寄　書

漢字のネモニックコード化への試み*

木　沢　誠**

漢字を含む日本語文を情報処理機械で取り扱おうとするとき，まず問題になるのが入出力装置で，字種数の多いことがその原因であることはいうまでもない．とりわけその人間くさい部分，すなわち，原始データの入力を行なう装置は，とかく取り残されがちな気配である．

現在使用されている漢字テレタイプ，またはモノタイプの鍵盤さん孔機が別に悪いということはないが，打鍵に際して，欧文タイプやカナタイプのような盲打ちができないこと，操作速度の遅い人間が専用する装置が，複雑大形で高価，かつ持ち運び不便であることなどのハンディキャップは，もう少し何とかならないかと感ぜられる．

ここに紹介するのは，その打開策としての筆者の思いつきである．

この着想は，簡単にいえば，平生用いる漢字をカタカナ4字によってコード化し，カナモジ用鍵盤さん孔機によって，テープさん孔しようというものである．日本語の文中には，漢字とカナとがまじっているから，その変わり目に SO や SI と同様の性質の漢字シフト，またはカナシフト（実際にはひらがなシフトとカタカナシフト）の機能コードを入れ，漢字に対しては，その字に対する4字のコードを，カナに対しては，対応するカナの鍵をそれぞれ打鍵するのである．人間が直接操作する装置としては，本質的に欧文用の鍵盤さん孔機程度のもので間に合うし，カナタイプに習熟した人ならば，漢字のコードさえおぼえていれば，漢字鍵盤の場合のように，文字の発見のために眼を原稿から離し，かつ時間を費やすこともなく，連続して盲打ちで打鍵できるはずである．

このようにして作成した紙テープは，適当な処理機械にかけて，処理に便利なコードに変換して使用すればよい．自動機械で処理できる部分は高速でもあるし，多少経費がかかること以外は，さほど心配する必要はあるまい．さん孔の校正には文字ディスプレイ装置も使用できよう．

この方法で一番問題になるのは，いうまでもなく，漢字にどのようなコードを与えるかということである．このコードは平生用いる漢字について容易に記憶でき，打鍵中になかば反射的に思い出せるものでなければならない．そのような趣旨から，次のような法則を設定する．

〔**基本法則**〕 漢字を表わすコードのカタカナ4字のうち，前半の2字には，その漢字の音読みの最初の2字を，後半の2字には，その漢字の訓読みの最初の2字をそれぞれあてる．ただし

（1） 濁音字は次の（2）の場合を除き，濁点を取り去って清音字として取り扱う．

（2） 音読みまたは訓読みが，濁音字1字のみの場合に限り，濁音字を使用する．この場合，濁点を清音字の次の1字として数える．

（3） 音読みまたは訓読みが，清音字1字のみの場合には，第2字目はスペース（以下△印で表示する）と考え，これに該当する文字コードを入れる．

（4） ○ュウという音は，○ウと書き換えて取り扱う．

たとえば，“山”という漢字の音読みと訓読みとは，それぞれ“サン”および“ヤマ”であるから，コードは“サンヤマ”となる．また，“情”（音ジョウ，訓ナサケ），“五”（音ゴ，訓イツツ）のコードは，それぞれ“シヨナサ”，“ゴイツ”である．

この基本法則で万事を律しきれれば，話しは簡単であるが，実際には種々の問題が発生する．そこで，まず当用漢字を取り上げて，実際にコード化を適用してその問題を考えることにしよう．

当用漢字1,850字のうち，当用漢字音訓表の中で，その音読みと訓読みとを両方もっているものは969字

---

* Toward the Mnemonic Coding of Chinese Characters, by Makoto Kizawa (Electrotechnical Laboratory)

** 電気試験所

である．これらについては，基本法則に従って一応コードを作ることが容易である．ここで

〔附則 1〕 同一の漢字について，定められた法則に従い，普遍性の高い知識によってコード化をするとき，2種以上のコード付けが可能のときには，識別性を損わない限り，同一の漢字に2種以上のコードが存在することを妨げない．

と定めれば，2種以上の読み方のある漢字について，コード化に迷う心配が少ない．このコードは入力専用であるから，同一の漢字に多種のコードが存在しても，それによって他の漢字とまぎれない限り，コード変換のプログラムに関係するのみで，利用上の支障はない．

これら969字のコード化において問題になるのは，2字以上に対してコードが同じになる次の各組である．

①旗，機（キ△ハタ）	②共，供（キヨトモ）
③型，形，傾（ケイカタ）	④肩，堅（ケンカタ）
⑤行，興（コウオコ）	⑥高，耕（コウタカ）
⑦志，試（シ△ココ）	⑧使，仕（シ△ツカ）
⑨収，修（シウオサ）	⑩上，植（シヨウエ）
⑪小，緒（シヨオ△）	⑫床，所（シヨトコ）
⑬振，震（シンフル）	⑭送，贈（ソウオク）
⑮倉，蔵（ソウクラ）	⑯中，仲（チウナカ）
⑰頭，当（トウアタ）	⑱道，導（トウミチ）

これらのうち②，⑤，⑩，⑪，⑫，⑰については，特例として，次記のカッコ内の読み方をコード化の対象から除くことにすれば解決できる．

②供（トモ）	供→キヨソナ	共→キヨトモ
⑤行（オコナウ）	行→コウイク	興→コウオコ
⑩上（ウエ）	上→シヨカミ	植→シヨウエ
⑪小（オ）	小→シヨコ△	緒→シヨオ△
⑫床（トコ）	床→シヨユカ	所→シヨトコ
⑰頭（トウ）	頭→ズアタ	当→トウアタ

③，④，⑥，⑦，⑧，⑬のように，訓読みが3字以上のものがあれば，特則として，その第1字と第3字をとって（傍点は変えた字）

③型→ケイカタ　形→ケイカチ　傾→ケイカム
④肩→ケンカタ　堅→ケンカイ
⑥高→コウタカ　耕→コウタヤ
⑦志→シ△ココ　試→シ△コロ
⑧使→シ△ツカ　仕→シ△ツエ
⑬振→シンフル　震→シンフウ

のように区別しよう．

⑭，⑮，⑱については，特則として音読み中の濁字を採用して

⑭送→ソウオク　贈→ゾオク
⑮倉→ソウクラ　蔵→ゾクラ
⑱道→トウミチ　導→ドミチ

のように定めよう．また⑨，⑯は音読みの表現を変えて

⑨収→シウオサ　修→シユオサ
⑯中→チウナカ　仲→チユナカ

のようにできる．

①が残ったが，後述の附則2ともにらみあわせ，機が木ヘンであることから最後にキの字を入れて

①旗→キ△ハタ　機→キ△ハキ

とでもして区別することにしよう．

結局，特例として読み方を制限するもの6字，特則によるもの 13 字であるが，この程度の例外は記憶にさほど負担となるとは思われない．

次に，当用漢字音訓表の中に音読みがなく，訓読みのみがある漢字は 30 字である．これらは原則としてコードの最初の2字をスペース（△△）とすれば解決するが，（畝，瀬）の組合せのみは，識別のためやむを得ず音訓表にない音読みを前者に適用して

畝→ホ△セ△　瀬→△△セ△

とする．また，沖，貭，貝，株，津，矢などの字の音読みは音訓表にないが，やや知識のある人は知っているので，次のように2とおりのコードにする方が実用上便利であろう．

沖→△△オキ，チウオキ　貭→△△オソ，グ　オソ
貝→△△カイ，ハイカイ　株→△△カフ，シユカフ
津→△△ツ△，シンツ△　矢→△△ヤ△，シ△ヤ△

当用漢字のうち上に取り扱った 969＋30＝999 字を除く 851 字については，音訓表に音読みしか与えられていない．これらは一般に多数の字が同音になる場合が多く，たとえば，音読みがカンで訓読みの与えられていない字は，刊官棺館簡閑完感憾看歓観勘漢乾患喚敢款監艦鑑寛緩環還甲と，実に 27 字にも及んでいる．したがって，訓読みからくるコードを△△としては区別ができない．しかし，ここで当用漢字音訓表による拘束を緩和して，高校卒程度の学力で，通常事実上知っていると思われる，または覚えてもさほど不当でない訓読みを採用すれば，ここに挙げた 27 字のうち，次の 18 字についてはコード化ができる．

棺（ひつぎ）→カンヒツ　館（やかた）→カンヤカ
閑（ひま）　→カンヒマ　完（まったく）→カンマツ
憾（うらみ）→カンウラ　歓（よろこぶ）→カンヨロ

漢（あや）→カンアヤ　乾（かわく）→カンカワ
患(わずらう)→カンワス　喚（よぶ）→カンヨフ
敢（あえて）→カンアエ　艦（ふね）→カンフネ
鑑（かがみ）→カンカカ　寛（ひろし）→カンヒロ
緩(ゆるやか)→カンユル　環（わ）→カンワ△
還（かえる）→カンカエ　甲（かぶと）→カンカフ

このほか，看と観もミルという訓読みができるが，区別ができないので一応ここからははずしておく．

これと同様の方法でコード化できる当用漢字は 480 字余ある．ただし，華看観処仁定貞附の 8 字には特則が必要である．残る約 370 字に対しては，観点を変えて次の附則を適用する．

〔**附則 2**〕 適当な訓読みがなく，基本法則によって解決しがたいものに対しては，訓読みの代わりに，部首の呼び名を用いる．たとえば

刊（刀部）→カンカタ　官（ウ部）→カンウ△
簡（竹部）→カンタケ　感（心部）→カンココ
勘（力部）→カンチカ　款（欠部）→カンアク
監（皿部）→カンサラ

ただし，部首名といっても，たとえば医（酉部）や厳（口部）のように，字体の変化によって古来の部首からはずれてしまっているものもあるので，古来の部首にこだわらずに，現在の字体によって，部首に準ずるものを適当に定めればよい．また，識別能力を増すために，日（ニチ），頁（オカ），阝(右，オサ)，酉(サケ）などのような若干の特則を設けた方がよい．

この方法に対して，特則により識別しなければならないのは（塔堂），（俳倍）の 2 組 4 字と“刑”である．

とにかく，このようにして当用漢字 1,850 字のネモニックコード化ができる．

同様の方法によって，人名用漢字別表の 92 字，および当用漢字補正試案により追加する 28 字（うち 4 字は人名用と重複）についてもコード化ができる．この際特則による識別を要するものは(佳嘉)，(宏弘浩)の 2 組 5 字である．

当用漢字以外の漢字に対しては，まだよく検討してないが，基本原則と附則 2，時には若干の特則の適用によって，かなりの程度までコード化が可能であろうと予想される．この際，そのコード化の基底となる読み方を知るのにやや高い学力を要し，したがって，記憶しやすさがそこなわれるおそれはあるが，しかし，そもそも当用漢字以外の漢字は，なるべく使用しないことを主旨とするならば，時としてやむを得ず使用したとしてもその頻度は低く，そのたびにコード表を繰っても，全体から見れば大した支障にはならないであろう．

もう一つの考え方は，打鍵されたデータはいずれにせよ校正を必要とするから，打鍵のときにすぐにコードのわからない文字は，とりあえず活版印刷のゲタに相当するもので代用して，後刻校正の際に直せばよい．要は最初から完全を期すことではなく，全体として完全な状態に早く到達させることである．

カタカナの字種数は 46 で，これに濁点とスペースとを加えると 48 になる．$42^2=2304$，$48^3=110592$，$48^4=5308416$ などから考えると，数千種の漢字のコード化にカナ 2 字では不足し，少なくも 3 字は必要で，ネモニックにするためにコードの種類の使用効率が非常に低くなることを考えれば，4 字（1 字 6 ビットとして 24 ビット）まで使用するのはやむを得ないと思う．ちなみに漢字の音読みの種類はきわめて限られており，15,000 字程度を収容する漢和字典によっても現代の発音では 314 とおりに過ぎず，これから作られるコードの第 1，2 字の組合せは 220 とおりしかない（コードの第 2 字はイウキクチツヤユヨンの 10 字およびスペース，ならびに濁点の 12 とおりのうちの一つしかない）．それに，ヲ，ン，濁音，スペースが第 3 字に位置し得ず，かつ濁音字は 20 字しかないことと，△△の場合とを考慮すれば，この方法で作りうるコードの種類は，例外の特則を作らない限り，たかだか $221(44\times47+20)=461448$ とおりである．

実際の鍵盤ではカナが 1 段にはいりきらず，たとえば，ヌ，ロなどがそれぞれネ，ルなどと同じ鍵になることがある．これに伴って多少特則を必要とすることがあるかも知れない．

漢字のコード化法としては，他にも考えられると思うが，日本人で高校卒程度以上の学力を基準とすれば，主として読み方に基づく上記の方法は反射的に近く敏速で，かつ特に新たに暗記する事項が少ないなどの利点があり，実行に本質的な障害は少ないのではないかと思う．

この着想は 5～6 年前からもっていたのであるが，世間一般に日本語文の機械処理の気運が低かったために，発表する勇気が出せないままとなっていた．漢字処理特集号を機に，しるしてご参考に供する次第である．

最後に余興としてコードによって次の言葉を捧げよう．

シヨナサホウムクシヨトロリ△タマカクマナ
カイアウハンヨロサイトシ

（昭和 44 年 3 月 28 日受付）