が取れるだろう，と考えていた [5] のである．

この提案に，中国は反対であった．中国は既に GB 2312 の拡張を進めており，1988 年 7 月実施の GB 8565 において，13 〜 15・90 〜 94 区に漢字を追加収録していた．しかも，GB 8565 での追加に加え，さらに 12 区にも漢字を追加した文字コード [3] を，CCITT (現，ITU-T) の Recommendation T.101 に含めるよう，提案していたのである．すなわち，GB 2312 の 16 区以降だけでは，既におこなっている他の国際提案に比べて，非常に不十分な文字コードとなりかねなかった．また，中国国内では 1988 年 3 月に，現代漢語通用字表 [4] が発布されていた．この 7000 字のうち，407 字が GB 2312 には収録されておらず，その意味でも GB 2312 では不十分だったのである．

しかし，中国はこのような事情はおくびにも出さず，当時のアメリカやカナダの代表者が主張していた漢字統合を，支持する立場を取った．すなわち，ドラフト第 1 版では，例えば「一」という漢字ですら，2020306C と 2020D23B と 2020ECE9 の 3ヶ所に収録されてしまう．これは，検索上も非常に不便だし，そもそも文字の一意な符号化という文字コードの基本理念に反するので，一つの符号位置に統合すべきである，という立場である．結局，1991 年 6 月におこなわれた投票の結果，ISO 10646 のドラフト第 1 版は否決され，ドラフトの改訂がおこなわれることとなった．

ドラフト第 2 版では，Plane，Row，Cell の値の制限はなくなり，00 〜 FF の全ての値が使用できることになった．Group に関しては，00 〜 7F のみが使用できることになっており，実質上 31 ビットの文字コードとなっている．BMP は，Group 00 の Plane 00 に移され，BMP の 00004E00 〜 00009FFF に漢字が収録されることが決まった．漢字を統合するかどうかは，日中韓の判断に委ねられ，1991 年 7 月，CJK-JRG (China, Japan, Korea-Joint Research Group) の最初の会議が開かれた．

この段階で，中国は既に統合漢字表を作成しており，それを CJK-JRG に持ち込んだが，日本はこれに反発，漢字統合はルール・ベースでおこなうことを主張し，漢字の形による統合ルールを提示した [3]．このルールによ

* 京都大学人文科学研究所附属漢字情報研究センター

[1] 1981 年 5 月実施，2001 年 3 月廃止 (GB 18030 に移行)．94 × 94 の文字コードで，16 〜 87 区に簡化字を収録していた．

[2] 1974 年 9 月制定．1987 年の改正で 94 × 94 の文字コードとなった．16 〜 40 行 (符号位置に対して，区点ではなく行列という名称を用いている) にハングル，42 〜 93 行に漢字を収録している．

[3] 現在「ISO-IR 165」と呼ばれている．

[4] 正式名称は，现代汉语通用字表．中国における現代漢字の規範化すなわち一種の漢字制限を目しており，情報処理，機械処理，印刷出版事業の発展に応えるべく，通常用いられる漢字 7000 字を収録している．

れば，例えば「饮」と「飲」と「飲」のような部首の新旧の違いにあたるものは，統合しないことになっていた．これに対し中国は,「饮」と「飲」は簡化字と繁体字の関係にあるので，分離するのが適当だが,「飲」と「飲」は単なる書体差だから統合すべきだ，と主張した．そこで日本は，統合漢字の採録元となる各国の国内規格において，既に別々の符号位置を与えられている漢字どうしは，たとえ統合の対象であっても分離する，という原規格分離を条件に，中国の主張を受け入れた．この結果,「飲」と「飲」は統合の対象ではあるが，原規格 (統合漢字の採録元) の 1 つである JIS X 0208 において，それぞれ 16 区 91 点と 61 区 27 点という別々の符号位置が与えられているので，統合をおこなわない，ということになったのである．

また，日本は，日本からの原規格を JIS X 0208 と JIS X 0212 の 2 つに限定することを表明，他の国にも，原規格および収録漢字の範囲を明確にするよう迫った．この際，日本のメーカー外字や人名用漢字許容字体表 (前編 4 章参照) を採録すべきではないか，との意見があったが，日本は国内規格のみに固執し，これらの意見を受け入れなかった．中国は，形の上では原規格の範囲を国内規格に限定したかのように見えたが，実は原規格に含まれていない漢字の収録を勝手におこなっていたことが，最近になって判明している [4,6]．韓国にとっては，原規格の限定はそう問題ではなかったが，原規格分離は多少ややこしい問題を含んでいた．KS C 5601 の漢字は，読みのハングル順に並べられており，例えば「類」には複数の読みがあることから，55 行 30 列 (読みは𠮟) と 75 行 26 列 (読みは𤋮) の 2ヶ所に収録されていたのである．原規格分離に完全にしたがうことにすると，全く同じ形の「類」が，複数収録されなければならない．しかも，KS C 5601 には，このような漢字が 262 組 530 字もあるのである．結局，韓国は，KS C 5601 にダブって収録されている漢字に関して，原規格分離を要求しなかった．この結果，CJK-JRG が 1992 年 3 月に完成した CJK 統合漢字 20902 字には,「類」は 1 字しか含まれず，しかも「類」と統合されていた．

ところがこれに対し，アメリカの代表者が異を唱えた．確かに漢字統合という側面からは,「類」を複数収録することがあってはならないから，CJK 統合漢字 20902 字を，そのまま ISO 10646 の 00004E00 ～ 00009FA5 に収録すべきである．しかしながら，既存の各国規格とのラウンド・トリップ・コンバージョンも何らかの形で保証すべきではないか，という意見である．ラウンド・トリップ・コンバージョンの保証とは，すなわち，KS C 5601 で書かれたデータを ISO 10646 に変換し，それをまた KS C 5601 に再変換した場合に，完全に元に戻る必要がある，という考え方である．しかもアメリカの代表者は，1992 年 6 月に，Unicode 1.0[2] という 16 ビットの文字コードを発表していた．そこでは「類」は 985E に収録されていると同時に，F9D0 にも収録されていた[1]のである．Unicode は，ISO 10646 の BMP と互換である，というのがウリであったから，アメリカの代表者としては，何としても ISO 10646 の 0000F900 ～ 0000FA2D に，互換漢字を収録する必要があったのである．

結局，1993 年 5 月に制定された ISO 10646 では，CJK 統合漢字 20902 字が 00004E00 ～ 00009FA5 に収録されると同時に，互換漢字 302 字が 0000F900 ～ 0000FA2D に収録された[2]．この結果，JIS X 0208 や GB 2312 とのラウンド・トリップ・コンバージョンは，CJK 統合漢字での原規格分離によって保証されるが，KS C 5601 とのラウンド・トリップ・コンバージョンは，互換漢字によって保証される，というヤヤコシイことになってしまったのである．例えば,「禄」と「祿」は本来統合されるべきものだが，JIS X 0208 に対する原規格分離によって，00007984 と 0000797F という別々の符号位置が与えられている．その上，KS C 5601 とのラウンド・トリップ・コンバージョンのために，0000F93C にも「祿」が収録されているのである．

さらに，ISO 10646 の翻訳規格である JIS X 0221 (1995 年 1 月制定) には，日本の独自規定[3]である，日本文字部分レパートリが盛り込まれた．日本文字部分レパートリは，JIS X 0221 の文字 (すなわち ISO 10646 の文字) のサブセットであり，この規格を日本で実装する際に，どのような部分実装を推奨するかを規定したものだと言える．日本文字部分レパートリは，基本日本文字集合，追加非漢字集合，追加漢字集合，補助漢字集合，その他の漢字集合と，互換用全角英数字集合，互換用半角片仮名集合，の 7 つのサブセットからなっている．このうち，基本日本文字集合は，JIS X 0208 の全文字に ASCII を追加した 6884 字[4]であり，追加漢字集合は，JIS X 0212 策定の際に，8 つ以上の漢字表に含まれていた漢字を中心とする 918 字[5]，補助漢字集合は，JIS X 0212 の残りの漢字 4883 字である．すなわち，日本文字部分レパートリというのは，JIS X 0208 と JIS X 0212 の各文

[1] KS C 5601 との変換においては，985E が 55 行 30 列，F9D0 が 75 行 26 列の「類」にそれぞれ対応していた．

[2] 「類」は，0000985E と 0000F9D0 に収録され，Unicode と互換となった．なお，0000985E の「類」は,「類」と統合されている．

[3] 2001 年 4 月改正の JIS X 0221 では，日本文字部分レパートリは，規定から参考に格下げとなった．

[4] ASCII と JIS X 0208 でダブっている文字は，互換用全角英数字集合に追い出されている．また，当時の JIS X 0221 の附属書 1 には「BASIC JAPANESE (基本日本文字集合) は，JIS X 0201 及び JIS X 0208 で規定された図形文字のうち，全角英数字及び半角片仮名を除いたものとする」とあったが，この記述は誤りであり，2001 年 4 月に全面改正された．

[5] 当時の JIS X 0221 の解説では，なぜか 890 字となっていた．

字を JIS X 0221 から選び出したものに過ぎず，日本での漢字使用の実態に即したものではない．例えば，人名用漢字許容字体表の「巣」や吐噶喇列島の「噶」は，いずれも JIS X 0221 に収録されている[6]にもかかわらず，日本文字部分レパートリでは，その他の漢字集合に入れられてしまっているのである．

以上，述べてきたとおり，ISO 10646 の制定は，日本の漢字を国際規格に収録する絶好のチャンスであった．これに対し，日本が取った態度は，国内規格の JIS X 0208 と JIS X 0212 に固執し，実際に収録すべき漢字の調査も追加要求もおこなわない，というものであった．この結果，ISO 10646 の漢字部分は，中国にいいようにされてしまい，それが現在も続いている，というのが正直な印象である．

10. MIME と ISO-2022-JP

インターネットにおける電子メールのフォーマットは，RFC 822 (現，RFC 2822) によって定義されている．しかし，RFC 822 は，文字コードとして 7 ビットの ASCII を想定しているため，それ以外の文字コード，特に 8 ビットの文字コードによるテキストは，電子メールでは直接送受信することができなかった．この問題を解決するために，1992 年 6 月に RFC 1341 (現，RFC 2045) として提案されたのが，MIME (Multipurpose Internet Message Extensions) である．

MIME の最大の特長は，メールヘッダに Content-Type フィールドを追加し，メール本文にどのようなものが含まれているかを，明示できるようにしたことである．メール本文の文字コードは，Content-Type 中の charset パラメータで明示されると同時に，もし 8 ビットコードであれば，Content-Transfer-Encoding フィールドに 8bit を指定すればよいことになっている．Content-Transfer-Encoding には，他に quoted-printable と base64 が指定可能であり，これらはいずれも 8 ビットデータを，7 ビットに押し込めるための形式である．これらにより，8 ビットデータを 7 ビットの通信路に流す場合でも，Content-Transfer-Encoding が 8bit から quoted-printable もしくは base64 に自動変換されて，ビット落ちなしに送受信できるようになっている．

また，メールヘッダに対しては，RFC 1341 と同時に発表された RFC 1342 (現，RFC 2047) によって，ASCII 以外の文字列を扱う手法が提案された．「=?」と「?=」の間に，文字コードと変換形式 (Q あるいは B) および変換後の文字列を埋め込むことで，ASCII 以外の文字列を英数字と若干の記号で表現する手法である．これによって，メールヘッダの From フィールドや To フィールドに，ASCII 以外の文字を使うことができるようになったのである．

これを受ける形で 1993 年 6 月，JUNET コード (前編 7 章参照) は RFC 1468 となった．ただし，文字コードの名前は，JUNET コードではなく，ISO-2022-JP と名づけられた．JP ドメイン[1]における ISO 2022 準拠[2]の文字コード，という意味だったが，JUNET コードとは違い，ISO-2022-JP は ISO 2022 準拠ではない．JUNET コードでは「1B 24 42」を，JIS X 0208 (旧称，JIS C 6226) の 1983 年版 (前編 6 章参照) への切り替えとしていたが，ISO-2022-JP ではこれを，JIS X 0208 の 1990 年版 (前編 8 章参照) への切り替えとみなしてもよい，ということに変更してしまった．すなわち，本来「1B 26 40 1B 24 42」であるべきところを，「1B 24 42」としてしまったために，ISO-2022-JP は，ISO 2022 非準拠となってしまったのである．

さらに 1993 年 12 月には，JUNET コードに，JIS X 0212，GB 2312， KS C 5601 (現，KS X 1001) などへの切り替えエスケープシーケンスを加えた 7 ビットコード ISO-2022-JP-2 が，RFC 1554 として提案された．ISO-2022-JP-2 は「1B 24 42」を，あくまで 1983 年版の JIS X 0208 への切り替えとみなしており，ISO 2022 準拠ではあるが，JIS X 0208 の 1990 年改正で追加された「凜」と「熙」は使用できない[3]．1997 年 11 月に RFC 2237 として提案された ISO-2022-JP-1 [4]も，ISO 2022 準拠ではあるが，やはり「凜」と「熙」は使用できない．

現在，日本の電子メールで用いられている文字コードは，JUNET コード以来の歴史から，ISO-2022-JP がほとんどである．ただし，JIS X 0212 の文字を含むような電子メールにおいては，「凜」と「熙」の事情もあってか，ISO-2022-JP-1 や ISO-2022-JP-2 はほとんど用いられず，もっぱら 8 ビットコードの UTF-8 (次章参照) が用いられるようである．

11. ISO 10646 の Amendment

ISO 10646 には，1993 年 5 月の制定後，多くの Amendment が追加されており，文字コードとしては，次々に変更が加えられている．

Amendment 1 (1996 年 10 月制定) である UTF-16 は，31 ビットコードである ISO 10646 を，16 ビット環境に

[6]「巣」は 00005DE2 に，「噶」は 00005676 に収録されている．

[1]日本国内のメールアドレスの多くは，トップドメインが junet だったが，1990 年 4 月に全て jp に切り替えられた．その後，1991 年 10 月に，JUNET はネットワークとしての活動を，事実上終了した．

[2]すなわち，JIS X 0202 準拠．なお，ISO 2022 と JIS X 0202 (旧称，JIS C 6228) の関係は，前編 2 章参照．

[3]JIS X 0208 の「凜」と「熙」は確かに使用できないが，KS C 5601 の「凜」と「熙」(符号位置はそれぞれ 55 行 47 列と 93 行 87 列) は使用できる．

[4]JUNET コードに，JIS X 0212 の「1B 24 28 44」を加えた 7 ビットコード．

マッピングするための方法であり，16 ビットコードである Unicode にどのように変換するかを，示したものであるといえる．ISO 10646 の 00000000 ～ 0000D7FF と 0000E000 ～ 0000FFFF は，UTF-16 では，0000 ～ D7FF と E000 ～ FFFF にマッピングされる．また，ISO 10646 の 00010000 ～ 0010FFFF は，UTF-16 では，D800 DC00 ～ DBFF DFFF にマッピングされる．ISO 10646 の 00110000 以降の文字は，UTF-16 では使用できない．

Amendment 2 (1996 年 10 月制定) である UTF-8 は，31 ビットコードである ISO 10646 を，8 ビット環境に完全にマッピングするための方法である．マッピング方法は非常に複雑で，00000000 ～ 0000007F は 00 ～ 7F に，00000080 ～ 000007FF は C2 80 ～ DF BF に，00000800 ～ 0000FFFF は E0 A0 80 ～ EF BF BF に，00010000 ～ 001FFFFF は F0 90 80 80 ～ F7 BF BF BF に，00200000 ～ 03FFFFFF は F8 88 80 80 80 ～ FB BF BF BF BF に，04000000 ～ 7FFFFFFF は FC 84 80 80 80 80 ～ FD BF BF BF BF BF に，それぞれマッピングすることになっているが，2 バイト目以降には 80 ～ BF しか用いられないことを理解すれば，機械的には読みやすい形になっている．しかも，ISO 10646 の 00000020 ～ 0000007E は，ASCII の 20 ～ 7E を収録したものであることから，UTF-8 は ASCII と上位互換な 8 ビット 1 ～ 6 バイトコードであるといえる．

ISO 10646 における最も大きな変更は，ハングルの符号位置の移動である．1993 年 5 月の制定時には，ハングルは 00003400 ～ 00004DFF に 6656 字が収録されており，このうち 00003400 ～ 00003D2D の 2350 字は，KS C 5601 の 16 ～ 40 行と全く同じ順序に並べられていた．ところが KS C 5601 の附属書 3 には，現代ハングル 11172 字を全て符号化可能な 8 ビット 2 バイトコード Johab が示されて[1]おり，これとの対応が問題になったのである．そこで韓国は，ハングルの符号位置を 0000AC00 ～ 0000D7A3 に移動して，そこに現代ハングル 11172 字を全て収録するよう要求した．この要求は結局，1996 年 8 月に承認され，1997 年 11 月に Amendment 5 として制定された．非常に大きな非互換変更が，ISO 10646 に対してなされたのである．

ハングルが移動した後の空き領域に対しては，漢字が追加収録されることが，1996 年 11 月に決定された．ただし，BMP への漢字収録は，この 00003400 ～ 00004DFF を最後にするという条件付きで，である．これに対し，日本はメーカー外字 658 字を提案したのみだった．しかも，提案はメーカー外字の一部にとどまっており，例えば「堅」(「野」の異体) のような，人名での使用頻度が高く，しかも複数のメーカーが実装しているような外字が，提案から漏れてしまっている．あるいは，10^{24} を意味する「秭」のように，メーカー外字に含まれていなかった漢字は，この提案には含まれていない．結局，中国 (および台湾) から提案された漢字を中心に，6582 字の CJK 拡張漢字が，1998 年 4 月に確定し，これを BMP の 00003400 ～ 00004DB5 に収録した Amendment 17 が，1999 年 7 月に制定された．BMP への漢字追加はこれが最後であり，これ以後の漢字収録は，Group 00 の Plane 02 におこなうことになったのである．

なお，これらの Amendment を全て含める形で，ISO 10646 は，2000 年 10 月に改正されており，これに合わせて，JIS X 0221 も，2001 年 4 月に改正[2]された．

12. JIS X 0208 の 1997 年改正

1997 年 2 月，JIS X 0208 の 3 度目の改正[3]がおこなわれた．改正の中心となったのは，1994 年 4 月に発足した符号化文字集合調査研究委員会 WG2 である．改正の主眼は，JIS X 0208 の各符号位置が，どういう文字の情報交換を意図しているかの明確化であり，収録されている 6879 字それ自体については，一切変更をおこなっていない．明確化の柱は，収録文字の明確化，字体の合併の明確化，符号化方法の明確化，の大きく 3 つである．

収録文字の明確化に関しては，徹底した典拠調査を，WG2 はおこなっている．行政情報処理用標準漢字選定のための漢字使用頻度および対応分析結果，という，JIS X 0208 のいわば原典 (前編 3 章参照) を発掘したのは，WG2 の最大の成果であるといえよう．あるいは WG2 は，1995 年 8 月時点の国土行政区画総覧に，加除式出版除去分 33000 ページを加除して，1972 年 11 月時点の国土行政区画総覧を復元し，ここから JIS X 0208 の各文字がどのように収録されたかを明らかにしている．

字体の合併の明確化に対しては，包摂規準が規定された．包摂規準は，1978 年制定の作業時におこなわれた字体の合併を中心に，その後の JIS C 6226 ～ JIS X 0208 規格票の字体変遷を加味したものである．これにより，例えば 20 区 54 点は，規格票に示されている「間」のみならず,「閒」をも包摂していることが，明らかとなった．各符号位置が表す字体の範囲は，これ以前は合併という形で例示的にしか示されていなかったが，包摂規準を規格に含めることで，初めて全てが明示されることになったのである．

符号化方法の明確化としては，日本語版 EUC (前編 7 章参照)，シフト JIS (前編 5 章参照)，ISO-2022-JP が，初めて規定として盛り込まれた．ただし，日本語版 EUC は，国際基準版・漢字用 8 ビット符号という名

[1] 1987 年改正以前の KS C 5601 は，94 × 94 の文字コードではなく，16 ビットコードとして Johab を規定していた．

[2] 正確には，JIS X 0221 を廃止し，新たに JIS X 0221-1 を制定，という形が取られた．

[3] JIS X 0208 (旧称，JIS C 6226) の制定については，前編 3 章参照．2 度の改正については，前編 6 章・8 章参照．

称になっており，JIS X 0201の片仮名を含んでいない．また，ISO-2022-JPは，本来7ビットコードであるべきところを，誤って8ビットコードとして規定してしまった．この結果，一部の電子メール用ソフトウェアが，ISO-2022-JPのContent-Transfer-Encodingを8bitとしてしまう，という事態が起こっており，その意味では，符号化方法の明確化に失敗している．

JIS X 0208の1997年改正は，あくまで規格の明確化を主眼としたものだったが，これにより，JIS X 0208という規格の持つ問題点が，浮き彫りになった．特に，JIS X 0208は本来，日本の地名や人名に用いる漢字を収録しようと意図していたが，その意図が十分に実現されていないことが明らかになった．これが，JIS X 0208の拡張規格の開発，という方向に向かっていくのである．

13. JIS X 0213の開発

1996年7月，符号化文字集合調査研究委員会WG2は，JIS X 0208を拡張する規格の開発を表明した．現代日本語文脈で安定して用いられる文字を，できうる限り収録した文字コードJIS X 0213の，開発開始である．JIS X 0213は，JIS X 0208の拡張規格であることから，JIS X 0208の1997年改正で掲げられた3つの柱を，それぞれ拡張することとなった．すなわち，収録文字の拡張，包摂規準適用除外の拡張，符号化方法の拡張，である．

収録文字を拡張するにあたって，WG2は，日常生活に用いられる文字，地名，人名を中心に，用例の収集をおこなっている．日常生活に用いられる文字としては，1997年度および1998年度の小学校・中学校・高等学校の文部省検定済教科書の全ページを，委員全員がよってたかってチェックし，異なり字数にして漢字3241字，非漢字1417字を得ている．地名としては，国土行政区画総覧の除去分33000ページのチェックに加え，国土地理協会や国土地理院の協力で，漢字477字を得ている．人名は，当時のNTT電話帳の全データから，漢字3241字を得ている．これらに加え，実際の文字の使用例を広く一般から募り，最終的には，異なり字数にして漢字13083字，非漢字2654字の，JIS X 0208未収録文字を収集している．

包摂規準適用除外の拡張に対しては，多数の要望が寄せられていた.「巣」と「巢」,「漢」と「漢」,「間」と「間」などは，JIS X 0208では包摂されていて書き分けられないのだが，これを書き分けることができるようにしてほしい，というものである．これに対しWG2は，人名用漢字許容字体(前編4章参照)と常用漢字表の括弧書きの字体に限って包摂規準を曲げることを決定し，それ以外の漢字に関しては，包摂規準の変更を基本的にはおこなわなかった．この結果,「巣」と「巢」あるいは「漢」と「漢」は，JIS X 0213ではそれぞれ別々の符号位置[1]に収録され，書き分けることが可能になった．つまり，常用漢字および人名用漢字と人名用漢字許容字体は全て，JIS X 0213で書き分けることが可能となったのである．しかし,「間」と「間」は，これらを一文の中で使い分ける例[2]が，著名な文芸作品[1]中に発見されていたにもかかわらず，人名用漢字許容字体でも常用漢字表の括弧書きの字体でもなかったため，JIS X 0213においても書き分けることができなくなってしまった．

符号化方法の拡張については，シフトJISと日本語版EUCをそれぞれ拡張した，Shift_JISX0213とEUC-JISX0213が，附属書として規定に盛り込まれた．これらの拡張のために，WG2は，JIS X 0213の構造を以下のように定めた．基本構造は，94 × 94の文字表を2面とする．1面は，94 × 94の全てを使用するが，JIS X 0208の上位互換となるように文字を配置する．2面は，1・3～5・8・12～15・78～94区の26 × 94=2444字の部分だけを使用し，残りの部分は空きとする．このようにしておいて，EUCの文字集合0にASCII，文字集合1にJIS X 0213の1面，文字集合2にJIS X 0201の片仮名，文字集合3にJIS X 0213の2面を配置したものを，EUC-JISX0213とする．こうすれば，EUC-JISX0213は，日本語版EUCの拡張であると同時に，EUC-JP [3]とはコードの上では衝突しない．しかも，EUC-JP用に書かれたソフトウェアを，簡単にEUC-JISX0213に流用できる．さらにShift_JISX0213では，1面1区1点～1面62区94点を `81 40`～`9F FC`，1面63区1点～1面94区94点を `E0 40`～`EF FC`，2面1区1～94点を `F0 40`～`F0 9E`，2面3区1点～2面5区94点を `F1 40`～`F2 9E`，2面8区1～94点を `F0 9F`～`F0 FC`，2面12区1点～2面15区94点を `F2 9F`～`F4 9E`，2面78区1点～2面94区94点を `F4 9F`～`FC FC`に対応させることで，シフトJISの拡張を実現している．なお，JIS X 0202準拠のエスケープシーケンス[4]を用いて，JIS X 0213の1面・2面およびASCIIを切り替える7ビットコードISO-2022-JP-3も，附属書に盛り込まれており，ISO-2022-JPからの移行が，多少は考慮されている[7]．

ここまでの作業の後，WG2は，非漢字659字，第3水準漢字1249字，第4水準漢字2436字を選定し，JIS X 0208の6879字と合わせて，11223字の文字集合[5]を完成した．これと並行してWG2は，これら11223字のうち，ISO 10646に含まれていない文字全てを，ISO

[1]「巣」は1面33区67点に,「巢」は1面84区8点に,「漢」は1面20区33点に,「漢」は1面87区5点に,それぞれ収録された．

[2][1]の68ページ15行目に「かやうに閒覩の態度で，有と無の間に逍遙してゐるのだらう」とある．

[3]日本語版EUCの文字集合3に，JIS X 0212を追加した，8ビット1～3バイトコード．

[4]JIS X 0213の1面と2面を，JIS X 0202で用いるための代表的なエスケープシーケンスは，それぞれ「ESC 2/4 2/8 4/15」と「ESC 2/4 2/8 5/0」である．

[5]1999年7月の時点では11225字としていたが，翌月「M̀」と「m̀」の2字を削除した．

10646 の BMP に追加するよう，国際提案することにした．というのも，この時点での ISO 10646 の実装の多くは，1998 年 7 月に発売された Microsoft Windows 98 日本語版を含め，UTF-16 の BMP 部分 (別名，UCS-2) のみに限定されており，BMP 以外の部分はほとんど実装されていなかったからである．しかし，ISO 10646 の漢字に関しては，CJK 拡張漢字 6582 字 (1998 年 4 月確定) が最後の BMP 追加ということになっており，この時点での漢字追加は，もう BMP 以外の部分におこなうしかなかったのである．そこで WG2 は，1 字でも多くの漢字を BMP に追加すべく，奇策を用いることにした．常用漢字表の括弧書きの字体や人名用漢字許容字体は，互換漢字への追加を提案することにしたのである．例えば，「漢」と「漢」は，ISO 10646 では 00006F22 に統合されているので，JIS X 0213 とのラウンド・トリップ・コンバージョンを保証するために,「漢」を互換漢字に追加する，という提案である．しかもこのようにすれば，例えば「類」は，既に互換漢字の 0000F9D0 に収録されている[6]ので，追加提案の必要すらなくなるのである．最終的に WG2 は，非漢字 128 字，漢字 303 字，互換漢字 60 字[7]を，ISO 10646 に追加提案することにした．

このように，多大な労力を傾注して開発された JIS X 0213だったが，最終審議の段階になって，とんでもないドンデン返しが待ちうけていたのである．

14. JIS X 0213 の制定

原案作成委員会が作成した規格原案は，日本工業標準調査会の最終審議を受けた後，主務大臣名で制定される．これが JIS 制定の大きな流れである．符号化文字集合調査研究委員会が作成した JIS X 0213 原案も，1999 年 9 月に開催された日本工業標準調査会情報部会での最終審議の後，JIS X 0213 として制定される予定であった．

ところが，この情報部会において，日本 IBM と日本電気が，JIS X 0213 の制定に猛反対したのである．反対理由は大きく 2 つである．1 つは，シフト JIS の空き領域には，各社がメーカー外字を入れて拡張しているので，JIS X 0213 が Shift_JISX0213 を規定すると，それとの間で文字化けが発生し，結果的にユーザーの資産が守れない，という主張である．もう 1 つは，文字コードのサポートという点では，各社とも ISO 2022 から ISO 10646 に移行しており，今さら ISO 2022 準拠の文字コードを JIS として制定すべきではない，という主張である．主張それだけを聞けば，もっともなものであり，賛同も多く得られることだろう．

しかし，ここまでの解説を読んできた読者ならば，これらの主張に疑問を感じるはずである．文字コードのサポートが，本当に ISO 10646 に移行しているならば，なぜ日本電気は,「堅」など自社の JIPS 拡張文字を，ISO 10646 に追加提案しなかったのか．ユーザーの資産を守る，というならば，ISO 10646 の BMP に追加提案すべきだったはずなのに，なぜその努力を怠ったのか．あるいは，なぜ日本 IBM は，IBM 拡張文字の「𤋮」を，ISO 10646 に追加提案しなかったのか．しかも，IBM 拡張文字と ISO 10646 との変換表において,「𤋮」を 000053DD の「𤋮」に誤変換[1]し，この結果，発生する文字化け[2]で，ユーザー資産を破壊し続けているのは，なぜなのか．ISO 10646 だユーザー資産だと，もっともらしい理屈をつけてはいるが，つまるところ，Shift_JISX0213 が規定となった場合，それをサポートするのが面倒だ，というだけのことではないのか．

実際，1999 年 9 月の情報部会は，紛糾して収拾がつかなくなり，JIS X 0213 は継続審議となった．そして翌月の情報部会で，Shift_JISX0213，ISO-2022-JP-3，EUC-JISX0213 の各附属書を，規定から参考に格下げする，という形で，何とか JIS X 0213 は，審議を通過することになった．この結果，JIS X 0213 を，規定を満たす形で実装するには，国際基準版・漢字用 8 ビット符号[3]などの方法しかなくなってしまった．すなわち，Shift_JISX0213 や ISO-2022-JP-3 だけを実装した場合には，規格違反となってしまうのである．しかも，1999 年 10 月の情報部会は，今後の日本の文字コードに関しては，ISO 2022 から ISO 10646 に移行する，ということを確認した．つまり，各国がそれぞれに開発した文字コードを持ち寄る，という ISO 2022 のやり方を放棄し，各国がそれぞれに文字を持ち寄る，という ISO 10646 のやり方に移行する，ということである．言いかえれば，この JIS X 0213 を最後に，日本では独自の文字コードを開発せず，全てを国際規格の ISO 10646 (とその翻訳規格の JIS X 0221) に委ねる，ということを，規格の最終審議をおこなう情報部会みずからが，宣言したわけである．JIS X 0213 は，2000 年 1 月に制定され，日本独自の文字コード開発は，ここに幕を閉じた．

[6] そもそもは，KS C 5601 (現，KS X 1001) とのラウンド・トリップ・コンバージョンを保証するためのものである．

[7] 「概」と「概」は，ISO 10646 において 00006982 と 000069EA に分離されているにもかかわらず，WG2 は「概」を誤って互換漢字に追加提案してしまった．

[1] 「𤋮」は，キュウという音で「急いでいく」意．これに対し，ISO 10646 の 000053DD「𤋮」は，コウという音で「つつしみ告げる」意．これらは，形も音も義も異なる全くの別字である．

[2] この文字化けは，Microsoft Windows 98 日本語版のワードパッドで，FA8F の「𤋮」を 72 ポイントで入力し，フォントを MS P ゴシックから MS P 明朝に切り替えるだけでも，発生が確認された．

[3] EUC-JISX0213 から，文字集合 2 の JIS X 0201 片仮名を除いた符号．

15. そして最新文字コード事情

その後，JIS X 0213の漢字は，全てISO 10646に収録されることが，国際会議の場で決定された．ただし，BMPへの追加収録は,「漢」のような互換漢字に限られ,「堅」や「秤」はGroup 00のPlane 02への収録[4]である．JIS X 0213の非漢字に関しては，その多くがBMPへの収録を予定されているものの，例えば「æ̀」は,「æ」と「`」の合成にすべきだ，という意見のために，現時点では収録を阻まれている[5]．なお，2001年11月に発売されたMicrosoft Windows XP日本語版では，これらGroup 00のPlane 02の漢字を含め，UTF-16で表すことのできる全ての文字を処理する仕掛けが搭載されている．したがって，BMP以外の部分に収録される予定の漢字も，フォントさえ実装されれば，今後は徐々に使えるようになっていくだろう．

日本国内では，2000年12月に国語審議会(現，文化審議会国語分科会)が答申した表外漢字字体表[8]に対し，文字コードをどのようにすべきか，ということが目下の関心事である．表外漢字字体表は,「法令，公用文書，新聞，雑誌，放送等，一般の社会生活において表外漢字を使用する場合の字体選択のよりどころ」として，1022字を示したものだが,「将来的に文字コードの見直しがある場合，表外漢字表の趣旨が生かせる形での改訂が望まれる」と，文字コードへの反映にはっきりと言及している．しかも，表外漢字の字体に関する問題は

(1) 一般の書籍類や教科書などで用いられている「鷗」や「濵」がワープロ等から打ち出せないこと
(2) 仮に「鴎」と「鷗」の両字体を打ち出すことができたとしても，どちらの字体を標準と考えるべきかの「字体のよりどころ」がないこと

の2点にまとめられる，というのである．日本のワープロは，東京芝浦電気(現，東芝)の「JW-10」(1979年2月発売)以来，基本的にJIS X 0208 (旧称，JIS C 6226)に準拠してきた．したがって，表外漢字の字体に関する問題は，すなわちJIS X 0208の字体に関する問題だ[10]，ということになる．

ならば，JIS X 0208に例示されている字体を，表外漢字字体表の印刷標準字体に合わせるように，JIS X 0208を改正すれば，問題はなくなるのか．残念ながら，それはあまりに無謀であり，かつ危険な選択である．一例を挙げよう．JIS X 0208の17区19点の例示字体は「嘘」である．これに対し，表外漢字字体表の印刷標準字体は「嘘」である．そこで，JIS X 0208の17区19点の例示字体を,「嘘」に変更したとしよう．JIS X 0208としては，この例示字体を変更したところで，包摂規準の範囲内だから，特に問題はないように思える．しかし，ISO 10646 (およびその翻訳規格のJIS X 0221)との関係を考えると，そうは問屋がおろさない．ISO 10646では,「嘘」と「嘘」は，それぞれ別の符号位置である00005618と00005653に，収録されているのである．つまり，JIS X 0208の17区19点の例示字体を変更すると，JIS X 0208とISO 10646との対応が変わってしまう．これは，ISO 10646上でJIS X 0208の文字を扱っているOS [1]にとっては，致命的ともいえる変更である．そのようなOSでは，JIS C 6226の1983年改正(前編6章参照)と同様に，新たな文字化けを発生させることになるからであり，したがって一部のメーカーは，変更に追随しない可能性もあるだろう．そのような意味で，JIS X 0208の改正は，あまりに無謀で危険な選択である．

では，JIS X 0208の17区19点の「嘘」はそのままにして，別の空き領域に「嘘」を追加するのは，どうだろう．表外漢字字体表の印刷標準字体1022字のうち，JIS X 0208の例示字体とは違うもの，およびJIS X 0208に収録されていない「鄧」や「藤」を，空き領域に追加するのである．しかしこれも，JIS X 0208への文字追加と考えられるJIS X 0213に対し，日本工業標準調査会情報部会が示した反応を考えると，現実的な選択とは思われない．しかも，情報部会にとっては，ISO 2022準拠の文字コードの開発は，もはや終了しているのである．

では，情報部会の言うとおり，今後はISO 10646のサポートだけをおこなうのならば，表外漢字字体表への対応は，どうなるだろう．これならば話は簡単で，表外漢字字体表1022字はISO 10646に全て含まれているので，実装上の細かい字形差を除けば特に問題はない，ということになる．できることといえば，せいぜい，JIS X 0221の日本文字部分レパートリに，表外漢字字体表を反映させることぐらいである．しかし，日本文字部分レパートリは，2001年4月改正のJIS X 0221で，参考に格下げとなった．その上，日本文字部分レパートリには，常用漢字表も人名用漢字別表も人名用漢字許容字体表も，全く反映されていないのである．そんな状態の日本文字部分レパートリに，表外漢字字体表を反映させたところで，大した意味は無いように思える．

あるいは，規格の改正をおこなわず，表外漢字字体表1022字と文字コードとの対応だけを示す，という視点からならば，JIS X 0213が俎板に上ってくる可能性はある．というのも，JIS X 0213には，表外漢字字体表1022字が全て含まれているからである．しかしそれも，「包摂規準の範囲内で」というただし書きが付くものであり,「JIS X 0213の例示字体は変更しない」という条件付きである．なぜなら，JIS X 0213の例示字体を変更

[4]「漢」は0000FA48,「堅」は0002131B,「秤」は00025771に収録予定である．ちなみに「叛」も，中国からの提案で，00020AF3に収録予定である．また,「類」が00029516にも収録予定であり，0000985Eの「類」との混乱が予想される．

[5]ちなみに「ǽ」は，既に000001FDに収録されている．

[1]現在のMicrosoft Windows日本語版は，全てこのやり方である．

するとなると，JIS X 0208 と同様,「噓」と「嘘」に代表される ISO 10646 との間の問題が発生するからである．しかも，表外漢字字体表と JIS X 0213 との対応を示したからといって，JIS X 0213 の制定に反対した各メーカーが，表外漢字字体表のために改めて JIS X 0213 を実装する可能性は，極めて低いと言わざるをえない．

残念ながら，これが，日本における最新文字コード事情である．読者諸氏においては，さぞガッカリしたに違いない．しかしながら，規格の制定や改正というものは，権謀術数と妥協の渦巻く，極めて人間臭い作業なのである．しかも，ある規格を制定あるいは改正したからといって，それに携わった者の腹のうちは，それぞれに異なっているのも，また事実である．本稿で述べてきたのも，文字コード事情のある一側面であって，他の人から見れば，また全く違う側面がありうるはずである．読者諸氏においては，まずは規格そのもの，それもハンドブックではなく規格票そのものを，解説も含め，スミからスミまで熟読するよう心がけられたい．そうすれば，世の中に流布する上スベリで煽動的な文字コード論に，振り回されるようなことにはならないはずである．

参 考 文 献

[1] 夏目漱石: 草枕; 新小説，第 11 年，第 9 巻，本欄，pp.1-144，1906 年 9 月．

[2] The Unicode Consortium: The Unicode Standard — Worldwide Character Encoding, Version 1.0; Addison-Wesley, June 1992.

[3] 小池建夫: 統合化 CJK 漢字集合の実際; しにか，Vol.4, No.2，pp.37-43，1993 年 2 月．

[4] 安岡孝一，安岡素子: 「吶」はなぜ JIS X 0221 に含まれているのか—Unicode 幽霊字研究—; 情報処理学会研究報告，Vol.97，No.80 (人文科学とコンピュータ研究報告 No.35)，97-CH-35-9，pp.49-54，1997 年 8 月．

[5] 和田英一: Unicode は好きですか？; 情報処理，Vol.39, No.4，pp.321-323，1998 年 4 月．

[6] 安岡孝一，安岡素子: 文字コードの世界; 東京電機大学出版局，1999 年 9 月．

[7] 安岡孝一: JIS X 0213 の符号化表現; 人文学と情報処理，No.26，pp.9-17，2000 年 4 月．

[8] 国語審議会: 表外漢字字体表; 国語審議会答申，2000 年 12 月．

[9] 小林龍生，安岡孝一，戸村哲，三上喜貴 [編]: インターネット時代の文字コード; bit 別冊，共立出版，2001 年 4 月．

[10] 小林一仁: 漢字政策と「表外漢字字体表」; Science of Humanity Bensei，Vol.31，pp.16-27，2001 年 4 月．

最 新 参 考 規 格

- ANSI X 3.4: Information Systems — Coded Character Sets — 7-bit American National Standard Code for Information Interchange (7bit ASCII); March 1986.
- ISO/IEC 646: Information Technology — ISO 7-bit coded character set for information interchange; January 1991.
- ISO/IEC 2022: Information Technology — Character code structure and extension techniques; December 1994 (corrigendum April 1999).
- ISO/IEC 10646-1: Information Technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane; October 2000.
- ITU-T Recommendation S.1: International Telegraph Alphabet No.2; March 1993.
- ITU-T Recommendation T.101: International Interworking for Videotex Services; November 1994.
- JIS X 0201: 7 ビット及び 8 ビットの情報交換用符号化文字集合; 1997 年 1 月．
- JIS X 0202: 情報交換用符号の拡張法; 1998 年 1 月．
- JIS X 0208: 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; 1997 年 1 月．
- JIS X 0212: 情報交換用漢字符号—補助漢字; 1990 年 10 月．
- JIS X 0213: 7 ビット及び 8 ビットの 2 バイト情報交換用符号化拡張漢字集合; 2000 年 1 月．
- JIS X 0221-1: 国際符号化文字集合 (UCS)—第 1 部: 体系及び基本多言語面; 2001 年 4 月．
- RFC 1468: Japanese Character Encoding for Internet Messages; June 1993.
- RFC 1554: ISO-2022-JP-2: Multilingual Extension of ISO-2022-JP; December 1993.
- RFC 2045: Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies; November 1996.
- RFC 2047: MIME (Multipurpose Internet Mail Extensions) Part Three: Message Header Extensions for Non-ASCII Text; November 1996.
- RFC 2237: Japanese Character Encoding for Internet Messages; November 1997.
- RFC 2822: Internet Message Format; April 2001.

著 者 略 歴

安岡 孝一（やすおか こういち）

1965 年 2 月 18 日生．1988 年 3 月京都大学工学部情報工学科卒業．1990 年 3 月京都大学大学院工学研究科情報工学専攻修士課程修了．同年 4 月京都大学大型計算機センター助手，1997 年 8 月同助教授．2000 年 4 月京都大学人文科学研究所附属漢字情報研究センター助教授，現在に至る．京都大学博士 (工学)．1996 年 10 月から 2000 年 3 月まで符号化文字集合調査研究委員会 WG2 委員．2000 年 9 月いきなり二女の父となり，2000 年 12 月から情報処理学会文字コード標準体系専門委員会委員．2001 年 6 月から符号化文字集合調査研究委員会委員．文字コードに関する著書に [6,9] がある．電子情報通信学会会員．yasuoka@kanji.zinbun.kyoto-u.ac.jp