

文字コード標準体系専門委員会報告書

2002年3月

社団法人 情報処理学会

情報規格調査会

All Rights Reserved, Copyright © Information Processing Society of Japan

目次

第1章 はじめに	4	4.3.5 評価	26
1.1 文字コード標準化活動の背景	4	4.4 構造化4バイトコード方式	27
1.2 第1ステージ委員会の提言	5	4.4.1 構造化4バイトコードによる方法	27
1.3 JTC1総会における日本の提案	6	4.4.2 異形字の拡張方法と任意の異形字を代表 形とする包摂関係の規定	29
1.4 第2ステージのための準備委員会の調査・検 討	6	4.4.3 4バイトコードの枝番号で規定した94個の 異形字を8836字に拡張する方法	29
1.5 第2ステージ委員会の目的と具体的考え方	6	4.4.4 任意の異形字に対する代表形と包摂関係 の規定方法	30
1.5.1 新字種の標準化	7	4.5 文字図形構成要素組合せ方式	30
1.5.2 異形字の標準化	7	第5章 今後の課題	32
1.5.3 その他の方針	7	5.1 運用方法についての課題	32
第2章 提言サマリ	8	5.1.1 枝番付与の方法 (制度・仕組み)	32
2.1 提言	8	5.1.2 枝番の利用 (データの生成)	33
2.2 新字種提案制度	8	5.1.3 送信と受信の関係	33
2.2.1 提案文字検討委員会	8	5.1.4 書体との関係	34
2.2.2 作業委員会	8	5.1.5 データとしての扱い	34
2.2.3 文字データベース	9	5.1.6 不適切な運用の防止	35
2.2.4 文字提案法	9	5.2 枝番方式の運用に関する考慮点	35
2.3 異形字アーキテクチャ	9	5.2.1 行政分野と学術分野の相違	35
2.3.1 異形字標準化の効果	9	5.2.2 行政分野において予想される粒度の相違	35
2.3.2 符号枝番方式	10	5.2.3 学術分野における粒度の相違	36
2.3.3 フォント枝番方式	10	5.3 符号枝番方式における運用上の要件と課題	36
第3章 漢字等標準化のための新字種提案制度の 創設	12	5.3.1 要件	36
3.1 背景	12	5.3.2 符号枝番方式における運用上の課題	37
3.2 提案概要	12	5.4 国際整合性について	37
3.3 提案内容	12	5.4.1 既存システムと国際的ネットワーク環境 との相互互換性	37
第4章 異形字アーキテクチャ	15	5.4.2 異形字における言語・文化依存性 (特に漢 字に関して)	38
4.1 異形字のアーキテクチャのまとめ	15	第6章 委員意見	40
4.1.1 提案解説	15	6.1 北澤誓四郎 (日本経済新聞)	40
4.1.2 境界	18	6.2 千田昇一 (NTT)	41
4.1.3 異形字に関する要求の合意の必要性	19	6.3 秦恒平、加藤弘一 (日本ペンクラブ)	41
4.1.4 その他の条件	19	6.4 宮澤彰 (国立情報学研究所)	44
4.1.5 本委員会外での動向	20	6.5 佐藤敬幸 (CICC)	45
4.2 符号枝番方式	21	6.6 Martin J. Dürst (慶應義塾大学)	45
4.2.1 概要	21	第7章 おわりに	47
4.2.2 符号枝番方式の構造	21	付録1 標準化動向	48
4.2.3 異形字機能コード	21	付1.1 国内標準化動向	48
4.2.4 異形字番号	22	付1.2 国際標準化動向	48
4.2.5 異形字指定の具体例	22	付録2 文献	50
4.2.6 符号枝番方式の効果	22	付2.1 論著	50
4.3 フォント枝番方式について	24	付2.1.1 1970年代以前	50
4.3.1 背景	24	付2.1.2 1980年代	51
4.3.2 問題意識	24	付2.1.3 1990年代	52
4.3.3 フォントによる字体の相違	25		
4.3.4 第一ステージにおけるフォント枝番方式 の提案	25		

付2.1.4 2000年代	57	付4.1 第1回委員会	71
付2.2 論文	58	付4.2 第2回委員会	71
付2.2.1 1980年代	58	付4.3 第3回委員会	71
付2.2.2 1990年代	58	付4.4 第4回委員会	71
付2.2.3 2000年代	61	付4.5 第5回委員会	72
付2.3 報告書等	61	付4.6 第6回委員会	72
付2.4 規格・標準	62	付4.7 第7回委員会	72
付2.5 ホームページ	64	付4.8 第8回委員会	73
付2.6 ガイドブック・マニュアル類	65	付4.9 第9回委員会	73
付2.7 新聞記事	65	付録5 委員会議事録および関連報告書	74
付2.8 字書・辞典	66	付5.1 委員会中間報告書	74
付2.9 パソコンソフト等	68	付5.2 委員会議事録	74
付録3 用語	69	付5.3 準備委員会報告書	74
付3.1 本委員会として厳密な定義を定めずに用 いる用語	69	付5.4 準備委員会議事録	74
付3.2 本委員会として定義を試みる用語、及びそ の定義	69	付5.5 第1ステージ委員会報告書	75
付録4 委員会配布資料一覧	71	付録6 委員会名簿	76
		付6.1 文字コード標準体系専門委員会	76
		付6.2 符号化文字集合検討準備委員会	77

第1章 はじめに

1.1 文字コード標準化活動の背景

漢字を含めた文字コードの国際標準化作業は、国際標準化機構（ISO）と国際電気標準会議（IEC）のもとに作られた合同の技術委員会JTC1（Joint Technical Committee No. 1）で進められている。そして文字コードに関する技術的な審議は、その下の専門委員会SC2が担当している。現在のところSC（サブコミッティ）は36までであるが、SC2が情報技術の国際標準化の中でも最も古い歴史をもつ。SC2の幹事国業務は日本（情報処理学会情報規格調査会）が受持つとともに、議長も日本から出している。

我が国における文字コードの標準化の歴史をたどってみると、わが国でコンピュータが使用されるようになり、コンピュータ上で漢字を使用するための「漢字コードの標準化」が試みられたのは、1978年のJISにまとめられたのが初めてである。これは78JISと呼ばれて、普及しだした初期の大型コンピュータに実装されるとともに、中国や韓国等の漢字使用国にも注目され、お手本になったそうである。

続いて1980年代のワークステーションの時代から1990年代になるとパソコンが普及し始めて、コンピュータの専門家だけでなく世界中の一般ユーザ間でもコンピュータネットワークによる電子メールが広まりだした。そのようなタイミングで、JTC1では世界中の文字を集め、コンピュータの上で自由に世界中の言語を表現できるようにする、文字コードの標準化の動きが始まった。1993年に制定されたISO/IEC 10646-1:1993という国際標準がその成果であり、わが国でも翻訳されてJIS X 0221:1995となった。この標準化には日本を初めとする漢字使用国がかなり貢献した。この10646には日本の既存のJIS漢字がすべてそのまま含まれている。また、欧米のラテン文字ばかりでなく、アジアのさまざまな国の文字が既に入り、世界中の少数民族の文字や古代文字なども標準化されつつある。

ISO/IEC 10646は多オクテットコードであるが、8ビット系の文字コードとしてISO/IEC 8859シリーズがあり、主としてラテンアルファベット文字が標準化されている。

1990年代後半に入ると、パソコンとインターネットの爆発的な普及で、使用可能な漢字についてのニーズが顕在化してきた。従来のJIS漢字では不足しているので、大幅な拡充が求められた。そこで国際標準ではISO/IEC 10646-1のための漢字の追加が拡張A（6582漢字）として実現した。これを含めてISO/IEC 10646-1:2000がまとめられた。引き続いて、わが国でも翻訳されてJIS X 0221-1:2001として制定された。次の段階として、すべての漢字の元となっている康熙字典にあるほとんどすべての漢字を中心とした標準化が拡張B（42711漢字）として大規模に進められ、ISO/IEC 10646-2:2001が制定された。これは、我が国で制定されたJIS X 0213:2000のほとんどの文字を含んでいる。

最近では、情報処理学会試行標準という新しい制度が始まり、五つのWGが活動している。その中でもWG2は、今昔文字鏡に含まれるすべての漢字集合と漢字の分解法を対象としている。WG5では8千字程度の基本的な漢字の集合を国際的に標準化する試みが進められている。

上で述べたような文字の不足問題に対して、国内だけでなく国際的な状況も考慮した文字コード体系全体の標準化の方針を検討したのが、1998年11月から1999年8月まで開催された文字コード体系検討専門委員会である。この委員会は第1ステージ委員会とも呼ばれ、その報告書には、それまで文字コードの標準化の歩みと問題点、今後の方針がまとめられている。そこでその概要をまとめ、今回の第2ステージの委員会の検討へつなげるように説明したい。

1.2 第1ステージ委員会の提言

第1ステージでは、漢字の追加要求は、使用者の範囲の増加やコード化すべき対象の増加などに従って際限無く続くという仮定から“青天井問題”と呼んでいた。さらに漢字の追加要求を検討すると、要求の大半は、既存漢字の形の変形であることも認められており、これを“異形字”問題と呼んだ。このように、漢字の要求は、新字種の追加要求と、異形字の追加要求とに分割して対応する必要が認められている。このような検討に基づいて第1ステージでは委員会の結論を下記のような五つの提言にまとめた。

- 提言1 文字コード標準の国際的分散開発の提案を行う。これには各種ユーザ団体との共同開発を行う協調体制を含む。
- 提言2 新字種を国際標準に登録するための新しい制度を提案する。ISO/IEC 2022系のISO 2375に対応するISO/IEC 10646系の登録制度を目指す。
- 提言3 多様な異形字に対応できる文字コード標準のアーキテクチャを提案し異形字の標準化の実現を図る。
- 提言4 異形字を扱える文字コード標準のためのコンピュータモデルの国際提案を行う。
- 提言5 市場性に関する考え方は情報技術の一般的な標準の意味と異なり、少数民族の文字や博物館と図書館などで必要な文字などの公共性を重視した特別な意味をもつ。

このような提言に至った審議の概要は第1ステージの報告書第3章にまとめてある。具体的な検討事項は第4章にまとめてある。主な内容は以下の通りである。

- a) 文字コード概念の整理。
 - 文字および文字コードに関連する用語の整理・統一。
 - 文字コードおよび文字コード標準に関する概念、考え方の整理・統一。
 - 文字コード標準化の国内外の動向に関するレビュー。
- b) 文字コード標準のあり方、適用範囲の明確化。
 - 標準化が望まれる文字セットの整理と分類。
 - 標準の適用範囲、公的標準、工業標準、ボランティア標準等の役割分担。
 - 多様なニーズに応えることのできる文字コード標準体系のあり方。
- c) 後の文字コード標準体系に関する提言。
 - 将来の国際文字コード標準体系のためのアーキテクチャの提言。
 - 既存の標準体系の改良および適切な運用に関する提言。
 - 今後の標準作成プロセスのあり方に関する提言。

委員会の運営は公開を原則とした。従って、委員会の審議の過程で使用した資料についても原則として公開し、議事録もホームページで公開している。

1.3 JTC1総会における日本の提案

1999年11月初旬に韓国ソウルで開催されたJTC1総会には、上記の文字コード体系検討委員会の提言に基づいて、日本からSC2の活動方針に関するコメントとして以下のような提案を行った。

SC2の標準化活動の透明性および信頼性の確保として、古典などの専門領域に関する膨大な文字の標準化は関係団体との協調体制で臨むこと、少数民族のスク립トの標準化に関してはネイティブの参加を原則とすることなどを提案した。また、SC2の標準の市場性は他のIT標準とは性格が異なり、公共性や少数民族への配慮が必要なことを提案した。

これらの提案は多くの国の賛同を得ることができて、レゾリューションも満場一致で採決された。

1.4 第2ステージのための準備委員会の調査・検討

第1ステージの経験に基づいて、第2ステージの委員会での検討方針や検討範囲などを事前に調査するために、情報規格調査会では平成11年12月に符号化文字集合検討準備委員会を組織した。

この背景には、漢字不足に対応する従来の議論が、単なる個々の漢字談義に流れ、具体的な漢字ニーズの把握や、最終的に必要な漢字をどう規格化するかという議論までなかなか進むことが出来なかった反省がある。具体的には、総務省や法務省における人名や地名に関する不足漢字の内容や対応に関する実態調査、今昔文字鏡における日本の古典における国字の調査、NTT電話帳における人名と地名の調査を準備委員会で担当を分担して実施し、面談による聞き取りや、一部では資料の提供を受け内容の検討を行った。

準備委員会の報告書は平成12年10月にまとまった。これは本報告書の付録に添付してある。そこでの結論は以下のように五つの検討課題としてまとめられている。

- a) アーキテクチャ提案の具体化（枝番方式を含む。）
- b) 文字コード標準の市場性についての詳細な検討
- c) 文字フォントの確保と交換の手段の詳細な検討
- d) 標準化プロセスの改善の議論
- e) 用語集の完成

このような準備委員会の検討に基づいて第2ステージ委員会の設立が平成12年11月の情報規格調査会の役員会で認められ、12月から会議が開始した。

1.5 第2ステージ委員会の目的と具体的考え方

本委員会の目的は、第1ステージの提言の中で未達成のものを再検討し、準備委員会の検討結果を参考にしながら、今後の文字コード体系のありかた、標準化の方針を検討することである。

第2ステージの委員会の出発点として、標準化された漢字の不足状態についての認識の共有がある。第1ステージの委員会の時点では主として使用できる漢字はJIS X 0208あるいはJIS X 0221が主なもので、さまざまな文字が足りないと言われていたが、その後、国際標準では拡張Aと拡張Bが実現している。その結果、7

万字程度の漢字が目に見える形で標準化されている。また、国内ではJIS X 0213の標準化が完成している。第1ステージの出発点と比べると今回は標準化された漢字の数は大幅に増えている。しかし、漢字の標準化では、漢字は際限無く増えるという意味で「青天井」といわれており、ある時点で十分ということは無いと考えられる。

漢字不足への対応手段として、現在の文字コードはそのままにしておいて、いわゆる外字での対応がある。しかし、それでは十分に使用者の期待に応えられないと考えられる。そこで、不足している漢字を、既存の漢字の微小な変形である異形字と、まったく新しい漢字という意味で新字種と呼んで2種類に分類し、それぞれの標準化について検討した。

1.5.1 新字種の標準化

新字種の要求は、その全体像を十分に把握することは困難であり、標準化過程のある時点からは、新字種の漢字の要求に基づいた対応をせざるを得ないと思われる。そのような要求は使用の場面や個人的な要求、あるいは時期によって変化するので、明確にこれで十分と言い切ることは難しい。わが国では国字に新字種が多数あると思われる。

1.5.2 異形字の標準化

異形字の標準化にも上記の新字種とまったく同じことが言えて、人名などの個人的差異が、新字種よりもさらに顕著に表れると思われる。したがって、単一の包摂というようなまとめ方は難しいと考えられる。また、異形字の標準化については市場性と標準化の必要性を幅広く検討する必要がある。異形字については新しいアーキテクチャの検討の必要性が認められた。

1.5.3 その他の方針

第1ステージでは、国内の漢字問題と国際標準化問題を平行して検討したが、第2ステージでは国内の漢字問題を検討し、国際提案へ向けての検討を行った。

文字フォントについては、個々のユーザが持つのではなく、情報の送り手が文字フォントを同時に送る方式が実現しつつある。フォントサーバに集中的に蓄えられたフォントをすべてのユーザが共用することも考えられる。そのような環境における文字フォントの確保と交換方式、さらに文字コードの在り方を検討する必要があった。

現在のISO/IEC 10646の開発でも、収容文字数が7万字に近くなると、文字の同定やユニフィケーション作業が膨大な手間になり、品質の確保が課題である。従って、標準化プロセスの改善と品質の確保が緊急の課題である。文字コードの対象が専門的、学問的になるに従って分散開発・分散保守の必要性がますます高まると考えられた。

第2章 提言サマリ

2.1 提言

ここでは本報告書で行う提言を要約して説明する。国際標準には既に7万字あまりの漢字が標準化されているが、さらに新しい漢字（新字種）の標準化のニーズがあることと、既に標準化されている漢字のバリエーション（異形字）が多く存在して実際に使用されており標準化のニーズがあることとに対応して2つの提言を行う。提言の具体的な説明は2.2と2.3にあり、詳細については第3章と第4章にまとめてある。それらに関係する今後の課題については第5章にまとめてある。

提言1 新字種提案制度の創設

公的標準で採録候補となる文字を提案する制度の創設。一般から提案された文字の審査のための委員会の設立と文字データベースの作成の提案。委員会での検討法や文字データベースの特徴などの詳細を整理。

提言2 異形字アーキテクチャの標準化

既に標準化されている文字を代表字として、その異形字を枝番で表現するための新しいアーキテクチャ。具体的には符号枝番方式とフォント枝番方式を提案。このような方式の利点と課題を整理。

2.2 新字種提案制度

本制度は、符号化文字集合に関する公的標準へ新規の文字を提案する制度である。これは従来のように担当委員会がみずから新しい文字を調査収集する方法を変更し、広く多くの人々が新しい文字を自由に提案できるようにするもので、提案された文字は検討された後に公的標準へ採録手続きが進められていくものである。

2.2.1 提案文字検討委員会

提案文字を検討し、文字データベースを維持管理するために提案文字検討委員会を設置する。

本検討委員会は、国語や文字に関して権威のある公的組織内に常設する。同時に事務局をこの組織に委託する。この委員会は、学識経験者、文字規格関係者、ユーザ団体、関連行政機関などから構成する。検討委員会は、後述の作業委員会に作業指示を行い、作業結果を審議する。

2.2.2 作業委員会

検討委員会の下に作業委員会を設置し、提案された文字の具体的な検討と文字データベースの維持管理作業を行う。作業委員会は、検討委員会委員の一部と実務担当で構成し、あらかじめ検討委員会により決められた検討条件で、提案さ

れた文字の文字データベースへの登録や、符号化文字としての適不適などを検討する。

2.2.3 文字データベース

文字データベースは、JISへの採録済み文字、国際規格採録済み文字、従来のJIS等開発時に提案・検討されたが採録されなかった文字、本検討委員会に提案された文字などを網羅し、常時最新のものに維持する。文字データベースには、部首・画数・読み・各種規格符号位置などのほか、典拠・用例、あるいは採録不採録の理由などのデータをできるだけ収集する。

2.2.4 文字提案法

候補文字を提案するものは個人または団体とするが、それぞれの応用分野などを代表する団体などが望ましい。提案は常時受理され、特定の受理期間は設定しない。

作業委員会は、あらかじめ設けられた検討条件により提案文字を一定期間ごとに検討し、検討委員会がその結果を審議する。検討委員会は承認した提案文字をJIS等の公的標準候補文字としてプールする。

2.3 異形字アーキテクチャ

異形字を扱う方法として、可能性を広く挙げると、以下のような方法が考えられる。

- 従来のアーキテクチャの元に、異形字に新しい文字コードを割り当てて次々に収録していく方法
- 新しいアーキテクチャの元に、既存の文字を代表字とし、その代表字の文字コードに異形字ごとに枝番を付す方法
- 異形字をコード化せずに、絵として直接送信する方法

また、第4章で詳しく述べるように、約83万字まで拡張可能な「構造化4バイトコード方式」と、文字を構成要素に分解してバリエーションを詳細に表現できる「文字図形構成要素組合せ方式」の提案もある。本章では上記2番目の枝番を使用する方法を主に、要約して説明する。

枝番を使用する異形字アーキテクチャを用いる方法の一つである符号枝番方式では、現在、フォントを作成して使用している文字をほとんどカバーする形で多く異形字を集めて、国内や国際で共通して使用できるように枝番を割り当てる。それに比べて、フォントを決めてから枝番を指定するフォント枝番方式は、すべての枝番を共通に使用するのではなく、ユーザ（グループ）ごとに枝番を決めることを許すので、異形字の違いの細かさに関する粒度の違いや個人の好みを反映しやすくなる特徴をもつ。

2.3.1 異形字標準化の効果

従来の文字コードに登録された文字だけでなく、異形字の標準化を行うことによって、次のような大きな効果が期待される。

- 不特定多数間で異形字を含む情報交換が可能

- 文字コード規格が早期に安定化
- 文字化け/字体相違の減少

2.3.2 符号枝番方式

異形字を具体的に指定した情報交換が行えるようにする仕組みで、異形字は異形字番号で管理され、コードに付随して使用される。

アーキテクチャは次の構造による。

代表字コード + 異形字機能コード + 異形字番号

従来の漢字コードに続けて異形字機能コードを付加し、更にその後ろに異形字番号を付けて前記コードの文字に対する異形字を具体的に指定して伝送する。異形字を指定しない場合は従来と同じコードのみでよい。異形字番号には異形字を区別する以外のほかの情報は含めない。ただし、異形字番号管理データベースには検索を初めとしてほかの用途に便利となる情報が必要である。

異形字機能コードはISO/IEC 2022系コードとISO/IEC 10646では異なった方法になる。

a) ISO/IEC 2022系コード

エスケープシーケンスにより、ESC + 「異形字を意味する機能コード」で表現される。具体的な機能コードはISO規格に基づいて決める。

b) ISO/IEC 10646

特定コードを異形字機能コードとして使用する。特定の位置だけがグラフィックスでなく異形字ファンクションという扱いとなる。このコードの直前の文字に対し、後続するオクテットが異形字番号を表す。

異形字番号は次の2通りに分けて考えることができる。

a) 漢字コード毎に連続した番号を使用

漢字コードの一つのコードポジションに対応した複数の異形字に対し、それぞれ個別の番号が与えられる。異形字番号だけでは異形字が特定できず漢字コードとの対応で意味をもつ。

b) ISO/IEC 2022系コードとISO/IEC 10646とで共通の異形字番号を使用

ISO/IEC 2022系コードとISO/IEC 10646とで共通にするにはキャラクタ表現は適当でないので、16進2バイトを用いる。

運用上の要件として下記のような項目が考えられる。

a) 異形字は管理台帳を作って登録し、同一文字の異形字に一連の番号を付加する。異形字管理簿には属性情報（部首、字素、読み、画数など）を付加するが、登録管理機関が必要になる。

b) 異形字判別規準の設定を行う。

規準作りが必須であり、字形と図形との区別を行う。あくまで情報交換ツールとしての異形字管理の規準である。

2.3.3 フォント枝番方式

前述の符号枝番方式では、現在、フォントを作成して使用している文字をほとんどカバーする形で多く異形字を集めて、国内や国際で共通して使用できるように枝番を割り当てる。それに比べて、フォント枝番方式は、すべての枝番を共通

に使用するのではなく、ユーザ（グループ）ごとに枝番を決めることを許すので、異形字の違いの細かさに関する粒度の違いや個人の好みを反映しやすくなる。

フォント枝番方式のアーキテクチャとその動作は、符号枝番方式とまったく同じである。文字コードにフォント枝番を与える時のデータの形は、次のようにISO/IEC 2022系とISO/IEC 10646に分けて考えられる。

- ISO/IEC 2022系のような比較的符号表いっぱい文字が割り付けられている場合は、新しいエスケープシーケンスを規定して行うか、枝番バイトを追加し、多バイト文字コードにすることが考えられる。
- ISO/IEC 10646のように符号位置に十分余裕がある場合は、枝番専用（各文字共通）の符号位置をとる方法、8ビットの枝番を指定する使用方法などが考えられる。

上述のように、フォント枝番方式と符号枝番方式の運用方法はほとんど同じで、フォントアクセスの方法も原理的にはフォント枝番方式と符号枝番方式とで差はない。そのため、アーキテクチャという目でみると、フォント枝番方式は、ユーザ定義型の符号枝番方式ということもできる。しかし、その背景にある考え方はまったく異なっている。

また、送受信間のフォント同期の方法も課題なので詳細に検討することが必要である。

第3章 漢字等標準化のための新字種提案制度の創設

本制度は、符号化文字集合に関するJISなどの公的標準へ新規の文字を提案する新しい制度である。これは担当委員会がみずから新しい文字を調査収集する従来の方法を変更し、広く多くの人々が新しい文字を自由に提案できるようにするもので、提案された文字は検討の後にJISなどの公的標準へ採録手続きが進められていくものである。

3.1 背景

JIS X 0213の発行により、JISに採録された文字は1万字を越えた。また国際規格 ISO/IEC 10646もその採録文字は7万字を越えようとしており、これで十分ではないかという意見もみられる。既採録文字の多さ、未採録文字の見出し難さから、新規採録候補文字を収集することが効率的に行えなくなって来ている状況である。にもかかわらず、JISや国際規格に採録されている文字では、漢字を中心にしてまだまだ十分ではないという意見も根強い。

現実には、JIS X 0213発表時点で、その採録文字1万余文字の1%に当たる約100個の漢字が、7万余文字採録されている国際規格ISO/IEC 10646に採録されておらず、国際的に新規採録提案をした事実が、「規格文字はまだ足りない、だがその不足文字を特定するのは困難である」ということを端的に物語っている。

従って、今後ユーザの要求に応えるには、新規採録候補文字の収集を継続し、JIS等の公的標準へ採録する必要がある。しかし、その収集作業はますます困難になり、現在の方法ではユーザのニーズに十分に答えられない状態である。

この問題を解決するには、JIS等の公的標準採録候補文字を収集する作業を従来からの「規格原案作成委員会が行う」などの方法から、下記に提案するようなユーザ中心の効果的な方法に変更する必要があると考える。

3.2 提案概要

本委員会は、JIS等の公的標準採録候補文字の提案制度の創設を提案する。この制度は、提案文字検討委員会（仮称、以後、検討委員会と呼ぶ）を公的機関内に常設委員会として設置し、採録希望候補文字をもつ個人もしくは団体は候補文字を検討委員会に、必要な付帯データとともに提案するものである。

提案された文字は、検討委員会での調査検討ののち、必要事項を満たせば、原則的にJIS等公的標準採録候補文字としてプールされる。プールされた文字は、次のJIS等公的標準作成時に、採録文字候補として扱われる。

検討委員会は審査に必要な文字データベースを作成して公開維持管理を行う。この文字データベースは検討委員会における検討作業、および、提案者が参照などに利用することを目的とするが、汎用の公的データベースとして公開し、文字コード開発以外の広範囲な利用にも供する。

3.3 提案内容

提案内容の詳細は以下の通り。

a) 提案文字検討委員会

- 提案文字を検討し、文字データベースを維持管理するために提案文字検討委員会を設置する。
 - 本検討委員会は、国語や文字に関して権威のある公的組織内に常設する。同時に事務局をこの組織に委託する。
 - 委員は、学識経験者、文字規格関係者、ユーザ団体、関連行政機関などから選任する。
 - 検討委員会は、後述の作業委員会に作業指示を行い、作業結果を審議する。
- b) 作業委員会
- 検討委員会の下に作業委員会を設置し、提案された文字の具体的な検討と文字データベースの維持管理作業を行う。
 - 作業委員会は、検討委員会委員の一部と実務担当で構成する。
 - 作業委員会は、あらかじめ検討委員会により決められた検討条件で、提案された文字の文字データベースへの登録や、符号化文字としての適不適などを検討する。
 - 検討条件は公開し、必要に応じて見直しを行う。
- c) 文字データベース
- 文字データベースは、JISへの採録済み文字、国際規格採録済み文字、従来のJIS等開発時に提案・検討されたが採録されなかった文字、本検討委員会に提案された文字などを網羅し、常時最新のものに維持する。
 - 文字データベースには、部首・画数・読み・各種規格符号位置などのほか、典拠・用例、あるいは採録不採録の理由などのデータをできるだけ収集する。
 - 文字データベースには、字体判断の基礎となる、統一されたデザイン方針による例示字形を用いる。各種のデザインの混用は、この文字データベースの性格から望ましくない。字体の揺らぎの影響を受けにくく、また、字体の揺らぎを明示しやすいものが望ましい。
 - 文字データベースは、検討委員会や作業委員会の作業用だけでなく、新規文字提案者による既存・提案済み文字の確認作業にも使用できるように、広く一般に公開する。
 - 文字データベースは各種の文字研究などの基礎データとして広く利用されることを想定することが望ましく、可能ならば、公的基本データの性格を持つことが期待される。
 - この文字データベースで使用される例示字形のための文字字形データは、広く一般に公開され、巨大文字集合印刷のための最低限の社会資産として自由使用を許されることが望ましい。このために、文字字形データに対する知的財産権の対応にも留意する必要がある。
- d) 文字提案方法
- 提案者：候補文字を提案するのは個人または団体とするが、それぞれの応用分野などを代表する団体などが望ましい。
 - 提案時期：提案は常時受理され、特定の受理期間は設定しない。
 - 作業委員会は、あらかじめ設けられた検討条件により提案文字を一定期間ごとに検討し、検討委員会がその結果を審議する。

- 検討委員会は承認した提案文字をJIS等の公的標準候補文字としてプールする。
- 実際の公的標準へ向けての作業は、別途設けられるJIS等の公的標準作成組織が行う。

e) 費用

- この制度は公的資金を中心とした運営が望ましい。
- 特に、文字データベースの構築、フォントの作成には、多額の初期投資が必要になると想定される。

第4章 異形字アーキテクチャ

4.1 異形字のアーキテクチャのまとめ

本章では、異形字WGに寄せられた異形字アーキテクチャの提案をまとめる。

一般に文字コードに漢字が不足だという声とともに提案されるのは、いわゆる内字の領域の拡張提案であるが、現在の文字コードは内字領域を十分確保できるので、今回は内字のための領域拡張方法よりは、異形字の「要求の特徴にあせた対応するには、どう拡張すべきであるか」という視点で各提案を検討する必要がある。

4.1.1 提案解説

「空き領域が十分あるならば、それに必要な文字を追加していけばよい」という意見がある（従来延長型）。つまり、枠組みは従来そのまま、必要な文字を内字領域に追加するものである。それを補う形で、図形交換を積極利用する案もある。

これに対して「そもそも追加要求は、新字種と異形字に分けられるので、それぞれに別の方法を考えるほうがよい」という声もある。本委員会第2ステージの狙いは、単純な従来延長型でない対応の可能性を探ることである。本委員会では、この狙いに基づき、「新字種は従来延長型で対応するとしても、異形字には新字種の場合と異なる性格があるので、異形字の性格により適した対応方法を考えてみよう」という前提のもとに、異形字アーキテクチャWGを設置した。

WGに寄せられたアーキテクチャの提案は、表面的には相当に異なっている内容も必ずしも排他ではなかった。さらに、従来のアーキテクチャとも整合的であった。そこで、これらを、従来のアーキテクチャ（現行の文字コードに基づく従来の対応方法）との相互の関係を含めて整理することができた。これを図1に示す。

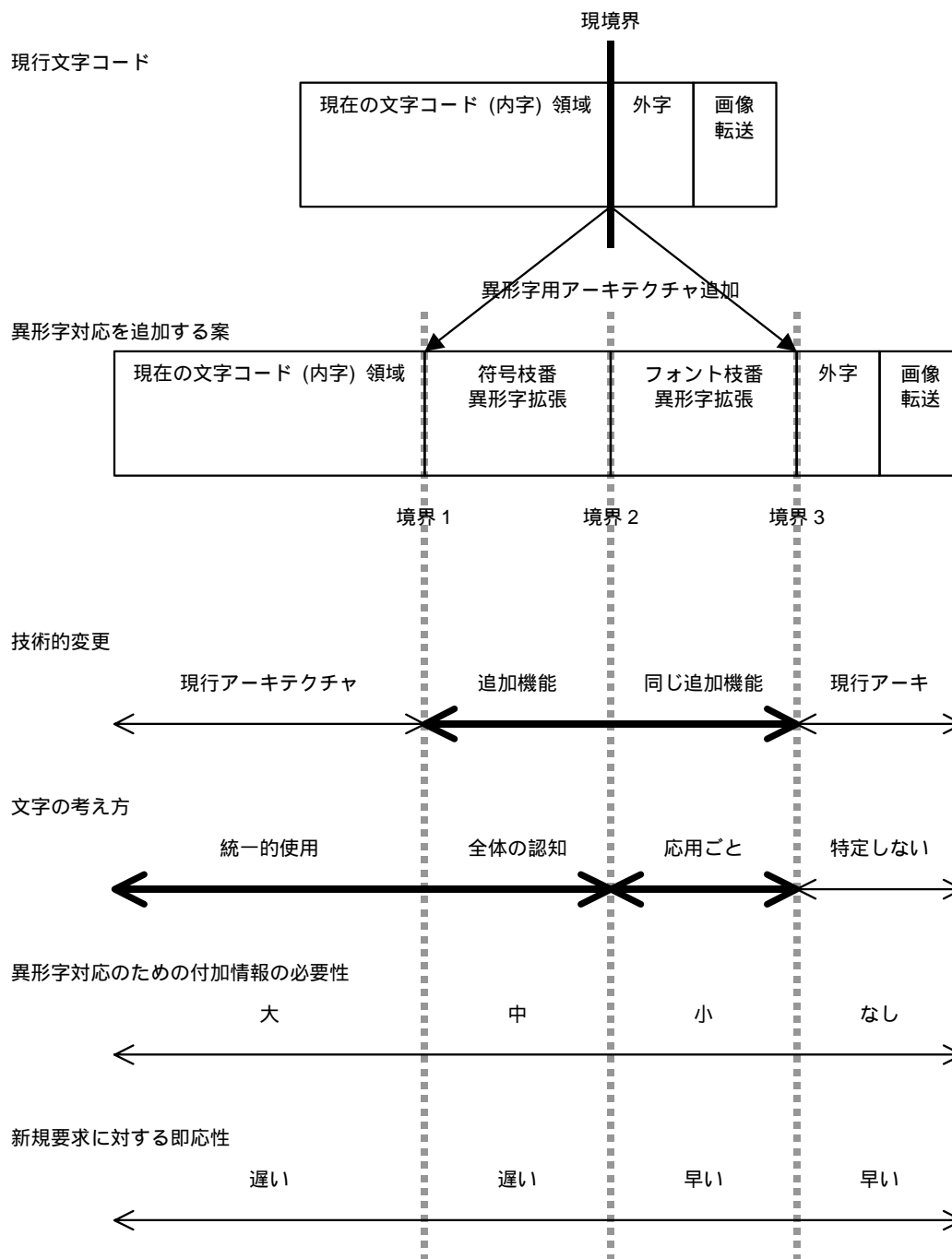


図1 異形字アーキテクチャのまとめ

異形字アーキテクチャWGに寄せられた提案、情報の内から、符号枝番異形字拡張とフォント枝番異形字拡張という代表的な二つの考え方を取り上げ、従来の文字コード規格との関連を述べる。

符号枝番異形字拡張アーキテクチャの動作は、別途詳細が説明されており、フォント枝番異形字拡張も動作は同じである。

内字部分と外字・画像転送部分に関するアーキテクチャは従来と同じで、異形字拡張部分のための拡張アーキテクチャは、符号枝番とフォント枝番について技術的にはまったく同じもので実現可能である。

具体的な提案としては、符号枝番方式は(広い範囲での社会的合意を前提とした固定的・共通的な異形字をサポートすることを前提としているので)正規の規格に

よる異形字内字、フォント枝番部分は（それぞれの用途や、使用者グループ内だけの利用を前提としているので）正規の規格の範囲外である異形字外字のように見える。

逆に、文字を選択するという観点からは、社会的合意（全ての使用者の間で同じ認識を共有している）による独立した文字の識別が必要であるという点で、従来の公的規格（内字）と符号枝番は共通である。すなわち、符号枝番を利用するためには、今までの符号化作業と同様に、異形字を認知し、公認し、規格化するというプロセスが必須である。（異形字に対しても、社会的な共通認識が成立するという前提がある）。

また、国内規格の国際規格との整合性という観点から、国際的な合意を取り付けることも必要となろう。

一方、本報告書で説明されているフォント枝番方式では、ある種のグループ内での合意さえあればよいという前提に立っている。（すなわち、異形字の要求にはいろいろな考え方があり、共通認識を形成することは現実的でないという前提に対応している）。また、異形字の必要性や、その包摂関係の認識が、個人やグループあるいは応用毎、さらには書体間でも異なるという認識がある。

このように、符号枝番とフォント枝番とは、同一の技術的アーキテクチャによる形の上では、異形字内字・外字のように見えるが、異形字の要求の認識がまったく異なる。

異形字アーキテクチャWGでは具体的な提案はなかったが、内字に文字を追加し続ける従来延長型で、符号の間の異形関係を、別途データテーブルを使用して対応する案もある。この案も含め、データテーブルの規模は、内字だけの場合が一番大きく、符号枝番では枝番がその情報を一部肩代わりし、フォント枝番ではかなり多くの情報を（応用別情報などまで含んで）枝番そのものが含んでいる。特に、内字だけの場合、応用別情報を個々にデータテーブルに入れる必要がある。このように、補助情報をどのような形でどこに持つかが、それぞれの方式で大きく異なっている。

調査などの結果、異形字の要求が文字コードの開発時点で全て把握されるならばどのような方式を取ろうとも問題は生じない。しかし、仮に異形字の要求が随時現れるとすれば、対応に必要な時間が問題になる。対応時間については、外字や画像転送による対応が一番即応性があり、フォント枝番、符号枝番、の順で対応に要する時間が長くなり、従来延長方式が一番長いと想定される。

どこまで符号化するか、どこは符号化しないかの境界については、現在の方法の「現境界」は、かなり厳密な設定が必要であり「なぜ、この文字は符号化し、これはしないか？」などのコメント、あるいは包摂規準などの粒度の定義に対する異論が（たとえ、内字拡張を続けたとしても）絶えないことが予想される。これに対して、異形字アーキテクチャを導入すると、境界1、境界2、境界3に対する厳密さの要求は軽減され、また、特に境界2は、粒度の認識の差をかなり吸収できることが期待できる。

注) 上記の図表は、いままで議論されたさまざまな異形字のニーズと特徴を、全て同じ重みで対応しようとするとの前提で、作成してある。したがって、このまま採用したとすれば、結果的に「やりすぎ」の重いシステムになる可能性も否定できない。

4.1.2 境界

以下に、上記の図表に基づき、各境界についてやや具体的に検討する。

境界1は、従来の公的符号化文字集合規格と、それ以外のさまざまな文字情報交換方式との境界である。従来の公的符号化文字集合の内字を拡張し、現境界と境界1を拡張する（右へ動かそうとする）ためには、従来は異形字として包摂されていたもののなかから、採録すべき文字が事前に特定でき、その包摂関係（粒度）も共通で、その数はあまり多くないという前提が必要である（管理上の問題が大きい）。特に、ISO/IEC 2022系の符号は、空き符号位置が少ないので、現境界や境界1を内字拡大方向（右方向）に動かそうとしても、動く余地は少なく、現実的には、巨大スペースのある符号系だけに適用可能である。ISO/IEC 2022系において実現することにこだわれば、内字拡張方法には自ずから限界がある（符号枝番には、この空き地問題への回答という面もある）。なお、内字を狭くするというのは、現符号の変更なので、検討しない。

また、現国際規格では、漢字統合規則との関係で、異形字の追加には、かなり抵抗があることが予想される。

粒度などに関する国際合意も課題になる。

境界2は、アーキテクチャ的には共通である符号枝番方式とフォント枝番方式との境界である。これについては、もし、多くの異形字が事前に特定でき、かつそれぞれ独立し（包摂の重なりがなく）、粒度にもコンセンサスが得られると仮定すると、符号枝番方式が適用できる範囲が広がり、境界線を右方向（拡大方向）に移動させる力が働く。異体字の多くが、事前には特定できず、かつ、包摂の範囲や粒度の合意形成も困難と判断されると、境界2を左方向（フォント枝番へ拡張）に動かし、符号枝番を抑制する方向に働く。

事例1: 異形字の塊のような人名用漢字について考えると、もし、全ての関連する役所の間人名漢字についての合意が得られるとすれば、境界2は右へ動き、合意が形成できない場合は、符号枝番の領域は限定的になり、フォント枝番を拡大しようと左へ動かそうとする。

事例2: 人名漢字について、仮に、いわゆる誤字俗字は使うべきでないと考え抑制すると、人名用異形字は固定的になる。しかし、実際に使用されているのだから誤字俗字であっても符号化するという結論になると、応用分野ごとに個別の対処をすればよいとか粒度の問題であるとかの議論になりやすい。

境界3は、枝番方式によって異形字を特定する方式と、従来の外字や画像転送による送り手と受けての間だけの合意のみを前提とする方式との境界である。

符号枝番方式で、ほとんどの異形字要求に対応でき、残りは非常に特殊と考えれば、境界2と境界3は一致しようとする。また、フォント枝番がある程度進むと、境界3は右に動こうとする。

しかし、時間的な即応力を持つ外字・図形転送の需要は、軽視できないので、結局これらは残ると思われるが、この需要を満たすためには、もっと、誰にも容易に使えるようにする必要がある。

とは言え、送受信者相互や、公的に合意された図形を送受信双方で保持しておき、さらなる画像指定をするような画像転送や外字方式は、すでに特定されたものは符号化できるという観点を勘案すると、採用すべきではないと思われる。外字や画像転送の利用用途は限定的なものとなる。

事例3: 国語審議会答申がかなりの抑制力となることを前提とすると、異形字を多用する応用は限定的になり予測可能となる可能性がある。

事例4: 学術や古典など、変動的な要因の高い、異形字源が、一般応用から隔離された独自の世界を作るか、逆に汎用性を要求するかの判断で、突発外字や、図形転送のニーズは大幅に異なり、境界3に大きな影響を与える。

上記のような事例を勘案すると、これらのさまざまな方式を適用するには、次の段階において、どの方式をどこまで採用するか（しないかの）という判断が必要となろう。

4.1.3 異形字に関する要求の合意の必要性

前述のように、今回検討したアーキテクチャ案は、いままでいろいろ述べられた異形字の性格すべてに可能な限り対応する、性格の軽重を考えないという、“なんでもあり”の案であり、思考のスタートポイントというべきものである。この案をこのまま実現することは現実的ではない。

そこで、上述のアーキテクチャを念頭に置きながら、もう一度、異形字の要求を、現時点でどう見るか？ 将来は、どうなるだろうか？ というところで最後の詰めが必要であろう。

実際には、今の時点と将来の異形字の要求を、われわれがどう考えるか、それはなぜか、というコンセンサスが必要である。このコンセンサスは、まさしく意思決定であり、事前にかなり広範囲な議論が必要となろう。しかし、今後の議論においては、今までの議論を繰り返すのではなく、本委員会での到達点を前提とした上で、上記のアーキテクチャ案をどうまとめていくかという目的を明確にした検討が必要であると思われる。

この議論の中には、以下のような点が含まれる必要がある。

- ISO/IEC 2022系のJISは、今後も必要か？ これによって、現状拡張が可能かどうか決まる。内字領域がもうほとんどとれないのに、内字拡張を考えても意味がない。
- 内字拡張に、国際的合意がとれるか？
- 異形字の要求（特に粒度）は、各国同じか？ 同じでないとするれば、国別要求にどう応えるか？
- 特殊な古典・仏典の要求を一般用と同じに考えるか？

この、次の段階の議論とコンセンサスの必要性が、異形字アーキテクチャWGからの最重要なメッセージである。

4.1.4 その他の条件

フォント枝番方式では、情報の送信側が指定したフォントを、受信側が使うという条件を満足する手段を提供する必要がある。この要求は、文字コードが巨大になれば、普通の内字でも必要になるし、符号枝番方式でも結局必要になるので、今後重要な技術課題となる。そのため、どれかの方法の特有の要求ではなく、いずれにしても挑戦しなければならない課題なので、今回は議論しなかった。

文字データ列は、出来るだけ一様である方が好ましいので、多様な様式の情報データ列に入るようなことは避けたい。

寄せられた提案の中で、異形字相互間情報の整理に関するものは、異形字情報のための付加情報を定義する場合に参考になる。

寄せられた情報の中で、異形字の定義法に関するものは、異形字の分析や、特定の方法として参考にしたい。

4.1.5 本委員会外での動向

本委員会において異形字問題として取り上げた問題については、本委員会の外でも、いくつかの動きが見られる。以下に、JISテクニカルレポートとUnicode Technical Committeeにおける議論の二つを取り上げて概説する。

(1) TR X 0047:2001 XMLによる画像参照交換方式

このTRは、XMLを用いて共有される画像データへの参照情報を交換する言語を規定したものである。この方式を用いて、画像データとして漢字の図形表現を直接交換することも可能である。しかし、先にも述べたように、公的に合意された図形を送受信双方で保持しておき、さらなる画像指定をするような画像転送や外字方式は、すでに特定されたものは符号化できるという観点を勘案すると、採用すべきではないと思われる。このTRで規定している方式を用いるに当たっては、送り手側と受け手側で共用する画像データ情報をどの範囲で、どのように規定するか、という運用上の問題を慎重に検討する必要がある。

TRの詳細については、下記のURLをご参照いただきたい。

http://www.y-adagio.com/public/standards/tr_lsi_xml/toc.htm

(2) UnicodeのVariations Selectorについての議論

Unicode Technical Committeeは、Unicode 3.2において、合計256個のVariation Selectorを規格化している。

(http://www.unicode.org/unicode/reports/tr28/#13_7_variation_selectorsを参照。)

これらのVariation Selectorsは、当初、古モンゴル文字において同一文書内に出現する意味的には同一でグリフの異なる文字を表現するために導入された。その後、数学用記号における、意味的には同一で、学会等の表記規定などによる字形の相違を吸収するためにも利用されるようになっていく。

また、UTCに対しては、1998年に、Sun Microsystems、Justsystem、Apple Computerから、“Plane 14 Variant tag”というプロポーザルが出されている。これは、元来は、漢字の異体字処理を前提としたものであったが、その後は、さまざまな事情により、UTCにおける議論の進展はない。

しかし、2001年12月のIRG東京会議において、Unicode Consortiumより、現状でのVariation Selectorsの説明が行われ、IRGにおいて、漢字にVariation Selectorsを漢字に適用可能か、可能であるとすると具体的にはどのようなものになるかの検討を開始して欲しい、との要請がなされた。

UTCは、以前から漢字の追加が際限なく行われることへの危機感を持っており、この対案の一つとしてVariation Selectorsを用いることへの期待感は大い。

今後、符号枝番方式の適用を検討する上で、Variation Selectorsを巡るUTCの動向には注視が必要である。

4.2 符号枝番方式

4.2.1 概要

情報交換において異形字を具体的に指定する仕組みである。現在の漢字コード規格は包摂規準によって一つのコードに複数の字形が存在するが、それらの字形を具体的に指定する方法がない。また人名用には包摂範囲を超えてそれ以上に異形字を区別して使いたいというニーズがある。本案は、これらのニーズに応え、コードによる文字指定を拡張して、字形レベルまで指定できるようにする方法を提供するものである。

異形字は異形字番号で管理され、コードに付随して使用される。図形とみなされるもの（例：印影、百寿/百福など掛け軸のデザイン文字等）は指定の対象外である。

4.2.2 符号枝番方式の構造

次の構造による。

代表字（親字）コード + 異形字機能コード + 異形字番号

従来の漢字コードに続けて異形字機能コードと異形字番号を付加して前記コードの文字に対する異形字を具体的に指定する。ここでいう代表字（親字）とは、学術的な代表字形ということではなくそのコードポジションにおける文字のことである。

異形字を指定しない場合における情報交換では、従来と同じ漢字コードのみによる。

情報交換においては、異形字番号以外に他の情報は付けない。また異形字番号にも異形字を区別する以外の他の情報は含めない。ただし、異形字番号管理データベースには管理および検索などのため他の情報も必要となろう。

4.2.3 異形字機能コード

ISO/IEC 2022系コードとISO/IEC 10646では異なった方法になる。

a) ISO/IEC 2022系コード

- ISO/IEC 2022に基づくエスケープシーケンスによる：
ESC+異形字を意味する機能コード。
- 具体的な機能コードはISO規格に基づいて決める必要があり、別途専門家による検討を経てISOへの登録が必要である。
例えば現在使用していない(?)と思われる2/6、2/7など。場合によっては2バイトになる。

b) ISO/IEC 10646

- 符号表上の特定位置コードを異形字機能コードとして使用
BMP（基本多言語面）における特定の位置だけがグラフィックスでなく異形字ファンクションという扱い。例えばR領域の一つのコードポジション
このコードの直前の文字に対し、後続するオクテットが異形字番号を表す。

- 機能コード位置はISOおよびUTCで具体的に決めてもらう必要がある。(UTF-16で拡張される第14面における異形字タグとの関連の明確化を含む。)

4.2.4 異形字番号

異形字番号は、以下のように決める。

- 異形字番号は、漢字コード毎に連続した番号とする。
 - 漢字コードの一つのコードポジションに対応した複数の異形字に対し、それぞれ個別の番号が与えられる。
 - 異形字番号だけでは異形字が特定できず、漢字コードとの対応で意味をもつ。異形字番号が相違する同一字形の存在がありうるし、あるコードの異形字の一つが別のコードの代表字形と同じということもありうる。
- ISO/IEC 2022系コードとISO/IEC 10646とで共通の異形字番号を使用する。
 - ISO/IEC 2022系コードとISO/IEC 10646とで共通にするにはキャラクタ表現は適当でないので、16進2バイトを用いる。
 - 1バイトコードの制御符号部分および10646のサロゲートペアと一致する部分を避ける。
例えば2121～277Eの範囲を使う(564字分で不足なら範囲の拡張は可能)。

戸籍簿上の異体字を対象とするレベルなら2桁の数で足りると考えられるが、一応余裕をみて数百字表現可能とする。

4.2.5 異形字指定の具体例

管理台帳番号Xページ

管理上の代表字： 辺

台帳上の異形字番号： [2121] [2122] [2123] [2124] [2125] ...
 辺 邊 邊 邊

異形字指定 (X 0208での例)

辺： [4A55] ESC Fi [2124]
 または [6E34] ESC Fi [2124]
 または [6E35] ESC Fi [2124]

Fi：異形字機能コード

[XXXX] は 16進コード

[4A55] は 辺、[6E34] は 邊、[6E35] は 邊

4.2.6 符号枝番方式の効果

- 不特定多数間で異形字を含む情報交換が可能となる。
 現在、コード規格で包摂されている異形字や、包摂されていないそれ以外の異形字を使うためには、外字処理によるかあるいは異形字のイメージデータを貼り付けて送る以外に適当な手段がない。しかし外字を含む情報は相手に正しく伝えられず、情報交換には不向きである。またイメージデー

タの転送では、検索に不便であり、文字の大きさを揃えるための処理が必要になる場合があるなど、扱いも煩雑である。

異形字に枝番を付して標準化することで、コード規格では包摂されている異形字を具体的に指定して情報交換することができ、また包摂されていない異形字も新たに異形字として登録することにより、不特定多数間での異形字の情報交換が可能となる。

- b) 現行コード方式で文字を追加するより早期実用化が可能。

異形字は、それが出現した時点で随時登録する方式がとれるので、異形字を交換する具体的な方式が決まれば技術的には比較的早期に実用化が可能であり、コード規格へ追加されるのを待つより確実に早く使える。(運用面での検討が今後必要であり、それについては後で述べる。)

- c) 相手に対応異形字フォントが実装されていなくても意図は伝わる。

符号枝番方式による情報交換では、受信側に当該異形字フォントが実装されていない場合は当然その異形字は表示されないが、漢字コードによって代表字は明確であり、指定された異形字を知りたい場合は、別途管理データベースに照会することによりどのような異形字かを知ることができる。(運用上そのようなようになることが期待される。)

また、このような方式が確立すれば、異形字フォントオンライン販売業者等の出現も期待でき、現在の外字処理による方法より遙かに簡便に多くの異形字を扱える環境の出現が期待できる。

- d) 文字コード規格が早期に安定化。

現在のコード方式は青天井問題を抱えている。符号枝番方式の普及で異形字の標準化が進めば、文字コード規格に登録する文字は未登録の新規文字に限られるため、規格化の改正期間が短縮でき、規格として早期に安定化できる。

- e) 文字化けおよび字体相違の減少。

文字コード規格は包摂規準の範囲内で字形にゆれがあり、フォントベンダ毎に字体が相違しても不思議ではない(デザイン差は当然であるが、ここでいう字体とは、デザイン相違を含まないいわゆる字体のこと)。したがって、同じ文字がフォントの違いにより文字化けまたは字体相違になる場合がある。(かつてベンダ間の字体相違が大問題だった時期があったが、平成書体の字体をフォント設計の規範とする暗黙の了承ができて、現在ではベンダ相違による字体問題は少なくなっている。)

情報交換では送り手からの情報は受け手に対して変化せずに伝送されることが重要であるが、現行のコード規格の下では包摂の揺れの範囲で変化することは防ぎきれない。

符号枝番方式が普及すれば文字化けは極小化でき、情報交換の際の字形のゆれも防げる。もちろん現行機と新規格適用機が市場に併存している限り文字化けは完全には防げないが、少なくとも現在より相当減少することは間違いない。

- f) 一般の字形に対する意識の向上で問題発生が減少。

更に、異形字の標準化により、副次効果として一般の字形/字体に対する意識が向上すると思われ、文字化けや字体相違等に関しては実用上問題にならないくらい改善することが期待できる。

4.3 フォント枝番方式について

4.3.1 背景

文字コード標準体系検討専門委員会第1ステージ報告書では、2.4.7において、異形字問題への対応として、二つの異なる方向性を持った解決方法についての提案を行った。このうち、2.4.7(1)「線引きが可能と合意された場合」については、本報告書においても、いくつかのアーキテクチャの具体的提案がなされている。しかし、2.4.7(2)「線引きが困難と判断された場合」については、特に具体的な提案がなされなかった。

ここで、議論の枠組みを再度確認するため、2.4.7(2)において、示されたフォントレベルでの枝番対応（以下フォント枝番方式）について、やや詳細に述べる。第1ステージ報告書から引用：

2.4.7 提案

(2) 「線引き」が困難と判断された場合

- (a) フォントの中の必要な「字体」対応位置に、「枝番」をもうけ、基本「字形」だけではなく、異なる「字形」を直接指示できる様なアーキテクチャを考える。
- (b) この枝番は、フォントに共通である必要は必ずしも無い。
- (c) これでも「線引き」は必要であるが、精度の要求が格段に甘くなる。
- (d) 「枝番」指示の具体的方法は、符号に付加する別途の符号でも制御シーケンスでも良い。この方法、線引きなどはステージ2以降で検討する。
- (e) 枝番部分のフォントの供給方法は、システムの今後の発展を考慮し、また、市場への参入障壁を下げることを念頭において、これもステージ2以降に検討する。例えば漢字全体のフォントの提供は大変な投資を伴うので、特定の字形1個でも供給可能なような方法やしくみを検討する必要がある。

4.3.2 問題意識

精緻な包摂基準を定めて、文字の典拠や用例を調べても、それからもれる異形字の要求があることにはかわりがない。

また、“何でも符号を与える方式”や“コードに枝番などを与える方式”は、そのやりかたのいかんにかかわらず、要は（形はどうあれ）符号を与えて、全員共通に使用するというにはかわりなく、“認知された異形字を含む文字集合”ができるだけで、漢字問題の基本の“あれが認知されたなら、これも認知すべきだ”という個人の好悪価値観のレベルに至る議論には、応えていない。単に“線引き位置”が変わっただけで、新たな“線引き問題”が発生するだけではないか？

特定の異形字の必要性は、最終的には、使用者の主観的な問題が中心になり、結局、必要な異形字集合の一般的コンセンサスを得ることは、本字集合のコンセンサス作りより困難ではないか？

さらに、その異形字の包摂の範囲は、さらに主観的で、それぞれの意見により重なり合い、それぞれ、独立した固有の包摂の範囲をもった異形字集合のコンセ

ンサスを作ること無理ではないか? となると、登録制度ですら、実現困難にならないか?

むろん、あるレベルまでの異形字集合のコンセンサスを得ることは出来、それをもとに、符号枝番を考えることは可能であるが、それで異形字の要求が終わることは無いだろう。

もし、詳細に規定した異形字コードを採用したとして、フォントデザイン上、それをそれぞれ区別して、全てのデザインに適用することは、きわめて困難であると同時に、特定のフォントが無いことを容認したとしても（文字コードはデザインを特定しないため）運用上の混乱をまねく。

逆に、伝達される情報の内容が大幅に変わるわけではないから、異形字は特定の条件を満たしたときに正確に意図通りに伝われば十分ではないのか?

実際には、特定の応用の枠組み内（例えば、戸籍情報システム）を前提とした正確な情報交換や、印刷物では活字を指定して印刷するということが行われているのではないのか?

デザインごとに、異形字が決まる可能性が高い

4.3.3 フォントによる字体の相違

明朝体	楷書体	ゴシック体	明朝体	楷書体	ゴシック体
絢	絢	絢	葵	葵	葵
綾	綾	綾	茜	茜	茜
維	維	維	飯	飯	飯
緯	緯	緯	饨	饨	饨

図2 フォントによる字体の相違例

上記の例でもわかるように、書体によって伝統的に部首の作り方が異なる場合がある。外字処理においては、名刺等で多用される楷書体において、明朝体のような系偏を要求される例、明朝体に4画の草冠を要求される例が多々見られる。

4.3.4 第一ステージにおけるフォント枝番方式の提案

このような視点から、第一ステージのフォント枝番方式では、以下の提案をしている。

- a) いっそ異形字に文字コードを与えることをあきらめて、文字コードに同一文字コードに対応するフォント枝番を追加したデータを交換する。

- b) フォントは、同一コードの枝番に異形字をもっている
- c) この枝番異形字は、フォントごとに異なっても良い。必要な場合の情報交換はフォントを指定することで行われる。
- d) フォントの指定が無い場合や、フォントが無い場合の対応は別途考えるが、基本的には、枝番情報が無視され、本字が使用される
- e) 理論的には、コンセンサスを得られる部分を符号枝番、それ以外をフォント枝番とすることも考えられるが、処理とデータ形式が煩雑になるので、統一してある。

A 社銀行協会用ゴシック体フォント

符号	代表字	枝番 1	枝番 2	枝番 3	枝番 4
1111	辺	辺	邊	邊	・	・	・	・
2222	崎	埼	崎	碕	寄	・	・	・
3333	小	・	・	・	・	・	・	・
4444	齊	齋	齋	齊	・	・	・	・
⋮	・	・	・	・	・	・	・	・
⋮	・	・	・	・	・	・	・	・

B 社文芸作品用明朝体フォント

符号	代表字	枝番 1	枝番 2	枝番 3	枝番 4
1111	邊	邊	邊	・	・	・	・	・
2222	崎	埼	崎	碕	寄	・	・	・
3333	小	・	・	・	・	・	・	・
4444	齊	齊	・	・	・	・	・	・
5555	齋	齋	・	・	・	・	・	・
⋮	・	・	・	・	・	・	・	・

図3 フォント枝番方式の例

4.3.5 評価

この方式は、

- a) データの形式は、いわゆる符号枝番と同じであり、特にフォント枝番だからという特殊な形式は不要、
 - b) フォントごとに、自由に異形字を選定追加可能、などの利点があるが、
 - c) データ処理上は、必要があればフォント名フラグが必要、フォントごとの同定が必要なので、データベースの相互運用が困難、の欠点がある。
- 希望的に考えると、
- a) 応用別にその応用専用フォントの開発が進む、

- b) 業界標準として、共通部分が作られる、
 - c) 結局、どこかに落ち着く、
- などが考えられる。

最大の利点は、本字を符号化すれば、その異形字処理は文字コード開発の手を離れることと、ユーザやフォントメカが自由に出来ることであり、共通利用の異体字を含む文字セットを、その利用者グループの自己責任で、即応体制が取れることである。

これに、よって、全国区レベルのコンセンサスと周知徹底のための時間遅れ無駄を排除することになる。

なお、文字の属性データは別途持つ必要がある。(これは全面符号化方式と同じで、自由度が高い)。

逆に、符号化枝番など、文字の属性情報の一部をコードに取り込む方法は、狙い通りの使用法の時は便利で有効であるが、汎用性と柔軟性に欠けるので多様な異形字の要求には対応が難しいと考える。

この提案は、多様な異形字の要求に対して、精緻かつ共通の分析的対応は困難だという判断のもとに、符号化文字集合としての対応を放棄し、“勝手にどうぞ”の範囲を提供したものであると言えよう。

4.4 構造化4バイトコード方式

4.4.1 構造化4バイトコードによる方法

情報交換用漢字符号JIS X 0208は、多言語機能、文字種の拡張機能、規格の安定性に問題がある。また、JIS X 0208は、漢字配列や検字手段として重要な、漢字の読み・部首・総画数などの属性情報や、代表形（親字）と異形字の関係が明確に規定されていない。そのほか、JIS X 0208では、漢字符号に与えている機能が文字の識別に限定されている。これらの問題を解決するためには、漢字符号に実装する機能や情報の複合化と、情報を任意に追加できる開構造形式の符号を作ることが重要である。4バイトコードの構造化は、以下に示した事柄を実現するために導入した。

4バイトコードの構造化によって実現できる機能と情報

- a) 約83万字の見出し漢字（代表形）と異形字、漢字の読み・部首・画数・言語名に対する符号化規準の統一。
- b) 利用者が文字集合を任意に選択し、二次符号を与える機能。
- c) 利用者規定の文字集合間で相互に情報交換を行う機能。
- d) 中国・韓国漢字など多言化機能。
- e) 任意の異形字を使った代表形と包摂対象の規定。
- f) 符号間への漢字の追加機能と、規範の維持。
- g) 既存の漢字符号との併用機能。
- h) データベースを使った漢字情報の一括管理。
- i) 4バイトコードと実装した情報のデータベースの編成方式と内部符号の変更からの独立機能。

4バイトコードを3バイトと1バイトの二つの要素に分け、3バイト部分に2バイト8836字の領域を94個埋め込む。4バイトコードの二つの要素に基本機能や情報を対応させたものを論理符号とする。4バイトコードの3バイト部分で約83万字の見出

し漢字を規定し、1バイト部分は枝番号とし、異形字など追加される可能性の高い異形字を置く。

論理符号は、論理符号として下位構造に置いた2バイトコードに対する問題点を解決するための機能を与える。拡張機能は、4バイトコードの3バイト部分と、2バイトコードとの関係を使って規定する。最後に、論理符号は、データベースの編成方式や内部符号の変更から独立させるため、二つの構造の下位に内部符号を置いた3層構造を作る。図4は、4バイトコードに設定した要素と構造および構造間の関係を示した3層構造のモデルである。

以下に、4バイトコードを構造化することによって導出できる機能や情報を示した。

- a) AとBの関係から導出できる機能
 - 2項関係を使った異形字、各言語対応漢字、属性情報の統一規定。
 - 大規模な文字集合と符号間への文字の追加機能の併存。
 - 外部情報に対する検索キー機能。
- b) AとDの関係から導出できる機能
 - 規範としての大規模な文字集合と目的や専門分野別文字集合の規定。
 - クライアントサーバシステムでの情報運管理の実現。
 - 論理符号の内部符号からの独立。
- c) AとBとDの関係から導出できる機能
 - 任意の異形字に対する代表形と包摂関係の規定。
 - 複数の繁体字に対する簡体字を代表形とする関係の規定。
 - 日本と中国漢字の対応関係の規定。

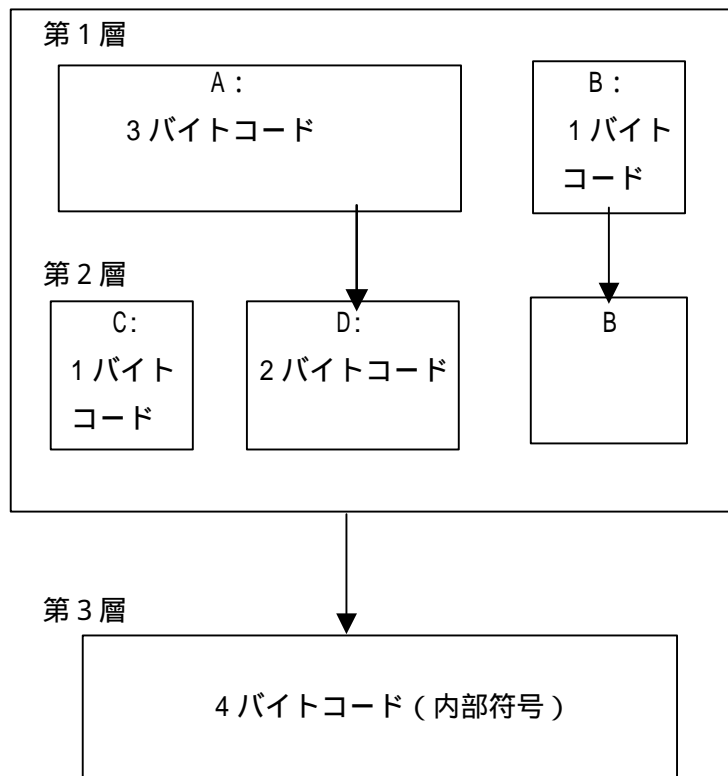


図4 3層構造で表した4 バイトコードのモデル

4.4.2 異形字の拡張方法と任意の異形字を代表形とする包摂関係の規定

用語や用字の総索引の作成では、システムに入力する漢字が実装されていない場合、類似した字形で代用することがある。また、入力できない漢字のほかの漢字字形で代用する場合には、研究者の主観が入ることが多い。このようにして作られた総索引は、データで同じ代表形が使われている場合も単純に比較・照合することができない。

二つの索引で代表形と包摂関係が異なる事例	代表形と包摂対象になった異形字
代表形が同じで、包摂対象が異なる場合	剣 * 劍劔劔 劔 * 劔劔
代表形は異なるが、包摂対象が同じ場合	剣 * 劔劔 劔 * 劔劔

この問題を解決するためには、利用者が総索引で使う代表形と包摂関係を任意に指定できること、索引利用者に対しては包摂関係が明確にされていることが重要になる。そのほか、中国語処理で複数の繁体字を一つの簡体字に変換し、再変換する処理や、日本と中国の漢字を対応させる場合にも、変換前・後の字形の関係が記録されていることが必要である。以下、4バイトコードの枝番号1バイトで規定する94個の異形字を8836字に拡張する方法と、8836の異形字から任意に代表形と包摂対象を指定し、情報交換を行う方法を述べる。

4.4.3 4バイトコードの枝番号で規定した94個の異形字を8836字に拡張する方法

異形字は、出現頻度の高いものを4バイトコードの枝番号1バイトで94字を規定し、出現頻度の低いものを2バイト8836字部分に置く。この二つの文字集合を「A」、

「AB」とする。異形字「AB」の先頭位置にある異形字を副見出しとし、「A」で規定した異形字と同じ字形を置く（副見出し：類似した異形字を細分類したもの）。1バイト[A]と2バイト[AB]の「A」は同じバイトとする。

異形字「A」と「AB」に対する代表形は、4バイトコードの「XXX.A」の3バイト部分「XXX」で規定する。「AB」には、4バイトコードの3バイト部分で規定した代表形や、中国（簡体字・繁体字）・韓国語の類似字形も登録されているものとする。「AB」で規定した異形字は、4バイトコードと独立した面で管理する。データで「AB」を引用する場合は、4バイトコード、JIS X 0208、シフトJISコード、拡張UNIXコードとの混在を考慮し、漢字の読み、部首、画数などの属性情報と同様に扱う。以下は、「A」と「AB」の関係と、4バイトコード対応の漢字を代表形に使った包摂関係のモデルである。

代表形と包摂対象の変換表

XXX.A	<AB>.<CC>
XXX	4バイトコードの3バイト部分で規定した見出し漢字。
A	4バイトコードの枝番号で規定する94個の異形字。2バイトコードで規定する異形字に対する副代表形。
AB	利用者が代表形として指定した2バイトコード対応の異形字。
CC	包摂対象になる2バイトコード対応の異形字。
DD	利用者規定の2バイトコード、拡張UNIXコード、シフトJISコード。

4.4.4 任意の異形字に対する代表形と包摂関係の規定方法

代表形と包摂対象を符号化する場合は、8836字の異形字から代表形<AB>と包摂対象<CC>を指定する。異なる言語で使われている漢字との対応は、<CC>を<AB>に変換する。見出し漢字と異形字の包摂関係の対応は、4バイトコード「XXX.A」と「AB.CC:」を結合し、変換表「XXX.A.AB.CC」を作る。次に、見出し漢字・異形字の包摂関係と、利用者規定の二次符号や、拡張UNIXコード、シフトJISコードとの対応を取るため、符号変換表「XXX.A.AB.CC」に拡張UNIXコード、シフトJISコードなどの2バイトコード「DD」を付加した「XXX.A.AB.CC.DD」を作る。

データの送付は、最初に相手先クライアントへ変換表「XXX.A.AB.CC.DD」を送る。次に、2バイトコード「DD」で作成したデータを「XXX.A」に変換した後、相手先へ送る。受信側では、「XXX.A」形式で送られてきたデータを、事前に送られた変換表「XXX.A.AB.CC.DD」を使い、「DD」へ変換する。この時、送られたデータに対して、送信側に実装されている文字集合の規模が大きく、受信側クライアントの規模が小さい場合、サーバからオンデマンドで「XXX.A」対応の文字フォントをダウンロードする。受信側で包摂関係を確認する場合は、符号変換表を表として印刷する。

4.5 文字図形構成要素組合せ方式

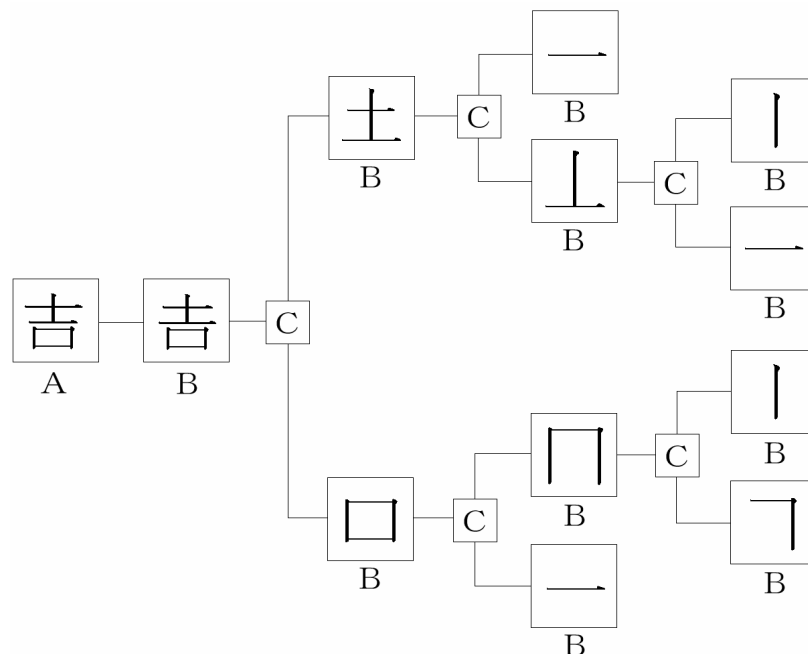
- a) 情報用交換符号として文字コードとともに、文字に対応する抽象的な図形表現に対して付与された番号（以下、文字図形コードとする）を使用する。

- b) 文字コード、文字図形コードが割り当てられていない異形字を対象とする時は、文字コードと文字図形コードとともに文字図形構成情報コードを使用する。

文字コード+文字図形コード+文字図形構成情報コードの形を基本とし、文字図形コードおよび文字図形構成情報コードは従来のコード体系と共存可能なエスケープシーケンスなどを用いて、文字コードに付随させて実装を行う。

文字図形コードは抽象的な図形表現に対して付与されている番号である。文字図形構成情報コードは、文字を段階的に分解し、構成要素となる文字図形の上下、左右などの位置関係を示す情報、接触、非接触などの接続を示す情報、図形の相対的な大きさや長さなどの図形的関係を示す情報をコード化したものであり、階層構造順に列挙して使用する。(図5参照。)

文字要素による文字図形構成情報を使用することで、フォントが用意されていない異形字を取り扱うことも可能である。



- A: 文字コード
- B: 文字図形コード
- C: 文字図形構成情報コード

本概念図でのコードは「A+B+CBBCBBCBCCBCCBB」となる。

図5 文字図形構成要素組合せ方式の概念図

第5章 今後の課題

5.1 運用方法についての課題

異形字対応のための文字コードのアーキテクチャについては、文字コードに異形字の情報（何らかの枝番）を付加して異形字を特定する方法を提案した。異形字の情報の表現形式についての最終的な詰めなど細部の検討は必要だが、この方向については、異形字対応のための文字コードの拡大の手段として、コンセンサスが形成された（または形成されつつある）と言える。

しかし、そのアーキテクチャをどう使用するかという側面については、議論が深まっていない上に異論も多く、コンセンサス形成には今後の議論が必要な段階にある。ここでは、どう使用するか、つまり運用方法について、今後の課題の整理を試みる。

今までに議論に上った運用方法についての話題は、いくつかの、相互に関係はあるがほぼ独立した課題に分割することが出来ると思われる。これには、例えば、以下のようなものが含まれる。

- 異形字をどう収集し、それにどんな考え方で枝番を付与するのか、その枝番を、データとしてどう生成し、どう利用するのか。
- 送信者と受信者の関係はどう構築するのか。
- 意図した字形を極めて正確に受信側で表現して欲しい、という要求と、書体の選択・設計の自由度との関係をどう考えるのか。
- 表示・印刷の場面での異形字の扱いと、テキストデータとしての扱い（例えば検索など）とでは、理想型が一致しない面があり、要求に多様性があるが、これにどう対処するのか。
- そもそも、送信者側の意図通りに受信者側で表現できること（例えば、必要なフォントファイルが受信者の手元に存在すること）をどう担保すべきか。

なお、このような様々な視点からの運用の検討にあたっては、上記以外に検討すべき視点がないかどうか、十分に議論する必要がある。運用方法に関する課題は、いずれも絶対的な正解のない課題であると思われる。最終的には、異形字に対するマーケットニーズをどう捕らえるか、利用者と提供者の両者を含むコンセンサスをどのように作るのかが最重要課題であろう。つまり、具体的な要望の形を取っていない「次のニーズ」に対する先見的なコンセンサスを形成し続ける必要がある。

ひょっとすると、アーキテクチャが決まれば、異形字の問題は、市場ニーズの把握の問題になり、それに対応するための周辺技術の課題になるかもしれない。文字コードの課題という時期は終了するかもしれない。

5.1.1 枝番付与の方法（制度・仕組み）

枝番を付与すべき対象の異形字をどのように収集するか、どのように管理するか（または管理しないか）については、まとまった議論を行えなかった。

アーキテクチャの検討では、異形字として、どちらかという一括りにして扱っていたが、実際に必要と言われている異形字の中には様々な性格の違いがある。例えば、以下のようなものである。

- 日常生活で一般的に用いられる、汎用性も高く共通的と考えられる異形字。
- 戸籍、住民基本台帳、その他の人名・地名などのための、ある種の権威や法的根拠があり、時間と費用をかければ収集可能であることが確実であり、かつ、その後の管理も可能と考えられる異形字。
- 古文書、古典、仏典、あるいは芸術作品のための、例示は可能であっても全貌を前もって捕らえることは困難と考えられる異形字。

これらの、様々な性格の違いに均等に対応するためには、制度の最初期はどうか、最終的には、新字種の提案制度の場合と同様に、提案制度に頼らざるを得ないと考えられる。

提案制度と一口で言っても、アーキテクチャの議論の過程で話題になったものの中には、提案されれば何でもそのまま採用する自由登録制度に近いものから、予め審査条件を決めておき要件を満たすかどうかを審査した上で採用する高管理型提案制度まであった。その他にも要件しだいで内容がかなり変わってくるが、提案制度による方式は、利用可能な異形字を一元的に管理するという点は共通である。

これに対して、フォント枝番方式が一つの運用方法として想定しているような、管理を行わず、送信者と受信者の合意（暗黙の了解を含む）に任せるという考え方もある。この場合、不特定多数を相手とする、いわゆる開かれた情報交換の場面での異形字の一元的な利用は、標準化組織などがイニシアティブを取るやり方ではなく、市場がde-facto標準を選択することに任せればよいという意見もある。

5.1.2 枝番の利用（データの生成）

枝番を付与する形式が標準化され、個々の異形字に対する枝番が決定したとしても、その枝番情報を含むデータを生成し利用する手段がなければ、異形字を扱えるようになったとは言えない。データを、どう生成し、どう利用するかという点についても、現状では温度差がある。

ある意見では、異形字が必要な時には、異形字登録簿を参照すればよいとされている。これは、紙の台帳、または電子的に配布される一覧表形式のイメージデータのようなものに、検索に便利な付加情報を加えたものと思えばよい。

具体的な提案ではないが、別の意見では、IME（パソコン等でキーボードから日本語入力を行う際に利用する、かな漢字変換ソフトウェアの類）の機能を拡大することを想定しているように思われる。

登録簿を参照するのであれば、データ生成を行う機器にはデータに含めようとしている異形字の文字フォントが無くてもいいように思える。他方、IME方式では、少なくとも変換操作を行っている時点では、データ生成機器で文字フォントが利用可能になっている必要があるように思える。

5.1.3 送信と受信の関係

不特定の送受信者間で、多数の枝番に対応した情報交換を期待通り行うために、システムをどう構成するかという点も、重要な運用の問題である

異形字を使用したいという要求の中には、送信者の意図した字を正確に受信側が再現したいというものがある。ある種の閉じたグループ内であれば別だが、不特定多数の間での情報交換において、このことをどう担保するのかということは、

技術的にも、制度的にも、自明ではない。例えば、登録制によって随時新しい異形字が追加され続けるような状態で、しかも莫大な数の異形字に枝番が割り当たると考えれば、登録された全ての異形字（全ての文字の全ての枝番）のフォントが、予め受信者側の機器に用意されていると想定することは現実的ではないと思われる。

いわゆるブロードバンドインターネットのような環境を前提に、情報本体に一律に添付するという意見や、フォントサーバのようなものを用意して受信側で必要になった時点で動的に取得するという意見もあるが、費用負担の問題や、知的財産権の問題など、解決しなければならない課題が多い。

逆に、受信者側の機器に指定された異形字のフォントが用意されていなかったときは「無いなりの動作」をすればよいという意見もある。この場合も、どういう動作が「無いなりの動作」としてふさわしいのかは自明ではない。いくつかの異形字情報の符号化方式の提案では、その方式の利点として「特定の異形字のフォントが用意されていなければ、基本の文字のフォントで代替することができる」ことが主張されている。しかし、異形字の要求の中には、例えば「特定の文学作品の特定の箇所では、ある特定の異形字が必要で、JISの字で代用したのでは著者の意図が伝わらない」のような主張もある。この種の要求に応えるためには、特定の異形字のフォントが無いときの「無いなりの動作」は、「表示できない文字の印」を表示したり、そもそもその作品の表示そのものがエラーになったりすることが望ましいという意見もある。

5.1.4 書体との関係

現代的な情報機器では、様々な書体が用意され、使い分けられるようになっている。異形字が登録された場合、その枝番は特定の書体に関するものなのか、任意の（登録の際に「見本」として使われたもの以外の）書体に関しても適用されるものなのか、などの点も今後検討が必要である。

ある異形字を「元の字」と区別している特徴は、任意の書体で表現できるわけではない可能性もある。また、表現できないわけでもなくとも、特定の書体のデザイン原則に反するかも知れない。

5.1.5 データとしての扱い

異形字を扱うという要求は、しばしば、文字の表示印刷の正確さの確保の問題として語られる。しかし、文字データの用途は、表示印刷だけではない。表示印刷以外の応用で、異形字アーキテクチャをどのように運用するのか、についても今後検討が必要である。

異形字を指定する枝番が付与されたテキストデータと、異なる枝番が付与されたテキストデータ、または枝番が付与されていないテキストデータがあったときに、それらのデータをどのように処理するのが望ましいのかは自明ではない。可能性としては、異なる形であっても「同義の字」として同一データとして扱いたい場合もあれば、全く別のデータとしたい場合もあるだろう。中間的なケースとして、「類似のデータとして扱いたい」などの場合も考えられる。（例えば、データベース検索において、検索条件で指定された通りの異形字情報を持ったも

のがあればそれだけを検索結果とするが、一件もヒットしなければ（またはヒット数が少なければ）異形字情報が異なるものもヒットさせる、など。）

5.1.6 不適切な運用の防止

ここまでの議論で明らかになる点の一つは、たとえアーキテクチャに合意が得られたとしても、それだけでは、その方法を実装することができるとは言えないという点である。つまり、運用方法に関する検討が行われて、これについての結論が得られる前に「とりあえず出来るものを実現する」ようなやり方で実装すると、結局、本来利用者が求めていたものとは懸け離れたものになる恐れがある。十分に検討されていない不適切な運用であっても、一度広まってしまうと固定化される危険が大きい。その意味でも、運用方法に関する課題を、早急に解決する必要がある。

特に、米国から提案され、ISO/IEC 10646に追加されたVariant Selector文字を安易に漢字に適用するような拡張のアプローチは危険であると考えられる。アーキテクチャの一案として検討する価値はあるだろうが、運用方法の対応を同時に検討した上で、漢字への適用が適切かどうかを決める必要がある。

5.2 枝番方式の運用に関する考慮点

枝番方式を運用する際に考慮しなければならない重要な点として、適用分野が異なる場合、そこで必要とされる枝番の粒度 (granularity) をどう捉えるか、という問題がある。

以下、分野が異なった場合に、必要とされる粒度が異なること、ある分野で峻別することが求められている異形字が、他の分野にとっては同一視することが求められる場合があることを、本委員会ではたびたび議論の対象となった行政分野、学術分野を例としてあげて検討する。

5.2.1 行政分野と学術分野の相違

行政分野においては、圧倒的に人名・地名などの固有名詞の共時的な差異を明確にすることを求められる場合が多い。

それに対して、学術分野では、歴史的な文書を扱う局面が多々あり、異形字関係を通時的に把握する必要がある。当然ながら、人名・地名などの固有名詞を通時的に区別したり同一視したりする必要性も多々存在する。

本ステージでは、これら行政分野と学術分野において、異形字アーキテクチャがどのような用いられ方をし、どの程度の粒度が必要とされているか、についてまで議論を深めることが出来なかった。今後、異形字アーキテクチャを適用、運用していく際には、これら分野の相違による粒度の違いについても、充分留意していく必要がある。

5.2.2 行政分野において予想される粒度の相違

行政分野内においても、必要とされる異形字粒度には、さまざまな差異が予想される。

巷間、固有名詞異形字に関わる議論でしばしば言及される例に、土吉と土吉を区別した際の税金納付通知と郵便為替による原稿料送付の話題がある。

税金納付通知の土吉と土吉の区別が間違っていたとき、名前が違っているので同一人物ではない、という理由で受け取りを拒否できるか云々。

一方、同一人物に郵便為替で原稿料が送られてきたとき、名前が違っているという理由で受け取りを拒否するかどうか云々。

この例は、あくまでも思考実験でしかないが、ほとんど瞬時に、同一人物においても、文脈が異なれば、自らの名前に関わる異形字の粒度についても、異なる要求が存在しうる、ということを示している。

行政分野における現在進行中の問題を考える際も、住民基本台帳、戸籍、自動車運転免許証、旅券（旅券には、漢字による姓名の記載はない）などで、どのような局面においてどのような粒度で異形字を峻別する必要があるのか、また、同一視する必要があるのか、について、運用面での検討をする必要があるだろう。

5.2.3 学術分野における粒度の相違

学術分野内でも、同様の問題の存在が予想される。

研究分野と研究方法の相違により、求められる粒度は大幅に異なる。

ISO/IEC JTC1/SC2/WG2/IRGにおいて審議が開始された統合漢字拡張Cのレパートリー候補として、韓国から提案されている文字には、高麗大蔵経に現れる多くの異形字が含まれている。ことの正否は措くとして、仏教研究では異形字の微細な差異に注目する場合が多々見られる。

一方、国語・国文学の研究においては、翻刻されたテキストの字形には拘泥する必要がなく、必要に応じて適宜原資料もしくはその写真製版版を参照すればよい、という議論をする研究者がいる。（豊島正之『JIS漢字批判の基礎知識』

<http://sharetext.honya.co.jp/archives/jiskanji/mt-gene.pdf>、小林龍生「要求する側の責任ということについて」（『電腦文化と漢字のゆくえ』平凡社、1998年所収）参照）

これらの例だけからでも、研究分野においても、複数の異なるアプローチがあることは明白であり、あるアプローチからの粒度要求が、他のアプローチからの異なる粒度要求を阻害することがないよう、運用方法についてさらに検討する必要があるだろう。

5.3 符号枝番方式における運用上の要件と課題

5.3.1 要件

a) 異形字は登録制

- 異形字は管理台帳を作って登録する。
- 同一文字の異形字に一連の番号を振る。
- 登録字形と異形字番号およびその関連情報は公開する。Webにより誰でも見られる環境を整備する。
- 異形字管理簿データベースには属性情報（部首、字素、読み、画数など）を付加し、属性情報による検索等の便に供せるようにする。
- 登録管理機関が必要になる。

b) 異形字登録規準の設定

- 異形字登録に関する規準作りが必須。
- 規準では、字形と図形との線引きおよび登録済異形字と登録候補異形字との同定/非同定の判別規準を明確化する。
- あくまで情報交換ツールとしての異形字管理の規準である。
- 基本的には俗字や誤字も異形字として登録を可能とする。

5.3.2 符号枝番方式における運用上の課題

a) 登録機関の問題

- 運用上の課題のひとつは登録機関をどこに置くか、その運営費をどうするかという問題である。
- ISO/IEC 2022系では各国国内規格毎に字形登録を行うという方法があるが、4.2にて述べたようにISO/IEC 10646と共通の字形番号にするためには国際的な登録機関が必要となる。
- あるいは、字形番号の番号帯をC、T、J、K、Vなど言語スクリプト毎に分けるという考え方も出てくる。このようにすればそのスクリプトを標準的に使う国毎の登録でよいことになる。
- 日本での登録機関はできれば政府またはその外郭団体等の公共機関が望ましい。
- 運営費は、登録と台帳閲覧を有償にしてそれを原資にするという考え方もある。しかし、それだけで賄えるのかどうかという危惧もある。

b) 登録基準の問題

- 大きな課題のもう一つは登録規準に関してである。
- 例えば変体仮名のような場合に、それが字形か図形かをどのように判別するかという問題がある。
- また明朝体等における筆おさえ（起筆）の有無をどうするかという問題もある。これに関しては人名表記等で筆押さえの有無を別字形として扱うニーズがある以上、別々に字形番号を付与できることでよいと思う。
- さらに外国文字をどうするかという問題がある。例えばなべぶた（一）の上は中国文字では一般に点であり、日本文字では棒になる。台湾ではお月さんの月と肉づきとで字形を厳密に区別する。また台湾ではしんによりの形は活字体であっても日本の筆記体のような形にしなければならない。
- ISO/IEC 10646ではもともと文化や言語が異なる文字を字形が似ているかどうかで統合している。ISO/IEC 10646の代表字をベースとするなら、5欄表記のスクリプト別に字形番号帯を分けて登録するという考え方も出てくる。

5.4 国際整合性について

5.4.1 既存システムと国際的ネットワーク環境との相互互換性

本委員会では、行政、学術・学芸などのさまざまな分野で、符号化文字集合JISやISでは対応できない、いわゆる外字処理、異形字処理といったニーズに対して、

従来の符号化文字集合規格としてどのように対応するか（新字の追加規格化提案制度）、従来の符号化文字集合規格では包摂規準やUnification Ruleに照らして対応が困難な問題への対処方法（異形字問題）に分けて、審議を進めてきた。

その結果、異形字問題に関しては、広い意味での枝番方式の採用を提言するに至った。

しかるに、このいわゆる異形字問題についても、従来の一般的な対処方法としては、既存のJIS、ISをベースとしながら、ユーザ定義領域を活用して個々の異形字に対して個別に符号を付与する、という方法が多く採られてきた。

外部システムとの相互運用性、JIS、ISとの整合性を考慮しなければ、上記のような個別的な対処方法でも、今後も問題が顕在化することはないとも考えられる。

しかし、行政領域についても、学術領域についても、今後、国の内外を問わず、既存データの公開、ネットワークを介した相互運用性の社会的ニーズがますます増大することは明白である。

既存のUCSをベースとし、従来のUnification Ruleとの整合性を保った異形字タグアーキテクチャの標準化は、このような従来のプロプライアタリーなシステムとオープンなネットワーク環境との間の相互運用性確保という観点からは、非常に有効であると考えられる。

なお、この場合のUCSとの相互運用性については、JIS X 0213とUCSとの整合性確保の過程で明瞭になったように、いたずらにaround the world trip conversionを求めるのではなく、ある既存システムからUCSへ、UCSから既存システムへといったround trip conversionのレベルに留めることが、規格の国際的円滑運用の観点からは、重要であると考えられる。

Around the world trip conversionを担保するためには、異形字タグアーキテクチャを採用した場合でも、個々の異形字の採録に当たっては国際的な合意形成が不可欠となるが、次項に述べるように、異形字の分別基準は、言語・文化、適用領域などにより大幅に異なることが予想される。統一したアーキテクチャの採用という視点と、個々の異形字にどのように枝番を振り、どの範囲で情報交換を行うか、という運用の問題は、別個に検討すべきであると考えられる。

5.4.2 異形字における言語・文化依存性（特に漢字に関して）

上項でも述べたように、何らかの形で、異形字アーキテクチャを標準化したとしても、そのアーキテクチャを用いて、どのようなプロセスで個々の異形字に枝番を付与するか、という運用上の問題がある。

ISO/IEC 10646-2の統合漢字拡張B策定の際、各国/地域から提案された字形のどれを規格票の例示字形として示すか、という点については、各国/地域の言語・文化的差異が如実に現れた。

符号化文字集合としての字体のUnification Ruleに関する現状での合意を前提とするにしても、国や地域によって、字体、字形、Glyph、デザイン差などの区分け認識は異なっており、ある国/地域にとっての異なる異形字として区別するニーズと別の国/地域にとっての区別しないニーズが、相反するケースが生じること、容易に予測できる。

また、目を国内に転じて、行政領域における異形字識別の必要性一つ取ってみても、より細かい異形次式別を必要とする局面とより緩やかな識別を必要とす

る局面がさまざまな度合いで存在する。行政領域と学術・学芸領域でのニーズの差が存在することは、述べるまでもない。

したがって、異形字アーキテクチャによる枝番の付与と運用に関しては、下記のような異なったレベルでの対応が不可欠となろう。

- a) 国際的に合意のとれた汎用異形字番号（要実現可能性の検討。）
- b) 各国/各地域内での相互運用性を保証する共通異形字番号。
- c) 適用分野、適用目的を限定し、関係者内での合意のみを前提とする私用異形字番号。

今後、異形字アーキテクチャの国内規格化、国際規格提案を行う場合は、上記のような問題点に鑑み、アーキテクチャと運用形態を峻別することに留意すべきである。

第6章 委員意見

ここでは、委員やオブザーバの方々のご意見を自由に書いていただいて、順不同に載せています。委員会で言い逃したこと、報告書に書かれていないが特に言いたいこと、報告書への意見として特に残しておきたいことなどを集めて記載しました。

6.1 北澤誓四郎（日本経済新聞）

一般レベルのワープロユーザの立場から意見と感想を述べます。

「新字種提案制度」は必然です。7万字を超える漢字に文字コードが付与された現在、担当委員会が自ら漢字を調査収集するという方法が限界にきているのは誰の目にも明らかです。また、「文献、資料など文字情報を原典に忠実に（文字コードを介して）電子的に記録・保存し、交換可能にもする」という前提を置くなら、「枝番方式」は非常に魅力的な手法に思われます。採否はともかく、客観的評価が可能にだけに、話が迷走する恐れも少ないでしょう。ただ、「すべての文字情報を……」という前提には、多少の違和感を覚えています。

報告書は第5章「今後の課題」で運用方法に言及していますが、“実行機関”が気掛かりです。今後、新たな異形字の発生頻度が高いと予想され、対応を迫られる分野は、（住民基本台帳など行政関係は国に整備義務があるので、ひとまずおくとすれば）報告書も指摘するとおり、文化・学術分野と見られます。報告書が登録管理機関を含め「提案文字検討委員会」について『国語や文字に関して権威のある公的組織内に常設する……学識経験者、文字規格関係者、ユーザ団体、関係行政機関などから構成する』としているのも、こうした認識に基づいていると理解しました。これに『公的資金を中心とした運営が望ましい』とすれば、既存の組織ですぐ頭に浮かぶのは、その機能を国立国語研究所に持たせることです。しかし、所管官庁である文部科学省の情報インフラ整備への貢献実績や当事者意識の薄さを考えると、いささか心許ないと言わざるを得ません。あるいは「書籍・文献の保存と提供」という需要面からは、国立国会図書館が中心となった何らかの組織、というのもあるかもしれません。

実務能力からすると、経済産業省系の団体、例えば日本規格協会といったところが候補に挙げられます。しかし、「提案文字検討委員会」には新字種（「符号化枝番方式」の管理も含む）の採否の判断などが求められます。第1ステージの初期に線引き問題にからみ「誰が決めるのか」で議論の応酬があったと記憶しています。ことが「工業規格」からすぐれて「文化・国語政策」の問題になってきている状況からすると、実務者としては気が重いでしょう。

「パソコンで文章が書ける」仕組みが、一般ユーザにいかに知られていないかを改めて噛みしめました。もちろん、ほとんどの一般ユーザーには必要のないことです。しかし、「パソコンの文字が足りない」と言うような人なら、ある程度の知識は不可欠です。本委員会の会期中も時折、漢字情報処理に関する記事が新聞に掲載されました。しかし、相変わらず、「字形の違い」や「収録文字の多寡」にばかり目が向いた過去の「文字不足談議」のパターンが多かったように思います。認識不足の責任の一端は新聞社にもあると自戒する次第です。

6.2 千田昇一 (NTT)

従来の文字コード標準化では、国内利用者からの要望を国際規格に反映するため、各種の努力がなされてきており、かつ十分な成果を出しているにも関わらず、結果的には国内利用者に満足いただけていない面があるというのも事実である。

たとえば、日本国内からの文字の不足を指摘する声に対しては、そのほとんどを国際規格に反映させることに成功しているにも関わらず、文字の不足を指摘する声はなくなるようには思えない。これは、標準化に関わる側からの説明が不十分で、一般の方の誤解を払拭できていないということかも知れないが、従来の文字コード標準化活動においては、利用者からの要望を聞いているようで、実際にはその要望の表面だけしか理解せず、その要望の本質部分を理解していなかったという面もあるのではないかと思う。現実には文字の不足を訴えてきた人は、実際には、その人を含むコミュニティが関心を持ついくつかの字だけに対しコミュニティの中だけでの解決策でも十分だったかもしれないのに、標準化のプロセスに当てはめられたとき、この一部のコミュニティの関心事でしかなかったことを、グローバルに共通の関心事として認定させるという途方もない労力を必要とする活動に対応付けざるを得なかったということではないだろうか。その結果、標準化活動関係者の努力が報われることがなく、かつ一般利用者の不満は解決されないという状況が継続していたものと思われる。

たとえば、今回の委員会で議論された「規格が制定されることとその規格に沿った製品が供給されることのギャップ」などは、一般の方々への周知が不十分だった事項の代表例と言えるのではないだろうか。文字コードについて、標準化で対応できる範囲と標準化だけでは対応できない範囲を明確にし、標準として規定すべき事項と標準では規定すべきでない事項があること、あるいは、標準で規定するための社会コストを配慮しても規定することが適切な事項と標準以外の方策を模索すべき事項があることなどが認識されたことは、大きな成果であったと思われる。この認識のもと、今回の委員会で提案された、「提案制度に基づく新字種追加」や「フォント枝番方式」等は、標準化活動と一般からの要望との間のギャップを埋めるための重要な一步と思われる。特に、異型字の枝番を「フォント」ごとに独立に管理運用するという考えは、「文字」に対する認識がコミュニティごとに差異があるという現実認識に立脚した考えであり、必要以上のことまでを標準で規定するという誤りを是正するという意味でも画期的な提案と思われる。

6.3 秦恒平、加藤弘一 (日本ペンクラブ)

まず、異形字の必要を認め、異形字をふくむ情報交換の技術的基盤について真摯な議論のおこなわれたことに敬意を表します。

われわれは「払い」や「止め」といった文字の顕微鏡的な細部に拘泥しているわけではなく、微細な違いの異形字は枝番やフォント名、グリフ名のような補助的手段によって表現するという本委員会の結論に、基本的に賛同するものです。

しかし、常用漢字表前文に「字体は文字の骨組み」とあるように、文字の骨格にかかわるような字体差まで補助的手段にまかせる解決法はあってはならないと考えます。

たとえば、JIS X 0208および、その拡張版であるJIS X 0213では「間」と「閒」を同一の符号に包摂していますが、このような包摂はとうてい容認できません。

また、JIS X 0208では「飲」と「飲」に別の符号をあたえて区別していますが、「蝕」、「館」、「飢」なども正字と通用字体を区別できてしかるべきです。現状の文字コードは粒度（包摂の範囲）が大きすぎる上に、右の「食」偏の字の例に明らかなように不揃いです。

ISO 10646は粒度がJIS X 0208やJIS X 0213よりも細かい点は歓迎しますが、粒度の不揃いも、一層はなはだしくなっています。

ユニコードが新たな標準になりつつある以上、ユニコードの粒度をもっと細かい粒度に合わせて均一化することを要望します。

フォントの実装についても、つよい要望を書きます。

「日本ペンクラブ」ではインターネット上に「電子文藝館」を開設し、島崎藤村初代会長をはじめとして、物故会員および現役会員の代表作を無償公開していますが、MacOS9までのマッキントッシュはJIS X 0208の範囲でしかフォントを搭載していないために、ユニコード使用に踏みきることができません。

アップル社は早くからユニコード対応をうたっていたにもかかわらず、最近の機種までJIS X 0208の範囲のフォントしか載せていませんでした。これはあまりにも不誠実な態度ではないでしょうか。

アップル社は旧機種ユーザーのために、機械の不得意な人でも容易にユニコード・フォントがインストールできるような手段を、無償で講ずべきです。

また、マイクロソフト社とアップル社の間で、搭載するグリフに数千字の異動があるのも、インターネット上で表現をする者にとって重大な障害となっています。BMPに含まれる漢字2万7千字をすべて搭載するよう、強く要望します。

(加藤)

ついでながら、「日本ペンクラブ電子文藝館」はすでに百人百編ちかい作品を公開し、「招待席」の森鷗外、幸田露伴、樋口一葉ら、「物故会員」の谷崎潤一郎、与謝野晶子、吉川英治ら、そして志賀直哉、川端康成、井上靖、遠藤周作ら「歴代会長」さらに多彩な現代各種受賞作家等の秀作力作を含んでいますが、それらの、「発信」に於いてでなく、広範囲な不特定機器による国内外「受信」に於いて、信じられぬほど多くの化け文字を露呈し、余儀ない苦しい対応を日々に迫られています。

こんなことで、良いわけではないのです。

表現に於いて「双方向で安定した正確な受発信」でなければ文化的に意味がない、「いくら文字セットに文字を満載し標準化してみても」、受発信間で確実に再現し合える搭載漢字数が極端に限定されてある限り、こういうウェブ事業を現実に起こした場合は無残なことになるだろうと、発言し続けました。だからといって例えば「電子文藝館」様の文化事業自体を不用とか時期尚早とか批判し非難することは出来ぬ道理であります。

文字の搭載・実装にどれほど隘路があるかを承知の上で、これこそが「表現者」団体の痛く不足・不十分を感じている困った現実であり、前段に述べた「要望」が、単なる口舌でないことを強く広く訴えたいのです。

われわれに何が出来るのか、今後をよく考慮し、他団体とも協調したいと願っています。

(秦)

幹事団注記

上記日本ペンクラブのご意見の中には、アップルコンピュータ社の文字コード対応に対するご意見が含まれていたため、この点に限定して、幹事団の判断で、同社に対してコメントを求めた。以下は、ご回答いただいたコメントである。（【】内は、メールでいただいた引用を含む回答を読みやすくするための幹事団の注記）

まずは、石崎委員長様からのご対応に感謝します。その上で、Appleとしての見解を下記に述べさせて頂きたいと思います。

【上記ペンクラブ意見中、MacOS9以前の文字コード対応に関して】

まず、これは符号化方法ではなくグリフレパートリーについてのご要望であると理解させて頂いた上で、話を進めさせて頂きたいと思います。

Unicodeへの対応は弊社の大きな目標の一つであり、その旨を公式に表明し続けてきたのは事実です。但し、その完全な実装については次世代オペレーティングシステムであるMac OS Xに対して行う旨、当初より併せて申し上げておりました。Unicodeへの完全な対応を実現するためにはオペレーティングシステム根幹からの対応が必要でしたが、Appleはかつて会社としての決定を下すに当たって、Mac OS 9までの旧オペレーティングシステム環境を大きく改築する事よりも、この全く新しくかつ旧Mac OSを大きく上回るポテンシャルを持つオペレーティングシステムであるMac OS Xに対して、全開発リソースを投入するという選択を行いました。これにつきましては、一重にご理解をお願いするばかりです。

【上記ペンクラブ意見中、BMPに含まれる漢字の搭載に関して】

これは、旧Mac OSと現在のWindows環境の差分であるJIS X 0212の収録文字数についてご指摘頂いていると理解させて頂いてよろしいでしょうか。Appleは、日本語での情報交換における新しいスタンダードとして、JIS X 0213をグリフレパートリーとしてMac OS Xにおいて既に採用しており、この段階において差分の一部は吸収しています。さらに、Microsoft様も今後リリースされるバージョンのWindows XPで、JIS X 0213を同じ意味でサポートする旨発表されていますので、ここにおいて両者の情報交換はUnicode及びJIS X 0213をベースにして行う事が可能になる旨お知らせしたく思います。

また、BMP中の漢字フルサポートにつきましては、日本ペンクラブ様からのご要望として承らせて頂いてよろしいでしょうか。かつ、これは同時にMicrosoft様を始めとする各コンピュータベ

ンダ様へのご要望でもあるべきと思いますが、必要であればこの辺りについてももう少しお話をお聞かせ頂ければと思います。

6.4 宮澤彰 (国立情報学研究所)

文字に対してまじめな態度で、データベース作成およびサービスを行っている立場からの意見です。

大胆に要約すれば、我々の立場からは、文字コードが安定していることが最も望ましく、文字の追加や変更は最小限（字数よりは頻度として）にとどめられるべきであるということです。さすがに当委員会内では、表せない文字があることがけしからんという単純粗暴な不足論は、あまり聞かれなくなりましたが、社会的にはまだまだある程度の説得力をもっています。しかし、文字が多くなることによって迷惑をこうむる（製品の質が悪くなるか、質を保とうとすればコストが上昇する）人々の声は、あまり大きく聞こえません。この点について意見を述べます。

最初に異形字（この委員会での用語の）の追加に関してです。新たな異形字が（内字での拡張で）文字コードに追加されたとします。仮に「劔」の字が新たに追加されたと仮定しましょう。（これは実はJIS X 0208にすでにあるものですが。）この異形字を使いたかった人は喜ぶでしょう。しかし、文字についてまじめなデータベースサービス業者は、このことによって次のような対応を取らなければなりません。最初に、「劔」「劔」「劔」などの字が、これまで蓄積されたデータベース中のどこに現れるかを調べます。次いで、その各々の現れについて、「劔」として記録されるべきであったかどうかを調査しなければなりません。この調査は、当然ですがデータベース中の記録だけを見ても意味がなく、データベースを作成したときのソースに戻って調査する必要があります。最後に「劔」として記録すべきであった文字の場合、データベースの修正をすることになります。この作業は、これまで蓄積したデータベースに百万件、千万件単位のレコードがあった場合、大変な労力となることは容易に想像していただけるでしょう。

文字にまじめな態度のデータベース作成・サービス業者は、ここで判断しなければなりません。今述べたような大変な労力をかけてこの作業を行うか、この文字が文字コードに追加される以前のデータについては、旧コードによる作成といて済ませるかです。後者の立場を取った場合、データベースの一貫性がなくなることにより、質の低下ということになります。自らに起因するものではなく、他者が異形字を区別したいという要求をしたことによって起こったわけです。

実は、データベースサービス業者に限らず、テキストデータを保有している人は、すべて潜在的にこの危険にさらされているわけです。文字コードに異形字が追加されたとき、保有しているテキストファイルを自分で調べるか、売られているものであれば新コードによる版を購入しなおすか、自分には区別がなくてよいのでそのまましておくかです。ところが、自分には区別がなくてよいといっても、システム上で区別されてしまえば、新しいテキストファイル「劔」を含むものと一緒に検索するわけに行かなくなるのです。この点、自分には関係ないと思っ

ているところで、保有しているファイルの有用性の低下が起こるわけです。異形字の追加でない新文字種の追加は、これより「罪が軽い」といえるでしょう。しかし、影響がないわけではありません。文字にまじめなデータベース作成者は、何らかの方法で「外字管理」をしてその情報を保持しています。たとえば、

私用文字コード領域にそれらの文字を割り当てて、その文字に関する情報を辞書などの形で別途保持する方法、外字であることを示すいわゆるゲタ文字にして、それがどういう字であったかを別途記録する方法などがあります。このような「外字管理」をしている場合、新文字種の追加がおこると、追加された新文字種の表と、保持している私用外字表とを比較して文字を同定し、私用文字から新たに追加された「内字」へ文字コード変換をすることになります。この作業は、異形字の場合と異なり、原則としてデータベース作成時のソースに戻ることは必要なく、変換表さえ確定すればあとは自動的な変換作業となります。とはいえ、データベースの規模が大きい場合、サービスを中止することなくこれらの作業を行うには十分な準備が必要で、とても日常的に行えるものではありません。

この点、新文字種の提案制度が、文字の同定という「外字管理」の最も大変な部分をパブリックにサポートするものとして働くようになれば、個々の事業者の負担を減らすことができ、歓迎するものです。

新文字種の追加にしる異形字の追加にしる、我々のようなユーザにとっては(そしてテキストファイルを保持しているすべてのユーザにとって)、データをカレントに保つための余計な労力を要求されることがらです。当然、これらは広い意味でのコストになります。もちろん、私用の範囲でなくより広いユーザに正しいテキスト情報を頒布することができるようになるというプラス面はあります。したがって、文字コードの拡張を全面否定するものではありません。ただこれらのマイナス面を少なくするために、文字コードの変更は必要最小限度の回数にとどめ、なるべく長期間安定させることが望ましいと考えます。

6.5 佐藤敬幸 (CICC)

文字コード委員会第二ステージを振り返ると。成果として、漢字の追加要求を新字種と異形字に分離した議論が違和感なく行われるようになったことと、新字種の要求への対応方法が見えてきたことだと思います。

他方、異形字については、まだまだ迷走状態で、コンセンサスには遠い状態だと感じています。

まず異形字の要求の実態についてのコンセンサスが十分形成されていないこと、つまり、ほとんど異形字の要求についての議論までにはいたらず、各自の意見表明に終わってしまったと感じています。この点の合意にはまだまだ真剣な議論が必要ではないでしょうか。そのためには少しばかり工夫が必要なようです。たとえば「文字化け」があって困ると言うだけでは問題は解決しません。もし、「包摂の範囲内でも文字の字形を特定したい応用がある」問題なのか「送受信者間の文字コードの同期」問題なのかなどに絞れば検討課題となります。

現状は、まだ群盲象をなでている状況のようで、全体像を構成するための努力が必要なようです。

6.6 Martin J. Dürst (慶應義塾大学)

適切で迅速な情報交換・情報処理のためには文書に使われる文字を特定し、それぞれに番号(文字コード)を振る必要があります。その場合にはいわゆる「包摂」基準を一般的によく行われる処理に合わせないといけません。現在のJIS X

0208で昔から使われ、最近、より明確にされた包摂基準がISO 10646でも殆ど変更なしに使われているのは日本からの大切な貢献の一つといえます。

しかし特に漢字の場合には長い歴史とその膨大な数のため、たまにしか出てこない新字種、外字、異体字、異形字、俗字、誤字など、標準に入っていない字(形)・入れにくい字(形)もたくさんあります。特に地名・人名、歴史的資料、文学などでは、このような字(形)に対応できないといけません。しかし用途や個人によって必要な字(形)はいろいろ異なります。そのような多様な要求に対応できないといけません。

本報告書の第4章において重点が置かれている「枝番方式」などではこのような多様な要求には対応できません。正式な字の標準化に必要な時間より登録に必要な時間が少し短くなるだろうが、日々加速する情報社会には対応できないでしょう。登録機関も必要になってきますが、そのための運用費は誰が支出するのでしょうか。誰がどの様な基準で運用するのでしょうか。また登録されないものにはどのように対応すればよいのでしょうか。最近の規制緩和などの流れに逆らっているのではないのでしょうか。もっと簡潔かつ柔軟で適切な方法はないのでしょうか。

「枝番」を使わない提案はもう既にいくつか仕様化されています。

- 1) JIS X 4052:2000 (日本語文書の組版指定交換形式) の <CH> 要素
- 2) JIS TR X 0047:2001
- 3) W3C Scalable Vector Graphics (SVG) 勧告の altGlyph 要素
(<http://www.w3.org/TR/SVG/text.html#AlternateGlyphs>)
- 4) WAPForum Pictogram 仕様書
(<http://www.wapforum.org/what/technical.htm>から参照)

これらの仕様の共通点は次の通りです。

- a) 符号化のレベルではなくて、タグ付けのレベルで使われています。
- b) 字(形)を特定するには URI (ウェブ・アドレス) が使われています。
- c) URI を参照すると字(形)についてのデータ(画像データ、書体の輪郭など)が得られます。現在は表示用のデータが主ですが、発音、意味、使い方、他の字(形)との関係などのデータも扱うことが可能です。
- d) 基本的には新字種にも特定の字形にも対応できます。
- e) 字形の場合にタグの内容として元の字を入れることによって、検索もできますし、万が一、字形データが見つからなかったらフォールバックとして元の字を表示できます。新字種でも単なる画像にフォールバックができます。

これによって一時的な実験用の字(形)から、幅広い分野で共有される字(形)まで様々な要求に柔軟に対応できます。新しい字(形)の追加コストはそれを利用したい人が負担すれば済みます。同じ字(形)が何回も定義される恐れや、新字種を標準に追加した後のデータの更新の問題もありますが、字(形)の関係の情報を更新可能にすることで解決できます。

「枝番方式」をベースに具体的な議論に入る前にタグ付けによるやり方をもっと検討すべきです。そうしないと最近のウェブ技術やインターネット技術に適していない、柔軟性が欠如したものになりかねません。

第7章 おわりに

2000年12月にスタートした第2ステージの文字コード委員会は9回の開催を持って終了した。第1ステージの文字コード委員会は1998年11月にスタートしたので3年余りにわたって文字コードについて検討してきたことになる。特に第2ステージの委員の多くは第1ステージに引き続いて委員をお願いした方がほとんどなので、長期間のご貢献に改めて感謝したい。お忙しい中を熱心に参加していただいた。

第1ステージの委員会では、文字コードの標準化についていわゆる青天井問題としてとらえて、青天井に増える可能性のある文字の標準化について提言を行った。その5つの提言の一部はJTC1の総会で日本から提案し賛同を得た。第1ステージ委員会の終了後の状況の変化を踏まえて始めた第2ステージの文字コード委員会では、国内の漢字問題を中心に検討した結果、本報告書の第2章にあるように二つの提言にまとめることができた。

提言の1番目に付いては、その趣旨を生かした活動が2002年度から始まるとうとしている。また2番目の異形字のアーキテクチャについては、国内ばかりでなく国際的な動きもあるが、異形字は文字の微妙な違いを扱うだけに、関係者の合意を得るのが難しい面がある。今後の標準化の進め方に注意を要する課題である。

これらの標準化は、ニーズを満たすための文字コードを増やす方向での検討である。しかし、一方では日常生活に必要な文字（漢字）はかなり少なく済むという考えがある。これは、情報規格調査会の名誉会長の和田弘先生が従来から主張されているものである。そのような考えに沿った検討は、2001年12月に正式にスタートした情報処理学会試行標準制度で検討されており、東京学芸大学松岡氏のBUCSを中心とした8000字程度の漢字集合として2002年3月に試行標準として成立している。5月以降の国際会議にも報告して漢字使用国の賛同を得ていく予定である。

本委員会の審議を進めて行く上では、SC2専門委員会のメンバーを中心とした準備委員会による支援が重要な役割を担った。また2001年秋からは異形字WGを組織して、異形字アーキテクチャや異形字の標準化の効果や問題点について検討し、率直な意見が多く出され本報告書に反映された。これらの委員会に参加した皆さんに感謝したい。

第1ステージの報告書の五つの提言の中では、第2ステージの活動によって既に実現できたか実現の方向に動き出しているものが多いが、残された課題としては国際的な登録制度と文字モデルの提案がある。これらは今後の国際的な標準化活動に柔軟に対応する必要がある。従って、次の段階の文字コード委員会に期待したい。

最後に、本委員会にオブザーバとして参加した関係者の方々、一般参加として熱心に委員会に加わっていただいたの方々、また、事務局として多くの事務処理に努力していた情報規格調査会の三田事務局長、木村氏に感謝したい。

付録1 標準化動向

1999年8月以降の文字コード規格関連の動向としては、下記のもものが挙げられる。

付1.1 国内標準化動向

- a) 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 0213の制定と国際整合

平成12年1月20日に制定された。この制定に伴い、国内標準と国際標準の整合性を確保するため、情報処理学会情報規格調査会SC2専門委員会を通して、ISO/IEC JTC1/SC2/WG2およびISO/IEC JTC1/SC2/WG2/IRGに対し、

- CJK Compatibility IdeographsへのJIS Compatibility Ideographsの追加
- 統合漢字拡張Bへの追加提案
- 非漢字のBMP領域への追加提案

などを行い、2001年中には、ほぼ所期の目標を達成した。

- b) 国際符号化文字集合 (UCS) 第1部：体系及び基本多言語面 JIS X 0221-1:2001 (ISO/IEC 10646-1:2000) の制定

平成13年4月20日制定。表題の通り、ISO/IEC 10646-1:2000の翻訳規格である。漢字に関しては、統合漢字拡張Aが含まれており、従来の統合漢字と併せて約28000字が採録されている。

- c) 国語審議会答申「表外漢字字体表」とJIS

平成12年12月8日に国語審議会から答申された「表外漢字字体表」を受け、その趣旨を日本工業規格の符号化文字集合 (JIS文字コード) に反映することを目的として、財団法人日本規格協会符号化文字集合調査研究委員会において、平成13年5月以来検討を続けている。2002年1月15日より一ヶ月間、考え方の公開レビューを行った。2002年度以降、何らかの形でJISへの反映が行われることとなろう。

付1.2 国際標準化動向

- a) ISO/IEC 10646-1:2000の制定

ISO/IEC 10646-1:1993の改訂版に相当。1993年度版以降に追加された追補がすべて反映されている。翻訳規格は、JIS X 0221-1:2001。

- b) ISO/IEC 10646-2:2001の制定

UTF-16によって利用可能となる16面のうち、1面 (記号類及びIdeographs以外のスクリプト)、2面 (Ideographs)、14面 (スクリプト、記号類以外の機能符号など) を制定。

漢字関係では、統合漢字拡張Bが2面に制定された。

- c) ISO/IEC 8859の新しいパート

2001年に、

- ISO/IEC 8859-11:2001 Information technology — 8-bit single-byte coded graphic character sets — Part 11: Latin/Thai alphabet
- ISO/IEC 8859-16:2001 Information technology — 8-bit single-byte coded graphic character sets — Part 16: Latin alphabet No. 10

が追加された。

付録2 文献

ここでは、当委員会の活動にあたって参考としたものを含め、関連する内容を含む文献を一覧として示す。これは、第1ステージ委員会で作成した文献リストを元に、内容を追加・更新して作成した。特に第2ステージ委員会の主要な課題であった漢字の異形字に関連する文献を多数追加している。

リストの作成にあたっては、文字鏡研究会に協力をいただき、備考についても提供を得た。

付2.1 論著

付2.1.1 1970年代以前

- 1) 橋本進吉：文字及び假名遣の研究，橋本進吉博士著作集，3，岩波書店 (1949)
- 2) 橋本進吉：國語音韻の研究，橋本進吉博士著作集，4，岩波書店 (1950)
- 3) 倉石武四郎：漢字の運命，岩波新書 (1952)
- 4) 大野晋：上代假名遣の研究 日本書紀の假名を中心として，岩波書店 (1953)
- 5) ムーアハウス著，ねずまさし訳：文字の歴史，岩波新書 (1956)
- 6) さねとうけいしゅう：中国の文字改革，くろしお出版 (1958)
- 7) 魚返善雄：日本語心得帳，大修館書店 (1961)
- 8) 築島裕著：平安時代の漢文訓讀語につきての研究，東京大學出版會 (1963)
- 9) 藤堂明保：漢字の起源，徳間書店 (1965)
- 10) 西田龍雄：生きている象形文字 モソ族の文化，中公新書 (1966)
- 11) 貝塚茂樹編：古代殷帝国，みすず書房 (1967)
- 12) 藤堂明保ほか：言語，中国文化叢書，1，大修館書店 (1967)
- 13) 方言・釋名・說文解字・說文解字繫傳通釋・玉篇・廣韻，四部叢刊初編，5 經部，台湾商務印書館 (1967)
- 14) 小松茂美：かな その成立と変遷，岩波新書，青-679，岩波書店 (1968)
- 15) 大矢透：仮名遣及仮名字体沿革史料，勉誠社 (1969)
備考：原明治42年
- 16) 白川静：漢字，岩波書店 (1970)
- 17) 藤枝晃：文字の文化史，岩波書店 (1971)
- 18) 白川静：金文の世界，東洋文庫，184，平凡社 (1971)
- 19) 杉本つとむ：漢字入門 干禄字書とその考察，早稲田大学出版会 (1972)
- 20) 中田祝夫編：音韻史・文字史，講座国語史，2，大修館書店 (1972)
- 21) 白川静：甲骨文の世界，東洋文庫，204，平凡社 (1972)
- 22) 芝川町郷土史研究会編：異体文字集，私家版 (1973)
備考：文字鏡研究会蔵
- 23) 石丸久：私の漢字教室 悪いカンジいいカンジ，学芸図書 (1973)
- 24) 上田正昭編：日本古代文化の探究 文字，社会思想社 (1975)
- 25) 柴田武ほか：日本語と国語学，岩波講座日本語，1，岩波書店 (1976)
- 26) 渡辺茂：漢字と図形，NHKブックス，日本放送出版協会 (1976)

- 27) 白川静：漢字の世界1，東洋文庫，281，平凡社（1976）
- 28) 白川静：漢字の世界2，東洋文庫，286，平凡社（1976）
- 29) 橋本萬太郎ほか：音韻，岩波講座日本語，5，岩波書店（1977）
- 30) 千野栄一ほか：国語国字問題，岩波講座日本語，3，岩波書店（1977）
- 31) 矢島文夫：文字学のたのしみ，大修館書店（1977）
- 32) 鈴木修次：漢語と日本人，みすず書房（1978）
- 33) 鈴木修次：漢字 その特質と漢字文明の将来，講談社現代新書，497，講談社（1978）
- 34) 樺島忠夫：日本の文字，岩波新書（1979）
- 35) 佐藤喜代治：日本の漢語，角川書店（1979）
- 36) 小松英雄著：いろはうた 日本語史へのいざない，中公新書，558，中央公論社（1979）

付2.1.2 1980年代

- 37) 橋本萬太郎ほか：世界の中の日本文字 その優れたシステムとはたらき，弘文堂（1980）
- 38) 大原信一：漢字の移り変わり，東方書店（1980）
- 39) 貝塚茂樹ほか：中国の漢字，日本語の世界，3，中央公論社（1981）
- 40) 小松英雄：日本語の音韻，日本語の世界，7，中央公論社（1981）
- 41) 西田龍雄編：世界の文字，講座言語，5，大修館書店（1981）
- 42) 倉石武四郎：漢字・日本語・中国語，くろしお出版（1981）
- 43) 築島裕：仮名，日本語の世界，5，中央公論社（1981）
- 44) 尾崎雄二郎編：訓読説文解字注，東海大学古典叢書，東海大学出版会（1981～）
備考：（現在の所、八冊目まで刊行）
- 45) 原田種成：漢字の常識，大修館書店（1982）
- 46) 西田龍雄：アジアの未解読文字，大修館書店（1982）
- 47) 築島裕：仮名，日本語の世界，5，中央公論社（1982）
- 48) 中田祝夫：日本の漢字，日本語の世界，4，中央公論社（1982）
- 49) 藤堂明保：漢字の過去と未来，岩波新書（1982）
- 50) 丸谷才一編：国語改革を批判する，日本語の世界，16，中央公論社（1983）
- 51) 説文会編：説文入門，大修館書店（1983）
- 52) 鈴木修次：漢字再発見，PHP研究所（1983）
- 53) 西田龍雄：漢字文明圏の思考地図，P H P（1984）
- 54) 阿辻哲次：漢字學 説文解字の世界，東海大学出版会（1985）
- 55) 樺島忠夫ら編：事典・日本の文字，大修館書店（1985）
- 56) 杉本つとむ編著：漢字入門 干禄字書とその考察，新装・改訂増補版，早稲田大学出版部（1985）
- 57) 江守賢治：解説字体辞典，三省堂（1986）
- 58) 沼本克明：日本漢字音の歴史，国語学叢書，第1期，10，東京堂出版（1986）
- 59) 築島裕：歴史的仮名遣い，中公新書，810，中央公論社（1986）
- 60) ガウアー著，矢島文夫ら訳：文字の歴史 起源から現代まで，原書房（1987）
- 61) 呉惠霖著，高畑常信訳：木簡手帖，木耳社（1987）

- 62) 小林芳規：角筆文献の国語学的研究，汲古書院（1987）
- 63) 白川静：文字逍遥，平凡社（1987）
- 64) 佐藤喜代治編：漢字講座，全12巻，明治書院（1987～1989）
- 65) K.F=パップ著，矢島文夫ら訳：文字の起源，岩波書店（1988）
- 66) 此島正年博士喜寿記念論文集刊行会編：国語語彙語法論叢，此島正年博士喜寿記念，桜楓社（1988）
備考：伊藤雅光「コンピューターによる『今昔物語集』本文の自動分割の問題点」、小倉肇「パソコンによるKWIC漢字索引作成の試み」などを収録。
- 67) 白川静ほか：文字の文化学，季刊 iichiko, No.9, iichiko（1988）
- 68) 野村雅昭：漢字の未来，筑摩書房（1988）
- 69) 阿辻哲次：図説・漢字の歴史，大修館書店（1989）
- 70) 加納喜光：漢字の常識・非常識，講談社現代新書（1989）
- 71) 岸俊男編：ことばと文字，日本の古代，14，中央公論社（1989）
- 72) 原田種成：漢字小百科辞典，大修館書店（1989）
- 73) 小林芳規：角筆のみちびく世界，中公新書，810，中央公論社（1989）
- 74) 松丸道雄ほか：中国古文字と殷周文化，東方書店（1989）
- 75) 陳原著，松岡榮志編訳：ことばの社会機能 五個の漢字をめぐる，凱風社（1989）
- 76) 尾崎雄二郎：漢字の年輪，角川書店（1989）
- 77) 馮志偉編：現代漢字和計算機，北京大学出版社（1989）

付2.1.3 1990年代

- 78) G.ジャン著，矢島文夫監修：文字の歴史，白水社（1990）
- 79) エツコ・オバタ・ライマン：日本人の作った漢字 国字の諸問題，叢書ことばの世界，南雲堂（1990）
- 80) 田嶋一夫：最新JIS漢字辞典，講談社（1990）
- 81) 李学勤著，小幡敏行訳，佐野光一訳：中国古代漢字学の第一歩 古代文字学入門，凱風社（1990）
- 82) 鈴木修次：漢字情報力再発見，創拓社（1990）
- 83) [特集] 漢字とコンピュータ，しにか，Vol.1, No.2, pp.6-47，大修館書店（1990）
- 84) [特集] いま漢字の規格化を問う，しにか，Vol.1, No.7, pp.8-48，大修館書店（1990）
- 85) 森博達：古代の音韻と日本書紀の成立，大修館書店（1991）
- 86) 加納喜光：漢字の博物誌，大修館書店（1992）
- 87) [特集] 古典とコンピュータ，しにか，Vol.3, No.2, pp.8-55，大修館書店（1992）
- 88) 阿辻哲次：教養の漢字学，大修館書店（1993）
- 89) 阿辻哲次：漢字のベクトル，筑摩ライブラリ，筑摩書房（1993）
- 90) 高崎一郎編：正字と新字，私家版（1993）
- 91) 国立歴史民俗博物館編：科学の目でみる文化財歴博フォーラム，アグネ技術センター（1993）
- 92) 三上吉彦，池田巧，山口真也編著：電腦外国語大学，技術評論社（1993）
備考：Unicode普及以前の工夫を網羅。

- 93) 馬淵和夫：五十音図の話，大修館書店（1993）
- 94) [特集] 漢字コードの国際標準化，しにか，Vol.4，No.2，大修館書店（1993）
- 95) 阿辻哲次：漢字の字源，現代新書，講談社（1994）
- 96) 阿辻哲次：漢字の文化史，NHKブックス，日本放送出版協会（1994）
- 97) 加藤弘一：石川淳 コスモスの知恵，筑摩書房（1994）
備考：文字コード問題の本。
- 98) 河野六郎：文字論，三省堂（1994）
- 99) 紀田順一郎：日本語大博物館，ジャストシステム（1994）
- 100) 国士舘大学文学部考古学研究室編：正倉院文書異体字集成，国士舘大学文学部考古学研究室（1994）
備考：大日本古文書1～5巻巻末所載の「異体字一覧」を分類して読みその他の注釈を加えたもの。
- 101) 勝村哲也，丹羽正之：漢字典，京都漢字研究会（1994）
- 102) 箭内敏夫：電脳辞書の国語学 ワープロ日本語変換の徹底検証，おうふう（1994）
備考：技術評論社のコンピュータ雑誌ざべに連載されたFEP（IME）の登録辞書に関する比較検証。
- 103) 東野治之：書の古代史，岩波書店（1994）
- 104) おうふうコンピュータ資料センター研究所編：都という文化，歴史文化研究，2号，おうふう（1995）
- 105) 加納喜光：見て味わう漢字の満漢全席，徳間書店（1995）
- 106) 河野六郎，西田龍雄：文字鼻眞，三省堂（1995）
- 107) 蔵中進：則天文字の研究，翰林書房（1995）
- 108) 中嶋幹起：コンピューターによる北京口語語彙の研究第1冊資料編，アジア・アフリカ基礎語彙集，28，内山書店（1995）
- 109) 矢島文夫監修：人間と文字，モリサワ（1995）
- 110) 林四郎，松岡榮志：日本の漢字・中国の漢字，三省堂（1995）
- 111) [特集] 正しい漢字とは何か，しにか，Vol.6，No.5，大修館書店（1995）
- 112) パソコンを使う日本語研究，日本語学臨時増刊号，日本語学，14巻，8号，明治書院（1995）
備考：専門分野に関わる網羅的な紹介に欠いていた当時、画期的な特集。
- 113) Ken Lunde著，春遍雀來，鈴木武生訳：日本語情報処理，ソフトバンク / O'Reilly&Associates（1995）
備考：漢字文化圏の文字コードについて網羅的にまとめた初の書物。（原1993）
- 114) 安永尚志：文化系のための情報検索入門 パソコンで[漱石]にたどりつく，セミナー：原典を読む，8，平凡社（1996）
備考：情報検索の視点からパソコン利用法をわかりやすく解説。脚注も豊富で著者のサービス精神が垣間見える。
- 115) 伊藤英俊：漢字文化とコンピュータ，中公PC選書，9，p.180，中央公論社（1996）

- 備考： NECで開発に携わっていた著者によるわかりやすい文字コード問題の解説。
- 116) 遠藤織枝： 中国の女文字，三一書房（1996）
- 117) 堅田直： 情報考古学 パソコンが描く古代の姿，ジャストシステム（1996）
- 118) 言語学林1995-1996編集委員会編： 言語学林1995-1996，三省堂（1996）
備考： 井口三重「パソコン利用による糸魚川言語地図作成」などを収録。
- 119) 国文学研究資料館： 第1回シンポジウムコンピュータ国文学講演集，国文学研究資料館（1996）
備考： 資料館が蓄積してきたコンテンツと経験を背景にその年度の雰囲気伝える貴重な資料。
- 120) 西野嘉章編： 歴史の文字 記載・活字・活版，東京大学コレクションIII，東京大学出版会（1996）
- 121) 中野洋： パソコンによる日本語研究法入門 語彙と文字，笠間書院（1996）
備考： 付録FDDに独自開発のMS-DOSプログラムを収録し、その使用法を解説したもの。DOSがわかる読者には今でも有用。
- 122) 内田慶市，野原康宏： マックで中国語，ひつじ書房（1996）
- 123) 府川充男： 組版原論：タイポグラフィと活字・写植・DTP，太田出版（1996）
- 124) 文化庁文化政策課編： JIS情報交換用漢字符号漢字字体関係参考資料集，文化庁（1996）
- 125) [特集] 漢字入門Q&A，しにか，Vol.7，No.6，大修館書店（1996）
- 126) [特集] 国境をこえる日本語の条件」，GLOCOMNEWSLETTER，Vol.5，No.1，学校法人国際大学グローバルコミュニケーション・センター（1996）
備考： 芝野、豊島、笹原、高田、古瀬、家辺の各氏による対談。
- 127) 歴史の文字 記載・活字・活版，東京大学総合研究博物館（1996）
- 128) [特集] 文字コード現状と未来，人文学と情報処理，No.10，勉誠社（1996）
備考： 1997
- 129) 「人文科学とコンピュータ」シンポジウム（主催国文学研究資料館第2回シンポジウム「コンピュータ国文学」/科研費重点領域研究「人文科学とコンピュータ」テキスト処理研究班）於機械振興会館（1996）
- 130) (社)日本文藝家協会： 国語審議会会長への要望書（1997）
- 131) M.ベンゼ著，草深幸司訳： 情報美学入門 基礎と応用，勁草書房（1997）
- 132) 安永尚志： 1996年度「人文科学とコンピュータ」シンポジウム予稿集，研究成果報告書，3，国文学研究資料館（1997）
備考： 重点領域「人文科学とコンピュータ」テキスト処理研究班と資料館の第二回シンポジウムとを兼ねた会合。
- 133) 加藤弘一，吉目木晴彦，坂村健： TRONSHOW'96再現シアター「ネットワーク時代～日本語が危ない」，TRONWARE，vol.44，パーソナルメディア株式会社（1997）
- 134) 芝野耕司編著： JIS漢字字典，日本規格協会（1997）
- 135) 石川忠久，松岡榮志： 漢字とコンピュータ，p.187，大修館書店（1997）

- 136) 太田昌孝：いま日本語が危ない 文字コードの誤った国際化 ， p.221，丸山学芸図書（1997）
備考：文字コード問題の本。
- 137) 大島正二：<辞書>の発明 中国言語学史入門，三省堂（1997）
- 138) 中村康夫著，国文学研究資料館編：国文学電子書齋術 コンピュータに何をさせるか，セミナー原典を読む，10，平凡社（1997）
備考：データベースを作成する場合の留意点・問題点を資料館でのデータ設計の経験から抽出して解説。逆の視点から読めば検索するときのコツやデータベースの癖の読みとり方も見えてくる。
- 139) 張静賢編，松岡榮志訳：現代中国漢字学講義，p.245，三省堂（1997）
- 140) 藤本和貴夫，木村健治編：言語文化学概論，大阪大学出版会（1997）
備考：細谷行輝「言語とコンピュータ」、今井光規「コーパス言語学」、渡辺秀樹「CD ROM版の利用法」などを収録。
- 141) 文化庁文化部国語課編：字体・字形差一覧，漢字字体関係参考資料集，文化庁（1997）
- 142) 文字鏡研究会編：文字鏡研究会会報，1，文字鏡研究会（1997）
- 143) 李範文，中嶋幹起編著：電脳処理西夏文雑字研究，不二出版（1997）
- 144) じんもんこん，重点領域「人文科学とコンピュータ」研究情報誌，4，重点領域「人文科学とコンピュータ」事務局（1997）
備考：現在、ネット上で重視されている研究はこの科研費プロジェクトに関係した物が多い。別途、CD-ROMでの成果報告書が刊行されたこともあり、その中には近藤みゆき・近藤泰弘による和歌のデータベースCD-ROMもある。
- 145) 島田雅彦ほか：電脳売文党宣言，アスキー出版（1997）
備考：文字コード問題の本。
- 146) [特集] ネットワーク時代と日本語の「危機」，三田文学，Vol.76，No.48，pp.84-131，三田文学会（1997）
- 147) (社)日本文藝家協会：「漢字を救え！文字コード問題を考えるシンポジウム」全記録（1998）
- 148) ルイ＝ジャン・カルヴェ著，矢島文夫ら訳：文字の世界史，河出書房新社（1998）
- 149) 阿辻哲次：中国漢字紀行，大修館書店（1998）
- 150) 阿辻哲次：文字，三省堂（1998）
- 151) 安永尚志：国文学研究とコンピュータ，勉誠社（1998）
- 152) 横山詔一，笹原宏之，野崎浩成，エリクロング：新聞電子メディアの漢字 朝日新聞CD-ROMによる漢字頻度表 ，国立国語研究所プロジェクト選書，1，三省堂（1998）
備考：実証的な統計的検証。
- 153) 漢字文献情報処理研究会編：電脳中国学，好文出版（1998）
備考：CD-ROM付き。
- 154) 小林芳規：図説・日本の漢字，大修館書店（1998）
- 155) 赤間世紀：自然言語・意味論・論理，共立出版（1998）
- 156) 中嶋幹起ほか編著：電脳処理西夏文字諸解対照表(稿)，不二出版（1998）
- 157) 鳥井弘之：国際標準と日本の企業・産業，ENGINEERS（1998）
- 158) 藤田，河原：国際標準が日本を包囲する，日本経済新聞社（1998）

- 159) 平井昌夫：[復刻] 國語國字問題の歴史，三元社（1998）
- 160) 平凡社編：電腦文化と漢字のゆくえ 岐路に立つ日本語，p.303，平凡社（1998）
- 161) 和田英一ほか：Unicodeは好きですか，インタラクティブエッセイ，情報処理学会誌，Vol.39，No.4，pp.321-325（1998）
- 162) [特集] 文字，ユリイカ，Vol.30，No.6，pp.64-303，青土社（1998）
- 163) どこへいく漢字，This is 読売，Vol.9，No.9，pp.108-143，読売新聞社（1998）
- 164) [特集] 文字コード入門，季刊モジカ，No.11，p.25，文字デザイン研究会（1998）
- 165) <コンピュータ国文学の展望> 情報流通のための漢字用語定義集試案：改訂増補版，第3回シンポジウムコンピュータ国文学講演集，国文学研究資料館（1998）
- 166) [特集] 文字，ユリイカ，1998年5月号，青土社（1998）
- 167) [特集] コンピュータと文字，本の窓，1998年9・10月合併号（1998）
- 168) [特集] 地球言語の時代，SClaS，1998-10-16（1998）
- 169) 日本語の文字と組版を考える会第八回公開セミナー「文字コード」（1998）
- 170) 日本語の文字と組版を考える会会報，8（1998）
- 171) 国際標準化Q&A，工業技術院標準部（1998）
- 172) [特集] 漢字という文明，国際交流，78，国際交流基金（1998）
- 173) Ken Lunde: CJKV Information Processing, O'Reilly&Associates (1999)
備考：現在最も詳細で正確な文字コード関係の解説本。邦訳は未刊。
- 174) 安岡孝一，安岡素子：文字コードの世界，東京電気大学出版局（1999）
備考：やや専門的な各国文字コード解説。
- 175) 加藤友康：WWWサーバによる日本史データベースのマルチメディア化と公開に関する研究，東京大学史料編纂所（1999）
備考：科研費報告書
- 176) 神田範明，丸山一彦著：文科系のためのコンピュータ活用入門：実践編，同文館出版（1999）
備考：ストーリー仕立てにするなど良く工夫されており、半期の教養課程教科書として有用。
- 177) 川俣晶：パソコンにおける日本語処理/文字コードハンドブック，技術評論社（1999）
備考：CD-ROM付き。詳細・正確でバランスの取れた文字コード関係の解説本。
- 178) 池田証寿編：古辞書とJIS漢字，1，北海道大学文学部言語情報学講座（1999）
備考：毎年出ている。非常に細かい作業の積み重ねが特徴。
- 179) 池田証寿編：古辞書とJIS漢字，2，北海道大学文学部言語情報学講座（1999）
- 180) 中嶋幹起編：電腦処理御製増訂清文鑑第9冊，言語文化接触に関する研究，14，不二出版（1999）
- 181) 二階堂善弘，千田大介，池田巧編：コンピュータで中国語 Win&Mac，しにか，1999年5月別冊，大修館書店（1999）

- 182) 文字鏡研究会編：パソコン悠悠漢字術 今昔文字鏡徹底活用 ， 紀伊國屋書店 (1999)
備考：今昔文字鏡の無償簡易版を付録CD-ROMに収録し、その利用法を解説したもの。Mac用フォントも収録。
- 183) 文字鏡研究会編：文字鏡研究会会報，2，文字鏡研究会 (1999)
- 184) 平尾良光，山岸良二編：古墳・貝塚・鉄器を探る，文化財を探る科学の眼，4，国土社 (1999)
備考：古墳・貝塚・鉄器を探る画像処理、地下遺構探査、酸素同位体比、微量元素分析、年輪年代など。
- 185) [特集] 漢字WAR，InterCommunication，Vol.8，No.1，pp.50-112，NTT出版 (1999)
- 186) [特集] 中国語電腦事情，しにか，Vol.8，No.4，大修館書店 (1999)
備考：このほかにも時々単発でパソコン関連記事が載る。
- 187) [特集] 電腦東洋学，人文学と情報処理，24，勉誠出版 (1999)
備考：毎号、参考になる特集多し。
- 188) [特集] 日本の漢字を考える，しにか，Vol.10，No.7，大修館書店 (1999)
- 189) 第62回研究セミナー報告，京都大学大型計算機センター研究開発部 (1999)

付2.1.4 2000年代

- 190) 影浦峽：計量情報学 図書館/言語研究への応用，丸善 (2000)
- 191) 加藤弘一：電腦社会の日本語，文春新書，94，文藝春秋 (2000)
備考：文字コード問題を俯瞰できる。正誤表が
<http://www.horagai.com/www/moji/nihon/correct.html>にある。
- 192) 漢字文献情報処理研究会編：漢字文献情報処理研究，創刊号，好文出版 (2000)
備考：漢字文献情報処理研究会の機関誌。
- 193) 漢字文献情報処理研究会編：電腦国文学，好文出版 (2000)
備考：CD-ROM付き。
- 194) 山田奨治：文化資料と画像処理，勉誠出版 (2000)
- 195) 池田証寿編：古辞書とJIS漢字，3，北海道大学文学部言語情報学講座 (2000)
- 196) 中村康夫著，国文学研究資料館編：古典研究のためのデータベース，原典講読セミナー，5，臨川書店 (2000)
備考：データベースの具体的操作法についても図入りで解説。
- 197) 東野治之編：金剛寺本遊仙窟，塙書房 (2000)
- 198) 日本学術会議歴史学研究連絡委員会，歴史情報専門委員会：シンポジウム「歴史資料におけるデータベースの標準」記録集，日本学術会議歴史学研究連絡委員会・歴史情報専門委員会 (2000)
備考：日本学術会議創立50周年記念
- 199) 文字鏡研究会編：パソコン悠悠漢字術2001 今昔文字鏡徹底活用 ， 紀伊國屋書店 (2000)
- 200) 文字鏡研究会編：文字鏡研究会会報，3，文字鏡研究会 (2000)
- 201) 電子メディア時代の文章法，國文學臨時増刊号，45巻，3号，學燈社 (2000)
備考：文章作法と新しいメディアのとの関係を論じる。

- 202) [特集] 漢字もの知りQ&A, しにか, 2000年6月号, 大修館書店 (2000)
- 203) 第60回 (平成12年前期) 全国大会記念セッションイベント報告集, 情報処理学会 (2000)
備考: 公開の場で大規模文字セットの三大プロジェクトが同座したのはこれが初めて。なお、前年3月に、東京大学史料編纂所主催「歴史学のためのウェブサイト第2回経験交流会 歴史学系データベースと文字コード」が行われており、この様子は人文学と情報処理25号に掲載。
- 204) 清水哲郎: 図解でわかる文字コードのすべて, 日本実業出版社 (2001)
- 205) トニー・グラハム著, 乾和志, 海老塚徹訳: Unicode標準入門, 翔泳社 (2001)
- 206) 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)
- 207) 杉本つとむ: 漢字百珍 日本の異体字入門, 八坂書房 (2001)
- 208) 池田証寿編: 古辞書とJIS漢字, 4, 北海道大学文学部言語情報学講座 (2001)
- 209) 府川充男, 小池和夫: 旧字旧かな入門, シリーズ日本人の手習い, 柏書房 (2001)
- 210) 文字鏡研究会編: 文字鏡研究会会報, 4, 文字鏡研究会 (2001)
- 211) [特集] 中国語電脳Q&A, しにか, 2001年5月号, 大修館書店 (2001)
- 212) [特集] 漢字おもしろ探偵団, しにか, 2001年6月号, 大修館書店 (2001)
- 213) [特集] 漢字のなりたちQ&A, しにか, 2001年12月号, 大修館書店 (2001)
- 214) フォント開発・普及委員会: 平成書体, 財団法人日本規格協会 (2002)
備考: 平成3書体のデザインコンセプト及び開発文字数一覧表等収録。
- 215) 三上喜貴: 文字符号の歴史アジア編, p377, 共立出版社 (2002)

付2.2 論文

付2.2.1 1980年代

- 216) 斎藤秀紀: 漢字コードの拡張法に対する試案, 国立国語研究所報告, 83, pp.57-103 (1985)
- 217) 斎藤秀紀: 漢字情報データベース, 国立国語研究所報告, 94 (1988)

付2.2.2 1990年代

- 218) 斎藤秀紀: 漢字コードのメタコード化の方法, 情報処理学会第46回全国大会講演資料集(1), pp.23-24 (1993)
- 219) 笹原宏之: JIS漢字と位相, 日本語学, 12巻, 8号, pp.75-85 (1993)
- 220) 斎藤秀紀: 1字体に1符号を対応させる漢字符号化の方法, 計量国語学, Vol.19, No.5, pp.233-233 (1994)
- 221) 斎藤秀紀: 大漢和辞典の検字番号に基づく構造化4バイトコードの提案, 情報処理学会論文誌, Vol.35, No.6, pp.1119-1126 (1994)
- 222) 池田証壽: 篆隸万象名義データベースについて, 国語学, 178, 武蔵野書院 (1994)

- 223) 福田雅史：諸橋大漢和辞典初版縮写版 修訂版間の差，国語学，178，武蔵野書院（1994）
- 224) 漢字字書データベースの作成とその利用（plain text版），情報処理語学文学研究会会報，15（1994）
- 225) 斎藤秀紀：構造化4バイトコードによる1字体1コード化の有効性，計量国語学会第50回全国大会講演資料集(3)，pp.201-202（1995）
- 226) 笹原宏之：JIS漢字における国字・新字・異体字の典拠 JIS漢字幽霊文字研究序説，文化女子大学紀要 人文・社会科学研究，3，pp.1-15（1995）
- 227) インターネットを利用したの学術情報交流 JIS漢字関連情報の収集とその公開（plain text版/html版），人文学と情報処理，No.8（1995）
- 228) 斎藤秀紀：漢字情報と文例情報を結合した日本語データベースの構築，情報処理学会研究会報告 96-CH-30-7，Vol.96，No.42，pp.35-40（1996）
- 229) 斎藤秀紀：JISに無い字をどうあつかうか - 具体的事例報告，人文学と情報処理，No.10，pp.22-25，勉誠社（1996）
- 230) 笹原宏之：「JIS X 0208」における音義未詳字に対する原典による同定「標準コード用漢字表（試案）」と「国土行政区画総覧」，国語学研究と資料，20，pp.1-17（1996）
- 231) 片岡裕：国際化・多国語化の基礎と実際，bit，3～4月号，共立出版（1997）
- 232) 小林龍生：文字コードについての真の国際化とは，bit，11月号，共立出版（1997）
- 233) 松岡榮志：マルチリンガル文書処理と漢字，bit，11月号，共立出版（1997）
- 234) 伊藤英俊：コンピュータから見た日本語，日本語学のみかた。，アエラムック，30（1997）
- 235) 加藤弘一：文字コード問題，この一年，本とコンピュータ，1997年夏号（1997）
- 236) 宮澤彰：漢字情報処理とUCS，情報の科学と技術，47（1997）
- 237) 斎藤秀紀：構造化4バイトコードによる多言語表現法の提案，言語処理学会第3回発表論文集，pp.185-188（1997）
- 238) 坂村健：「漢字の国」を襲う米国産文字コード，SClaS，1997年5月，朝日新聞社（1997）
- 239) 坂村健：文字コードは文化の担い手，日経パソコン，1997.11.17号（1997）
- 240) 笹原宏之：字体に生じる偶然の一致 「JIS X 0208」と他文献における字体の「暗合」と「衝突」，日本語科学，1，国立国語研究所（1997）
- 241) 芝野耕司：JIS漢字の拡張計画，標準化ジャーナル，Vol.27（1997）
- 242) 芝野耕司：文字コードの体系化と漢字文化，朝日新聞セミナー，1997年12月（1997）
- 243) 小林龍生，松岡榮志，小町祐史：マルチリンガルと文字情報処理，bit，10～12月号，共立出版（1997）
- 244) 川上進：日本語と文字コードのあいだで，本とコンピュータ，1997年夏号（1997）
- 245) 前田年昭：われ発見せり 文字コードをめぐる迷信，ユリイカ，第29巻，第14号，p.334（1997）
- 246) 島田雅彦：帝国主義者ネットを諭す，ワイアード，3月4日，pp.53-59，DDPデジタルパブリッシング（1997）

- 247) JIS漢字コード, 日経コンピュータ, 1997.6.7 (1997)
- 248) 伊藤英俊: JIS漢字コード化の裏事情, 月刊望星, 東海教育研究所 (1998)
- 249) 斎藤秀紀: 漢字の属性情報に対する符号化法の提案, 計量国語学, Vol.21, No.4, pp.131-144 (1998)
- 250) 笹原宏之, 横山詔一, 野崎浩成: 異体字に対する好みとなじみの相関関係, シンポジウム人文科学における数量的分析, 3, 文部省統計数理研究所 (1998)
- 251) 芝野耕司: JIS漢字の歴史, 第8回公開セミナー (テーマ: 文字コード) 資料, 日本語の文字と組み版を考える会 (1998)
- 252) 芝野耕司: JIS漢字批判に答えて, 第8回公開セミナー (テーマ: 文字コード) 資料, 日本語の文字と組み版を考える会 (1998)
- 253) 小宮山博史: 美しい書体はいらないのか 「六万四千漢字」発表会で考えたこと, 技術と人間, 1998年8・9月号 (1998)
- 254) 松岡榮志: 「漢字の危機」は杞憂にすぎない, 中央公論, Vol.113, No.4, pp.272-282, 中央公論社 (1998)
- 255) 松岡榮志: 電脳時代でも漢字は滅びない 「ユニコード・ファシズム論」の誤解, 日本の論点'99, 文藝春秋 (1998)
- 256) 松岡榮志, 伊藤英俊, マルティン・デュールスト: 「鼎談」コンピュータ時代の漢字, 国際交流, 78号, 国際交流基金 (1998)
- 257) 青木, 斉藤ほか: 国際標準とその必要性, 標準化と品質管理, 51巻, 10号 (1998)
- 258) 前田年昭: 文字文化を否定する「GT明朝プロジェクト」 歴史を捨てた「文字文化の継承」を批判する, 技術と人間, 1998年8・9月号 (1998)
- 259) 池田証壽: 新旧分離字と新旧包摂字 JIS漢字の所拠漢字表と包摂規準, 訓点語と訓点資料記念特輯, 訓点語学会 (1998)
- 260) 豊島正之: JIS漢字批判の基礎知識, 第8回公開セミナー (テーマ: 文字コード) 資料, 日本語の文字と組み版を考える会 (1998)
- 261) 豊島正之: 「原文に忠実な翻刻」をめぐる, 第8回公開セミナー (テーマ: 文字コード) 資料, 日本語の文字と組み版を考える会 (1998)
- 262) 鈴木一誌: 批評を封殺する透明性 「六万四千漢字」発表会来賓挨拶 15分版, 技術と人間, 1998年8・9月号 (1998)
- 263) 鈴木一誌: 「テキストは文字の集合」か? 「六万四千漢字」発表会来賓挨拶 30分版, 小特集「文字コード問題」とは何か, 文藝, 1998年冬号, 河出書房新社 (1998)
- 264) 鈴木一誌: 「六万四千漢字」への批評, あるいは問いかけ, 日本語の文字と組み版を考える会 (1998)
- 265) Unicodeを巡る放談会, ASCII, Vol.22, No.12, pp.477-481 (1998)
- 266) Windows98と日本語の問題を考える, ASCII, Vol.22, No.5 (1998)
- 267) 漢字文化は危なくない ユニコード国際カンファレンスから, ASCII, Vol.22, No.6 (1998)
- 268) GT明朝, プリンティングインフォメーション, 1998年3月号 (1998)
- 269) 加藤弘一: 多言語テキスト処理はどこまで可能か, 季刊本とコンピュータ, 春号, トランスアート (1999)

- 270) よしおかひろたか：文字コードが変わる！，日経ソフトウェア，Vol.2，No.2，pp.106-111，日経BP社（1999）
- 271) 古家時雄：今昔文字鏡の開発と普及の経緯，人文学と情報処理，24，勉誠出版（1999）
- 272) 前寺正彦：文字鏡フォントをUnicodeにマッピングしたエディタについて，月刊TransTECH，2000年1月号，翔泳社（1999）
- 273) 谷田貝常夫，谷本玲大：文字鏡の基本方針，人文学と情報処理，25，勉誠出版（1999）
- 274) 谷本玲大：和漢比較文学の典拠論におけるインターネット活用，人文学と情報処理，24，勉誠出版（1999）

付2.2.3 2000年代

- 275) 谷田貝常夫，古家時雄，谷本玲大：今昔文字鏡の概要，全国大会記念セッションイベント報告集第60回，（平成12年前期），情報処理学会（2000）
- 276) 谷本玲大：大規模漢字システム，漢字文献情報処理研究，1，好文出版（2000）
- 277) 横山詔一，笹原宏之，エリクロング，谷本玲大：研究ノート 新聞漢字調査の現状と将来，日本語科学，9，国書刊行会（2001）

付2.3 報告書等

- 278) 朝日新聞東京本社印刷局編：活字使用度数調査・熟語使用度数調査，朝日新聞社（1950）
- 279) 日本新聞協会：新聞活字字体統一に関する資料，日本新聞協会（1959）
- 280) 国立国語研究所：電子計算機による国語研究，国立国語研究所報告，31，国立国語研究所（1968）
備考：刊行年を見ると、その進取性は驚異的。
- 281) 杉本つとむ編：異体字研究資料集成（1期・2期），雄山閣出版（1973～1995）
- 282) 国立国語研究所：現代新聞の漢字，国立国語研究所報告，56，国立国語研究所（1976）
- 283) 国文学研究資料館研究情報部情報処理室編：国文学研究資料館におけるコンピュータ及び漢字システム，国文学研究資料館報告，1，国文学研究資料館（1978）
- 284) 国立国語研究所：雑誌用語の変遷，国立国語研究所報告，89，国立国語研究所（1987）
備考：国立国語研の雑誌語彙研究のこの時点での集成整理版
- 285) 国語関係訓令・告示集，文化庁（1991）
- 286) 国文学研究資料館：国文学データベース研究集会報，4，国文学研究資料館（1995）
備考：翌年から「シンポジウムコンピュータ国文学」に改称し発展的に継続。
- 287) 国語に関する世論調査，文化庁（1995）
- 288) 第20期国語審議会：新しい時代に応じた国語施策について，文化庁（1995）

- 289) 日本工業規格(案) JIS X0208-1996 付属書1・2 公開レビュー資料, INSTAC (1995)
備考: 3500円で実費頒布されたJISの原案資料。
- 290) 第20期国語審議会報告書, 文化庁 (1996)
- 291) 第21期国語審議会第1回議事要録, 文化庁 (1996)
- 292) 国語審議会答申・建議集, 文化庁 (1996)
- 293) JIS情報交換用漢文字符号 (第1分冊), 漢字字体関係参考資料集, 文化庁 (1996)
- 294) 文化庁文化部国語課編: 漢字字体資料集, 国語施策沿革資料, 11-12, 文化庁 (1996~1997)
- 295) 国立国語研究所: 現代雑誌九十種の用字用語全語彙・表記, 国立国語研究所言語処理データ集, 7, 三省堂 (1997)
備考: FD版
- 296) 文化庁文化部国語課編: 漢字出現頻度数調査, 漢字字体関係参考資料集, 文化庁 (1997)
- 297) JIS情報交換用漢文字符号 (第2分冊), 漢字字体関係参考資料集, 文化庁 (1997)
- 298) 字体・字形差一覽, 漢字字体関係参考資料集, 文化庁 (1997)
- 299) 漢字出現頻度数調査, 文化庁 (1997)
- 300) 国語に関する世論調査, 文化庁 (1997)
- 301) 国文学研究資料館: 第3回シンポジウムコンピュータ国文学講演集, 国文学研究資料館 (1998)
- 302) 第21期国語審議会: 新しい時代に応じた国語施策について, 審議経過報告, 文化庁 (1998)
- 303) 平成10年度国語施策懇談会, 文化庁 (1998)
- 304) 第21期国語審議会報告書, 文化庁 (1998)
- 305) 漢字字体資料集諸案集成1, 国語施策沿革資料, 11, 文化庁 (1998)
- 306) 漢字字体資料集諸案集成2, 国語施策沿革資料, 12, 文化庁 (1998)
- 307) 多言語情報処理環境技術成果報告書, 1997年度, CICC (1998)
- 308) 沖縄の歴史情報研究, 沖縄の歴史情報研究事務局 (1998)
備考: 科研費報告書
- 309) 国文学研究資料館: 第4回シンポジウムコンピュータ国文学講演集, 国文学研究資料館 (1999)
- 310) 谷田貝常夫, 谷本玲大: ネットワーク・コラボレーションによる大規模フォントコンテンツの移植と応用 情報基盤整備の試み, 人文科学とコンピュータ研究報告, 99-CH-44-1, 情報処理学会 (1999)
- 311) 文化庁文化部国語課編: 明朝体活字字形一覽1820年~1946年(上・下), 漢字字体関係参考資料集, 大蔵省印刷局 (1999)

付2.4 規格・標準

- 312) JIS X 0201:1997 7ビット及び8ビットの情報交換用符号化文字集合
- 313) JIS X 0202:1998 情報技術 文字符号の構造及び拡張法
- 314) JIS X 0208:1997 7ビット及び8ビットの2バイト情報交換用符号化文字集合
備考: 古辞書にまで遡って用例を調べた付属書7はJIS規格としても異例。

- 315) JIS X 0212:1990 情報交換用漢字符号 補助漢字
- 316) JIS X 0213:2000 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
- 317) JIS X 0221-1:2001 国際符号化文字集合 (UCS) 第1部: 体系及び基本多言語面
- 318) 標準情報 (TR) TR X 0047:2001 XMLによる画像参照交換方式
- 319) ISO/IEC 646:1991 Information technology — ISO 7-bit coded character set for information interchange
- 320) ISO 1073-2:1976 Alphanumeric character sets for optical recognition — Part 2: Character set OCR-B — Shapes and dimensions of the printed image
- 321) ISO/IEC 2022:1994 Information technology — Character code structure and extension techniques
- 322) ISO/IEC 2022:1994/Cor 1:1999
- 323) ISO 2033:1983 Information processing — Coding of machine readable characters (MICR and OCR)
- 324) ISO 2375:1985 Data processing — Procedure for registration of escape sequences
- 325) FCD 2375 Information technology — Procedure for registration of escape sequences and coded character sets (SC2N3592) (2002)
- 326) ISO/IEC 4873:1991 Information technology — ISO 8-bit code for information interchange — Structure and rules for implementation
- 327) ISO/IEC 6429:1992 Information technology — Control functions for coded character sets
- 328) ISO 6936:1988 Information processing — Conversion between the two coded character sets of ISO 646 and ISO 6937-2 and the CCITT international telegraph alphabet No. 2 (ITA 2)
- 329) ISO/IEC 6937:2001 Information technology — Coded graphic character set for text communication — Latin alphabet
- 330) ISO/IEC 7350:1991 Information technology — Registration of repertoires of graphic characters from ISO/IEC 10367
- 331) ISO/IEC 8859-1:1998 Information technology — 8-bit single-byte coded graphic character sets — Part 1: Latin alphabet No. 1
- 332) ISO/IEC 8859-2:1999 Information technology— 8-bit single byte coded graphic character sets — Part 2: Latin alphabet No. 2
- 333) ISO/IEC 8859-3:1999 Information technology— 8-bit single-byte coded graphic character sets — Part 3: Latin alphabet No. 3
- 334) ISO/IEC 8859-4:1998 Information technology — 8-bit single-byte coded graphic character sets — Part 4: Latin alphabet No. 4
- 335) ISO/IEC 8859-5:1999 Information technology — 8-bit single-byte coded graphic character sets — Part 5: Latin/Cyrillic alphabet
- 336) ISO/IEC 8859-6:1999 Information technology — 8-bit single-byte coded graphic character sets — Part 6: Latin/Arabic alphabet
- 337) ISO 8859-7:1987 Information processing — 8-bit single-byte coded graphic character sets — Part 7: Latin/Greek alphabet

- 338) ISO/IEC 8859-8:1999 Information technology — 8-bit single-byte coded graphic character sets — Part 8: Latin/Hebrew alphabet
- 339) ISO/IEC 8859-9:1999 Information technology — 8-bit single-byte coded graphic character sets — Part 9: Latin alphabet No. 5
- 340) ISO/IEC 8859-10:1998 Information technology — 8-bit single-byte coded graphic character sets — Part 10: Latin alphabet No. 6
- 341) ISO/IEC 8859-11:2001 Information technology — 8-bit single-byte coded graphic character sets — Part 11: Latin/Thai alphabet
- 342) ISO/IEC 8859-13:1998 Information technology — 8-bit single-byte coded graphic character sets — Part 13: Latin alphabet No. 7
- 343) ISO/IEC 8859-14:1998 Information technology — 8-bit single-byte coded graphic character sets — Part 14: Latin alphabet No. 8 (Celtic)
- 344) ISO/IEC 8859-15:1999 Information technology — 8-bit single-byte coded graphic character sets — Part 15: Latin alphabet No. 9
- 345) ISO/IEC 8859-16:2001 Information technology — 8-bit single-byte coded graphic character sets — Part 16: Latin alphabet No. 10
- 346) ISO 9036:1987 Information processing — Arabic 7-bit coded character set for information interchange
- 347) ISO/IEC 10367:1991 Information technology — Standardized coded graphic character sets for use in 8-bit codes
- 348) ISO/IEC 10367:1991/Cor 1:2001
- 349) ISO/IEC 10538:1991 Information technology — Control functions for text communication
- 350) ISO/IEC 10646-1:2000 Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane
- 351) Draft ISO/IEC 10646-1:2000/Amd.1, Mathematical symbols and other characters (SC2N3588) (2001)
- 352) Draft ISO/IEC 10646-1:2000/Amd.2, Limbu, Tai Le, Yijing and other characters (SC2N3584) (2001)
- 353) ISO/IEC 10646-2:2001 Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 2: Supplementary Planes
- 354) Draft ISO/IEC 10646-2:2001/Amd.2, Aegean, Ugaritic, and other characters (SC2N3585) (2001)
- 355) ISO/IEC TR 15285:1998 Information technology — An operational model for characters and glyphs
- 356) IPSJ-TS 0005:2002 符号化文字基本集合
- 357) The Unicode Consortium: The Unicode Standard, Version 3.0, Addison-Wesley (2000)

付2.5 ホームページ

- 358) Jim Breen: ジム・ブリーンの日本ページ,
<http://www.dgs.monash.edu.au/~jwb/japanese.html>

- 359) Ken Lunde: Ken Lunde's Home Page (小林劍のホームページ),
<http://www.ora.com/people/authors/lunde/>
- 360) ライン・ラボ: 東亜文字処理ライン・ラボのWebページ,
<http://www.linelabo.com/>
- 361) 安岡孝一: <http://www.kudpc.kyoto-u.ac.jp/~yasuoka/>
- 362) 国文学研究資料館: <http://www.nijl.ac.jp/>
- 363) 山本太郎: <http://www.kt.rim.or.jp/~tyamamot/charcode/>
- 364) 大原望: 和製漢字の辞典, <http://member.nifty.ne.jp/TAB01645/ohara/>
- 365) 池田証壽: <http://member.nifty.ne.jp/shikeda/>
- 366) 日本語の文字と組版を考える会: <http://www.pot.co.jp/moji/>
- 367) 国語審議会: http://www.mext.go.jp/b_menu/shingi/12/kokugo/index.htm
- 368) 加藤弘一: ほら貝の参考文献一覧,
<http://www.horagai.com/www/moji/bunken.htm>
- 369) 青空文庫, <http://www.aozora.gr.jp/>
- 370) 歴史フォーラム, <http://www.nifty.ne.jp/forum/frekishi/index.htm>
- 371) 勝村哲也ほか: e漢字, <http://nohara.u-shimane.ac.jp/ekanji/>
- 372) 京都大学人文科学研究所附属漢字情報研究センター:
<http://www.kanji.zinbun.kyoto-u.ac.jp/>
- 373) 文字鏡Net (今昔文字鏡), <http://www.mojikyo.gr.jp/>
- 374) LightHouse Home Page, <http://www.l-h.co.jp/>
- 375) Unicode Consortium: <http://www.unicode.org/>

付2.6 ガイドブック・マニュアル類

- 376) 人名用漢字と誤字俗字関係通達の解説 (新版), 日本加除出版 (1991)
- 377) 標準校正必携 第7版, 日本エディタースクール (1995)
- 378) 吉田智子ほか: マルチリンガル環境の実現 X Window/Wnn/Mule/WWWブラウザでの多国語環境, プレンティスホール出版 (1996)
- 379) Windows NT漢字処理技術協議会: Windows NT拡張漢字処理仕様書第2.11版 (1998)
- 380) 川俣晶 + XKP普及部会: Windows NT日本語処理ガイドブック (1998)
- 381) 日本加除出版編集部編: すぐわかる正字・誤字俗字一覧, 日本加除出版 (2000)

付2.7 新聞記事

- 382) 斎藤秀紀: 文化の継承は4バイトコードで, YOMIURI ON-LINE (1998)
備考: <http://www.yomiuri.co.jp/bitbybit/bbb02/832501.htm>
- 383) 漢字不足に作家ら危機感, 朝日新聞, 1998-2-5, 朝刊 (1998)
- 384) 漢字が足りない!, 読売新聞, 1998-2-11, 朝刊 (1998)
- 385) 漢字が使いにくくなる?, 日本経済新聞, 1998-2-21, 朝刊 (1998)
- 386) 漢字コード「国際規格化」進む, 読売新聞, 1998-4-15, 朝刊 (1998)
- 387) 漢字不足解決へ新規格, 読売新聞, 1998-12-23, 朝刊 (1998)
- 388) 書き誤りも字体変える力に, 揺れる漢字 (上), 朝日新聞, 1999-7-8, 朝刊 (1999)

- 389) 文字コードへの懸念と挑戦, 揺れる漢字 (中), 朝日新聞, 1999-7-9, 朝刊 (1999)
- 390) 手書き文字 活字との違いは当然, 揺れる漢字 (下), 朝日新聞, 1999-7-10, 朝刊 (1999)
- 391) 小学校にて, にゅうすらうんじ 漢字の未来の物語 1, 朝日新聞, 2000-9-21, 夕刊 (2000)
- 392) 鉛筆の持ち方, にゅうすらうんじ 漢字の未来の物語 2, 朝日新聞, 2000-9-22, 夕刊 (2000)
- 393) 砂金探し, にゅうすらうんじ 漢字の未来の物語 3, 朝日新聞, 2000-9-28, 夕刊 (2000)
- 394) 四庫全書, にゅうすらうんじ 漢字の未来の物語 4, 朝日新聞, 2000-9-29, 夕刊 (2000)
- 395) 円マーク, にゅうすらうんじ 漢字の未来の物語 5, 朝日新聞, 2000-10-5, 夕刊 (2000)
- 396) 「漢字の国」のIT革命, 海外文化 中国, 朝日新聞, 2001-1-5, 夕刊 (2001)
- 397) 何が起きているか 6, ニッポンのことは第1部, 朝日新聞, 2001-1-15, 朝刊 (2001)
- 398) 藤 戸籍入力ミス, 読売新聞, 2001-7-14, 夕刊 (2001)
- 399) 漢字文化圏の未来 1, ニッポンのことは第4部, 朝日新聞, 2001-12-16, 朝刊 (2001)
- 400) 漢字文化圏の未来 2, ニッポンのことは第4部, 朝日新聞, 2001-12-17, 朝刊 (2001)
- 401) 漢字文化圏の未来 3, ニッポンのことは第4部, 朝日新聞, 2001-12-18, 朝刊 (2001)
- 402) 漢字文化圏の未来 4, ニッポンのことは第4部, 朝日新聞, 2001-12-23, 朝刊 (2001)
- 403) 漢字文化圏の未来 5, ニッポンのことは第4部, 朝日新聞, 2001-12-24, 朝刊 (2001)
- 404) 身近な漢字紙面に登場 日本新聞協会漢字使用制限を緩和, 読売新聞, 2002-1-26, 朝刊 (2002)

付2.8 字書・辞典

- 405) 干禄字書, 官版
- 406) 玉篇及原本零卷, リプリント版, 台湾: 藍燈
備考: 影印の合本。摛宋陳彭年等奉勅重修大庑益会玉篇影印、附:日本現存旧鈔本零卷。
- 407) 康熙字典, 殿版
備考: 無窮會蔵
- 408) 康熙字典, 安永本
備考: 文字鏡研究会蔵
- 409) 遊仙窟, 慶安五年本 (1652)
- 410) 伏見冲敬編: 書道大字典, 角川書店 (1974)
- 411) 龍龕手鑑, 覆刻日本古典全集, 現代思潮社 (1977)

- 412) 武部良明：日本語の表記，角川小辞典，29，角川書店（1979）
- 413) 児玉幸多編：くずし字解読辞典 普及版，近藤出版社（1979（原1975））
- 414) 江守賢治：楷行草筆順字体字典，三省堂（1983）
- 415) 山田勝美監修，難字大鑑編集委員会編：異体字解読字典，柏書房（1986）
備考：原題「難字大鑑」
- 416) 篠崎晃雄：実用難読奇姓辞典，増補新版，日本加除出版（1986）
- 417) 漢語大字典編纂委員会編：漢語大字典，四川辞書出版社・湖北辞書出版社（1986～1990）
- 418) 有賀要延編：難字・異体字典，国書刊行会（1987）
備考：原題「仏教難字大字典」（1986）
- 419) 漢語大詞典編輯委員会，漢語大詞典編纂処編纂：漢語大詞典，上海：漢語大詞典出版社（1988～1994）
- 420) 飛田良文監修，菅原義三編：国字の字典，東京堂出版（1990）
- 421) 日外アソシエーツ編集部：漢字異体字典，日外アソシエーツ、紀伊國屋書店（1994）
備考：Unicodeが国内で使用される以前の刊行。
- 422) 空海編：篆隸萬象名義（影印版），中華書局（1995）
備考：崇文叢書第1輯の縮影印
- 423) 高田竹山監修，法書會編集部纂：五體字類，西東書房（1996）
備考：原1916
- 424) 芝野耕司編著：JIS漢字字典，日本規格協会（1997）
備考：JIS第一水準、第二水準漢字の字典。いわゆるワープロ漢字字典としての使用以外に、付載されるNTT電話帳の人名の特殊な「よみ」や国土行政資料による地名の「よみ」情報は和訓資料としての価値も高い。規格の正式版にある付属書7も再録されている、
- 425) 中國漢語大詞典編纂處，（日本）禪文化研究所編纂：多功能漢語大詞典索引，上海：漢語大詞典出版社（1997）
- 426) 江守賢治：解説字体辞典 普及版，三省堂（1998）
備考：原1986
- 427) 組版工学研究会：和文電子活字見本帳，朗文堂（1999）
- 428) 田中慶太郎編：支那文を読む爲の漢字典，研文出版（1999）
備考：民国の「学生字典」文求堂蔵版（1940）を翻訳したもの。
- 429) ユニコード漢字情報辞典編集委員会編：ユニコード漢字情報辞典，三省堂（2000）
備考：ISO委員である松岡榮志氏が中心となって編纂。現在最も完備したUnicodeの漢字字典。
- 430) 日本タイポグラフィ協会編：日本のタイプフェイス — database，インプレス（2000）
- 431) 河野六郎，千野栄一，西田龍雄編著：世界文字辞典，言語学大辞典別巻，三省堂（2001）
- 432) 国立国会図書館編：国立国会図書館著者名典拠録 2000年版，紀伊國屋書店（2001）
- 433) 内藤記念くすり博物館：大同薬室文庫蔵書目録 附：館蔵和漢古典籍目録，内藤記念くすり博物館（2001）

- 434) 内藤記念くすり博物館：薬物名出典総索引 — 江戸・明治初期の薬物検索のための —, 内藤記念くすり博物館 (2001)
- 435) フォント開発・普及委員会：平成書体別冊, 財団法人日本規格協会 (2002)
備考：平成3書体の開発全文字 (基本ウエイト) を収録。

付2.9 パソコンソフト等

- 436) 天台宗典編纂所：天台電子佛典CD1, 天台宗典編纂所 (1997)
備考：仏典のテキストデータを収録したCD-ROM刊行物。併載されている漢文文献の索引作成用スクリプトは他の文書にも適用可能。
- 437) JEPA World Font CD '99, 日本電子出版協会 (1998)
備考：CD-ROM
- 438) 小口雅史編：日本古代土地経営関係史料集成 — 東大寺領・北陸編, デジタル古文書集, 同成社 (1999)
- 439) 文字鏡研究会編：今昔文字鏡 単漢字8万字版, 紀伊國屋書店 (1999)
- 440) 文字鏡研究会編：今昔文字鏡, 紀伊國屋書店 (1999)
備考：第二版
- 441) 旧かな・歴史的假名遣の日本語変換ソフト「契沖」, 有限会社申申閣 (2000)
備考：旧かな旧漢字変換を可能とするIME
- 442) エーアイ・ネット：今昔文字鏡 筆文字鏡 卷之一 楷書, 紀伊國屋書店 (2001)
- 443) 文字鏡研究会編：今昔文字鏡 単漢字10万字版, 紀伊國屋書店 (2001)

付録3 用語

文字や文字コードに関する情報処理用語は、その使用用途や分野・グループによっていろいろ微妙な差があり議論を混乱させる一因になっている。第1ステージでは、この点を考慮し、議論のための用語についても検討の対象とし、用語集を用意した。

第2ステージの委員会でも、用語に関する基本の考え方は同じである。本報告書でも、原則として第1ステージと同じ用語を用いている。

以下に、第1ステージで作成した用語集から、特に基本的と思われるものを再録しておく。

付3.1 本委員会として厳密な定義を定めずに用いる用語

- 文字
- 異体字
- 文字群
- script
- character
- repertoire
- unification

付3.2 本委員会として定義を試みる用語、及びその定義

- 符号化文字 (coded character)
文字の抽象的な情報・論理・意味などの概念で図形的な形を持たない。文字の符号を付与する対象で、処理・蓄積・検索などの対象となる。(ISO/IEC TR 15285。)
- 字体 (glyph)
文字の抽象的な形 (骨格) の概念で、文字の骨組みなどともいわれ、具体的に視覚化することは不可能である。(ISO/IEC TR15285、国語審議会資料などから。) 本委員会では、漢字部首も符号化文字に含まれる、という立場から、部首にも“字体 (glyph)”が存在すると考える。
- 字形 (glyph image)
印字・表示などの手段によってグリフ表現を表示 (presentation) することによって得られる“glyph”の可視化表現で、必然的に何らかの書体によってしか表示できない、一般的な意味で文字出力と言われるもの、あるいは、印字・表示・転写・手書きなどによって可視化された結果の文字。
- 書体 (typeface)
表示などに使用するため、統一的なコンセプトに基づいて作成された一組の文字などの意匠デザイン。英語で“typeface”はTimes Romanなど特定のデザインを指すことが多いが、日本語で“書体”は、明朝体などのデザイン類型を指すことが多く、日本語と英語の間に微妙な差がある。
- フォント (font)
ある書体でデザインされた字形の集合。ときに、この定義の“フォント”の中の特定の“glyph”に対応する部分だけを、その“glyph”の“フォン

ト” という様な使われ方をされることも多い。これはこの用語の誤用であるが、実質問題が起きるような使われ方をすることは少ないのでこれも容認する。

- Glyph variation
同一の “ coded character ” に対応する異なる “ glyph ” の集合。
- 異形字
“ 字体 ” 、 “ 字形 ” が異なるが、意味や読みなどが同じ漢字で、文中でそれらを入れ替えても善意の第三者にとっては同じ意味の文になる漢字。従来の符号化文字集合の枠組みの中では、包摂若しくはunifyされることもある。
- 用字
一つ以上の言語の表記の方法で使用する図形文字の集合 (JIS X 0221-1:2001。) 本報告書では、同じ意味で “ 文字群 ” という用語を用いる場合がある。

付録4 委員会配布資料一覧

付4.1 第1回委員会

資料1-0	議事次第
資料1-1	委員名簿
資料1-2	準備委員会報告書
資料1-3	文字コード標準化動向
資料1-3-1	国内標準化の動向
資料1-3-2	国語審議会資料
資料1-3-3	国際標準化の動向
資料1-4	委員会の審議内容について

付4.2 第2回委員会

資料2-1	議事次第
資料2-2	第1回委員会議事録
資料2-3	委員名簿
資料2-4	第1ステージ報告書
資料2-5	漢字標準化の進め方

付4.3 第3回委員会

資料3-1	議事次第
資料3-2	第2回委員会議事録
資料3-3	委員名簿
資料3-4	宿題解答
資料3-4-1	委員回答
資料3-4-2	委員回答
資料3-4-3	委員回答
資料3-4-4	委員回答

付4.4 第4回委員会

資料4-1	議事次第
資料4-2	第3回委員会議事録
資料4-3	事前準備資料
資料4-4	異形字問題のきっかけ
資料4-5	委員意見
資料4-6	委員意見
資料4-7	委員意見
資料4-8	委員意見
資料4-9	委員意見
資料4-10	「文字集合」、「文字コード」、「エンコーディング」さらに「実装」(解説)

付4.5 第5回委員会

資料5-1	議事次第
資料5-2	第4回委員会議事録
資料5-3	IRG香港会議報告
資料5-4	前回委員意見のまとめ
資料5-5	提案制度(案)
資料5-6	異形字問題にはいる前に
資料5-7	委員名簿

付4.6 第6回委員会

資料6-1	議事次第
資料6-2	第5回委員会議事録
資料6-3	中間報告書
資料6-3a	意見1
資料6-3b	意見2
資料6-4	異形字に関する宿題
資料6-4a	委員意見
資料6-4b	委員意見
資料6-4c	委員意見
資料6-4d	委員意見
資料6-4e	委員意見
資料6-4f	委員意見
資料6-5a	文字の要・不要の議論
資料6-5b	字体の揺らぎ範囲の異形字、包摂基準との問題
資料6-5c	異形字問題 (その2、古典の文字など)
資料6-5d	その他

付4.7 第7回委員会

資料7-1	議事次第
資料7-2	第6回委員会議事録
資料7-3a	SC2総会報告
資料7-3b	IRG会議情報
資料7-4-1	異形字WG議事録
資料7-4-1a	第1回
資料7-4-1b	第2回
資料7-4-2	異形字に関するまとめ
資料7-4-3	異形字アーキテクチャ提案
資料7-4-3a	委員提案
資料7-4-3b	委員提案
資料7-4-3c	委員提案
資料7-4-3d	委員提案
資料7-4-3e	委員提案

- 資料7-4-4 異形字標準化の進め方など
- 資料7-4-4a 委員意見
- 資料7-4-4b 委員意見
- 資料7-4-4c 委員意見
- 資料7-5 標準化セッションプログラム

付4.8 第8回委員会

- 資料8-1 議事次第
- 資料8-2a 第6回委員会議事録公開版
- 資料8-2b 第7回委員会議事録
- 資料8-2c 異形字WG議事録
- 資料8-2d 異形字WG議事録
- 資料8-2e 異形字アーキテクチャWG活動報告
- 資料8-3 IRG会議報告
- 資料8-4a 異形字標準化の全体像
- 資料8-4b 枝番方式
- 資料8-4c フォント方式
- 資料8-5a 異形字標準化の効果
- 資料8-5b 国際整合性の課題
- 資料8-6 最終報告書目次案
- 資料8-7 汎用電子情報交換環境整備プログラム

付4.9 第9回委員会

- 資料9-1 議事次第
- 資料9-2a 第8回委員会議事録（案）
- 資料9-2b 第5回委員会議事録公開版
- 資料9-2c 第7回委員会議事録公開版
- 資料9-2d 第8回委員会議事録公開版
- 資料9-2e 異形字WG（2月8日）議事録
- 資料9-2f 異形字WG（3月8日）議事録
- 資料9-3 情報処理学会試行標準制度について
- 資料9-4 文字コード標準体系専門委員会報告書（案）

付録5 委員会議事録および関連報告書

本委員会の議事録および関連報告書は、本報告書と同様に、インターネット上で公開されており、情報規格調査会のホームページから参照できる。以下にその参照アドレス (URL) を示す。

付5.1 委員会中間報告書

中間報告書 <http://www.itscj.ipsj.or.jp/domestic/mojicode/index.html>

付5.2 委員会議事録

第1回委員会 <http://www.itscj.ipsj.or.jp/domestic/mojicode/2giji1.htm>

第2回委員会 <http://www.itscj.ipsj.or.jp/domestic/mojicode/2giji2.htm>

第3回委員会 <http://www.itscj.ipsj.or.jp/domestic/mojicode/2giji3.htm>

第4回委員会 <http://www.itscj.ipsj.or.jp/domestic/mojicode/2giji4.htm>

第5回委員会 <http://www.itscj.ipsj.or.jp/domestic/mojicode/2giji5.htm>

第6回委員会 <http://www.itscj.ipsj.or.jp/domestic/mojicode/2giji6.htm>

第7回委員会 <http://www.itscj.ipsj.or.jp/domestic/mojicode/2giji7.htm>

第8回委員会 <http://www.itscj.ipsj.or.jp/domestic/mojicode/2giji8.htm>

第9回委員会 <http://www.itscj.ipsj.or.jp/domestic/mojicode/2giji9.htm>

付5.3 準備委員会報告書

報告書 <http://www.itscj.ipsj.or.jp/domestic/mojicode/001218a.pdf>

付5.4 準備委員会議事録

第1回準備委員会

<http://www.itscj.ipsj.or.jp/domestic/mojicode/jgiji1-991220.htm>

第2回準備委員会

<http://www.itscj.ipsj.or.jp/domestic/mojicode/jgiji1-991220.htm>

第3回準備委員会

<http://www.itscj.ipsj.or.jp/domestic/mojicode/jgiji1-991220.htm>

第4回準備委員会

<http://www.itscj.ipsj.or.jp/domestic/mojicode/jgiji1-991220.htm>

第5回準備委員会

<http://www.itscj.ipsj.or.jp/domestic/mojicode/jgiji1-991220.htm>

第6回準備委員会

<http://www.itscj.ipsj.or.jp/domestic/mojicode/jgiji1-991220.htm>

第7回準備委員会

<http://www.itscj.ipsj.or.jp/domestic/mojicode/jgiji1-991220.htm>

第8回準備委員会

<http://www.itscj.ipsj.or.jp/domestic/mojicode/jgiji1-991220.htm>

付5.5 第1ステージ委員会報告書

報告書 <http://www.itscj.ipsj.or.jp/domestic/mojicode/houkoku922.pdf>

付録6 委員会名簿

付6.1 文字コード標準体系専門委員会

委員長	石崎俊	慶應義塾大学環境情報学部
幹事	小林龍生	株式会社ジャストシステムデジタル文化研究所
幹事	佐藤敬幸	財団法人国際情報化協力センター国際情報化研究所
委員	伊藤英俊	日本電気株式会社NECソリューションズインターネットソフトウェア事業部フォントソリューショングループ
委員	北沢誓四郎	日本経済新聞社出版局編集部編集委員
委員	小池和夫	印刷史研究会
委員	斎藤秀紀	独立行政法人国立国語研究所名誉所員
委員	笹原宏之	独立行政法人国立国語研究所研究開発部門
委員	島倉達郎	株式会社東芝デジタルメディアネットワーク社技術・品質統括部
委員	関口正裕	富士通株式会社ソフトウェア事業本部
委員	千田昇一	日本電信電話株式会社情報流通プラットフォーム研究所アプリケーションプラットフォームプロジェクト
委員	高野郁子	株式会社三省堂出版局デジタル情報出版部
委員	田代秀一	独立行政法人産業技術総合研究所情報処理研究部門
委員	土居慶司	株式会社日立製作所情報コンピュータグループコンピュータ営業本部
委員	棟上昭男	東京工科大学メディア学部
委員	内藤勝	NTT番号情報株式会社情報開発部基本情報第1担当
委員	秦恒平	社団法人日本ペンクラブ
委員	古家時雄	株式会社エーアイ・ネット
委員	松岡榮志	東京学芸大学教育学部
委員	Martin J. Dürst	慶應義塾大学SFC研究所政策・メディア研究科
委員	宮澤彰	国立情報学研究所実証研究センター
委員	安岡孝一	京都大学人文科学研究所附属漢字情報研究センター
委員	山本容士	社団法人共同通信社情報システム局通信部
オブザーバ	氏原基余司	文化庁文化部国語課
オブザーバ	岡本好史	総務省行政管理局行政情報システム企画課
オブザーバ	小熊善之	社団法人日本文藝家協会書記局
オブザーバ	木戸達雄	経済産業省産業技術環境局標準課情報電気標準化推進室
オブザーバ	関達雄	財団法人日本規格協会情報技術標準化研究センター

付6.2 符号化文字集合検討準備委員会

委員長	小林龍生	有限会社スコレックス
幹事	佐藤敬幸	財団法人国際情報化協力センター国際情報化研究所
幹事	芳賀進	富士ゼロックス株式会社I&U開発部
委員	石崎俊	慶應義塾大学環境情報学部
委員	伊藤龍彦	日本ユニシス株式会社W2Kテクノロジーセンター ESソフトウェア室 (A10-S)
委員	織田哲治	日本アイ・ビー・エム株式会社ソフトウェア技術 推進 (LAB-SA4)
委員	島倉達郎	株式会社東芝デジタルメディアネットワーク社 技術・品質統括部
委員	関口正裕	富士通株式会社ソフトウェア事業本部
委員	千田昇一	日本電信電話株式会社情報流通プラットフォーム 研究所アプリケーションプラットフォーム プロジェクト
委員	永井裕司	経済産業省産業技術環境局標準課情報電気標準 化推進室
委員	中出秀樹	株式会社三菱総合研究所システム政策研究セン ター情報通信政策部
委員	宮澤彰	国立情報学研究所実証研究センター
委員	山本知	株式会社日立製作所システムソリューショング ループオープンプラットフォームソリューショ ン本部Windowsソリューションセンタ