

JIS 漢字批判の基礎知識

豊島正之 (東京外国語大学アジア・アフリカ言語文化研究所)

日本語の文字と組版を考える会 1998 年 3 月 8 日

- 1 JIS 漢字の理解
 - 1.1 「文字概念」の問題
 - 1.2 文献の符号化の問題
 - 1.3 包摂規準
 - 1.4 曖昧検索で包摂規準を代替する案
- 2 JIS 漢字の今後
 - 2.1 新 JIS 漢字策定のスケジュール
 - 2.2 関連 URL
 - 2.3 開発の理念
 - 2.4 開発の資料
 - 2.5 包摂規準の再考
 - 2.6 組版規則との関連
- 3 JIS 漢字批判の傾向と対策
- 4 引用文献
 - 4.1 一般文献
 - 4.2 JIS 批判関連文献

1 JIS 漢字の理解

1.1 「文字概念」の問題

1.1.1 旧 JIS の「文字概念」の語の問題

78JIS¹・83JIS²・90JIS³の解説は、規格が「文字概念を符号化している」としている。

この「文字概念」の語は、83JIS が字体置換え (所謂「拡張新字体」化)・区点位置の交換等を行ってもなお「別の文字概念を採用したわけではない」とした為に、78JIS での各区点位置の「文字概念」より遥かに粗く、異体・新旧字体をも同一文字概念とする迄に拡大解釈されて仕舞った。文字概念を符号化しているのなら、別の文字概念を採用せずに区点位置が交換可能とは思われない。「鶯・鶯」、「蠣・蛎」等までを同一文字概念と強弁する為には、勢い「文字概念」の語は、曖昧な字義の謂いとでもせねばならなくなった。

文字の符号化自体は、具象的な図形そのものではなく、その概念上の存在に対して行われるものであり、且つ、図形文字の概念は、(字義の差は図形差に反映されるという前提の下での) 図形概念に依存しているから、文字概念は、字義 (字の意味・用法) と字の図形概念とが相互に規定しつつ結ばれたもので、つまり、それ自体が記号である。

¹JIS C6226-1978 情報交換用漢字符号系

²JIS C6226-1983 情報交換用漢字符号系 (後に JIS X0208)

³JIS X0208-1990 情報交換用漢字符号

97JIS⁴の立場は、文字の符号化を(字義の支援を受けつつ維持される)図形文字の図形の抽象的概念の符号化とするものだが、これは、符号は「記号としての文字概念」に振られていると考える事と、さして径庭が無い。

1.1.2 符号化された文字概念に対する合意としての同定情報

符号化された図形文字の文字概念によって文字の情報交換を行なうには、当該図形文字の概念に関して、合意が無ければならない。つまり、その符号が「どの文字」を指しているのが混乱しない様な方策が必要である。

この為、その符号の示す図形文字の文字概念に、合意の出来る範囲を明確化して置く事が必要であり、「文字の同定情報」(字義の合意)と「字体の包摂規準」(図形概念の合意)が必要になる。

X0208:1997 附属書 6 が与える、部首・画数・音訓・字書参照等の情報は、掲出字形と同じく、「ほら、あの文字」と理解して貰う為の同定情報であり、字形・音訓などを規範として定めたものではない。

1.1.3 字義の合意

意味の指定は難しいので、字義の合意は、何らかの文献参照で替えられる事が多い。この場合、文献自体に就ての合意があれば参照底本とし易いので、著名な字典、法律、経典等を用いて、これらは既に本文と字体が定まっているという合意を利用する。「合意済の文献」の参照によって字義の合意が済んだ積りになっているだけなので、(異文などで)その本文が動いて仕舞えば、既に参照の意味が無い。

康熙字典はこの例で、安価な中華書局複製本(同文書局版の再複製)を「康熙字典」と称して済ませていられる(ISO 10646-1)ならばよいが、字体の一点一画を論じるには、康熙字典「内府本」(「武英」殿本)との差は、到底無視出来ない。内府本自体、(板本の通例として)刷りの途中での誤植・誤刻の訂正が歴然で、更にその覆刻(「おっかぶせ」)迄あるらしく⁵、「康熙字典内府本の字体」がメートル原器の如くに存在すると考えるのは幻想である。例えば、内府 A 本・境田本が区別に努める「あみがしら」(罔)と「よこ目」の区別は、既に内府 B 本・安永和刻本では喪われつつあり、諸橋大漢和辞典は両者を完全に統合している。

「本文と字体が定まっているという合意」を満たす文献を強権で実現した一例が唐の開成石経であるが、一般には、合意というより幻想に近づく事もある。

尚、字義を殆ど欠いた固有名詞では、その名前を持つもの(referent)の存在を参照対象とするしかなく、同定は時に困難で、「人の名前は本人が違うといったら文字は絶対違う」(坂村健,1997)と、同定を諦める見解も現れる事になる。

1.1.4 図形概念の合意

1.1.4.1 精度規定としての包摂規準

実際に存在する字形は、一回限りの存在であり、二つとして同じ字形は存在しない。こうした存在を符号化する為には、それらの図形的な差を捨象して、抽象化せねばならない。全ての字形(加算無限)に符号を振る符号化は、(理論的には可能だが)現実的ではなく、「見た目により同一形状の文字は統合する」(坂村健,1996b)のが、符号化の一般である。

例えば、どのような微細な文字区別を行おうとする符号化も、サイズの差や色の差までも区別しようとするものは聞かないし、多くの符号化は、明朝体活字のの様なデザインの差(常用漢字表解説)や、ローマン・イタリック・明朝・ゴチックの様な書体差は捨象し、それらを統合して符号化を行っている。

こうした「捨象」をどの程度まで行うかに就ての合意が「包摂規準」であり、工業規格としての符号化文字集合では、いわば情報交換の「精度」の規定であると見てよい。

1.1.4.2 包摂規準の設定

一般論としては、包摂規準は「字の表現する意味の差は図形的な差に反映される」という前提を満足すべく、字

⁴JIS X0208:1997 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

⁵東京都立中央図書館諸橋文庫の内府本 A(832/MW/2)と境田稔信氏尚蔵本は、最内匡郭のサイズが殆ど同じだが、諸橋文庫内府本 B(832/MW/2)は、縦横それぞれ 1cm 内外小さい。精査していないが、境田本は、字形は諸橋文庫 A 本に近く、文字の配置・配列は諸橋 B と同じである様で、境田本が誤訂を経た初版、諸橋 B 本がその覆刻ではないかとも考えられるが、尚精査を要する。この件に就ては、池田証壽(北海道大学)の精細な研究がある。<http://member.nifty.ne.jp/shikedada> を参照。

義が異なる字体間には差を出し、同義の字体は包摂する様に設定するのが、使い易い⁶。字義が同じで、字体を包摂しないもの(包摂出来ない程に大きく字体が異なるもの)が異体字である。

しかし、これでは、包摂規準の設定規準の判断を「字義の異なり」の判断に預けただけなので、実質的には作業が大変難しい。

X0208:1997の包摂規準は、78JIS策定時に存在していた明示的な包摂規準(「異体字の取扱いの原則」)、その後の改訂・過去20年間の実装実績等による暗黙の包摂規準などを再編集したものである。つまり、97JISは、これを新たに創り上げたのではなく、編集し、明確化しただけである。(規格票解説14[「JIS漢字字典」左338]を参照)。実際の作業では、「包摂規準を立てる規準」、「包摂規準を立てない規準」の両面のメタ規準を設け、極力合理的に説明可能な形で規準を立てる事としたが、それでも、包摂規準 ad hoc(特設)委員会⁷は80時間を越えた。(各委員の事前準備はその数倍から十数倍)。

人名などの固有名詞に於て、極めて精密な字体概念を主張する方がいるが、固有名詞の漢字と一般の漢字を断絶したものとして捉える(53)と、固有名詞は字義を殆ど欠いているから⁸、そもそも包摂規準の立てようが無く、即ち抽象化(符号化)は困難になる。ここで、字義(meaning)を参照(referent)で代替すれば、極端には、各referent(個人)毎に別の字体(ロゴ)を主張する事も可能になる。戸籍等に使用する文字に就ての通達等は、この為に特に要請された別の規準である。

特に問題なのは、固有名以外には用いられない漢字⁹で、一般の漢字に就ての部分字体に基づく包摂規準を援用する以外手が無さそうだが、「もぎき」は、難題の一例である。

1.1.5 「全ての漢字」のコード化

以上のように考えれば、「全ての漢字をコード化する」為には、先ず、何を以て「全ての漢字」と見なすか自体に合意が必要である。前述のように、図形特徴を全く捨象しない符号化等あり得ないから、これは、どの程度の差は捨象するかの規定、即ち包摂規準そのものである。

「すべての文字」という言い方があやういのは、単に数が多いからではなく、文字の用途と定義についての吟味をバイパスしていく可能性があるからです。ちょっとでも違う字は別コードにしましょう、というのは何も考えないのと同じことです。(家辺勝文、1995)

という評言は、至言である。

更に、(書体差までも含めた)あらゆる字形差を別コードポイントとしてコード化するとすれば、文字コードが字形デザインの自由と可能性を剥奪するものとなる虞れもある。山本太郎(1998)に専論がある。

1.1.6 誤字

誤字は、誤字である故に、字義・字体に関する合意が取れない。故に、誤字は符号化出来ない。「裕」の衣へんを示へんにした文字、などのメタな表現による符号化は可能だが、それは、合意出来ている文字を合意出来ている方法で変形する事を示すのみで、図形文字の文字概念として示されているのではない。(勿論、何が誤字であるのかの判断規準は、全く別の問題である)。尚、誤字は、実際に用いられるからこそ誤字なのであって、定義によって用例は必ず存在するから、用例の有無は問題にならない。

1.2 文献の符号化の問題

1.2.1 符号化されるのは解釈文字列である

文献は、解釈を経て釈文(解釈文字列)に翻刻され、その翻刻釈文が符号化されるのであって、当然、解釈者・翻刻の目的によって釈文も異なり、その結果は別の符号になる。

⁶この前提が満足出来ないものが「衝突」。「JIS漢字字典」コラム29(p.571)等を参照。

⁷1995年9月16日(土)10:00-18:30、9月17日(土)10:00-18:00、2月18日(土)10:00-18:00、3月17日(日)10:00-22:00、7月20日(土祝)10:00-21:30、7月21日(日)10:00-21:40、8月3日(土)10:00-23:40[最長記録]、8月4日(日)10:00-23:20。会議が台風に見舞われた事が何回もあり、暴風雨の中で会議を続け、食事に出たら店が殆ど臨時休業だった、という経験もした。委員の間には、「新JIS」策定で又ぞろ「嵐の ad hoc」があるのではという暗い噂がある。

⁸太田さんが太っている保証は無く、良子さんが良い子とは限らない。

⁹X0208の地名で有名なのは、娯(53-19、めおとすぎ、金沢市)。

実際の原文には、多くの字形衝突が生ずる(OCRはここで間違ふ)が、それらは釈文の解釈段階で殆ど解消され、一方、多くのパリアントは統合・整理される。(衝突解消・整理出来ていないものは、そもそも「釈文」ではない)。

文字の符号化は、この様に一旦解釈された文字列が符号化されるのであって、原文が常に同一の符号列に変換される事を意味しない。符号化はOCRではない。包摂規準が明朝体で示されるのも、解釈された文字列の表現として現代では最も一般的と思われるからであって、原文として一般的であるからではない。

「原文 釈文」の解釈は言語・学問の分野に属し、符号化・工業規格の分野ではない。字形を字体で解釈するのも言語の問題(「漢字の常識」)であって、符号化の問題ではない。同様に、書体間の対応付けも言語・デザイン(技芸)の問題である。

1.2.2 釈文が解釈により異なる例

長島弘明(1998)p.146は、上田秋成「雨月物語」の「白峯」を掲げ、これは、「JISコードの文字ではとても原文そのままの表示はできない」ものであるとされている(長島弘明,1998)。

次に、この部分を安永五年板本の行取りのままに改行して示す。(振り仮名を省略)。

- A : 中央公論社版「上田秋成全集」(1990) 巻7 p.226
- B : 坂村健(1996a)
- C : 長島弘明(1998)
- D : 木越治・高木元らによる電子化ファイル Ver 0.99 (1992.7)

これらの間に異文がある時は、(A/B/C/D)の形で示す。「≡ [棧]」等は、C(長島)がX0208に無いとしたX0208内の字。

1. あふ坂の関守にゆるされてより。(秋/≡/≡/秋) こし山の黄葉見過しが
2. たく。(浜/濱/濱/濱) 千鳥の跡ふみつくる鳴海がた。不(尽/盡/盡/盡)の高嶺の煙。浮嶋
3. がはら。清見が関。大磯小いその浦(≡/≡/≡/々)。むらさき艶ふ武(蔵/蔵/蔵/蔵)野
4. の原塩竈の和たる朝げしき。象(潟/≡/≡/潟)の(蚕/蚕/≡/蚕)が(筍/筍/筍/筍)や。佐野の
5. 舟梁。木曾の(棧/棧/≡ [棧]/棧)橋。心のとゞまらぬかたぞなきに。猶西の(国/国/國/國)の
6. 哥(枕/枕/≡/枕)見まほして。仁安三年の秋は。葭がちる難波を(経/経/経/経)
7. て。(須/須/≡/須)磨明石の浦ふく(風/風/≡/風)を身にしめつも。行(≡/≡/≡/々)讃岐の真
8. 尾坂の林といふにしばらく(≡/≡/≡/≡)を植む。草(枕/枕/≡/枕)はるけき(旅/旅/≡/旅)路
9. の(勞/勞/≡ [勞]/勞)にもあらで。観念修(行/行/業/業)¹⁰の便せし庵なりけり。

C(長島)は、蚕・枕・須・風・旅・(棧・勞は誤解)をX0208で表現不可能とする点で、最も微細な差を見ている。

長島(1998)は、これを含む例示に基づき、「JISコードを通して流通しているのは、単なる不完全な文字データもどきであって、文字データでも、ましてや文化情報でもない。」とし、「劣化を防ぐ筈のデジタル化が、より劣化したテキストデータしか生み出せない滑稽な仕儀となっている。デジタル資料の最大の長所の一つが、これまた文字コードの制限によって拒まれていくことになる。」(p.152)とする。

B(坂村)は、「せっかく文字資料を風化から守るためのデジタル・アーカイヴが、資料を虫食いにしてしまうことになる」(p.290)と説いた上で、TRONコード環境なら「完全に表現できる」と主張した翻刻なのだが、Cが要求する水準の区別が行なわれないばかりか、A(中央公論社)同様、常用漢字への置換すら行なった体のものである。

A(中央公論社)とD(木越・高木)は、「浦(≡/≡/≡/々)」「行(≡/≡/≡/々)」の踊り(「くの字点」)、「(≡/≡/≡/≡)を植む」の「つえ」¹¹の2字以外は全てJIS内字で翻刻しており、AもDも「より劣化したテキストデータ」とも呼ばれようが、Aは、別段文字コードの制限によって成ったものではなく、自ら「原則として、常用漢字表及び人名漢字表にあるものについては、その字体を採用し、その他の字についても、現在通行の字体を用いた。」(凡例6)という方針を立てた翻刻である。

C(長島)・D(木越・高木版)が一致して旧字体に翻刻した蔵・濱・盡・經・國・勞のうち、A(中央公論社)は全てを、B(坂村)は、蔵・經・國・勞を、常用漢字体で翻刻している。しかし、ここから、中央公論社版全集(A)が研究の役に立たないとする研究者は、いたとしても、極く少数ではなからうか¹²。

¹⁰「修業」は、C・D共通の誤植。

¹¹NTT電話帳に、名[(きょう)子(こ)]の用例がある。

¹²B(坂村版)釈文は、どこを見ても、この「白峯」冒頭一丁以外が見当たらず全文の通読が不可能なので、しばらく置く。http://www.um.u-tokyo.ac.jp/に雨月物語の画像があるが、釈文は無い。因みに、この雨月の画像は、間紙を入れずに撮影されている様である。

研究者が本文の研究を翻刻本で済ませる事は無く、特に「雨月物語」の様に、複製もあり、原本も比較的容易に見られるものは、必ずそれらと対照しながら見る。研究者が求める「原本に忠実な翻刻」とは、決して原本をそのままに引き写したものの事ではなく、原本の情報を伝えるべく適切に(明確化された方針によって)解釈された本文である。

練達の音楽家の手によって一つの楽譜原本から様々な演奏(interpretation=解釈)が生れるのと同じく、原本に対する解釈が、解釈者・解釈の対象(受け手)によって様々に異なるのは当然の事で、「原本に忠実な翻刻」が常にただ一つ存在するというのは、学芸に対する見方としては余りに狭量である。A(中央公論社版)本文は極力常用漢字で解釈し、D(木越・高木版)本文は所謂正字で解釈しようとした本文であって、言うまでもなく、それぞれ、用途に応じて極めて有益である。

1.3 包摂規準

1.3.1 世に行われる包摂

符号化とは抽象化であり、字義に支えられた図形概念の抽象化、即ち字体の包摂は、(上述の通り)文字の符号化には必須である。X0208の包摂は、コードポイントが不足するという理由で行われたものではない¹³。

字体包摂は、どの様な符号化でも、必ず行なわれている。

1. ASCIIでの'g'と'g'
2. NTT(日本電信電話株式会社)外字表
複数の字体が同一外字コードにまとめられており、区別が無い。
3. 日本文藝家協会(1997)「お寄せ頂いた文字による文字コード調査一覧」
(これは符号化に結びつくものかどうか不明であるが)採集文字の字形が諸橋大漢和の字形と異なる時のみ、その旨明記するとされている。ところが、明記されない字形でも、諸橋(修訂版本文見出し字¹⁴)と異なるものが少なくなく、それらの字形差は捨象されて同定されている事になる。

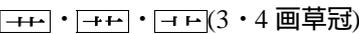
1.3.2 包摂の規準の共有

字体の包摂の規準(規則)は、情報を交換する当事者同士の合意(情報の共有)が必須である。どの程度の抽象化を意図しているのかに就て、両者が合意しなければ、そもそも情報交換にならない。

共有されねばならない情報は、公開されるべきである。従って、包摂規準は、規格の規定内に無ければならない。

1.3.3 字体包摂のレベル－ 列挙に替る表現法

X0208:1997は、字体包摂の規準を、列挙によって掲げた。

- a 「八」の第2画の始めに横線が入るか否か(「八屋根」)
- b  (3・4画草冠)
- c 
- d 
- e 区・區

X0208:1997は、b, c, dを包摂し、eは包摂せず、aはデザイン差として列挙から外した。(aは、常用漢字表解説に明示されており、規格は、当該解説自体を引用している)。

もし、包摂規準に就て、「xが包摂ならばyも包摂」、「yが分離ならばxも分離」の様な、「包摂のレベル」自体に合意があれば、包摂規準は、その一部を例示してレベルの合意を(メタに)示唆する事で替る事が出来る。例えば、「者」の一点の有無を包摂する位なら「しんによう」の一・二点も包摂するだろう、とか、草冠3・4画を分離するのであればも分離する筈だ、等の推論が可能なら、「草冠3・4画は包摂しますので、他も同様に、ひとつよろしく」で済ませられるかも知れない。実際、78JISの解説は、そうした「包摂のレベルの常識」を期待

¹³78JISが、使われている文字を出来る限り多く集めようとした事に就ては、芝野主査担当の「JIS漢字の歴史」を参照。

¹⁴諸橋大漢和は、普及版と修訂版、修訂版の各巻頭の部首画数順一覧と本文見出し字で、それぞれ字形が異なる事がある。

していたかに見える。97JIS が 185 包摂規準¹⁵を列挙してレベルの合意による表現を採らなかったのは、明確さを優先した為である。

1.3.4 文字コード規格・包摂規準は、言語規範ではない

文字コードの規格と、その字体包摂の規準は、言語としての文字に対する規範ではない。

当用漢字字体表は、言語としての文字に対する規範であった。一方、X0208 とその包摂規準は、文字を符号化して交換する際の抽象化に関する合意であって、文字自体に対する規範ではない。言語としての文字は、X0208 からは勿論自由である。

字体包摂の規準は、言語としての文字に基づいて情報交換の「精度」指定をしているだけで、字体のゆれの許容範囲を言語規範として定めているのではない。JIS S1005-1984「家庭用流し台・調理台・こんろ台」が日本人の身長・座高に基づいて調理台の高さを 800mm ± 2mm と定めた事が、規範として日本人の身長・座高を規定した事にならないのと同じである。

字体包摂の規準は、そもそも文字に対する規範ではない以上、漢字制限とも無縁である。

1.4 曖昧検索で包摂規準を代替する案

1.4.1 「何でもコード化・曖昧検索」の主張

ここに「何でもコード化・曖昧検索」というのは、下記の様な立場である。

1. コードポイントを沢山用意し、どんどん登録する。

- 一度すべての漢字をコード化して、然る後にその利用法を考えるという発想が必要である。これを小研究会では「総背番号制」と称し…(30)
- 存在する漢字は議論を待たずに入れてしまえ、ということなんです。全部入れてしまうんだから基本的に言うところ簡単なんです。だからできるだけ多くの漢字が集められればおしまいなんです。今のコンピュータは十分な余裕を持っていますから技術的にも全然問題ないんですよ。(24)

2. 検索用では、それらを適宜まとめて、曖昧検索する。まとめ方は、検索者が決める。

- 別の文字だと検索できないなんていうのは、いかにもコンピュータを知らない話で、要するに文字から文字へリンクを張っておけばいいわけでしょう。一個引いたら、これもありますよと並べてくれて、全部見ますかと聞かせてイエスかノーで答えればいい。(2)
- 最初から異体字に別々のコードポイントをわりあてて、検索はシソーラスで対応した方がすっきりします。入力の場合も 37 種類のワタナベが変換候補として出てきた方が楽ですし、辞書の学習で、次回から手間が省けます。(8)

3. 或るコードポイントを使ったからと言って、別のコードポイントの文字ではないという意味にはならない。

4. 曖昧検索の対象になるのコードポイントの全てを知る必要は全く無い。

例えば、今、仮に𪛗(71-73)と区別して𪛗もコード化したとして

- 𪛗を用いた場合でも、𪛗でないという意味にはならない。
- 𪛗を検索したい人は、𪛗も検索すればよい。
- 𪛗を検索していた人には、𪛗も検索して貰えばよい。

1. 𪛗がある事を知らない人は、どうやって𪛗を曖昧検索リストに加えるのか?

字書(シソーラス)を用意すればよい。

2. 曖昧検索の仕方は人毎に異なる筈だが、シソーラスは共用出来るか?

自前のシソーラスを作ればよい。

3. 他人が、それぞれのシソーラスに基づいて作ったデータが、自分のシソーラスで正しく検索出来るという保証はあるか?

シソーラスを含めて流通させればよい。

¹⁵正誤票で 186 に訂正。

1.4.2 「何でもコード化・曖昧検索」の評価

上記に示した様に、この立場で作られたコードポイントによって情報交換するには、シソーラスを伴った交換が必要である。即ち、「何でもコード化、曖昧検索」という考え方が X0208:1997 と異なるのは、情報交換の「精度」の規定が

X0208 : 固定的、 包摂規準として規格化され、共有を前提とした交換が可能
「何でも」 : 可變的、 シソーラス共有の為に、別途手段が必要

という点に尽きる。つまり、「何でもコード化、曖昧検索」コード系では、情報交換の当事者間で、検索精度の合意を取る為に別途手間が掛る、というだけの事である。

更に、情報交換のための曖昧検索リストの作成には、情報交換の当事者が知る該当コードポイント関連の他コードポイントの情報が、一致していなくてはならない。つまり、該当コードポイントに関連する[と情報交換の当事者が判断する]全てのコードポイントが、当事者に予め知られていなくてはならない。

例えば、仮に、共用シソーラスに[芎]・[芎]が載っていないとして、[芎]を入力した人が[芎]の字でも検索して欲しかった場合は、その旨、共用シソーラスに書き加えねばならない。

入力時に[芎]の字が共用シソーラスに無かった場合、それは、

- [芎]のコードポイントが無いのか
- [芎]のコードポイントはあるが、共用字書に無いのか

をまず調べ(この時点で全コードポイントの探索が必要)、後者ならば、自前のシソーラスに追加して、それを共用に供する作業が別途発生する。即ち、入力時に一々共用シソーラスの update が必要になる。

結局、「何でもコード化」方針でも、情報を共有・検索したいのであれば、どのみち、入力時に全コードポイントを知って、且つ、検索時の同一視のリストを立て、それを共有しなければならないのである。

まとめると、「何でもコード化・曖昧検索」方式とは、文字の情報交換に必須な包摂規準の共有を、取り敢えず忘れた振りをして置く、という方針である。

尚、検索などは考えず、単に望む字形が出ればよいのだという主張もあり得、この場合は曖昧検索すら不要になるが、ポイント数や書体、解像度(1インチ辺りのドット数)までも指定したコード化を行うならともかく、そうでなければ、出力された字形が「望む字形」なのかどうか(出力装置の適合性)の判断すら困難になる¹⁶。即ち、「シソーラス」は、曖昧検索の為に他コードポイントだけではなく、そのコードポイントを使う為の同定情報も必須である。

結局、「何でもコード化・曖昧検索」方式でも、それを実際に運用するためには包摂規準と同定情報が必要になる訳で、(包摂規準が共有済・固定的か、未共有・可變的かという点を除けば)、X0208:1997 の主張と大きく隔たるものではない。

1.4.3 X0208:1997 の包摂規準は多過ぎるか

「何でもコード化」方式の様な、包摂規準の逐次設定方式では、使わないコードポイントの包摂規準が見えないので、事前設定の包摂規準は確かに少なく見えるだろう。

例えば、X0208:1997 の 56 番 [卍]・[卍]が適用可能な区点は、実質「卍」(50-34)のみ、144 番 [𠄎]・[𠄎]は「𠄎」(38-52)のみ、181 番 [荊]・[荊]は「荊」(23-53)のみで、これらの区点を使わない人には関係が無い。

こうした規準を事前に共有する必要が無ければ、一見包摂規準が小さく、或は皆無である様に見える、というだけの事で、実は情報交換の為には、必要な包摂規準は共有しなければならない。

尚、世上の文字交換は、包摂規準も無く、全文字の熟知も無しに行なわれており、文字コードとてこれに同じであるという主張があるが、世の中の文字交換では、「漢字の常識」の形で暗黙に包摂規準が共有されている。(一部は、常用漢字表に明示されている)。これを無視し、全く当該字形のみによって同定・検索を行なっているかの様に考えるのは、文字の文化・伝統に気付かないだけである。

¹⁶ 「イエスカノーで答えればいい」(2) というが、どの様にしてイエス・ノーの判断をするのだろうか。

2 JIS 漢字の今後

2.1 新 JIS 漢字策定のスケジュール

「新 JIS 漢字」は、現行の JIS X0208:1997 「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」を拡張して、約 5,000 字の漢字・記号類を新たに追加するもので、1998 年 9 月～12 月に公開レビュー、1999 年 3 月に規格制定・公刊を予定している。

2.2 関連 URL

御意見、御要望、御質問の宛て先	mailto:new-jis@jcs.aa.tufts.ac.jp
JIS 漢字の歴史と新 JIS 開発計画	http://www.itscj.ipsj.or.jp/topics/jisx0208.html
新 JIS の開発計画	http://www.tiu.ac.jp/JCS/
新 JIS 開発の進捗状況	http://jcs.aa.tufts.ac.jp/new-jis/
JIS X0208:1997 公開レビューの記録	http://jcs.aa.tufts.ac.jp/x0208/

お寄せ頂いた御意見、御要望、御質問などは、上記 URL 又は公開レビュー資料・規格票解説等で、(お名前などを伏せて) 公表する事があります。御質問には、直ちにお答え出来るとは限りませんが、上記 URL 又は公開レビュー資料・規格票解説等には、回答に当たるものを掲載する様に努力します。

2.3 開発の理念

X0208:1997 を補うものとして、第 3 水準 (約 2000 字) 及び第 4 水準 (約 3000 字) の計約 5000 字の拡張文字集合を追加する。

今回追加する拡張文字集合 (新 JIS 漢字コード) は、JIS X 0208:1997 の文字集合を補い、JIS X 0208 が当初符号化を意図していた、現代日本語を符号化するために十分な文字集合を提供することを目的として設計する。そのため、JIS X 0208 と同時に用い、JIS X 0208 を補完するものとし、また、現状の使用環境で直ちに実装できるように、第 3 水準及び第 4 水準として追加する。

JIS X0208:1997 では、すべての区点位置についてその表す文字を再検討し、漢字については、字書典拠情報、字体・文字の包摂規準に基づく包摂の有無、実装字形の調査、及び可能な限り原典拠に遡った調査を加えて、文字の同定作業を行なった。この同定作業によって、X0208 の各区点の表す文字が明確化された。この結果、各区点ではどの文字が表せないかも明確になり、現行 JIS 符号化文字集合である JIS X 0201 及び JIS X 0208 では、現代日本語文の表記に必要な文字でも不足しているものが少なくないことを明らかにできた。

今回の第 3 水準及び第 4 水準の追加にあたっては、個々の図形文字の同定のための典拠情報を十分に与えることにより、実際の運用上のあいまいさを生じさせないことに留意した開発を行う。

符号化方法は、JIS X 0208:1997 に準じ、X 0208 が規定するすべての符号化方法で符号化可能な文字コードの開発を行う。

2.4 開発の資料

2.4.1 漢字

1. 資料を入手済で、資料間の一字ごとの照合も、ほぼ完了したもの

分野	数	NTT に無し	X0208 に包摂	典拠	用例
人名	3,246	-	220	NTT 八口ーページ (1997.2.14 現在)	有
地名	56	21		町字ファイル外字 (国土地理協会)	-
地名	183	125	14	町字ファイル書き換え字	有
地名	14	5	4	1 万分の一地図 (国土地理院,1995)	有
地名	138	72	12	2 万 5 千分の一地図 (国土地理院)	有
法令	5	3	0	現行法律・勅令・政令・省令外字	有
一般	3	3	0	部首字	有
一般	4	4	0	熟語専用字の一方だけ X0208 にあるもの	有
学術	5	3	0	文部省学術用語集	有
学術	16	9	0	切韻韻目 (鉅宋広韻による)	有
学術	1	1	0	韻鏡三十六字母	有
不明	360	83	39	IBM 選定文字	-
不明	4177	2539	679	JEF(富士通) 外字	-
不明	3566	2131	589	JIPS(日本電気) 外字	-
合計	7178	3932	966	(数字は異なり)	

2. 文献は入手済で、現在、外字を探索中のもの

- (a) 小・中・高等学校の検定済全教科書 (1,500 冊)。(参考資料として、各指導書用フロッピー)
- (b) 「新潮文庫の一〇〇冊」「明治の文豪」「大正の文豪」(新潮社御提供)

3. 資料は入手済で、他資料との照合が進行中のもの(五十音順)

- (a) 文脈・用例のあるもの
 - i. NTT 契約者名簿 (日本電信電話株式会社御提供)
 - ii. 弓道関係外字
 - iii. 大正新修大蔵経所収 漢字 (大蔵出版社御提供)
 - iv. 中国医学古典用外字 (日本内経医学会)
 - v. 日本国語大辞典 (新版) 外字 (小学館御提供)
- (b) 文脈・用例が無く、外字表だけのもの
 - i. KEIS 外字 (日立製作所御提供)
 - ii. 共同通信社外字 (共同通信社御提供)
 - iii. 「広辞苑」外字 (岩波書店御提供)
 - iv. 「大辞林」外字 (三省堂御提供)

4. 非漢字

- (a) 資料を入手し、追加すべき文字集合が既に明確になったもの
 - i. アイヌ語表記用の仮名
 - ii. 鼻濁音表記用の仮名
 - iii. 歯科記号
- (b) その他検討中のもの
 - JIS Z8202

2.5 包摂規準の再考

JIS X0208:1997 の包摂規準は、(1.1.4.2 p.3) に述べた様に、それ以前の(明示的・非明示的)包摂規準を再編集して明確化したものに過ぎないので、新しい文字集合である新 JIS に適用する包摂規準に就ては、必要な再検討を加える事になるであろう¹⁷。

2.6 組版規則との関連

上付き・下付き数字、縦中横など、組版技術によって実現されている項目もあるが、圏点など、どこまで文字コードで扱うべきかが未定の問題もあり、JIS X4051 組版規格の委員会(電子文書処理調査研究委員会 WG2、主査芝野耕司)との緊密な連絡を取りつつ、開発を進めている。

3 JIS 漢字批判の傾向と対策

A 一般に文書の電子化に就て

1. 出来る限り原文のままの電子化が望ましい
歴史的な文書も電子文書として扱いたい、という希望が第一にあります。この意味は、「この漢字は使おう、この漢字はひらがなで処理しよう」という判断をしてはならない世界にいる、ということです。できるかぎり、昔のままの形で、電子化していかなければ、ならない。(15) (1.2.1 p.3)
2. すべての漢字をコード化
一度すべての漢字をコード化して、然る後にその利用法を考えるという発想が必要である。これを小研究会では「総背番号制」と称し、この実現を柱に答申案を纏める事とした。…今後展開されるであろう[10]646 拡張言語面への準備としても、この際全漢字のコード化を行なっておくべきであると考えられる。幸いなことに、大漢和辞典(諸橋辙次編)及び[10]646 或は[X0]221 の配列順は共に、康熙字典に準拠しており、大漢和辞典の漢字番号がそのまま利用できるとの予想が可能である。…個人用では常用漢字だけの 1950 字から多い人で 5000、印刷業としては 3000 から 1 万までが実用範囲と考えられる。(30) (1.1.5 p.3)
3. すべての文字をコード化
「すべての文字」という言い方があやういのは、単に数が多いからではなく、文字の用途と定義についての吟味をバイパスしていく可能性があるからです。ちょっとでも違う字は別コードにしましょう、というのは何も考えないのと同じことです(何のために別コードが必要なのか、またどの程度の違いで別字になるのかといった考察)。(50)
4. 議論せずに全部の文字をコード化
存在する漢字は議論を待たずに入れてしまえ、ということなんです。全部入れてしまうんだから基本的に言うところこう簡単なんです。だからできるだけ多くの漢字が集められればおしまいなんです。今のコンピュータは十分な余裕を持っていますから技術的にも全然問題ないんです。それにすべてにオープンというのが TRON の真髄ですから、マイクロソフトがそれを使いたいと言えばいつでも出せるんです。JIS 委員会にも欲しいと言われればあげられます。(24) (1.1.5 p.3)
5. 全ての文献のすべての文字をコード化
文字学の世界では、諸橋大漢和の中にも、出典に難のある文字があるという指摘があります。康熙字典にも疑問がある。だから、付[ママ]号化基準としてそのまま採用できない、という議論が、パソコン通信で行われているのを見たんです。私は、この議論は変だと思いました。諸橋大漢和も康熙字典も、歴史的文献として収録できなくてはならないんです。…付号化は、個々のコンテンツより大きなものでなくてはならないのは当然なのに、付号化基準とコンテンツをいっしょくたに論じている。…収録できるものを選別するのは、妙な話です。(52) (1.1.5 p.3)
- (1.1.3 p.2)
6. 「区別される漢字」の基準
どんな漢字が他と区別され、独立した一つの文字として認識されるべきなのか、それを決める基準は何なのか、考えてみなければしょうがないわけですよ。そうすると、「活字」が一つの拠り所になるだろう。…これまで活字になったすべての文字は、コンピュータでも使えるようにすればよいではないか…手書きで、書き癖とかいうことならばフォントの問題でしようけれど、印刷されれば活字の問題であるんです。…漢字一つひとつに対する人それぞれの意識がですね、まさに活字に由来するものだとするならば、活字を守ればよいんです。(29) (1.1.4.1 p.2)
7. 使い分けを主張する人に依存
区別の必要を感じる人がある限り、極力微細な区別をしたコードとし、必要なしとする人は、それらの好きなものを使えばよい。(1.4.1 p.6)

B 一般に文字コードによる文献の扱いに就て

8. 見た目の差に依存
漢字文字属に対応する漢字を収容するスクリプト層では、これらの文字を一括して扱えるようにし、見た目により同一形状の文字は統合する。ただしエントリーポイントが 2 バイト空間に限定されず十分な余裕があるので、ここでの統合は Unicode や ISO10646-1 で行われたような文字数を大幅に削減するための無理な統合ではなく、見た目に見差があるものについては統合しない。(21)

¹⁷包摂関連 AdHoc 会議は 3 月中旬に開催予定。

こうした場合、「見た目の差があれば別、無ければ同じ」という方針は、実は操作不可能で、そもそも「同値」の文字は、見た目が同じに見えるものである。(逆に、大きく見た目の差があっても実は同値なのが、異体字)。又、そこに現れる「文字」は、当代の文字区別による。例えば

- 「恵」字が「智慧」にしか用いられず
- 「慧」「恵」字が「恩慧」「佛恵」等に用いられる

という場合、恐らく当代の文字区別では、「慧」と「恵」の位置が現代とは入れ替わっているのだと思われるが、この場合、字源に頼った「文字概念」は無意味。つまり、何が「区別できるし、区別したい漢字」なのかは、研究の前には分からないので、「見た目の差があるかどうか」すら、実は研究を経ねば不明である。「見た目の差があれば全部入れる」というのは、その必須の研究を無視している。

9. 人名漢字は礼儀作法

少なくとも私は、他人の姓名を表記するのに、先方が名刺などに刷っている字体を使うよう心がけています。...これは礼儀や社会慣習の問題だと思っています。...工業規格が一般に認められている礼儀や社会慣習、常識を守れなくしてしまうのはおかしいことです。(53)

(1.1.4.1 p.2) 「書いた通り」にも、抽象化がある。「先方が名刺などに刷っている」ものをこちらで書き直す段階で、既に字体の抽象化を経ているのは言うまでもない。極端な話、先方の字形をそのままスキャンしたとしても、サイズや解像度、色などの属性は継承されない。サインに至っては、絶対そのままには書かない筈である。尚、本当にそれが「礼儀」であるのなら、工業規格にかこつけてそれに従わないのは無礼というべきであろう。

10. 文字コードは自分自身の包摂規程が書けなければならない

複数の[包摂された]字形を区別できない JIS の漢字コードでは、自分自身(すなわち JIS の漢字の規格表)を書くことができない。「 と は字形が違うが両者には同じコードを割り当てる」という文章を表現できないからである。

(45) これが区別して表現出来るコード系は無い。(クレータのパラドクス)。

C 曖昧検索論 (1.4.1 p.6)

11. 類似文字の過剰と検索の困難

別の文字だと検索できないなんていうのは、いかにもコンピュータを知らない話で、要するに文字から文字へリンクを張っておけばいいわけでしょう。一個引いたら、これもありますよと並べてくれて、全部見ますかと聞かせてイエスかノーで答えればいい。(2) (1.4.2 p.7)

12. 一意の符号化が無ければ、検索は困難

ここまで無理して文字数を減らしている Unicode ですが、逆に類似文字の過剰という問題も指摘されています。例えば、日本人が同じ漢字だと思っていたものが他の国で別の文字になっているものとユニフィケーションされている場合、国内では同じ文字のコードが複数あるかのように見えます。ラテン文字でもよく使われるアクセント記号付きの文字は独立したコードが別にありますから、見た目は同じ文字でも単独のコードのものと、合成で作った文字と二種類の表現が可能になります。文字に対してコードが一意に決らないということは、データベース検索など多くの点で混乱のもとになるでしょう。(23)

13. 二重符号化では検索に問題が出る

EPWING 外字をつかっているものは、比較的影響はすくないものの、外字の中には「新拡張 JIS」で増える文字がかなり含まれているはずなので、外字と「新拡張 JIS」と二重にマッピングされる文字が出てきて、検索に問題がでてくるでしょう。(8)

D 同定忌避

14. 文字の同定

一定の規範性を強制していると思われる「康熙字典」と「大漢和」はとりあえずすべて採字すべきだと思っています。異体字の許容範囲の設定や、同定作業なんて、工業技術院が出来る(やっつけいい)ことではないと思います。専門に研究している学者の間でも意見が分れるのに... (32) (1.1.2 p.2)

15. 文字の付帯情報

文字を利用するには、その文字に関する付帯情報も必要になる。例えば、「森鷗外」の「鷗」という字に対しては、「鷗」と字形の異なる文字であるが、両者の字義(文字の元の意味)は同じであり、慣習的にも混在して使われているといった付帯情報が必要である。こういった付帯情報は、文字コード割り当ての際の同一性の判断に使われるほか、コード割当てがすんだ後も、その文字を「使いこなす」ために有用である。(45)

16. 過去の漢字と現代の漢字には関係が無い

包摂は人名字体についても適用されます。...豊島正之...は、...寛永版万葉集に出てくる地名「吉野」...の例などを根拠に挙げています。...寛永版万葉集の地名表記と現代の人名表記にどういう関係があるのか不明ですが...(53)

僅か 350 年前の「吉」字と現代の「吉」字の関係が不明だという主張は、漢字の伝統は既に完全に断絶していると主張するに等しい。

E 異体字

17. 「異体字」の語の誤用

「齋藤」に使われる「齋」の異体字には「齋」「齋」もある。(出典多数)

齋と齋・齋は、全く別義で、異体字ではない。(1.1.4.2 p.3) 異体字とは、字体を異にする同義の字であるから、そもそも包摂できない。尚、包摂される字体に小差を認めて「異体字」と呼ぶのは、一般的な用法ではないが、文字コード論では、誤ってこの意で用いる向きがある。

18. C6226 の異体字の採録基準

JIS で複数の字体が収められている漢字を大別すると(1)旧字体、(2)異字体、(3)俗字となる。常用字体に対応する形で旧字体が収められているのであるなら、それも一つの考えといえようが、実際にはそうっていない。「礼」に

対して「禮」はあるが「社」の旧字体はない。異字体の扱いはもっと不徹底で、そもそも沢山の異字体を温存させることで混乱を招いているうえに、(a) 第一、第二水準に振り分ける(鈎と鈎など)、(b) 同一水準内の離れた位置にコードを与える(組と組など)、(c) 同一水準内で隣合わせに配置する(穎と穎など)というように処理の仕方も一定しない。以下略 (47)

配列は「新字源」順である。その他は、「JIS 漢字の歴史」(芝野主査講演)を参照。

19. 新旧字体はフォントの差と見得るか

旧字体と新字体は今、字形が違うので別コードがふられていますが、これは、旧字体フォント、新字体フォントという形式を使って、字を区別し、区別した形で検索することも可能でしょう。その場合、字体の違いにもかかわらず、コードが一致しているのだから、フォントの差を無視した形で、文字を検索すれば、字体を無視した形での、ある単語が含まれている複数文献の全文検索だって簡単に、できるかもしれませんし、フォントの切替えて、旧字体と新字体が一気に変わってくると、それだけで、楽しいものがあります。(16)

20. 御破算

JIS 漢字は全部、総ざらいして、新しく作り直して、例えば、旧字体と新字体の区別は、文字コードではなく、何らかの別の方式を採用して欲しいなんて思っています。情報交換が必要な暗号は「文字」だけじゃない、という、複雑な文字体系を持っているのが日本語なので、それをうまく表現できる情報交換符号(暗号)が必要だろう、と。(16)

F 包摂規準

21. 包摂基準は社会に受け入れられたか

第1~4水準漢字の選定方針に、字体・文字の包摂という考え方がある。これはJIS漢字に特徴的なものである。これは、広く社会に受け入れられているのだろうか。JIS内で説得力を持っていても、外部からは批判も多々でている。それ以上に、まだほとんど知られていない、というのが現状であろう。くち「高」と、はしご「高」の例をとってみても、みな十分納得させられたとはとても思えない。「JISの出している文字の同定基準が実際の文字の需要と合っていないかった」というコメントもある(TRONWARE vol.40 p.39)。国語審議会にでも、お墨付きがもらえれば良いのだろうか。(26) (2.5 p.10)

22. 包摂された字体・文字は、区別したいときに区別できるか

「JIS漢字の拡張計画」の文中に、丸付き数字などのいわゆる“合成文字”は“合成文字”として生成できると漠然と信じられてきたが、実は生成不可能であった。これらの多くは、メーカー各社がJIS外字として独自に実装しており、情報交換の際の混乱の大きな要因となっている。」とある。同様に、第1~4水準で包摂された字体・文字は、区別したいときに区別できると、“漠然と信じられてきたが、実は不可能であった”という事態にならないだろうか。そして、“メーカー各社が独自に実装することで、情報交換の際の混乱の大きな要因”となる懸念はないのだろうか。(26)

対応規格を別委員会が検討中

23. 包摂された文字の字形利用

千葉県九十九里浜に面する蓮沼村。この村の「蓮」という表記は、しんによろが「点2つ」が正式。しかしJISコードの「蓮」は点が1つ。国内の自治体の表記さえまともにできない日本工業規格って何じゃね。(35)

蓮(47-01)は、一・二点を包摂しているのであり、一点と定めているのではない。(草冠は何故不問なのだろう)

24. 包摂規準を字体規範と見る誤解

形がいろいろ分かれた漢字に対しては、どこまでを文字のデザインとして認めるかの「包摂基準」を細かく例示した。たとえば、「叱る」もとは「口」偏に「七」。しかし七がヒになったような字も使われる。どちらも同じ字と見なすというのが「包摂基準」だ。(43)

(1.1.5 p.3) X0208は、デザインを「認める」か否かというデザインの規範ではない。尚、この記事は、包摂基準を「包摂基準」と一貫して誤記。

G 字体・字形・書体

25. 字体と字形

異体字フォントごと情報を送ることもできなくはないですが、すべてのソフトが約束通りにつくられる保証はありません。掟破りのソフトや、古いヴァージョンのソフトが介在すると、どんな文字かわからなくなる危険性があるのです。(13)

掟破り(規格不適合の)ソフトウェアで保証が得られないのは当たり前。

26. 字体と書体

結局、書体と字体を意図的に混同しているわけですね。(2) 規格票にはしていないと書いてあるんですが、書体を変更すると字体もいっしょに変わってしまいますから、混同しているのと同じです。(13)

変更して字体が変わるからこそ、別の書体と呼ばれるのである。

H 83JIS

27. 83JISは嘘字をこしらえた・鷗等は、JISの創作による嘘字である

i. 八三年の改正では、...思いっきり簡略化され、ウソ字といった方がよいような字形が創作された(9)

83JISの創作ではない。規格票附属書7272[「JIS漢字典」左286]等を参照。

ii. こういう類推文字、あるいは擬造文字を、JISはいっぱいこしらえた。...JISの作った字はパソコンやワープロにくみこまれる。これらの機械を作った文章を書く人は、無意識のうちにおうしたウソ字を使わされてしまう。のみならず、新聞がJISの字を使う。そしてしゃあしゃあと「鷗や洗はいちおうの市民権は得ている」と書いている。...自分[新聞]がJISの字を採用して市民権を得させてしまったんじゃないか。...なるほどパソコンやワープロは、「鉱工業製品」だろうが、なんでそんな工業団体に、勝手に文字を作る権限があるのだ。(31)

新聞の用例の方が遥かに古い。

- iii. 「83JIS の嘘字」(37)(豊島)
無知による誤記。(懺悔します)。
- iv. 豊島正之 JCS 委員は、鷗外自身の手書き文書の中に鷗の字のあることを例示しました。しかし、プライベートなメモ書きのようなものと、公表された作品の署名を一緒くたにして論じるのは、考えるまでもなくおかしいことです。(53)
78JIS 以前・朝日文字以前の用例として言及している事を席上明言しており、筆写字体と公表署名を「一緒くたにして論じ」た等という事実は無い。

28. 83 改訂
1983/1990 年の改訂の問題点について、(政府の名において)一度国民にお詫びしておくというか、何かその類の対応をきちっとしておくべきではないでしょうか。(49)
X0208:1997 解説で、一応対応。

29. 83 改訂の 97JIS による評価
九七 JIS 規格書では、過去の改訂の矛盾点が一つ一つ指摘されています。…一面で言うと JIS 規格というのは、その時々で方針が変わることになります。そして方針の度重なる転換は、混乱をもたらすのです。(53)
エイズ被害の様に一貫して誤りを押し通す事は、望ましくない。

I 漢和字典に無い字が X0208 にあるのはおかしい

- 30. X0208 には、漢和字典にも無い文字が沢山ある。これは、如何にいい加減に規格が作られているかの証拠である。
前半は正しい。後半は、日本で実際に使われている文字を採集しようとする姿勢が漢和字典に無かったからに過ぎない。漢和字典に無ければいい加減な文字だと言う発想は、漢和字典に対する過剰な期待である。
- 31. 「諸橋『大漢和』に載っていない漢字というのが、そもそも異様である。」(38)p.91
同上
- 32. 現代のコンピュータ社会で、まがりなりにも漢字が扱えているのは漢字コードのおかげなのだから、その制定過程はさだめし熟慮されたものと思われるだろうが、あにはからんや、どうもわけのわからないことが多いのだ。たとえば比較的使用頻度が高いはずの JIS 第二水準漢字の中に全く意味も出典も不明な物があるという。(38)p.84
意味不明の文字は、「JIS 漢字字典」が「誤掲出」とした堵(52-55)壘(52-63)岨(54-12)榜(59-91)蟬(74-12)のみ。この内、暗合用例すら欠くのは弼(55-27)のみ。「多い」とは言い難いのでは。
- 33. 昨年から今年にかけて、今日流通している漢字の収集に集中した結果、重複はあるが 13 万字に達した。これを整理するために漢字データベースの構築が急務になり、文字画像と、部首、画数、音訓、そして種々の既存のコード、さらには漢字を構成する要素によって検索可能な方式を、現在実現しつつあるところである。出典や引用もわかる範囲で入れる予定だが、既存の JIS の中には典拠はおろか音訓の読みさえもわからないものが入り混じっている。13 万字のデータを整理すると、6 万字から 8 万字の漢字文字種が析出できよう。その漢字辞典(コード表)を作るのが当面の目標である。(34)
同上

J 規範性・漢字制限論への反発

- 34. 文化の領域である文字を工業規格である JIS が決定する権利は無い
(1.3.4 p.6) 当用漢字表、常用漢字表、康熙字典、大漢和などは、規範的な文字表、すなわち、使用漢字の制限又はガイドを行ったり、漢字の字形までも何が正しいとする規範的な文字表だと思いますが、一方、JIS コードは、このような意味をもつ文字表ではありません。
JIS コードは、存在する文字がどのコードポイントすなわち、ビット組合せに対応するかだけを規定しています。ただし、この規定の中で、どの範囲の字形が特定のビット組合せに対応するかに関しては、これは JIS が決めることであって、文部省が決めることではありません。
すなわち、国語審議会では、この漢字は、このように書かなければならない、とするのに対して、JIS コードでは、このように書かれた漢字のコードは、これである、とするわけです。
繰り返しますと、文部省・国語審議会で規定する文字表は、文字を書く行為に対しての規範を与えようとしているのに対して、JIS コードは書かれた文字のビット表現を規定しているだけです。(芝野)
- 35. JIS は文字を制限している。
かぎかつこが一種類でいいなんて、誰に決める権利があるのか。(48)
新 JIS では、約物類も追加します。
- 36. 規範性
コンピュータで(自然)言語を処理・伝達するときは、文字とは、などと、その言語の規範などを決めないとできませんが、それをやるのは JIS ではなく、国語審議会というところが行くべきだという論は、よく見かけるものです。(1) (1.3.4 p.6)
- 37. 文化の破壊
情報革命が手に入れた世界の直接コミュニケーションの結果が、多様な文化を破壊することであってはならない。むしろ固有文化を充実し、情報発信のためのツールを強化することこそ豊かな地球につながる。(39)
- 38. 文字コード黒船観
第一水準、第二水準あわせて 7 千から 8 千字。約物、異体字、外字 [ママ] などの省略は著しいですね。それでも、メイドインジャパンといえるのでしょうか。黒船の来航と一緒に、コンピューターという外国からの圧力でやむを得ず作った、という感がありありですね。私の名前はしんにゅうの点が二つなんです、DTP にはそんな文字はありません。じゃ私は日本人と違うんですかと、聞いてみたいですね。正直言って今は、日本語コードが鎖国するのか、それとも日本国憲法のようにアメリカに「正しい日本語コード」を作ってもらおうのか、これは国論を二分してもおかしくない大問題でしょう。歴史的に見ると、国家が文字について統制しようとしたとき、ろくな結果を招いては

いない(48)

39. 国家権力の横暴

JIS コードは国家権力の横暴の所産です。

それまでは、活字の分野で丁寧につけてきて、約6万字以上の集積があったわけです。その事実すら知らされることなく、闇に葬られようとしている。字を葬るということは、言葉を葬るということと同じでしょう。(48)

府川充男氏の御教示によると、小宮山博史氏尚蔵の活字総数見本帖の漢字数の内、字数が明記されているものの最大は、23,758字(華英校書房16ポイント分合活字見本帖、1846年)、日本では、13,540(秀英舎七ポイント半仮名附活字見本帖、大正15(1926)年)、11,195(モトヤ商会華文書体見本(昭和30年代前半))、10,794(博文館明朝四号活字見本帖、大正3(1914)年)程度で、他は1万字未満である。「約6万字以上の集積」の典拠を知りたい。

K 公開レビュー

40. JIS は密室審議である

すべての人々と密接に関わる問題が、一部の専門家の手によって知らないうちに決定されていた、というのはこうして書くのも気が引ける陳腐な決まり文句なのだが、事実は決まり文句通りに推移してしまった。ここで一つ不審に思うのは、文字コードの決定プロセスにおいて、当然、国語学や文字学の専門家への説明、ヒアリングは行なわれたであろうに、どうしてこのような事態になるまで歯止めがかからなかったのか、という点である。…インターネットが各分野で今後どのような存在になるかということについて、十分な理解を得た上で、国語学者や文字学者へ文字コード規格案を説明したのだろうか。(51)

公開レビューを行っており、密室審議ではない。X0208:1997の公開レビュー審議の記録も公開している。(2.2 p.8)

41. 公開レビューなど知らない

原案の公開レビューが行なわれましたが、このレビューの存在自体、一般の人は知りません。どこでどんな風にレビュー実施の広報がなされたのかすら知らないのです。…工業技術院やJIS規格制定に関わった情報処理の専門家は、もう何年も前から、今日の事態を予測できたはず。なぜ自分達だけで情報を独占し、多くの人々に問題の所在を告知しなかったのか、残念でなりません。(51)

次項参照。

42. 公開レビューなど誰も関心が無い

今年(1966年)の5月に、来年(1997年)早々に改正されるX0208の公開レビューがあったんですが、参加者は10人不足だったそうです。(12)

X0208:1997公開レビューは、1996.11~1997.2に懸けて行われ、雑誌やネットワーク上の媒体で周知した。この結果、150人以上の方から資料の郵送が請求され(FTPはその数倍)、海外を含む34人の方からコメントの送付・メールを受けた。公開レビュー公開審議(1996.3.18)参加者は10人であった。

L 78JIS

43. 日下部表出自字

僕はJIS漢字表を見て改めて驚いたのは、伝統的に書いて美しい楷書のかたちを活字にしたのがずいぶん入っているんですよ。[挙例無し]…活字は活字として統一されていないといけないのに、JIS漢字には楷書がいくつも入っています。(25)

「日下部表」出自字の事かと思われる。「JIS漢字辞典」コラム28(555)を参照。

M 漢字が足りない・多過ぎる

44. JIS X0208 では文字が足りない

御意。新JISを策定中。

45. コストパフォーマンスが悪い

私は、今回JIS漢字の見直しをすることを、数年前から聞いておりました、ひそかな期待を抱いていたんです。改訂作業で本格的に再検討し、改めるべきは改めて、使う人にとって使いやすく便利な、いわばコストパフォーマンスのいいものができると期待していたんです。しかし、レビューを見ましたところ、どうもそうではないらしいということがわかりました。のっけから、「字の入れ替えや字体の変更はしません」と改訂の概要説明に書いてあります。(46)

入替え・字体変更は、83JISの悪夢の再現になり兼ねない為、採らないこととした。規格票解説1.3(「JIS漢字辞典」左325)を参照。

N 誤解と伝説

46. 情報処理学会コードの成立

私がある人から聞いた話では、「最初は複数の国語学者が入っていたため、途中から作業が進まなくなった(国語学者の方ごめんさい)。そこでいったん解散して、次は林大先生一人にしてやりなおしたら、うまくまとまった」とのことでした。(41)

実際の委員構成は、国語学・言語学5、理科系3(1970)、国語学・言語学5、理科系6(1973-1976)であった。

47. C6226は、人名で区別される字は包摂せずに区別した(富・冨、峰・峯など)(3)p.71

C6226-1978規格票解説の誤読。包摂しなかった例で人名にある例を示しただけ。(これが誤読であった事を著者も認めている)。

48. X0208規格票の印字字形は、標準字形として推奨されたものだ。

誤読(1.3.4 p.6)

49. 部首統合の誤解

78JISの漢字索引部首は、形が似ているからというだけで匚・匚を統合するなど、奇怪な部首分けをしている。(5)

当該部首統合に関しては、78JIS 以前から例はある (eg. 長澤規矩也) し、部首統合自体は、和玉篇や康熙字典等、各時代でそれぞれ行なって来た歴史がある。

O 新 JIS 関連

50. X0208/新 JIS の字の採録基準

現代日本語用の漢字なのか、日本で使われる漢字なのか。後者には、他漢字文化圏の固有名詞等、過去の文献の(現代日本に於ける)翻刻など、他言語(アイヌ語など)の表記などを含む。

後者を想定している。(2.4.1 p.9)

51. 外字を使い続けたい

外字を使うシステムがまだ多数存在する状況では、83JIS 登場時をはるかにこえる混乱が生じると予想される。...外字原則禁止からたった二年で、外字領域を使った文字拡張を強行するというのが無理な話なのだ。JIS X 0213 は、最低五年間凍結すべきだと考えるが、いかがだろうか。(11)

新 JIS は、X0208 を補い X0208 との共同運用のみを想定しているが、別の符号化文字集合であって、X0208 単独で運用するのは依然自由であるから、その様な混乱は生じない。

52. 国語審議会答申を待つべき

JCS 委員会は、「新拡張 JIS」による五〇〇〇字拡張を進めていますが、なぜ国語審議会が字体に関する答申を出すまで待てないのか。「新拡張 JIS」は外字に使われている空き領域を使って文字を増やすというトリッキーなことをやっていると、その点でも九九年施行は早すぎます。(13)

JIS は字体規範を決めるものではない。(1.3.4 p.6) 国語審議会答申が出れば、当然それに従う。国語審議会答申が出るまで、それらの文字の情報交換が不要ならばよいが、情報交換が必要であれば、工業規格は必要である。

53. 新 JIS と国語審議会との関係

今年から 2 年間かけて国語審議会が、JIS 漢字の字体と伝統的な字体との差異をどうするか審議するという。JIS 第 3・4 水準漢字制定も、今年から 2 年間。密接に関係するテーマであるし、どちらも公的存在である。はたして、連携して作業が行われるのだろうか。2 年後に、両者の結論がバラバラだと、社会はかえって混乱してしまう。(26)

文化庁の委員も参加されている。

54. 外字を使わせなくするのは犯罪行為である

「新拡張 JIS」の 1999 年施行を強行したら、国民生活に深刻な影響が出るのは火を見るより明らかです。薬害エイズに匹敵する犯罪行為と言ってもいいかもしれません。わたしたちは、このような性急な施行を黙認した工業技術院をはじめとする行政の責任を追求していくつもりです。(7)

規格開発作業が犯罪呼ばわりされた事を、ここに記録に留めて置く。

55. 新 JIS と補助漢字 (JIS X 0212)

第 3・4 水準漢字の選定では、補助漢字とあえて重複をいとわず符号化する、となっている。これで、事実上補助漢字は見捨てられてしまいます。このような方針をとる理由は何なのだろうか。(26)

第 3、4 水準漢字は補助漢字を捨てた上で、作成されるものと聞いています。現在、UNIX や TRON 等で利用されている補助漢字を捨てるのは、いわゆる旧 JIS、新 JIS 問題と似たような状況を招くと推測していますが、もし、そうなった場合、それに対する責任はどのようなものになるのでしょうか。(補助漢字を捨てるな、と言っているのでは、ありません。)(19)

「JIS 漢字の歴史」(芝野主査講演)を参照。

56. 新 JIS は使ってもらえるか

補助漢字はほとんど使用されなかった。今度の第 3・4 水準漢字が同じ運命をたどらないために、何か方策を考えてあるのだろうか。(26)

X0208 との共同運用・実装可能な形態を選択している。(2.3 p.8)

57. 新 JIS とユニコード (ISO/IEC 10646, JIS X0221?) との関係

● ユニコードの空き領域

第 3・4 水準漢字は、ユニコードに追加提案するようである。中国・韓国なども文字を追加していると聞く。はたして日本の第 3・4 水準漢字を追加する空き領域は、確保できるのだろうか。それとも、何らかの切替えをするのか。(26)

UCS(ISO/IEC 10646)への追加要求に就ては実装面から検討中。Unicode への追加要求の予定は(今の処)無い。

● ユニコードの中での二重コーディング

ユニコードは、第 1・2 水準、補助漢字を含んでいる。第 3・4 水準漢字がユニコードに追加されると、補助漢字の部分で二重にコードが割当てられてしまう。これをどうするのか。(26)

同上

● ユニコード系と第 1～4 水準系の並立

今後 JIS には、ユニコード系と第 1～4 水準系に 2 つの文字コード体系が存在するようになるが、無駄(または面倒)ではないのか。また、どのように使い分けたら良いのだろうか。(26)

P 批判とは言い難いもの

「地球に優しい」執筆活動を切望する。

58. 虚偽

JIS が、当用漢字以外は当方がひきうけましよう、諫言の諫は諫 [X0208 に両方あり] に、洗煉の煉 [83 置換え] は煉にと、全部変えてしまったのだ。...JIS が僥倖を<人堯> [X0208 に無し] 倖にする...よってたかっの文字いじりである。...すくなくとも JIS の類推文字は、一度すっかり御破算にすべきだ。(31)

虚偽である事の指摘(4)(44)にも拘らず、そのまま単行本化。

59. 虚偽
Q: JIS X0201:1997 が、1 バイト片仮名を廃止したというのは、ほんとうですか。A: ほんとうです。(5)
嘘です :-) 規格票を読まずに書かれたもの。(この時点で規格票は印刷中)
「過去の日本語を忠実にテキスト化する」も結構だが、まずは現代の日本語に就て、典拠を確かめ、原典を理解した上で、事の実否を確認してから執筆して頂きたい。
60. 虚偽
JIS X0208:1997 は、「区点によって符号化されるのは、無形の「文字概念」ではなく、文字の骨格としての「字体」であり、その「字体」に肉づけしたものが「字形」であると明確に定義し、字形のゆれの許容範囲を「包摂」として明示している。(7)
規格票を読まずに書かれたものか。(この時点で規格票は刊行済)。規格票・解説のどこにも、この様な文言は無い。
61. 虚偽
78JIS のマシンの「鷗外」は、JIS X 0221(ユニコード)のマシンでは「鷗外」になり、97JIS のマシンでは「鷗外」になるのか「鷗」になるのかはわからない。わからなくていいというのが、97JIS の立場です。ひじょうにわかりにくかったと思いますが、これが JIS コードの現状です。(笑)(13)
規格票を理解せずに書かれたものか。X0208:1997 3.1.3 で、どちらを選んだかを明示する事が要求されており、「どちらか分かる様に明示しなくてはならない」が 97JIS の立場である。
62. 虚偽
字体包摂が明示してあるのはここ [規格票附属書 7 区点位置詳説] に出て来る文字と、包摂基準の説明に例示されている文字だけで、「JIS 基本漢字」の六五〇〇余字すべてについて書いてあるわけではないんです。(13) なるほどね。あとは推して知るべしという方針ですか。(2)
規格票附属書 6 を見ずに為された発言か。因みに X0208 の漢字は 6355 字で「六五〇〇余」ではない。これも、規格票を見なかった証拠 :-)
63. 虚偽
芝野耕司委員長に直接おたずねしたところ、以下のことがわかった。... 新拡張 JIS の五千字追加という文字種... は... 少ないように思えるが、... 膨大な文献資料で文字同定・使用頻度調査をおこなった結果、現代日本語を表記するには、これで十分ではないかという結論に達したということである。... 日本語全体では使用頻度が低くとも、特定の分野で絶対に必要な文字というものはあるのである。使用頻度で文字を選定するという思想は、技術が未発達で、コンピュータの影響力が限られていた時代には有効だったと思う。(9)
78JIS も新 JIS も、頻度に基づく文字選定など行っていない。(芝野主査講演の「JIS 漢字の歴史」を参照)。恰も直接取材に基づくかの如き行文は、虚偽。
64. 虚偽
[ユニコードには漢字が] 字体などが混乱したまま取り込まれている。単独文字の「區」を、簡単に「区」と表記すると決めてあるため、「鷗」の「區」も「区」としているのがいい例だ。このため、森鷗外が、「鷗」で表記されてしまう。同様に「國」が「国」となるための混乱など不整合な点が多い。(54)
Unicode Standard/ ISO 10646-1/JIS X0221、どれも読まずに書かれたものか。行文は全て創造上の産物。
65. 虚偽
字体の差をデザイン差と同じようにあつかいなさいということです。... 字体差をデザイン差と同様に扱っていいのかという問題ですが、僕はおかしいと思うんです。(13)
規格票を読まずになされた発言か。
66. 虚偽
文字コード問題の不幸は、文字コード開発をサポートする研究機関がなかったことである。文部省傘下に国立国語研究所があるが、ここの幹部は伝統的に漢字制限論者が多く、現代日本語の用字調査は新聞の論説を調べた程度で、本格的な研究は、最近になるまでおこなわれなかった。(14)
雑誌 90 種調査(1962) 等、本格的な研究は昔から多数ある。
67. 典拠不明
何で日本はこんなに派閥が好きなの、という失望感をどうしても抱いてしまいます。[新 JIS 開発が] やっていることは、ほとんど TRON が東大で行っている作業と同じでしょう。どうして協力して行えないのか。村社会的感性ははやく捨てて欲しい。(18)
68. 典拠不明
JIS コードは、理系の技術者によって作られたという先入見は間違っていました。技術者の方々は、文化に対してもっと謙虚だと思います。JIS コードは、漢字制限論の系譜に連なる国語学者が確信犯的に作りあげたもので、ひじょうに根が深いのです。... JIS の文字コードの漢字の扱い方がでたらめなのは、理系の方が乱暴なことをやったんじゃないって、文系の確信犯がやったわけですね。(12)
典拠も掲げずに、林大(はやし おおき)氏など、JIS 初版制定に 9 年近く尽力されて 78JIS 初版を創り上げた方々を犯罪人呼ばわりとは、余りに非礼ではないか。
69. 典拠不明
JIS 側は、国民の大部分は異体字や旧字体を使いわけの能力がないから、これで [今の包摂規準で] いいとしている。(10)
著者に確かめたが、直接のソースは無く、論理(?) は下記の通り。
(1) 朝日新聞の記事で、こうした字体の差が使い分けられていないものがあり、それを芝野主査が指摘した。
(2) 朝日新聞の記者の国語力で使い分けの事が出来ないものは、国民の平均的国語力では使い分けの事が出来ない。
(3) 何故なら、(以下原文のまま)「朝日新聞は日本第二の発行部数の大新聞で、入社試験もひじょうに難しいと聞い

ております。記者&校閲部の国語力は、国民の平均的国語力を上回っていると考えるのが一般的な認識であるからである。

- (4) よって、主査が朝日新聞の記事で使い分けられていない事を指摘した事は、国民の大部分が使い分けの能力が無いと指摘したのと同じである。

70. 意味不明

こういう[3画・4画草冠を包摂する様な]なんでもありの規格では、學術分野はもちろん、電子署名でも大変な混乱がおこることが予想される。(10)

電子署名は暗号化の事だが、何か根本的な誤解がある様である。

71. おまけ(欠陥標準)

シフト JIS は爆発的に広まって、市場標準(ディフェクト・スタンダード)になりました。(53)

dē factō (ラテン語) は defect (英語) ではありません。

4 引用文献

4.1 一般文献

- 上田秋成全集編集委員会(1990) 上田秋成全集 巻7(中央公論社、長島弘明校訂)
- 坂村健(1996a) デジタル・ミュージアムと文字(「歴史の文字-記載・活字・活版」, 東大出版会)
- 坂村健(1996b) インターネット時代の文字コード-既存他国語処理の破綻と TRON-(TRONWARE36)(パーソナルメディア)
- 坂村健(1997) ネットワーク時代-日本語が危ない(TRONWARE44)(パーソナルメディア)
- 芝野耕司(1997) JIS 漢字字典(日本規格協会)
- 長島弘明(1998) 清書できないワープロ、劣化したデジタル資料(平凡社、1998)
- 日本文藝家協会(1997) お寄せ頂いた文字による文字コード調査一覧(文藝家協会ニュース 556 附録)(日本文藝家協会電子メディア対応特別委員会)
- 平凡社(1998) 電脳文化と漢字のゆくえ(平凡社)
- 家辺勝文(1995) Niftyserve fshtext:fshtext:19-175
- 山本太郎(1998) 文字コードに関する最近の議論について
(<http://www.kt.rim.or.jp/~tyamamot/charcode/onchar.html>)

4.2 JIS 批判関連文献

1. 文献は、左端の番号で引用する。
2. 署名の取材記事は取材者の文献と見なし、無署名取材記事は刊行者の文献と見なす。
3. 匿名座談会の発言は、刊行者の文献とする。
4. 署名付き座談会の発言は、発言者の文献とする。
5. 取材記事の引用部分を、被取材者の文献とする事は無い。但し、全面引用は、被引用者の文献(所謂「孫引き」とする事がある。

- 1 五十嵐和弘(1997) fpmc:8-102
- 2 池澤夏樹(1998) 平凡社(1998) 中の座談会発言
- 3 太田昌孝(1997) いま日本語が危ない(丸山学芸図書)
- 4 翁雅男(1997) JIS への誤解について(週刊文春 97-10-30)
- 5 加藤弘一(1997) ほら貝 1997Feb25 現在
- 6 加藤弘一(1997) ほら貝 1997Mar05 現在
- 7 加藤弘一(1997) ほら貝 1997Aug18 現在
- 8 加藤弘一(1998) ほら貝 1997Mar01 現在
- 9 加藤弘一(1997) 思想としての文字コード(三田文学 76-48)
- 10 加藤弘一(1997) 島田雅彦等(1997)p.219 署名コラム
- 11 加藤弘一(1997) 文字コード問題、この一年(季刊本とコンピュータ 1997 秋)(トランスアート)p.12-13
- 12 加藤弘一(1997) ネットワーク時代-日本語が危ない(TRONWARE44)(パーソナルメディア)
- 13 加藤弘一(1998) 平凡社(1998) 中の座談会発言
- 14 加藤弘一(1998) 平凡社(1998) 中の「用語解説」
- 15 児島秀樹(独楽)(1995) fshtext:19-148

- 16 児島秀樹 (独楽)(1995) ftron:11-50
- 17 児島秀樹 (独楽)(1996) freki:18-2290
- 18 児島秀樹 (独楽)(1996) freki:18-2799
- 19 児島秀樹 (独楽)(1996) freki:18-3594
- 20 坂村健 (1996) 世界規模通信基盤としての多国語処理 (http://http://www.personal-media.co.jp/vs/mltp96/keynote__j.html,1998.3.1 現在)
- 21 坂村健 (1996) インターネット時代の文字コード-既存他国語処理の破綻と TRON-(TRONWARE36)(パーソナルメディア)
- 22 坂村健 (1996) 今だからこそ多国語処理について話そう (<http://www.personal-media.co.jp/vs/mltp96/sf1.html>,1998.3.1 現在)
- 23 坂村健 (1996) 未来の文字コード体系に私達は不安を持っています (<http://www.personal-media.co.jp/vs/mltp96/body.html>,1998.3.1 現在)
- 24 坂村健 (1997) ネットワーク時代-日本語が危ない (TRONWARE44)(パーソナルメディア)
- 25 佐野光一 (1997) 座談会教室の漢字・街角の漢字 (下)(ぶっくれっと 122)(1997.2 三省堂) 座談会中の発言
- 26 柴田義之 (1996) ftron:11-389
- 27 芝野耕司 (1995) freki:19-378
- 28 島田雅彦等 (1997) 「電脳売文党宣言」(アスキー出版局)
- 29 島田雅彦 (1998) 平凡社 (1998) 中の座談会発言
- 30 ジャグラ「印刷文字問題研究会」(1996) 漢字を総背番号制に (月刊プリテックスステージ)1996-7(ニューブリントング刊)
- 31 高島俊郎 (1997) JIS とは何物だ?(コラム「お言葉ですが」週間文春 97-7-3p.134-135)(1998 刊単行本でも変更無し)
- 32 谷本玲大 (1995) freki:19-428
- 33 田村毅 (1995) 日本漢字に“電子の住所”(日本経済新聞 1995-12-18)
- 34 田村毅 (1996) 活字から電子文字へ (<http://www.um.u-tokyo.ac.jp/>)
- 35 堤靖彦 (1997) ftron:11-506
- 36 豊島正之 (1991) コードのはなし (情報処理語学文学研究会会報 9)
- 37 豊島正之 (1995) コードのはなし:追補 (情報処理語学文学研究会会報 17)
- 38 中西秀彦 (1997) 「印刷はどこに行くのか」(晶文社)
- 39 日本経済新聞メディア取材班 (1996) メディアの波頭 (日本経済新聞 1996-1-28)
- 40 プリテックスステージ (1996) 国際文字規格・ユニコードと国産 OS トロンへの期待 (月刊プリテックスステージ)1996-3(ニューブリントング刊)
- 41 古瀬幸広 (1995) fshtext:19-278
- 42 平凡社 (1998) 電脳文化と漢字のゆくえ (平凡社)
- 43 由里幸子 (1997) 揺れる漢字の字体 電子化で多様化と統一と (朝日新聞 1997.3.1)
- 44 前田年昭 (1997) 文字コードをめぐる迷信 (コリイカ 1997-11)
- 45 松為彰 (1997) BTRON サブプロジェクト (TRONWARE48)(パーソナルメディア)
- 46 松岡榮志 (1997) 座談会教室の漢字・街角の漢字 (下)(ぶっくれっと 122)(1997.2 三省堂) 座談会中の発言
- 47 水羽信男 (1995) fshtext:19-213(村田忠福「漢字コードについて」中国研究者のためのワープロ活用術 5 『蒼蒼』20号、1988年6月の引用)
- 48 文字デザイン研究会 (1995) 「文字に異義あり」アンケートを終えて 季刊モジカ創刊準備号 (文字デザイン研究会) 匿名座談会
- 49 家辺勝文 (1995) fshtext:19-125
- 50 家辺勝文 (1995) fshtext:19-175
- 51 吉目木晴彦 (1997) 日本語が消滅する日 (三田文学 76-48)
- 52 吉目木晴彦 (1997) ネットワーク時代-日本語が危ない (TRONWARE44)(パーソナルメディア)
- 53 吉目木晴彦 (1998) いま、何が、なぜ、問われているのか (平凡社 1998)
- 54 渡辺新平 (1996) コンピューターの漢字使用 読売新聞 1996-5-9

- fpmc, freki, fshtext, ftron などは、ニフティサーブのフォーラム
- URL は、特記しない限り、1998年3月1日現在の内容
- 「ほら貝」の URL は、<http://www.win.or.jp/~horagai/>だが、内容は引用当時のもので、現在は、それと明示する事無く書き換えられている可能性がある。

謝辞

草稿に対して有益な御助言を賜った芝野耕司氏、前田年昭氏、家辺勝文氏に、深く感謝します。