

国立国語研究所学術情報リポジトリ

漢字コードの拡張法に対する試案

メタデータ	言語: Japanese 出版者: 公開日: 2017-03-31 キーワード (Ja): キーワード (En): 作成者: 斎藤, 秀紀, SAITO, Hidenori メールアドレス: 所属:
URL	https://doi.org/10.15084/00001097

漢字コードの拡張法に対する試案

斎藤 秀紀

1. はじめに

国立国語研究所（以下国研）では、昭和24年の「現代新聞の用語の一例」から、昭和59年「高等教科書の語彙調査」にいたるまで、各種の用語用字調査を行ってきた。用語用字調査の本格的作業は、昭和28年に行なわれた「婦人雑誌の用語」が最初であるが、以後「総合雑誌の用語」、「現代雑誌九十種の用語用字」など、過去4回の調査が行なわれている。国研におけるコンピュータの導入は、昭和41年3月に開始される第5次調査「現代新聞の用語用字」のために計画された。

設置されたコンピュータは、内部記憶容量 20KC、磁気テープ装置 6 台、ラインプリンタ、紙テープ入出力装置、プログラム作成用タイプライタを基本構成とし、漢字データ作成・修正用にオフライン型漢字テレタイプ（漢テレ）3 台を導入した。これらのシステムは、昭和49年までの8年間、朝日・読売・毎日新聞朝夕刊3紙1年分、約300万語の用語用字調査を中心に各種 KWIC 用例集の作成、社会調査、心理実験のデータ集計に使用された。

コンピュータの更新は、昭和49年同54年の2回、機種の変更が行なわれ、昭和60年3月には、日本電気の大型システム（ACOS—S550）への切り換えが予定されている。また、高速漢字プリンタ装置についても、昭和50年に第1号機が導入され、昭和55年に、そして同60年に新コンピュータとともに更新される予定である。漢字プリンタは、当初漢テレの印刷装置の低速度を補うものとして導入されたが、汎用漢字処理装置としての性格を持つ、当時としては非常に特徴のあるシステムであった。設置されたシステムは、コンピュータとの性能上の違いから、スタンド・アローン形式をとったが、新機種ではシステム全体が大型化され、ソフトウェアも研究用に十分耐えられる機

能があるためオンライン処理が主体になる。

ACOS—S550の性能は、これまでに利用してきた中型コンピュータに比べ大型であり、多様化する国語研究に十分対応できる能力と機能を持ったシステムとなる。

一方、漢字コードについては、昭和53年に情報交換用漢字符号系共通コードとして「JIS—C6226」コードが制定され、コンピュータ処理用コードの統一への道が開かれた。これとともに、国研の漢字入力装置についても旧型の漢テレから、JISコードを利用できる新型漢字入力装置への切り換えを併行して行ってきた。しかし、すでに13年間蓄積保存されてきたデータは、磁気テープ500巻以上になり、これを同時にJISコードに変更するためには、コンピュータの性能から時間・費用の点で無理がある。そのため当面の処置として変換用プログラムによる段階的移行を行う方法をとってきた。当然これらのデータ変換は、物理的コード、6 bit から8 bit への変換、JISコードへの文字対応上の問題、そのほか本稿で述べる外字表現方法と配列規準が移行上の重要項目となる。

外字処理は、現在の入出力装置の多様化によって対応可能な漢字数も1万字以上になっており、外字に関する処置もメーカーによって基本的な扱い方が異なったものとなってる。国研では、この外字処理の基本を諸橋轍次編「大漢和辞典」の検字番号に合わせ、約5万字をすべてコード化し理論的に20万字まで拡張可能にしている。理論コード化法の特徴は、部首単位による浮動コード方式を採用しているところに特徴がある。

基本的な考え方は、漢テレ盤内字450字分を1単位とし、同一部首内でこの数字を越える場合、部首の分割を行うとともに擬似的に部首を増加させ、康熙字典の214部首に対し271部首を立てている。しかし、このコード化法は、初期に導入された漢テレの物理コードと対応する2文字の組合せによって表現しているため、装置の保守、動態保存の限界が、そのまま外字コード利用の限界になる。本稿では、現在までのコンピュータ導入の経過から、国研コードの外字のあり方、コードブックとして大漢和辞典利用の妥当性の確

認、新システムに対する基本コード表現方法について試案を述べる。試案は、一部新システムで実用化されることを前提にしているが、可能な限り現行の外字表現の考え方を踏襲する。これは、新・旧コード間の変換辞書の互換性を簡単にするため部首配列を基本とする共通インターフェースの設定、他の漢和辞書との照合の容易性、コード変換処理の高速化、コードの拡張性と文字追加に対応するためのものである。

試案に対する実用化は、国研全体のコンピュータ利用者・国語・漢字・中国語の各研究者による横断的な意見の交換が必要であり、慎重な検討が不可欠である。しかし、前述の通り漢テレの動態保存が不可能になりつつある現在、新システムへの移行にはコード変換処理が最重要項目となる。本稿で述べる範囲は、コードに対する基本的思想の段階に限ったが、可能な限り多数の研究者から種々の批判を受けることが国研にとって最も重要であると考えたためである。この点からも、昭和60年3月の新コンピュータ切り換えを機会に、国研における漢字処理、日本語処理、所内データの蓄積・利用方法のあるべき姿を基本的な検討事項として位置付けておくことが、新国研コード体系を考える上で重要である。

これは長期的な展望に立った国研コードに対する検討が、今後大型化していくであろうコンピュータに対し、共通に利用される機械処理用辞書、データベースの利用と日本語処理に適したデータ記録形式の検討など、研究の多様化に対応する上で中核になる事項であると思われるからである。以下、国研コード、JIS、日本電気から提供される JIPS コードの問題について比較検討を行い新コードに対する試案を示す。

2. JIS コードと国研コード

漢テレは、昭和40年11月に第1号機が導入され、現在もコード変換用テーブルの保守に利用するため2台が利用可能な状態にある。しかし、昭和55年1月に漢字入力装置 (NEAC—N6300—50N) を導入した時点で、実用段階での役割は、ほぼ終了したと言ってよい。また、コード体系もこの時点で完

全に JIS コードを基本とした処理体制に入った。

JIS コードは、昭和53年（1978年）に情報交換用漢文字符号系 C—6226 として、第1版が制定され昭和58年（1983年）に第1回の改正が行なわれている。JIS コード制定の動機は、これまでに各メーカーが独自に設定していた漢字コードを異った装置間でも利用できるよう、共通コードとして互換性を持たせ、データ処理の効率化を図る目的で制定されたものである。現在では、漢字処理を行っている大部分の利用者は、この JIS コードを使用しているものと思われるが、長期間のデータ蓄積と各種調査を目的とする国研では、他機関とのデータ変換用としての利用以外に、次のような問題を解決する必要がある。

- 1) 長期的なデータ保存に対するコードの安定性。
- 2) 調査対象の拡大と蓄積データの多様化による文字種の増加への対応。
- 3) 共通データの多様化と漢字圏に対する共通コード化の問題。
- 4) 漢字追加機能と漢字総合辞書への拡張性。

1)2)については、外字コードに対する表現方法が重要になるが、現在使用している国研外字コードの表現は、諸橋轍次編「大漢和辞典」をコードブックとして使用し、各々の漢字に付加されている検字番号を基準にコード化を行っている。検字番号とコード化の対応は、検字番号に対し漢テレ盤内漢字2文字を組合せ、5桁の10進表示を漢字表現に変えている。漢字2文字で表現された外字コードは、JIS 及び漢テレ盤内字また、単独の外字コードとして付加させ、入力データに対するコード変換表としての位置付けを与えてある。この処理は、印字に外字表現用理論コードから JIS コードへの変換処理が必須となるための処置である。しかし、現行の外字表現は、外字コードを示す◇記号とともに、物理コードの配列順序をもとにしており、理論コードは装置に依存した体系となっている。また、対応している漢字も部首漢字と対応していないため、直感的な把握ができにくいなど、コード化の手続きを複雑にしている面が多い。この二つの問題は、漢テレの装置の耐久度、利用

可能な状態の保持が、そのままコード体系の利用限界となることを意味する。特に、検字番号が部首順配列を基本としているため、物理的なコード体系の変化とともに部首順配列に対する基本配列順序を維持できない問題が生じる。そのほか、4 byte を漢字1文字に対応させているため、メーカー提供の標準パッケージの利用ができないなど、プログラム・データの標準化を進める上で問題が多い。同様に JIS 改正による文字の変更と移動、国研独自のコード体系の持ち方、さらに汎用漢字辞書への文字追加方法などに対する基本的考え方を持つことが重要である。これは、漢字コードに対するあり方が今後の国研の言語処理、データの長期蓄積と共通化に大きな影響を与えることになるためである。

2.1 JIS コードの問題点

1983年9月に JIS コードが改正され、1978年版に対し84区以後の領域使用と、294字の字形の変更が行なわれた。字数は第1水準2965字、第2水準3388字、記号・英数字・罫線など524字分で、旧版に対し4文字増加している。配列順序は、第1水準は代表音訓順、第2水準は部首順、また追加用自由領域に約1688字分の予備を残してある。

改正版では、294字の字形の変更のほかに従来自由領域であった84区に「堯、楨、遙、瑤」の4字を追加し、第2水準の部首順配列の基本性を崩すことになった。すでにこの区点に漢字を配当している利用者に対して問題を残すことになると思われるが、以下、JIS コードを利用する上での問題点を示す。

第一の問題は、漢字表の各配列が検字の容易性に重点を置いた結果、各水準間に統一配列法を規定できなかった点にある。また、各水準間に連絡情報がないため、統一配列順序を取り込みにくい環境を作っている。これは、漢和辞書が使用される部首用漢字が各水準間に分散され、第1水準に142字分(1, 犂, 内, 兪は脱落)、第2水準に68字分が配当され、大部分が音訓順に配列されていることから明らかである。

第二の問題は、JIS コード試案の段階で指摘されていたことであるが、自

由領域に対し利用者登録を許したことにある。JIS コード表制定の基本目的は、情報交換の標準化にあったが、結果として交換用コードとしての基本性と特性を崩すことになった。同様に、文字追加方法も利用者の責任において行なうとしたが、当初漢字配列はある統一思想のもとで決定されたとしても、長期にわたるコード追加が続く場合、逆に利用者にも種々の問題を抱えこませることになった。大部分は自由領域内における追加文字数の増加によって、コード配列の統一性を維持できなくなったことである。また、追加領域不足に対する対応法も重要な問題となっていた。

国研の場合、この問題を避けるため当面第2水準までを利用対象とし、基本的には新規の文字登録は禁止する。新規に出現した文字は、入力段階で国研外字で対応させ、これをもって保存用コードもかねるというものであった。この方式をとる限り、たとえ JIS の改正が行われても変換テーブルの一部修正ですべて対応可能となり、保存ファイルに対する保守が容易になる。さらに、理論コードとして部首順配列を国研外字の統一理論コードに位置付けることによって、漢字の持つ特性をコード変換処理に应用できることになり、各種辞書引き処理に対する基準配列の設定が一元化される。

しかし、いずれにしても、これらの問題は、JIS コード変換表が半固定で与えられているために生じたものである。漢字コードを使用する場合、人間・機械双方ともに検索の容易性が重要である。この操作を JIS 漢字表に対応させた場合、第1、第2水準の未定義文字は、自由領域を一種のオーバフロー領域として使用することになり、各水準からのリンク情報が含まれていない以上、自由領域は独立状態になる。同様に、追加可能な文字数も約1700字程度であり、追加文字をすべてこの領域に登録することには無理がある。

これは、漢字の順序性を示す理論コードとパターン格納位置を一致させている限り、自由領域についても JIS の各領域の問題をそのまま引き継ぐことになり、改正版でも結局この方法に従った形をとっている。固定化コード法のもとでは、特定コードの間に新たな漢字を挿入登録しようとする場合、現行では対応不可能である。また、漢字の順序性を維持していくためには、コ

ード体系の再割当が必要となり、これも現存の規格のもとでは、実現は非常に困難な問題である。実務的な処理面で生じる問題は、コード変換前と変換後に処理されたデータの間では、データ保存とデータ交換の点で整合性を欠いたものとなる。これらは、統一配列順序を保つコードの設定によって、避けることができるものと思われるが、いずれにしても漢字パターン格納位置と検字用コードを同一レベルにおいたために生じたものである。

コードの固定化法は、対象となる漢字世界と JIS 内で定義される漢字数が大きく異なる場合、漢字の事前選択と統一配列の維持を難しいものになっている。一例として、大漢和辞典の収容漢字数約 5 万字に対し、JIS で使用される字数は 7 bit 2 列を使用している限り、16000 字が利用可能最大字数である。しかも JIS コードとして定義されている符号領域は、1～94 の区点で示される 8836 字分であり、1975 年版「情報交換用符号の拡張法 C—6228」でも、複数バイトに対する拡張法は定められていない。

以下、この問題に対する解決案を国研で使用している外字表現とコード化法をもとに、JIS への対応について 試案を述べる。

現在使用されているコード体系は、1 byte 系、2 byte 系 (JIS コード) とともに、文字とそれに付加されるコードは、一義的に決定され、かつ固定化された体系を使用している。このコード化法は、対象となる字母が少なく文字コードの移動がない場合問題化することはないが、JIS の 範囲を越える外字表現を対象とし、漢字のように字母が多く時間とともに使用状態が変化する場合など、固定化されたコードは問題が少なくない。

漢字パターンと付加されるコードの管理は、コードブックによって管理され、検字を容易にするため、漢字配列はある定まった順序に配列されている。しかし、外字の場合コード設定当初は、漢字配列に対する基本方針が決まっているにもかかわらず、特定コードとコードとの間に挿入が発生した場合、コード順序の統一性を保つためには、コードの再付加、枝番号指定が必要となる。これはコードの二重化と桁数を増加させ実務上問題が多い。この場合、JIS 表に比べ十分に大きな 漢字表を事前に用意できれば、表面化することは

少なくなる。

例えば、大漢和辞典をコード表として使用することによって、検字番号をそのまま外字コードとして利用するなどである。また、日本で最大の収容字数を持つ、辞書をコードブック化することで、コードブック作成労力の軽減と結果として追加処理を少なくすることが可能となる。同時に、漢字コードの理論コードと漢字パターン格納用物理コードの分離を図ることで、外字処理における辞書側修正を少なくする効果が得られる。この方式によって、入力コードは、原稿に正確に対応するデータ作成とゲタ対応の漢字パターン作成を分離させることによって、少なくとも外字コードは大漢和辞典の範囲内で情報交換用として利用する場合従来の方法よりも有利となる。

検字番号の入力表現は部首表示と相対番号指示用の2文字で行い、従来の国研外字表現と同形式とする。

□ □ ← 相対番号指示文字
↑
_____ 部首表現文字

上記の表現形式によって、外字がどの部首に属するかは、一義的に決定でき、相対番号についても、簡単な演算で指示可能となる。この処理で部首文字の順序性を維持しつつ、文字指定を行うためには、部首文字の JIS コード内への登録が必要となるが、84区から94区の11区、そのほか自由領域内に展開させることによって、外字ファンクションシフト機能も、部首文字に含ませることが可能となる。同時に、この機能は、コンピュータ内での実質的シフト機能として、漢字パターンを ROM にロードする選択機能にも利用できることになる。

2.2 国研コードの問題点

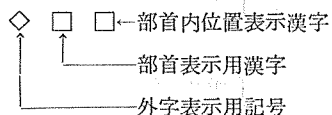
漢テレは、コンピュータ導入と同時に昭和41年に設置され、以来19年間各種の調査に使用されてきた。漢テレによる漢字表現は、6 bit を基本として2列12ditを漢字1文字に対応させている。これは当時のコンピュータのコード体系に従って設定したものであるが、12bit で表現可能な文字数4096字

に対し2400字分（漢字2110字）を内部コードとしている。

しかし、大量の用語用字調査を行う上で、2110字で表現できない漢字が出現する可能性があり、これに対応する特殊な漢字表現法が必要になる。この漢字の表現法を外字または盤外字と言っているが、国研ではコード化または解読処理をコードブックによって行ってきた。

コードブックは、大漢和辞典の検字番号を使用し、◇記号に盤内字2文字を組合せて表現している。外字コードに対する基本的な考え方は、漢テレの高出現頻度の漢字450字の順序性を利用し、変換表（表1）を使い、部首、部首内の相対位置をコード化する。同一部首で450字を越える漢字（艸では2172字）が収容されている場合は、450字を1グループとする擬似的部首を作り、グループ化された各々の部首は、独立した部首であるように部首用漢字を別に配当する。例えば、部首「一」から「龠」にいたる271部首に対し、計(0001)₈から合(2121)₈までの漢字を外字表現用コードとして対応させ、部首内の位置を示すコードは、計(0001)₈から力(3051)₈までの450字を相対番号表示用に当てている。このコード化法によって、外字表現されるコードは、コンピュータで分類した場合、大漢和辞典で定められた部首と同配列順になる。以下の例は、外字コードに対する理論コードの与え方を示したものである。

理論コードの与え方



1) 各部首の基準番号の求め方

- (1) 各部首の最初の漢字に付加されている検字番号の下2桁を切り捨てる。
また00の場合は100を引く。
- (2) 基準番号がグループ化（部首分割）されている場合、第2グループ以後の各グループは、先頭番号から1を引く。

例) グループ先頭字 一 人 倆

検 字 番 号 1 344 751

基 準 番 号 0 300 750

(3)一部首内の最大字数は 450 とし、これを越える場合は同一部首を擬似的に分割する。

2) 部首の求め方

(1) 271 部首に配当された漢字を表 2 から求める。

3) 部首内の位置番号の求め方

(検字番号－基準番号) MOD450＝N

(1) MOD450 は、450 の剰余を求めることを示す。

(2) 剰余が 0 の場合は、そのグループの最後の文字であることを示す。

(3) 求められた N 値を使い、表 1 から対応する漢字コードを求める。

4) 部首のグループ化例

部首の分割方法は「人」の部を例にとった場合、次のように行なう。検字番号の先頭の 344 「人」から 1335 の「儼」までの分割例である。

人の部	基準番号	300	対応漢字
	先頭の検字番号	344	人
		750	倅
		751	倆
		1200	儼
		1201	儼
	最後の検字番号	1335	儼
			第 1 グループ
			第 2 グループ
			第 3 グループ

表 1 理論コード用漢字表

部首	0	一	丨	丶	ノ	乙	丨	二	亠	人
0		1	2	3	4	5	6	7	8	9
1	建	計	形	型	経	芸	頭	決	結	月
2	投	見	件	県	伝	電	都	土	度	当
3	島	東	党	島	等	愛	悪	安	案	以
45	力			一	以	下	略			

表2 部首分割配当漢字（文献2より引用）

部首グループ (先頭字)	コード第1字 (グループ内 に共通)	グループ先頭 字の通し番号	グループの 基準番号	コード第2字 (先頭字に与え られたもの)
一	1 計	1		1 計
丨	2 形	67		67 声
丶	3 型	91		91 説
ノ	4 経	106	100	06 頭
乙	5 芸	161	100	61 横
丿	6 頭	224	200	24 等
二	7 決	247	200	47 数
一	8 結	286	200	86 赤
人 1	9 月	344	300	44 人
(𠂇) 2	10 建	751	750	1 計
(𠂇) 3	11 見	1201	1200	1 計
儿	12 件	1336	1300	36 員
入	13 県	1415	1400	15 電
八	14 伝	1450	1400	50 運
冂	15 電	1506	1500	06 頭
冂	16 都	1565	1500	65 成
丩	17 土	1607	1600	07 決
几	18 度	1737	1700	37 右
凵	19 当	1800	1700	100 両
刀 1	20 投	1845	1800	45 図
(𠂇) 2	21 東	2251	2250	1 計
力	22 党	2288	2200	88 折
勹	23 鳥	2493	2400	93 川
匕	24 等	2570	2500	70 政
匚	25 愛	2595	2500	95 船
匚	26 惡	2672	2600	72 製
十	27 安	2695	2600	95 船
ト	28 案	2774	2700	74 屋
冂	29 以	2836	2800	36 員
厂	30 世	2890	2800	90 設
ム	31 位	3057	3000	57 円
又	32 意	3115	3100	15 電
口 1	33 違	3227	3200	27 安
(哪) 2	34 育	3651	3650	1 計
(𠂇) 3	35 引	4101	4100	1 計
(𠂇) 4	36 員	4551	4550	1 計
曰	37 右	4675	4600	75 音
土 1	38 身	4867	4800	67 声
(中 略)				
龠	271 合	48882	48800	82 花

5) 外字コードの求め方

以下、「倆」検字番号 (751) について計算例を示す。

(1) 部首「人」の部を表 2 から「月」を求める。

(2) 部首内位置

基準番号 300

計算式 $(751-300) \text{MOD} 450 = 1$ (表 1 から「計」)

求めたコード ◇月計

以上、検字番号からの変換方法を示した。国研コード上の問題として、漢テレのハードウェア上の特性に強く依存していることはすでに述べたが、問題と利点を次の各項目にまとめることができる。

問 題 点

- 1) 検字番号の10進数値を漢字の物理コード順に対応させているため、外字表現に実際に出現した文字との連想関係がない。
- 2) 外字コード表現は漢テレのハードウェアに依存している。
- 3) 同一部首内の漢字が 450 を越えた場合擬似的部首が増加する。
- 4) メーカー提供の標準パッケージの利用と辞書類の直接利用ができない。
- 5) 一外字表現が、◇記号を含む漢字系 3 文字で表現されるため、印刷時の行末調整が必要になる。
- 6) 漢テレの動態保存が不可能になりつつある (部品、保守体制)。
- 7) 物理コードに 6 bit 2 列を使用しているため、JIS 7 bit との対応に整合性がない。
- 8) 外字表現に、漢テレの物理コードに対応した文字配列順序を使っているため、JIS 移行のさい、外字コードの部首順配列の基本性が失われる。
- 9) 検字番号の「//」点の付いた漢字をその部首の最後に登録したため、大漢和辞典の同一部首内総画順に配列の基本性を崩している。
- 10) 漢テレコード体系は、6 単位符号系の制御文字領域に抵触するため、

オンライン主体の処理に対応不可となる。曲 (0000)₈ ～建 (0017)₈ など
10文字分はビット列の変換が必要。

利 点

- 1) 擬似的な部首を含め、基準番号による部首別浮動コード方式を採用しているため、理論コードに対する拡張が容易。
- 2) 外字表現及び JIS 配列に対する統一理論コードとしての位置付けが可能。
- 3) 統一理論コードをもとに、総画・代表音訓、JIS の各配列による索引の作成と、他のコンピュータ用辞書との結合による情報の拡張が容易。
- 4) 大漢和辞典の検字番号を理論コードとして利用しているため、初期のコードブック作成が不用。
- 5) 印刷用漢字パターンが未定義である場合も、入力処理は明確なコード化が可能。
- 6) コード変換処理が必須となるが、JIS 改正、新規パターン登録時のコード対応が容易。
- 7) JIS コードと外字表現の併用により、入力・内部・印字可能なコードを実質的に JIS コードの世界へ拡張することが可能。

3. コード変換処理

3.1 入出力と内部コード

漢字コードは、入力時のコード、データ保存時の内部コード、また漢字プリンタなどの出力装置に対するコードが一致していることが必要である。JIS コードの主旨も、入出力及びデータ保存の三形態の中で共通コードの移動が、情報交換用コードとして重要な意味を持つことを前提に設定されている。しかし、コード表現の可能な組合せは、14 bit 表現である限り、16000 字が最大であるが、通信用ファンクションコード、EBCDIC コードの 1byte 系の併用を考えると利用可能な領域はさらに少なくなる。これは、長期的なデータ蓄積を考慮した場合少なからず不安がある。

そこで実際の運用面で与える影響を考えるため、データ入力、保存用内部コード、出力コードの3項目について関連性を述べる。なお、問題点を明確にするため、実際上の処理形態は、漢テレの利用が不可能になっている現状では、これより少なくなるが、入力装置に漢テレを使った場合、JISコードで入力した場合の各々について処理形態を概観する。

図1 入出力及び内部コード

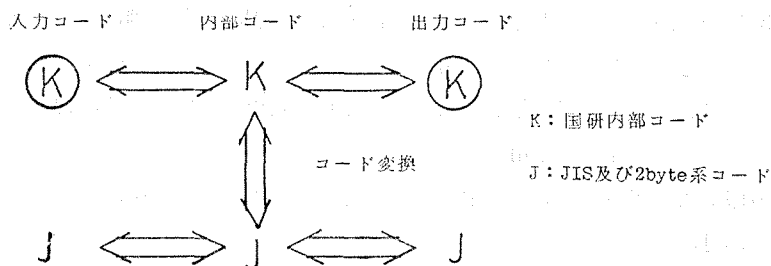


図1は、入出力及び内部コードに対するデータ相互の関係を示したものである。内部コードは、データ保存用コードと同様の意味を持つが、JISまたは、国研独自のコード体系のいずれかを意味する。この3項目の特性を調べることは、将来の国研コードの機能や問題点を探る上で有効な手段となる。これは、3項目の組合せは、8種の処理形態に分類されるが、この中の大部分は、昭和41年にコンピュータが導入された以後の各装置、コードの変遷を示しており実質的な処理の本質を含んでいると考えられるからである。

図1の関係は、コードの相互関係を表わしているが、入出力装置中のデータ記録装置も含めるものとする。

これによって、内部コードと装置そのものによって制約される問題の両面から検討が可能になる。また、現在生じている問題点の中心は、漢テレの利用と外字コード体系の新システムへの対応にあるため、表3に示した8例がすべての処理になる。この中には、過去の処理形態と実務上利用されない組合せも含んでいる。その点で、入出力処理上の「K」部分は、実際に機能し

表3 コードと入出力装置関係

	入力コード	内部コード	出力コード	使 用 例
1	K	→ K	→ K	入出力ともに漢テレコード（国研コード）を使用。初期の使用形態。
2	K	→ K	+ J	高速漢字プリンタ導入時の国研における第2期の使用例。入力装置は漢テレを利用。コンピュータ処理は国研コードによる処理。
3	K	+ J	+ K	国研コードで外注を含むデータ作成、オンライン処理後、保存データは国研コードで出力。国研コードとコンピュータ内部コードが異なるJIS制定前後の過渡期の使用例。
4	K	+ J	→ J	長期プロジェクトなどで、入力に国研コードを使用している場合。データ作成は外部コードか国研コード。コンピュータ処理用コード及び外部出力コードはJISコード。
5	J	+ K	→ K	JISコードによるデータ作成後、国研保存コードとの併合・保守、出力コードは国研コードで記録媒体に出力。
6	J	+ K	+ J	5)の処理と同様。ただし、出力は高速漢字プリンタを使用。
7	J	→ J	+ K	国研コードのJISコードによる保守。また国研コードで作成されたファイル、コード変換テーブルの保守など。
8	J	→ J	→ J	入出力及び内部コードともにJISコードを使用。JIS制定後の現行処理。

←：コード無変換処理
+：コード変換処理

ていないことになり、一部データ保存用の問題として残ることになる。

処理に関する代表的なものは「使用例」に示した内容になるが、表中の 8 例は、19年間に国研で使用してきたコンピュータの歴史を含んだ内容になる。1については、コンピュータ導入初期の漢テレとの関係であり、新聞の用語用字調査で使用された例である。2は、漢テレの印刷部の印刷速度の低速性を補う目的で導入された、高速漢字プリンタ使用の流れである。漢字プリンタ導入時の昭和49年当時は、JIS コード 制定以前であり出力コードはメーカ独自の 2 byte 系を使用している。漢字 プリントシステムは、漢字ディスプレイ装置によるデータ検索・修正処理、文字記号読取り装置によるターンアラウンド処理など、汎用漢字処理機能を持ったシステムとして使用された。

次に4は、現在でも一部この組合せの処理が行なわれているが、国研における長期調査の場合、中間にコンピュータ切り換えが入るため、入出力コードの維持が大きな問題となる。この形式は、初期データが漢テレで作成され、内部コードが JIS 化された処理例となる。

この形態は、オンライン処理によるデータ修正・検索処理に、JIS 対応の端末装置を使用した処理が可能になるため、データ作成を外注し JIS コードによる納入が行なわれた場合 8 と同等となる。そのほか 5～8は、入出力装置がすべて JIS に対応済みであり、外字処理以外直接問題となることはない。一方、出力装置が「K」の部分 5, 7については、3と同様、実務処理で利用されることのない形式であり、残る 6, 8が現在使用されているコード処理の主な流れとなる。

4 との関係は、4 が国研コードを JIS に変換しているのに対し、6 では国研コードの外字コードの理論性のみを維持し、物理コードは JIS である場合に限られる。その点で、この型は組合せとして實際上処理されることはない。しかし、ここで内部コード体系を国研独自のコード体系とする場合にも、利用可能な入出力装置は、すべて JIS 対応またはシステム内で閉じたコード体系を使用することになる。また、出力装置も JIS の範囲で印字可能な文字数が決定されるため、外字入出力問題は、JIS コードの制限下で特別処理の必

要が残る。

結局、JIS 世界ですべて処理する場合と JIS コードとの併用方式による国研内部コード処理の可能な 6, 8 が今後の検討課題となる。

3.2 コード変換処理の概要

コード変換に伴う処理上の問題点を概観した。次に漢テレコードから JIS コードへの変換、JIS コードによる国研外字表現（擬似コード）の必要性、またコード変換のためのテーブル構成について基本的な考え方を示す。コード変換は、盤内字コードである物理的コードと理論コードである外字コードの 2 種を対象とする。これは現行の国研コード体系が 6 bit 2 列を盤内字 1 文字に対応させ、さらに 2 次コードとして盤内漢字 2 文字を組合せ外字表現させているためである。2 種の入力データ形式は、漢字出力装置で対応可能な出力コードに変換しなければならないが、コード化とコード化された漢字の相互変換はコードブック及びテーブルによる方法をとった。

テーブルは、一次キーとなる物理的コードと入力データコードが一致していることが必須である。一次キーは、盤内字及び盤外字表現された国研コードとし、変換対象となるコードを各々一次キーに対応させコード変換用テーブルを作成する。この操作は変換対象となるコードが複数個の場合も同様である。テーブル作成用装置は、どのようなコード体系を持ったものでも基本的には問題ないが、主流となる入力データコードと一次キーとなるコードを一致させておく方がテーブル作成処理は容易になる。以下、現行の漢テレコードから JIS コードへの変換処理、入力装置に JIS コードを使用した場合の国研外字コードの位置付けについて問題点を示す。第一の点は、コード変換処理に伴うテーブル構成に関するものである。テーブル構成は次の 3 例が基本になるが、3) の場合 JIS コードについても 84 区以後に文字登録が行なわれているため、メーカーが異なるときにコード変換処理が必須となる。この処理は、記号系領域に独自に追加したものについても同様である。また後述するメーカーの JIS に対する対応が、かならずしも指定されたコード体系を使用し

ていないため、異機種間のデータ変換のさい2 byte 系相互のコード変換処理が必要になる。しかし、ここでは主に1)と2)について重点を置く。

コード変換処理

- 1) 国研盤内外字コード： JIS 系 2 byte コード
- 2) 国研擬似外字コード： 同 上
- 3) 2 byte 系コード： 2 byte 系コード

図 2 コード変換処理

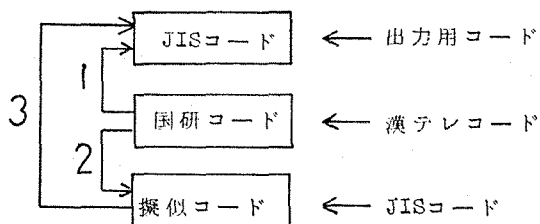


図 2 に示した、1～3 のデータ変換の流れは、テーブルについて擬似コード処理用と国研対 JIS 変換用の 2 種について、変換処理の流れを示したものである。テーブル作成作業は、現行漢テレコードによって行うが、物理的コードを JIS 化した擬似コードの作成は、1 で使用される変換用テーブルをデータとし、変換用テーブルを再交換することによって、JIS コード用テーブルへの対応が可能になる。

表 4 コードの種類

コード名	物理コード	外字コード
1) 国研コード	6 bit × 2 列	◇記号と漢字 2 文字の組合せコード列
2) JIS コード	7 bit × 2 列	特に定めず
3) 国研擬似コード	7 bit × 2 列(JIS)	国研外字コードに同じ

字種は JIS に準拠しているがコードは以下の種類がある。

メーカーの JIS 対応状態

1) 16進数(8080)を加え(2121)～(7E7E)領域から(A1A1)～(FEFE)領域へシフトして使用。

2) JIS 7 bit で表現された文字を、対応する EBCDIC 文字コードに変換後 2 byte で1 漢字に対応させる。

3) 字種、コードともに独自体系を使用。

4) JIS コードを使用。

図2で示したコード変換処理で、1は主に次の二種の処理に使用される。第一は高速漢字プリンタ(漢プリ)を使用するさい、PIF(Print Image File)を事前に作成しておくことが、漢プリ印字速度を低下させないための必要な処置となる。これは印字速度が高速であるため、ページ単位の処理を基本としているためである。また、高速印字に対応させるためには、データ読み取り速度の早い磁気テープ、磁気ディスクなどの媒体の利用が不可欠となる。

以上の二種の対応後、データは未登録文字以外の処理は、すべて直接印字可能な物理的コードに変換されている。

第二の問題は、データの照合、探索処理などの処理で、キー語の入力に対するコード変換上の問題である。オンライン処理の対象となるデータは、すでに JIS コードに変換されているため、国研外字は擬似コードになっている。この場合、入力パラメータやキーに国研コードを使用するさい、コード変換の問題が発生する。中規模以上では、応答時間の遅れと、ジョブ数の増加を招くことになる。逆に、漢テレコードを内部コードとして使用するさいも、これと同様の問題が生じるが、この場合処理上の漢字コードの順序性は維持される。

この点から、JIS コードまたは PIF でのデータ保存は不適当であると思われるが、前述の通り現行漢テレの保存状態のことがあり、特別の場合を省き当面は、JIS コードの保存もやむなしと考えざるを得ない。これは、実務的にも大部分の処理が JIS の範囲内で対応可能であり、擬似コードによる順序性は實際上、数の上で表面化することは少ないためである。しかし、長期のデータ保存を考慮した場合この問題は無視できないものとなることは確かで

ある。いずれにしても図2に示した3のルートが主流になり、1、2は、JISコード入力可能な装置の利用と新外字コード開発までの繋ぎとして臨時的な処置となる。

以上、コード変換処理に関するデータ変換の流れについて現状を示した。コード変換処理は、コードの標準化が進めば、当然なくなるものと考えられていたが、まだ外字処理の問題が新たに解決しなければならない対象として、重要な部分であることが明らかになった。また、中国を始めとして、他の漢字圏との共通コード化の問題も、いずれ話題としてのぼってくるものと考えられ、外字表現、内部コードについて総合的な判断が必要とされる状況であるように思われる。

3.3 新コード移行までの対応

新コード体系の実務段階での運用は、まだ、いくつかの実験を経なければならない。しかし、新機種導入まで、既存漢テレコードで作成されたファイルについて移行への対応問題があり、この処理はJIS対応漢字入力装置上で処理しなければならない条件が付けられる。そこで、新コード体系の具体的な案が設定されるまでの準備期間として、現行コード変換、表を中心にした暫定的移行処置とともに、新コード下での外字入力の方法、既存ファイルとの結合、コード変換とテーブル構造について基本的な考え方を示す。

暫定処理の一部には、新コード体系への理論的妥当性の確認を行うため、いくつかの試案と実験を含ませてある。実現方法は、テーブル構成の一部追加と漢字特性を利用したテーブル探索の高速化である。これによって、新コードの基本的な考え方が、この19年間使用された外字コードに対し、理論的背景と改良部分の正当性に十分答えたものであるか、また、新コード体系を長期間利用する上での基本的な設計基準をどこに置くのかなど、運用面からの確認を厳重に行うことが可能になる。同時に、旧コードは今後の国研コードを維持していく上で致命的な問題点を持った体系なのか、種々の面からの検討を行うことができる。

その点で、移行処置の中で処理の変更とともに、いくつかの実験を含ませておくことは、問題点を実務面で検証するのと並行し、移行に必要な各種情報を収集することが可能となる。これは、試案に対する問題点を明確にし、移行処置を容易にするために重要である。ここでは、主にコード体系について枠組設定上の問題解明に重点を置く。しかし、現行の問題点は、将来にわたって簡易化されるであろう字形への対応、利用される研究者の多様化と資料範囲の拡大による漢字への動的対応など、社会的要因に大きく影響を受ける面が多くなる。この問題は、長期のコード体系を維持する上で予測できない部分であり、漢字シソーラスへの発展、統一理論コードの設定、コードに対する漢字追加機能とともに直列的な配列から面への拡大が必要になる。これは、コードの構造を決定する上で考慮すべき必要最少限の要求事項である。以下、移行処理に対する具体的な試案と、実現方法に関する基本的な考え方を示す。

3.3.1 外字入力法の拡張方法

外字入力方法は、従来の漢テレ外字コード形式で入力する場合と、拡張機能として大漢和辞典の検字番号を直接入力する場合の併用方式をとる。この方式の導入によって、従来のコード化法に対し外字表現上の手続きに変更はなく、既存ファイルとの互換性も簡単なコード変換で対応可能となる。

検字番号を直接入力する場合、最大5桁の数字を使い漢テレ外字コード形式に変換する。この処理における2種の外字表現形式は、コンピュータ内部で同一表現形式に統一され、既存ファイル及び外字相互間の互換性には問題はない。

コード変換に対する基本テーブルは、表5に示した項目から構成されるが、従来の2進探索法によるコード変換、また、外字表現形式の入力について、当面、次の変換処理を試験的に取り入れる。この処理によって、コード変換処理の高速化と外字入力の多様性に対する二つの処理から、コード設定上の必要条件と情報が得られるものと思われる。特に、外字入力方式の多様化は、

将来の部首による検索の多様化，例えば文字構成要素からの検字と外字指定への道を開くために重要な項目となる。

変換テーブルは，実務で使用される基本テーブル表 8 も含め 3 種類のテーブルを使用する。この 3 種のテーブルは，表 1 の国研外字処理用理論コード表，表 8 の外字対 JIS コード変換表，そして表 5 の索引表である。

表 5 は，部首表現用漢字，対応する大漢和辞典の検字番号，第 3 項目は JIS-C6226 コード体系内での部首に再配列された後の部首の先頭位置を示したものである。

表 5 コード変換用部首索引表

No.	理論コード	大漢和検字番号	JIS 部首先頭位置
1	計 (一)	1	1
2	形 (丨)	67	21
3	型 (丶)	91	25
4	経 (ノ)	106	30
5	芸 (丨)	161	43
6	頭 (二)	224	52
7	決 (一)	247	58

— 以 下 略 —

検字番号から国研外字の求め方

- 1) 表 5 のコード変換部首索引表の第 2 項の検字番号，対応する JIS 部首位置及び部首表現用漢字を求める。
- 2) 入力された検字番号をもとに，部首基準番号を計算する。
- 3) $(\text{検字番号} - \text{基準番号}) \text{MOD} 450 = N$ を計算する。
- 4) N 値を使い表 1 から理論コード表現用漢字を求める。
- 5) 先頭に◇記号を付け，項目 1，4 で求めた漢字 2 文字を結合する。
- 6) 補遺版に登録されている漢字，及び，本文中の「'」，「''」のついた漢字は当面別処理とする。

以上が，検字番号による国研外字表現の求め方である。この処理の問題点

は、次の5項目が未解決のまま運用されることになるが、検字番号の入力方法の容易性と利点は、国研外字表現に対するコード化理論に比較し優れていると思われる。特に、コード化に専門的な知識を必要とせず、一般利用者が簡単に外字入力ができる利点、配当漢字の割当て状況がコンピュータによって、管理可能になる利点は無視できない事項となる。

問 題 点

- 1) 国研外字に対する問題点の解決が図られていない。
- 2) 過去に追加された新出漢字に対する対応が不可。
- 3) 入力処理に最大5桁の10進数の入力が必要になる。
- 4) 検字番号の「、」及び「、」に対する対応と登録済み漢字対応に不整合が生じる可能性がある。
- 5) 補遺版に登録されている漢字は特別処理が必要。

利 点

- 1) 国研対 JIS 変換表は変更不用。
- 2) 部首変換は表5の271部首について変換表で対応可。
- 3) 追加漢字に対する理論コード配当漢字の自動管理への道を開く。
- 4) コード変換処理の高速化が可能。
- 5) 外字コード入力及びコード化の一般利用者への開放が可能。
- 6) 85区または90区以降に一括登録するため、シフト記号の省略が可能。

3.3.2 外字入力と枝番号処理

外字コード指示に枝番号による漢字追加機能を持つことが重要であるが、入力方式の容易性の点で、検字番号の直接入力方式が従来の国研外字指示法に比べ優れていると思われる。これは専任オペレータ以外の一般利用者を想定した場合に使い易さが基本的な機能になる。外字コード構成は検字番号及び枝番号の2項目で指示し、整数部を検字番号に小数部を枝番号に対応させ

る。外字コードの判別は外字シフトキー指定によって行い、次に入力される7桁の数字を内部コードに変換する。また、整数部と枝番号の分離は、小数点・カンマなどの記号で対応させる。整数部の最大値は大漢和辞典の検字番号48902であるが、上限値は特に定めていない。同様に枝番号についても上限値は99まで指定可能であるが、この部分は使用ビットで表現できる最大値が15であるためこの数が事実上の上限となる。

以下、外字表示、内部コードへの変換方法と表示内容を示す。なお計算例は、内部コード14bitの第2byte目について処理を行った。この例では、上位3bitを整数部に、下位4bitを枝番号用とした。内部コード中の「V」記号は、仮想小数点の位置を示したものである。

例) 内部コード 111V0000 (10進数) 112 (2進数) 7 (10進数) 0 (2進数)

外字入力形式

◇ ×××××...×× 111V0000 112 7 0
外字シフト 検字番号 枝番号
キー

計 算 式

1) 整数部 = 検字番号 $\text{DIV} 2^n$

2) 小数部 = 検字番号 $\text{MOD} 2^n$

n : 枝番号表示用ビット数

DIV : 剰算関数

MOD : 剰余関数

計 算 例

内部コード 10進表示 1) 式計算値 2) 式計算値 + 入力検査番号

1) 111V0000 112 7 0 112 + 7.0 = 119.0

2) 001V1110 30 1 14 30 + 1.14 = 31.14

3) 111V1110 126 7 14 126 + 7.14 = 133.14

この計算式は、2進数内部コードを逆算によって入力検字番号を求めたが、

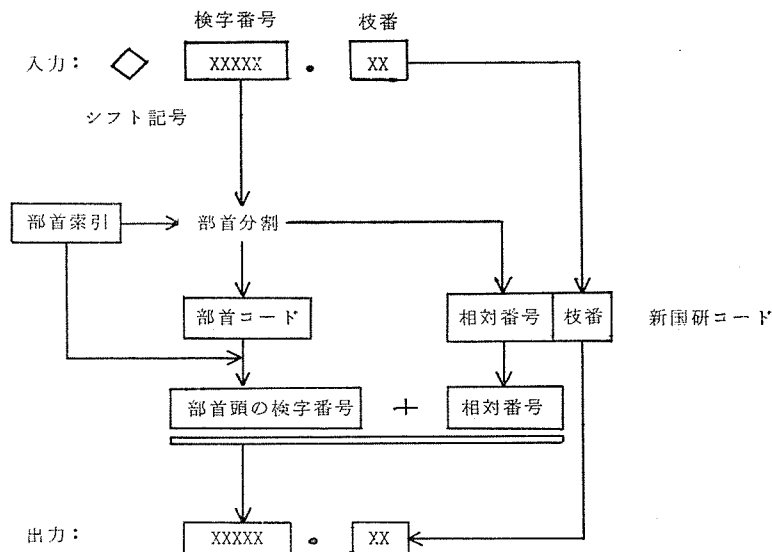
1)及び2)式によって求められた数値を結合することで、10進表示による検字番号の入力で対応可能であることを示した。変換された2進コードは、そのまま内部コードとなるが、表5を利用することで入力された検字番号がどの部首に属するか判断できることになる。この処理は、検字番号の入力に部首情報に対する指示を不要になる。

出力は、漢字ディスプレイ、プリンタなど文字の合成機能、指示可能なポイント数によって、制限を受けることが多いが、ホスト・端末間、特にディスプレイ上の表示を別処理としなければならない問題が生じる。しかし外字処理は、米国 Wang 社から発表されている「三角編号法」による入力方法を出力にも適用可能であると思われる、出力に対する文字合成への方向付けが重要である。この方式は、漢字入力に従来の、タブレット方式、カナ・漢字変換方式に対し、漢字の構成要素別に分解入力し、コンピュータによって内部コードに変換する。アルゴリズムは、漢字入力に対するものであるが、文字合成アルゴリズムは、出力処理に対しても応用可能であると思われる。

しかし、文字構成要素の結合による外字処理は、現行の出力装置の機能に制限を受け、対応可能な装置は少ない。実現性のある対応は、漢字プリンタのルビ文字機能を利用することによって、全角漢字1文字に対し、4桁の数字を天地左右に配分し合成出力する方式が考えられるが、この方式は、端末側の出力機能に制限され、漢字ディスプレイに対しては対応不可能である。その点で、外字出力形式は端末・ホスト両装置間に2種の異った外字処理を行うことになる。

漢字ディスプレイ側の処理は、従来の国研外字表示と同様に出力可能な漢字2文字の組合せで出力する、入力と同形式で検字番号を出力表示する、いずれかの方法で対応せざるを得ないものと思われる。画面表示上の点では、後者の方式が望ましいことになるが、外字表示にプリンタ同様統一性がないことは実務上問題が残ることになる。これも早急に解決しなければならない問題であり、今後の研究課題として重要である。図3は検字番号で外字入力を行う処理過程を示したものである。

図3 検字番号による入出力



3.3.3 外字コードの JIS コード内への対応

現在使用している漢テレコードは 6 bit 2 列を 1 漢字に対応させ、コード範囲も曲 (0000)₈～己 (7474)₈までを使用している。しかし、新コンピュータはオンライン処理が中心になるため、データ伝送には伝送用制御用コードの規格に合った処理を行わなければならない。この伝送路用コードは、JIS コードで決められており、理論コードについても、計 (0001)₁₆～県 (001D)₁₆の範囲がこの規格に抵触する。

また、JIS 漢字を外字理論コードとして使用した場合、現行部首の順序性を崩すことになる。そこで、理論コードは、現状と同様の機能を維持させ、従来通りの配列を維持するコード化法についていくつかの試案を述べる。

新コード体系については、文字追加機能とともに伝送路に抵触しないコード化法を考えなければならないが、当面これを次の方法で臨時処置として対応させる。この方法は、将来新コードへの移行のさい現行コードと臨時コ

ドとの複数コードの混在化の問題が生じる可能性があるが、基本は JIS コード対応であり、国研外字コードも JIS と同規格で利用できることになる。

移行処置では、部首順配列を維持したコード体系が必要な場合についてのみこの方法を用いる。しかし、新コードに対しても既存コードと同等の理論コード体系をとっているため、変換アルゴリズムは同一となる。コード変換範囲は、JIS 85区または90区以降に理論コード、「計」～「力」にいたる450字を配列する。また、外字シフト用記号は、この区区内のいずれかの位置に一括登録するか、90区以降のコードの重みをそのままシフトコードとする方法をとる。

以下、移行問題での試案に対する、理論コード配当例、利点と問題点について列挙する。

理論コード配当例

区	点1	2	3	4	5	～	94
90	計	形	型	経	芸		先
91	船				注		
92	町			合			非
93	飛						味
94	民			力			

問題点

- 1)旧漢テレコードで作成されたデータに対しコード変換が必要。
- 2)将来 JIS 改正で85区～94区に文字追加される場合への対応が未解決。
- 3)漢字パターンが ROM 化されている場合対応不可となる。
- 4)旧国研外字コード化法の問題が未解決のまま残る。
- 5)暫定処置によるコード体系の複合化。

利 点

- 1) 外字基本配列の理論順序である部首順配列の維持が可能。
- 2) JIS コード体系内で対応可能なため、伝送路及びシステム内のコード上の問題は解決される。
- 3) 従来の外字コード化法をそのまま利用可能。
- 4) 85区または90区以降のコードを外字判別用に使用可能。
- 5) 入出力装置に JIS 対応装置の利用が可能。
- 6) 外字入力方式との併用により、入力方法の多様化が可能。
- 7) 旧コードで作成されたコード体系間の変換が容易。
- 8) スタンドアローン用端末上での国研外字の利用が可能。

4. 新コード体系に対する試案

既存の外字コードの理論体系を中心に、新コードに必要な機能の枠組を決定する周辺問題を示してきた。本章では新コンピュータ移行に伴う暫定的な外字移行処理を含め、国研独自のコード体系設定の妥当性、大漢和辞典のコードブック利用と部首を中心とした統一配列順序の問題、コード化法とコードに対するビット構造などについて基本的な考え方を述べる。

新コード設計について留意した点は、ホスト・端末間の共通コード機能、伝送路とオペレーティング・システム規格への抵触問題、独立した内部コードとしての役割と JIS、または JIPS コードへの変換アルゴリズムのあり方、そのほか部首配列を基準に、面(Plane)、層(Layer)の擬似的立方体構造への展開方法とビット列への対応、折りたたみ処理の効果などである。

折りたたみとは、面に配当された部首を規準とした内部コードに対し、 $1/2n$ に細分化可能な構造形式を指す(図5)。また、内部コードを決定するに当たって基本としたのは、表6に示した大漢和部首順頻度表をもとに漢字を面に、枝番号を層に、部首を第4の次元として擬似的に対応させる方法である。

面は、現在は2048字分を各部首に配当させているが、面領域で大漢和辞典に収容されている最大部首収容字数の「艸」冠 2036 についても対応でき、214 部首のすべてを表現できることになる。これによって旧国研外字コード

でとっていた擬似的部首表現は不要となる。なお「艸」冠の漢字収容字数は表6で示した「部首頻度表」では、2172字となっているが、表7の補遺、ダッシュ、欠番を再整理すると実数は2036字である。補遺・ダッシュについては、この検字番号をすべて枝番号扱いとし、同一部首内の漢字は総画順に従う原則にもどすことによって、面への配当は問題となることはない。

以下、国研コードに対する基本的な考え方を列挙し、次に面・層、コード内部のビット構造、面の再分割とコード変換処理、内部コードへの展開方法のアルゴリズムについて述べる。

国研コードに対する基本的な考え方

- 1) JIS コードに対し、国研独自のコード体系を設定し、主に内部コードの構造化を図ること。
- 2) 内部コードは、JIS 漢字を包含し漢テレコード・端末・ホスト間のコード変換が容易な構造を持ち、JIS コードへの変換も可能なこと。
- 3) 内部コードは、4 byte とし、第1,2byte で部首を、第3,4byte で部首内相対位置を示すビット列とする。
- 4) 部首内番号は、各漢字間に追加機能を持ち、コード表現は3次元構造とする。層は追加用枝番号表示用とする。

Plane (面) × Layer (層) → 部首内漢字コード

- 5) 各面は、部首に対応させ、利用可能なコンピュータ記憶容量によって、細分可能なこと。また、層は計算による処理が可能なこと。
- 6) 旧コードとの互換性を容易にするため、漢字は部首順配列を基本とする。また、他の辞書との情報交換用に利用できること。
- 7) 漢字入出力表示は、当面 JIS コードの範囲で対応していること。
- 8) 漢字追加用枝番号は、木、網構造への拡張が可能なこと。枝番号の一部は、ポインタによる他の面への指定が可能なこと。
- 9) 枝番号は、ソート時に配列順序が変化しないこと。
- 10) 部首漢字は、JIS 85区または90区以降に一括登録し、両水準間の統一配列コードとして対応可能なこと。(脱字分4文字を含む)

- 11) 外字バッファ領域は、32面 214 部首を基本に置換アルゴリズムの適用が可能なこと。
- 12) 検字コードは、将来の部首再構成に対応できるよう、複数のキーから検索が可能なこと。
- 13) 内部コードは、伝送路規格に抵触しないこと。
- 14) ホスト・端末双方に同一コード体系が利用可能なこと。

4.1 ビット列構造と面・層の関係

コードを表現するビット列は、8 bit 2進表示コード 4列を基本とし、前2列は部首を後2列は部首を基準とした相対番号表示用とする。この4 byteコードは日本電気から標準漢字コード〔JIPS (E)〕として提供される 11525 字に、各々部首順配列後「大漢和辞典」の検字番号に従って新コードを配当する。新国研コードの先頭ビット列 2 byte は、部首表現用としてのものであるが、現行の標準的に使用されている 214 部首に対して、8 bit 分 256 個で対応させ残りの 1 byte は予備とする。また、相対番号表示は第 3 byte 以後の 2 byte で表わし、先頭バイト 8 bit を横方向に、第 4 byte の上位 3 bit を縦方向とする 2048 字を基本とした面構造を第一次単位とする。4 bit 目の 1 bit は当面これも予備とするが、将来枝番号用また面への拡張用ビットの利用が可能である。残る下位 4 bit は第二次単位として、漢字コード間に後で新出漢字を規定の配列順序で追加する場合の枝番号機能を持たせ、2048 を面とする 16 層の立方体コード体系を想定する。さらに、これを第三次単位として前述の部首を加え、全体として 4 次元構造を持ったコード体系と見なす。

図 4 相対表示用コード列構造

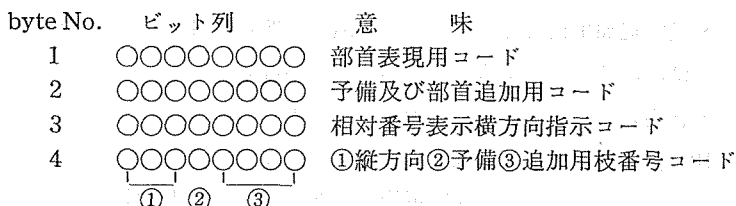
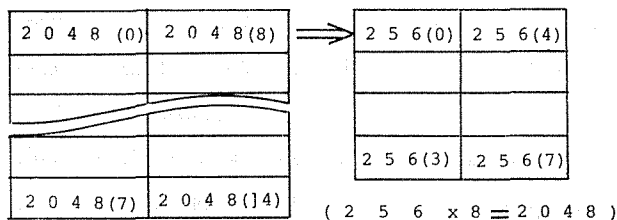


図5 ビット列の面・層への対応



上記のビット列構造は各々4 bit 単位に分割できるため、プログラム作成上、データ内容の16進表示のさい出力される数字との対応は完全に一致する。また図5に示した基本面2048は、ビット列256×8の11bitに分割使用することによって、次の4種の利用法に対応できることになる。

第一の点は、すべて基本面に対し 2^n の整数倍になっているため、ホスト用に使用される面と端末装置側で使用される面は、利用可能なコンピュータの内部記憶容量によって、常に最適領域指定が可能になる。この考え方は、コード定義領域に対し、基本的構造として本稿で提案したコード体系全体の共通した考え方となっている。第二点は、枝番号または面・層について、該当する各構造は「3.3.2 外字入力と枝番号処理」で示したMOD、DIV関数による計算法が利用できることである。

例えば、面を2048を基準とした場合、検字番号に対する該当部分を計算するために2048による剰余・商を求めることで、擬似的に増加させる場合の部首に対しても対応可能となる。同様に、図6で示した内部コード対JIPSコード変換用テーブルの分割処理も、基本面と枝番号を示す層によって表現されているため同じ方法で処理できることになる。これが第三の特徴である。第四は、16bit列が各々4 bit 単位に分離でき、16進表示によるプリンタ出力と16進直接入力が可能になる。

最後に、枝番号に対する新出漢字の配当問題があるが、扱い方を誤ると結局枝番号として付番された漢字間に再度追加しなければならない状態が生じ

る。従来よく使われてきた文字追加方法は、新出漢字を枝番号15個の二分の一の位置へ挿入、次の追加は前に登録された漢字を基準に、総画順序に従い、さらに二分の一の位置へ登録する方法である。枝番号に登録可能な数は $(00)_{16}$ を基本面に当てるため、最大15文字の追加が許されることになる。この枝番号の扱いは、他の方法も含めさらに研究の必要があるが、一例として15文字を越えた場合を想定すると、1 bit を別に定義できる「面」へのリンク情報として残しておくことも可能である。この場合利用可能な枝番号は1文字分少なくなるが、将来シソーラスなどの漢字と漢字間の関係を示す必要が生じるものと思われ、面に対する拡張用ビットの設定は必要な機能であると思われる。

4.2 コード変換処理と折りたたみ処理

新国研コードは国研用に独自に設定した内部コードであるため、JIS、JIPSコードとの互換性はない。これは、入出力可能な漢字はメーカーから提供される標準コードであり、日本電気からJIPSコードとして漢字11525字が提供され、印字処理を行うためには内部コードとして提供されるメーカーの標準コードに変換しなければならない。コード変換は「3.3.1外字入力 of 拡張方法」の中で述べた、表5の変換表を索引として使用し、大漢和辞典にありJIPSコードに登録されていないシステム外字は漢字パターン of 新規登録が行われるまですべて「ゲタ」扱いとする。

また、非漢字はJIPSコードを直接使用しJIPSコードと内部コードの併用方式を採用する。この処理によって外字コード入力指示は、大漢和辞典の検字番号、旧国研外字表現のいずれの方式も処理可能となる。旧コードとの互換性は、表5で示した「コード変換用部首索引表」によって変換処理 of 調整に使用でき、さらに旧コード of 擬似的部首271字分、新コードで使用する214部首の双方方向変換もこの処理の中に含まれる。

図6 新・旧コード変換表

JIPS コード	旧コード	新コード	その他のコード
漢	◇□□	××××	××××
ゲタ	◇□□	××××	××××

なお変換表の作成は、新システムの入力装置で作成可能であり、旧コードとの併合は現行の旧コード表をJISコード変換し、JIPSまたはJISコードをキーとした統一変換表を作成する。しかし、表5で示した索引を使用することによって検字番号から逆算しコード化することも可能になる。旧コードに対しては、従来の変換表を使用する場合と表5による変換の、いずれかの処理効率を測定した上で選択する予定である。

折りたたみ処理は、すべての変換表が主記憶装置内に読み込むことができない場合、折りたたみ処理によって新出漢字に対する枝番号の必要部分を計算により抽出させる。この方式は、大量データを対象とする一括処理では適当とはいえないが、端末装置などの入力操作には十分利用できるものと思われる。

この機能の他への利用法は、データ対象によって、大漢和辞典より収容字数の少ない小型辞書もコードブックとして使用できることである。例えば「新字源」の利用を想定した場合、面を縮小させる効果と検字番号の指示にも対応でき、これによってホスト・端末間の伝送路を中心とした処理と調査対象のデータに合わせて複数のコードブックの利用により、処理の効率化が可能になる。コードブックの作成は現在「大漢和辞典」と「新字源」について併行利用できるよう一部作業を進めている。

2種のコードブックを対象とする方法は、コードの二重化を招くおそれがあるが、現在の調査ではかならずしも5万字の世界を必要とすることは極めて少なく、その点で小型のコードブックとの併用方式を考慮しておくことが処理効率上有効であると思われる。併用方式の採用は、双方向の互換性を維持するため内部コードを中心に共通漢字に「新字源」であることを示す相対

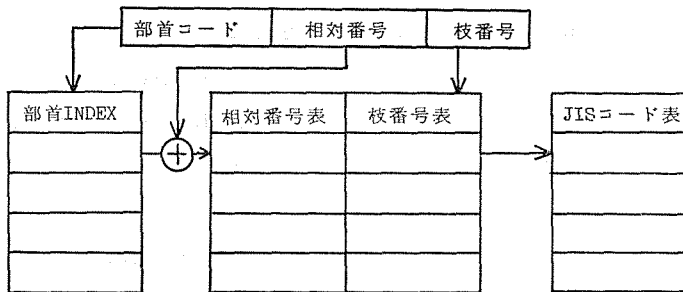
番号を付け、表5に示した部首用検字番号を付加することで対応可能となる（図6・表9参照）。

コード体系は、表5の変換表に新コードブックの検字番号の先頭位置を指定することで、内部コードを共通コードとした双方向のコード変換処理が可能になる。この処理は、新コードを別にコードブック別に設定する必要がなく、かつ複数のコードブック利用への道を開いたものとなる。また、他の辞書類についても、検字番号の付加された辞書類であれば同様に利用でき、写植などで作成された辞書の場合、内部コードとの結合は簡単になる。

さらに検字番号の複数指定の拡張は、現行の康熙字典で定められた部首から漢字利用の実態に合った新部首の導入の可能性を含めた処置にも繋がる。この機能は、漢字の持つ字形を中心とした検字がより使い易い方法であり、漢字の意味の実態から離れた基準より、実務面で妥当であると思われるからである。図7は国研コードまたは、作業の部首別コードブックを使用した場合の印字可能な文字への変換方法を示したものである。

図7 入力のためのコード変換手順

国研コード → JISコード



5. おわりに

本稿は、国研の外字コードについて、昭和54年（1979年）6月から、昭和

59年(1984年)7月までに、5回にわたって漢字プリンタ装置、入力装置導入とコード体系上の問題を部内報告「季報」に述べたことをもとにまとめたものである。昭和54年6月号では、高速漢字プリンタ装置導入(NEAC—5210D)に伴う、国研コード対日本電気コードに対する物理コードのコード変換上の問題。同年10月号では、新漢字入力装置(N6300—50N)導入とJISコード対応の問題。特に、漢テレコードの物理コードと外字表現との関係について整合性を述べ、国研の外字コードの基準に大漢和辞典をコードブックとして使用する場合の妥当性の検討。昭和58年(1983年)8月号と11月号では、JISコードを利用する上での問題点、各水準間の統一理論コードの必要性、自由領域への文字登録による併害の指摘と対応案を示した。さらに、昭和59年(1984年)7月号では、以上の経過をもとに、昭和60年3月に導入される新システム(ACOS—S550)への外字入出力方法、内部コード表現形式、新・旧コード及びJISコード間の関係について基本的な考え方と、国研独自の新コード体系の設定が必要であることを述べた。

また、この原稿を執筆している間に「JISC6228情報変換用符号の拡張法」の改正作業が進められていることを知った。改正前の1975年度版では、複数バイトに対する拡張法が定義されていなかったが、1984年版の改正案では従来のコード領域であるG0領域(2121)₁₆～(7E7E)₁₆にコードの拡張法が盛り込まれている。この図形文字集合の拡張に関する内容の詳細は、改正版の発表を待たなければならないがシフトコード法が使われるようである。

改正版による拡張コードは、現在のG0領域からG3領域までの4領域、35344字分が利用でき、改正案が制定された場合「4.1ビット列構造と面・層の関係」で述べた外字コードへのシフトは、JISによるシフトコードに従うことで、JIS規格内で対応可能となる。同様に、中文用コード「GB3212—80」コードについても、コード体系はISO規格に従っており拡張領域のいずれかをそのまま利用できる道がつけられるものと思われる。そのほか、新国研外字コードに対するJIS世界への対応も、基本的にはコード変換処理を前提にしているが、この場合C6226コードは外部へのデータ接続用であると

使用目的を絞ることで、より大きな開かれた漢字世界への対応が可能づげができる。現に調査ごとに大漢和辞典に収容されていない漢字が新出コードとして登録されている実情では、漢字コードに追加機能を持った独自コードを設定する以外に解決案はないと考える。

本稿で示したコード化法は、結局のところ文献 2 で示された松本案の枠組を出ることができなかった。ある意味では、単に JIS 領域への再配当案を提示したにすぎないとも言える。今後さらに時間をかけて試案の問題点を明確にし、実用に耐えられるコード拡張化を進めることが必要である。その点で、C6226及びC6228規格ともに、外字コードに対する統一理論コードへの対応が基本的に抜けているように思われる。特に、拡張法と漢字コード配列法との関係は不可分であり、統一的な配慮のもとでの検討が必要である。

また、現行の 7 bit 2 列のコード体系から、8 bit 2 列のコードへの拡張もこれらの検討事項の中に入れるべきと思われる。これは、本稿で述べた試案で、区点の中に再配列する方法をとったが、コード領域を広げることによって 2^n の整数倍、または折りたたみ方式の導入が可能になり、この方式が全体としてコード体系の構造化への道を開くことに繋がったからである。

例えば、第 1 byte 目の 7 bit を第一次の面とし、第 2 byte の 3 bit を層として、128 字の 8 層計 1024 字に対応させ、これをさらに 4 bit 分 16 層を加え 4 次元構造を基本とするコード体系を考える。この表現法が可能となれば、次の 3 種の利点が得られることになる。

第一は、すべて基本面に対し 2^n の整数倍になっているため、ホスト・コンピュータと端末用に使われる面は、利用可能なコンピュータの内部記憶容量によって折りたたまれた必要部分を指定できること。第二は、枝番号または面・層に該当するコード体系の構造について、MOD・DIV 関数の計算で目的とするコード位置を求めることが可能になること。第三は、検字番号に対する該当部首を計算するために、基準領域に対しその値の剰余、商を求めることで擬似的に増加させた部首に対しても、枝番号を求めたと同様の計算対応が可能になることである。これらの 3 点は、外字コード設定について同

一の思想と計算法の反復ですべて対応可能になることは考慮すべき点であると思われる。試案は、新機種の移行と旧コード体系が漢テレ保存上から維持不可能になったため、急ぎまとめたものであり基本的な部分で抜けがある可能性がある。これらの問題点のいくつかは、今後個々の実験によって結果の確認を行う必要があるが、新コードが長期間のデータ保存と新出漢字に対する対応を十分に機能させ、実用化を図るためには慎重な態度をとらなければならない。

最後に、新コードに対する試案を示す過程で、一橋大学教授松本昭氏には、旧コード設計のさいの基本的な考え方を伺った。また、言語計量研究部第二研究室長野村雅昭氏には旧コードに関する種々の資料の提供を受けた。そのほか、新システム移行と新コード設計に関し、日本電気の漢字コードとの整合性の確認は、日本電気官庁システム事業部第二システム部一杉宏一主任に色々の面でお世話になった。ここに記して謝意を表する。

(1984.10.31)

参考文献

- 1)『漢テレコードブック』,国語研究所,1967。
- 2)松本昭,国研用漢字テレタイプと同機利用の言語情報処理,国語研報告31, pp. 57-89,1968.3。
- 3)斎藤秀紀,言語処理におけるターンアラウンド・システム,国語研報告59, pp. 63-111,1976.3。
- 4)斎藤秀紀,多目的漢字入力システムの試案,国語研報告61, pp. 41-56,1978.3。
- 5)斎藤秀紀,分散システムへの試み,国語研報告67, pp. 73-88,1980.3。
- 6)斎藤秀紀,高田誠,日本語とパーソナルコンピュータ,電子通信学会誌4, Vol. 67, No. 4, pp. 418-423,1984.4。
- 7)斎藤秀紀,国研コードとJ I Sについて,国語研内部資料「季報」,1979.6~10。
- 8)斎藤秀紀,国研コードの外字コードについて(1)~(3),国語研内部資料「季報」,1983.3~1984.7。
- 9)野村雅昭,J I S-C 6226情報交換用漢字符号系の改正,標準化ジャーナル, pp. 4-9,1984.3。
- 10)田島一夫,日本語情報処理における文字セットコントロールシステム,情報管理,

Vol. 26, No. 7, pp. 554-567, 1983. 10。

- 11) 『情報交換用漢字符号系 J I S C 6226-1983』, 日本規格協会, 1984。
- 12) 『中華人民共和国国家標準 信息交換用漢字編碼字符集 基本集 G B 2312-80』, 技術標準出版社, 中国北京, 1981。
- 13) 『講習会テキスト J I P S の導入』, 日本電気講習会資料 C E -440-2。
- 14) 諸橋轍次編, 大漢和辞典, 1971。

表 6 部首頻度表（諸橋轍次編「大漢和辞典」第 3 印刷）

画数	部首	度数	画数	部首	度数	画数	部首	度数	画数	部首	度数
6	艸 ^{...}	2172	9	食 ^食	488	3	弓	223	4	殳	111
4	水 ^氵	1817	7	車	472	6	耳	220	8	長	110
4	木	1618	2	刀 ^刂	450	10	骨	218	13	鼠	109
3	口	1475	9	頁	439	12	黑	215	6	耒	103
4	手 ^扌	1322	6	革	404	2	力	211	3	尢 ^尢	101
4	心 ^忄	1274	8	阜 ^阌	400	3	大	210	4	片	97
6	虫	1181	3	巾	399	5	瓦	210	4	方	96
6	竹	1025	6	米	396	7	見	210	5	矢	95
2	人 ^亻	1008	3	山	366	7	角	206	9	面	95
7	言	1007	7	酉	363	3	口	202	4	月	91
8	金	986	8	雨 ^雨	362	10	鬼	201	6	缶	91
6	糸	923	4	文 ^父	351	3	尸	198	10	鬲	90
11	鳥	909	3	广	348	6	羊	196	4	斤	87
3	女	904	7	貝	345	4	戈	174	7	豆	87
6	肉 ^月	869	8	門	336	11	麥	172	2	儿	86
4	火 ^灬	839	7	走	335	2	厂	171	3	夕	86
3	山	820	4	牛	320	5	皿	167	5	矛	84
3	土	790	5	穴	320	7	豕	164	2	十	81
5	目	763	8	隹	289	7	豸	162	2	勺	78
11	魚	683	5	田	286	5	白	147	2	匚	88
6	衣 ^衤	681	5	示 ^礻	284	7	身	137	3	月	73
7	足	678	10	影	280	11	鹿	135	3	彡	76
5	疒	649	4	欠	266	6	虎	133	4	戸	74
7	辵 ^辵	626	3	彳	262	2	辵	130	4	爪 ^爪	72
5	石	614	4	歹 ^歹	255	8	韋	128	5	瓜	69
4	犬	592	4	毛	255	3	子	126	9	音	69
5	玉 ^玉	571	6	舟	249	5	立	120	1	一	68
4	日	558	9	風	242	2	又	122	3	少	67
10	馬	539	6	羽	249	5	皮	122	6	血	67
5	禾	513	15	鹵	238	6	臼 ^臼	116	12	黄	67
7	邑 ^阌	511	6	网 ^四	232	4	止	111	3	寸	66

画 教	部首	度教	画 教	部首	度教	画 教	部首	度教
1	乙	63	2	一	42	7	里	24
2	冂	63	7	赤	42	1	亅	23
2	几	63	3	勹	41	4	母	23
2	卜	63	11	麻	40	5	疋	23
6	舌	63	2	二	39	5	夂	23
6	行	63	4	文	38	5	用	22
7	谷	63	6	而	38	5	内	22
14	鼻	63	3	戈	36	5	聿	22
2	厶	62	6	色	36	14	齋	22
2	乚	61	2	入	35	4	爻	21
1	ノ	60	5	生	35	7	采	21
4	片	60	16	龜	35	17	侖	21
13	目	60	3	幺	34	3	夊	19
13	鼓	60	8	非	33	4	氏	18
3	𠂔	59	6	至	32	7	辰	18
11	鹵	59	9	首	32	3	廴	17
2	八	57	3	己	31	4	无	17
2	冂	55	4	比	31	6	舛	16
3	士	55	8	韭	31	1	丶	15
9	香	54	10	門	30	4	父	15
4	曰	53	2	乚	28	13	鼎	15
7	辛	53	3	工	28	8	隶	14
12	黍	51	6	老	28	9	飛	14
3	《川	50	7	臣	28	10	鬯	14
7	自	49	16	龍	28	4	牙	13
4	支	48	5	甘	27	12	𠂔	9
2	凵	45	6	而	27	5	玄	8
3	中	45	3	干	26	6	艮	6
4	斗	45	1	丨	25			
4	气	43	2	匸	25			
10	高	43	8	青	25			

表7 部首別度数 総数+補遺数+ダッシュ数-欠番数=部首実数

部首	総数	補遺数	ダッシュ数	欠番数	合計	部首	総数	補遺数	ダッシュ数	欠番数	合計
一	66	2			68	土	771	14	7	2	790
丨	24	1			25	士	53	2	1	1	55
丶	15				15	夕	17	1	1		19
ノ	55	6		1	60	乚	41				41
乙	63				63	夕	82	3	1		86
丿	23				23	大	205	5	1	1	210
二	39				39	女	894	7	3		904
ナ	58	3			61	子	124	3		1	126
人	992	10	6		1008	ㇿ	357	8	1		366
儿	79	6	1		86	寸	62	1	3		66
入	35				35	小	65	1	1		67
八	56	1			57	尢 ^允	91	10			101
冂	59	4			63	尸	196	2	1	1	198
冫	42				42	屮	44	1			45
勹	130				130	山	800	17	3		820
几	63				63	巛 ^川	45	3	2		50
凵	45	1			46	工	28				28
刀	443	5	3	1	450	己	29	1	1		31
力	205	3	3		211	巾	394	5			399
勹	77		1		78	干	24	2			26
乚	25	3			28	么	32	2			34
匚	77				77	广	345	3	1	1	348
匚	23	2			25	乚	17				17
十	79	1	1		81	井	73	4			77
卜	62	1			63	弋	36				36
卩	54	1			55	弓	216	5	2		223
冂	167	3	1		171	彡 ^彡	54	4	1		59
彡	58	4			62	彡	75		1		76
又	112	9	1		122	彡	258	1	3		262
口	1448	17	10		1475	心 ^忄	1235	26	14	1	1274
口	192	2	8		202	弋	166	6	3	1	174

部首	総数	補遺 数	ダッ シュ 数	欠番 数	合計	部首	総数	補遺 数	ダッ シュ 数	欠番 数	合計
戸	72	1	1		74	犬	580	8	6	2	592
手 _子	1293	15	17	3	1322	玄	7		1		8
支	47	1			48	玉 _王	550	21	2	2	571
支 _文	342	6	4	1	351	瓜	67	2	2	2	69
文	39			1	38	瓦	205	6	1	2	210
斗	45				45	甘	27				27
斤	86	1	1	1	87	生	33	1	1		35
方	96		1	1	96	用	20	2			22
无	17				17	田	271	13	3	1	286
日	545	8	6	1	558	疋	21	2			23
曰	52	1			53	疒	642	7			649
月	85	1	5		91	𠂔	21	1	1		23
木	1576	33	12	3	1618	白	145	2			147
欠	262	2	5		166	皮	118	4			122
止	106	4	3	2	111	皿	164	2	2	1	167
歹 _夕	254	1	1	1	255	目	741	23	3	4	763
殳	108	2	2	1	111	矛	83	1			84
母	22		1		23	矢	95	1		1	95
比	29	2			31	石	599	11	5	1	614
毛	254	1			255	示 _ネ	261	8	15		284
氏	17	1			18	内	22				22
气	40	1	2		43	禾	500	6	9	2	513
水 _子	1767	37	16	3	1817	穴	315	6	2	3	320
火 _{...}	803	30	10	4	839	立	120	6	5	5	126
爪 _爪	68	4	2	2	72	竹	991	33	9	8	1025
父	16			1	15	米	389	10	3	6	396
爻	21				21	糸	887	30	12	6	923
片	55	5			60	缶	91	2		2	91
片	96	1			97	网 _四	226	8	1	3	232
牙	13				13	羊	189	8		1	196
牛	312	8	2	2	320	羽	228	6	6	1	139

部首	総数	補 数	ダッ シュ 数	欠 番 数	合計	部首	総数	補 退 数	ダッ シュ 数	欠 番 数	合計
老	29			1	28	走	331	3	1		335
而	27				27	足	669	8	3	2	678
耒	101		2		103	身	138			1	137
耳	216	3	1		220	車	458	10	4		472
聿	21	1	1	1	22	辛	52		1		53
肉 _月	832	31	12	6	869	辰	18				18
臣	27	1			28	走 _レ	569	19	50	3	626
自	47	1	1		49	邑 _邑	494	13	5	1	511
至	31	1			32	酉	352	9	2		363
臼 _白	104	12	2	2	116	采	16	4	1		21
舌	61		2		63	里	21	3	2	2	24
舛	12	3	1		16	金	948	33	6	1	986
舟	246	3			249	長	108	2			110
艮	6				6	門	326	8	2		336
色	36				36	阜	389	5	7	1	400
艸 _艸	2036	101	40	5	2172	隶	14				14
虎	130	1	4	2	133	隹	273	15	1		289
虫	1160	22	4	5	1181	雨 _雨	354	6	4	2	362
血	65	2	1	1	67	青	21	1	3		25
行	62		2	1	63	非	33				33
衣 _衣	671	8	5	3	681	面	92	3			95
而	34	2	2		38	革	398	6			404
見	207	3	1	1	210	韋	128				128
角	202	4	1	1	206	韭	29	1	2	1	31
言	977	10	22	2	1007	音	68		1		69
谷	63				63	頁	423	12	5	1	439
豆	89			2	87	風	244	1	1	4	242
豕	162	3	1	2	164	飛	14				14
豸	160	2			162	食 _食	475	4	9		488
貝	337	7	3	2	345	首	29	2	1		32
赤	41	1			42	香	54				54

部首	総数	補遺 数	ダッ シュ 数	欠番 数	合計	部首	総数	補遺 数	ダッ シュ 数	欠番 数	合計
馬	526	8	8	3	539	黍	47	4			51
骨	215	2	1		218	黑	210	4	1		215
高	42	1			43	衛	9				9
髟	277	2	1		280	𥇑	58	2	1	1	60
門	29		1		30	鼎	15				15
鬯	11	3			14	鼓	60				60
鬲	86	4			90	鼠	108	1			109
鬼	198	2	1		201	鼻	62		1		63
魚	678	4	2	1	683	齋	23			1	22
鳥	894	15	2	2	909	齒	235	3	1	1	238
鹵	58	1			59	龍	29			1	28
鹿	131	4			135	龜	35				35
麥	170	2			172	龠	21				21
麻	39		1		40						
黃	65	1	1		67	合計	48,902	1,062	504	157	50,311

表 8 漢テレコード順序と JIS コードの関係

No.	外字表示漢字	16進表示コード	JIS 表示コード
000001	計	0 0 0 1	3 7 5 7
000002	形	0 0 0 3	3 7 4 1
000003	型	0 0 0 5	3 7 3 F
000004	経	0 0 0 7	3 7 5 0
000005	芸	0 0 0 9	3 7 5 D
000006	頭	0 0 0 E	4 6 2 C
000007	決	0 0 1 1	3 7 6 8
000008	結	0 0 1 3	3 7 6 B
000009	月	0 0 1 5	3 7 6 E
000010	建	0 0 1 7	3 7 7 A
000011	見	0 0 1 9	3 8 2 B
000012	件	0 0 1 B	3 7 6 F
000013	県	0 0 1 D	3 8 2 9
000014	伝	0 0 2 1	4 5 4 1
000015	電	0 0 2 3	4 5 4 5
000016	都	0 0 2 5	4 5 5 4
000017	土	0 0 2 7	4 5 5 A
000018	度	0 0 2 9	4 5 5 9
000019	当	0 0 3 1	4 5 7 6
000020	投	0 0 3 3	4 5 6 A
000021	東	0 0 3 5	4 5 6 C
000022	党	5 5 3 7	4 5 5 E
000023	島	0 0 3 9	4 5 6 7
000024	等	0 0 3 B	4 5 7 9
000025	愛	0 2 0 1	3 0 2 6

—以 下 略—

表9 コード変換用テーブル

文字	JIS	盤内	盤外	日立	NEC
一	306C	1923	0001/0001	(◆計計)	B0EC 243B
丁	437A	2823	0001/0003	(◆計形)	C3FA 2536
七	3C37	1935	0001/000E	(◆計頭)	BCB7 2441
万	4B7C	1521	0001/0017	(◆計建)	CBFC 2541
丈	3E66	181B	0001/0019	(◆計見)	BEE6 3147
三	3B30	1927	0001/001B	(◆計件)	BBB0 243D
上	3E65	1819	0001/001D	(◆計県)	BEE5 3145
下	323C	0407	0001/0021	(◆計伝)	B2BC 3A23
不	4954	1227	0001/0031	(◆計当)	C9D4 314C
与	4D3F	360E	0001/0033	(◆計投)	CDBF 3D50
丐	5022	0A0A	0001/0037	(◆計党)	D0A2 236A
丑	312F	0D29	0001/0039	(◆計島)	B1AF 4524
且	336E	0E28	0001/0209	(◆計以)	B3EE 2932
丕	5023	0A0A	0001/020E	(◆計世)	D0A3 6121
世	4024	020E	0001/0211	(◆計位)	C0A4 373E
丘	3556	2904	0001/0215	(◆計違)	B5D6 3327
丙	4A3A	3328	0001/0219	(◆計引)	CABA 334D
丞	3E67	3907	0001/0225	(◆計新)	BEE7 4525
両	4E3E	1823	0001/0237	(◆計水)	CEBE 3F57
並	4A42	1326	0001/030E	(◆計石)	CAC2 334F
个	5024	0A0A	0003/0337	(◆形政)	D0A4 236C
中	4366	0831	0003/0401	(◆形與)	C3E6 263A
卅	5025	0A0A	0003/0409	(◆形化)	D0A5 7021
串	367A	0A0A	0003/0413	(◆形加)	B6FA 4E35
、	5026	0A0A	0005/0431	(◆型鋭)	D0A6 236E
丸	345D	0708	0005/0437	(◆型先)	B4DD 2F31
丹	4330	2724	0005/0505	(◆型歌)	C3B0 3E45
主	3C67	1611	0005/0507	(◆型画)	BCE7 293F
井	5027	0A0A	0005/0509	(◆型回)	D0A7 5335
丿	5028	0A0A	0007/000E	(◆経頭)	D0A8 236F
乂	5029	0A0A	0007/0019	(◆経見)	D0A9 2460
乃	4735	3C38	0007/001D	(◆経県)	C7B5 4526
ノ	213A	0A0A	0007/0025	(◆経都)	A1BA
久	3557	071D	0007/0029	(◆経度)	B5D7 3323
之	4737	0E33	0007/0201	(◆経愛)	C7B7 4527
乍	4663	0A0A	0007/020E	(◆経世)	C6E3 5336
乎	3843	0A0A	0007/0211	(◆経位)	B8C3 4528
乏	4B33	3432	0007/0215	(◆経違)	CBB3 3651
乘	502B	0A0A	0007/021D	(◆経右)	D0AB
庸	6949	0A0A	0007/023B	(◆経井)	E9C9 3000
幸	502A	0A0A	0007/0301	(◆経云)	D0AA 7023
乘	3E68	1903	0007/0309	(◆経形)	BEE8 3445
乙	3235	2404	0009/031D	(◆芸横)	B2B5 282B
九	3665	1939	0009/0331	(◆芸声)	B6E5 2443
乞	3870	3218	0009/0337	(◆芸政)	B8F0 4529
也	4C69	3628	0009/0339	(◆芸勢)	CCE9 452A
乱	4D70	3737	0009/0423	(◆芸切)	CDF0 3E54

1バイトコード

2バイトコード

第2バイト

00 21 7E A1 FE FF(16進)

0 F

0 F

(JISC-6220)

第1バイト

G0 集合

G1 集合

65,536字 (256×256)

8,836字×2=17,672字 (94×94)

FE FF (16進)

第2バイト

21 7E

1 94点

A1 FE

1 94点

第1バイト

G0 集合

G1 集合

7E 94

FE 94

区

区

The diagram illustrates the system architecture, divided into two main horizontal sections by a dashed line. The top section contains two database boxes: '新MTデータ・ベース' (New MT Data Base) on the left and '現行MTデータ・ベース' (Current MT Data Base) on the right. The left database is associated with '規定されたフォーマット' (Specified format) and '国研コード (1種類)' (RIKEN code (1 type)). The right database is associated with '旧来のフォーマットコード4種類' (Old format code 4 types). Below these are two processing blocks: '④ コード変換機能' (Code conversion function) on the left and '③ コード変換' (Code conversion) on the right. The right block also includes '形式変換' (Format conversion) and is associated with '旧NEC(C5210D)' (Old NEC (C5210D)). Both processing blocks are connected to '国研:JIS対応表' (RIKEN: JIS correspondence table) cylinders. A '変換ツール' (Conversion tool) block is connected to the right processing block and the 'JIPsIE' cylinder. The bottom section features a 'JISコード' (JIS code) cylinder labeled '規定されたフォーマット' (Specified format), which is connected to a 'データ・ベース' (Data base) cylinder. The 'JIPsIE' cylinder is connected to the 'JISコード' cylinder. A power symbol is connected to the '形式変換' block, with the label 'N6300-50N系より' (From N6300-50N series).