字形検字方式による漢字入力の一方法

磯道義典（広島大学）
新谷順子（日立ソフトウェアエンジニアリング(株)）

1. はじめに

現在，さまざまな漢字入力法[1)2)3)4)5)]が考案され実用化されているが，次の様な利点を持つ字形検字方式による漢字入力の一方法を考案したので，報告する。

(1) 入力装置が小型である。
(2) オペレータの負担が軽い。
(3) 入力できる漢字数が比較的多い。

本論文では，まず字形検字方式を選択するに至った理由を述べ，その例を示した上で，コード作成の過程を説明する。

2. 字形検字方式

漢字を計算機で処理する場合，その入力装置として考えられるものに，まず，普通のＴＳＳ端末に備わっているキーボードがある。これは，一般に漢字入力に用いられているタブレットとかＯＣＲといった装置に比べて，小型で安価であるという利点がある。

常用キーを用いる場合はさらにどのように漢字を指定するかという問題が残る。一番手っ取り早い方法としては，テキストをすべて仮名文字化して入力し，漢字へ変換するのは計算機まかせという，仮名漢字変換方式がある。しかしこの方法は，日本語の特徴である同音意義語の処理のため，構文解析や意味解析にまで立ち入ったかなり複雑な変換プログラムが必要となる。

逆に計算機内での処理が最も簡単にすむのが，漢字に対してコードを割り当てそのコードを入力するというコード入力方式である。この方法はコードを暗記しなくては使えないという取っつきの悪さがあるが，一度コードを覚えてしまうとかなり高速の入力が可能になる。

他に，コード入力方式におけるコードを暗記する手間を省いた，対話型入力方式がある。コード入力方式のように漢字とそのコードが｜対｜に対応しておらず，コードに対して漢字が一意に定まらない時の最終決定は，改めてオペレータが入力するようになっている。この方法は初心者には使いやすいが，熟練してもあまり速度が上がらないことが多い。

この３つの方式のうち，本研究ではコード入力方式を考えた。何故なら，コードを暗記するというオペレータにとって一見不利な条件も長い目で見れば，入力速度が向上し負担が軽くなる[1)2)]ことにつながるからである。

さて，コード入力法における最大の問題は，漢字にどのようなコードを割り当てるかである。この漢字のコード化にも，ただ順に数字コードを割り当てるという実用にむかないものから，人間工学的にかなり考慮されている[1)2)]ライプット方式まで，いろいろな方法がある。本研究では入力対象として約４０００字種の文字を前提としていたので，比較的多い字種を入力でき，また速度の点においても期待できる字形検字方式をとった。

これは漢字をその構造に基づいてコード化するものである。コードを全部丸暗記しようとすると，人間が記憶できるコードの数には限界があるため，入力できる漢字数が限られてしまう。その点，字形検字方式では，コード化の規則さえ覚えれば視覚的にコードを

導き出すことができるので、比較的多い字数の入力が可能となる。また訓練すれば反射的にコードが出るようになり、入力操作の際原稿だけを見て打つタッチ打鍵に近づけることができる。[1,2]

3. 字形検字方式の例

字形検字方式では、漢字をその構造に基づいてコード化するわけだが、その方法にも2通りある。1つは、字根（漢字を構成している基本的な要素）とともにそれらの結合操作を指定する方法である。もう1つは、字根を並べてコード化する時に漢字の特定の位置から字根をひろってくる方法である。前者は、コードの長さが不定で、複雑な漢字ほどコードが長くなってしまうという欠点があるので、本研究では後者を採用した。

後者の方法において、字根の数と字根をひろってくる位置は研究によって異なり、それによって使用するキーの数や打鍵数が異ってくる。[6]例として、四角号碼法[6]と三角コード法がある。

四角号碼法とは、漢字がだいたい正方形で四つの角を有しているのに着眼した王雲五が考案したものである。［1928年］ まず10種類の基本字根を定義し、0から9までのアラビア数字で表現した。次に各漢字の四角の字根を左上、右下、左下、右下各角の順にひろっていき、4位の10進数で漢字をコード化した。四角号碼法の基本字形とコーディングの例を図1に示す。この方法は 異なる字で同コードになるものが多く、すべての漢字を入力するのは困難である。

これに対して、胡（L.Hu）らは三角コード法を提唱した。［1979年］ 胡はまず300個の基本符号を選び、99のグループに分け 01から99まで10進数2桁のコードを与えた。各グループ内の基本符号は皆類似した形を有し、しかも一桁目が四角号碼に関連している。次に、コード化の規則としてはZ形の角コード化原則を決定した。それは左から右、上から下、外から中へといった順序に従って3つの基本符号をとらえ、そのコードを配列するものである。すなわち6桁の10進数で漢字をコード化することができる。この方法では同コードで異った文字の場合が少なくなる。

コード	0	1	2	3	4	5	6	7	8	9
基本字根	工	一	丨	丶	十	扌	口	冂	八	小
		𠃌	丿	乚	乂					
		冫	亅							

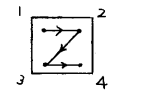
(1) 基本字根とそのコード

$$\overset{2}{2}\text{伏}\overset{3}{3} = 2323 \qquad \overset{0}{6}\text{説}\overset{8}{1} = 0861$$

$$\overset{3}{1}\text{汗}\overset{1}{4} = 3114 \qquad \overset{9}{2}\text{煇}\overset{7}{5} = 9725$$

(2) コーディング例

［図1］四角号碼法



(1) Z形の角コード化原則

[1] 馬 = 馬 35 00 00

[1][2] 什 = 亻十 22 40 00

[1][2/3] 法 = 氵土厶 33 41 74

[1/3][2] 郝 = 土阝小 41 15 91

(2) コーディング例

［図2］三角コード法

4. コード化の手順

本研究では次の手順で漢字のコード化を行った。まず漢字を構成する基本字形を26のグループに分け、A～Zの英字1字で表わした。そして各漢字から最高3つの基本字形を選び、書く順序に従ってそれらのコードを配列した。同コードとなる漢字を区別するためには数字，記号を補助的に用いた。その結果漢字約3000字を42キーの2～3打鍵で入力することができた。

なお、仮名、数字，記号，英字などはすべて2打ちとし，第1打目はキーボード上の文字に従い，第2打目は記号を用いた。

本コードの特徴を次にあげる。

(1)漢字の視覚的形を重んじている。

基本字形を26個のグループに分け3隊、筆画の形によって視覚的に分類した。

(2)良く使うものは短く，あまり使わないものは長いコードとなる。

漢字は2～3打鍵で入力するが，2打鍵となるのは基本字形1つで構成される漢字，すなわち比較的出現頻度の高いものとなっている。

(3)言語一般において頻繁に使うものは定形からはずれてよいが，使わないものは簡単な規則にのっとって形作られている。例えばbe動詞などは規則にのっとっていないが，あまり表われない動詞では3人称は-S，過去過去分詞は-ed(-en)となる。この点をコード化でも重視している。

コード化の規則は，基本字形1つで構成される漢字，2つで構成される漢字，3つ以上で構成される漢字のそれぞれで異なる。基本字形の数が少ないものほど，本来の英字コード以外に数字や記号を補助的に用いており，定形からはずれている。

5. 基本字形のグループ分けとそのコード

漢字はいくつかの基本字形から成る。例えば「胃」は田と月から成り，「漢」はシとサと果から成る。このとき，田，月，シ，サ，果をそれぞれ基本字形という。

基本字形はその筆順の第1画目～第3画目の形によって7グループに分けた。そのため主に左上の角あたりの形が類似した基本字形が同一グループに入る。

具体的な過程を示す。まず漢字の筆画はその形によって次の5種類に分類できる。[7][8][9]

「丶」，「一」，「丨」，「ノ」，「フ」

このことから基本字形を第1画目の形によって5つのグループに分けることができる。さらに第2画目，第3画目の形によって最終的に26個のグループに分けた。表1に各グループの筆画の形とそのグループに入る基本字形の例、そしてグループに与えられた英字1文字のコードを示す。

特に第1画目が「一」(よこ)，第2画目が「丨」(たて)のグループ(E～I)，第1画目が「丨」(たて)，第2画目が「フ」(かぎ)のグループ(M～P)に入る基本字形はいろいろな漢字に多く使用されているため，漢字コードを作成した時同一コードが多くなるのを避ける意味で細かく区分した。

各グループに与えた英字1文字のコ

ードは、漢字コードとして打つ時の指の動きを規定する。よって頻度の高いグループにはキーボード上の打ちやすいキーの英字を割り当てるというような配慮が必要である。しかし頻度を決定するのは入力テキストにおける漢字自体の頻度であり、これはテキストの内容や種類によっても大きく変動する。本研究では主にコードの覚えやすさに重点をおいたので、ここでは打ちやすさは考慮に入れずただ羅列的にアルファベットを割り当てただけである。しかし入力速度をあげるには、この基本字形のコードをいかに割り当てるかが問題となる。

〔表1〕 基本字形のグループとそのコード

コード	第1画	第2画	第3画	基本字形の例
A	・	・		丶冫氵ッ丬斗灬
B		一		亠古言亡云立产产户文亦主衣方广疒
C		｜ノ7		刂户州/丷䒑𦍌並羊必半首火米类/辶良礻心
D	一	二		平来/一二干开キ弋井夫未天末三丰夫
E		｜	一	十土士走龶戈赤
F				丁工下王正耳牙井上土/扌圭/弋寸
G			｜	廿卄廿甘其世丗皿臣
H			ノ	木才戈本
I			7	丙曰東束亩五亙巨瓦雨再両丙可吏事車
J		ノ		丆ナ厂大尤太犬不戊成百自而歹豕万
K		7		七匕云匚ラ戸亙与万夷声
L	｜	｜ノ7		｜小丬/卜非臣上止凸馬片/业川/小⺌
M		7		门几冖/丹/月且用目/内冂冗
N			二ノ	口中虫史串央呂回
O				日里甲申曳門禺果
P			｜7	巾田円皿曲凹冊由皮山出/水内
Q	ノ	・		ノハ㣺釆人入入食金メ谷父乎釆个
R		一		𠂉缶午牛缶矢失朱生年乍竹牛攵午
S				千壬我升禾毛手垂乗重夭牛
T		｜		亻隹斤丘斤氏身舟白臼爪白血鳥向豸禹
U		ノ7		彳彡/クケ乃九丸夕タ久夂儿几及勿乇风
V	7	・		乙くマ癶羽又叉ム叉瓜フレ丁
W		一		彐ヨ聿尹艮尸尺弓己巳串串民
X		｜		了子承阝巴尸也丑疋卫卩凵
Y		ノ		刀刃女力
Z		7		母糸乡幺鱼㔾巛

なお、同コードとなる漢字を区別するために補助的なコードを定めておく。これは基本字形1～2個で構成される漢字に用いられる。それは、前述の四角号碼と同様な部分字形を9種定義し、1～9の数字1字で表現したものである。(表2)

すでに述べた英字1字のコードは基本字形の左上の角あたりの形を示すものである。よってその不備を補うために、右下の角に注目してその字形によって四角号碼的な数字コードを与えてやる。

〔表2〕 補助コード

コード	部分字形
1	一乚
2	｜亅ノ
3	丶乀
4	十乂
5	扌卄
6	口
7	フ⺈フ
8	人八
9	小木

6. 漢字のコード化の規則

基本字形1つで構成される漢字（199字）は、原則として2打ちで入力する。1打目は、基本字形のコード（A～Z）であり、2打目に漢字の右下角の部分字形のコード（1～9）である。

例）　文　＝B4
　　言　＝B6

以上で同定しきれないものは3打目に記号（␣，；・：／）を任意に与えた。

例）　七　＝B1␣
　　立　＝B1，
　　主　＝B1；

コードの1部を表3に示す。

基本字形2つで構成される漢字（878字）は3打ちで入力する。1打目と2打目は、2つの基本字形のコード（A～Z）を書く順序に従って並べる。3打目は原則として␣（空白）だが、同コードの漢字がある場合には、漢字の右下角の部分空形のコード（1～9）を与える。

例）　童　＝BO␣
　　竜　＝BO1
　　音　＝BO6
　　課　＝BO9

ただしそれでも同定できなければ3打目に漢字の右下角でなく、2つの基本字形のうちどちらかの右下角に注目して数字コードを与える。

例）　冗　＝MU␣

[表3] 基本字形1つで構成される漢字とそのコード（1部分）

1st stroke	2nd stroke 1	2	3	4	5	6	7	8	9
A				斗					
B	七,立,主;		衣	文		言	方		赤
C	並,乏,	羊,州,	良	必	半	首	心	火	米
D	一,二,三;	平,千,	、	井,考;				夫,天,	来,未,末;
E	土,士,		走	十					赤
F	工,王,玉,正,	丁,耳,牙;	下	寸					
G	世,臣,					廿,甘,			
H		才		本					木
I	亜,瓦,五,互,巨/	可		史,更,再,	車,東,	西,画,	雨,甫,丙,両/		鬼,裏,
J	尤	不	戊,成,	丈		百	而	太,犬,	
K	七,互,			与		戸、	万	鬼	
L	上,止,	川	永,卜,			凹	片		小
M	且			丹		目	内,月,用,		
N		中	虫	史	串	口		央	、

(注) 漢字の右下の記号が3打目である。記号が書かれていない漢字は3打目は必要でない。

見 =MU1
肌 =MU7

コード表の1部を表4に示す。

基本字形3つ以上で構成される漢字(1888字)は、漢字から基本字形3つを選び出し、書く順序に従ってそのコードを配列する。よって英字3文

[表4] 基本字形2つで構成される漢字とそのコード(1部分)

1st stroke \ 2nd stroke	A	B	C	D	E	F	G	H	I
A		泣注誉	泌洋浪洲	活汗浅沫	汁壮	江汚汀	湛	栄	凍河浦酒
B	這	訪註	応志恋詳詠	評	庄計辛	対庁訂討証		床	庫諌
C	料為	粒	祥炎迷養道	字実	社	定宝灯守		宋	補煉
D	迂			春奏耕式					
E	过	坊裁	志南	坪	圭	寺	堪	栽	載
F			恥	扶持抹挟		式打	弄		拒捕珂
G		芳	芝芯	芋美	基	芽茸	華	某	苗苛
H		柱	様述	桟	杜戎	村柾李柑戒		材林	柄梗栖棟
I			恵悪速軒重票連酬栗			専雫酎		栗	輔
J			灰返逐威残辰		压在	耐左歪			奇
K		房			至	匡			
L	馬		業						
M		肪		肝		肘			同単
N	図	蚊跡虻	忠患	味囲践吐叶		呈団国吋虹		困呆	距圖
O	黒		悶関愚閑昧旦遇		早	是間旺	畢	閉閑	晒
P	辿		思畔	峡		町畏			
Q		忙	乏込悉	余銭鋳針釜		全飢鉦釘		釆鉢	合錬鋪
R	無	笠	迭	竿竺	牡		笹		矩策
S	科		秘秋	秤					
T	焦鳥島	位住依信	伴近進迫狼息	仁狭	仕什隼梟	代付射狂皇		休伐体集	何価伍他便師
U	冬	往	各条忽	行		征外		条	畠
V				発			弁	桑	司
W	尽		忌退			既	展		
X		防	隊						陳陣阿疎
Y		妨	娘忍辺妹				姫		
Z		絃紡	災巡			紅	紺		練

注) 漢字の右下の ␣(空白)および数字が 3rd stroke.

字のコードとなる。４つ以上の基本字形を持つ漢字では、どの３つを選ぶかという問題がある。主に次の３つのいずれかでコード化した。

(1)最初の３つの基本字形をとる。

例）  徴 ＝彳山王
山ＰＦ

(2)３打目に最後の基本字形をとる。

例） 懲 ＝彳山心
山ＰＣ

(3)途中をとばす

例）  戯 ＝广业戈
ＬＬＨ

以上で同定できないものは３打目に字形とは無関係な、他の漢字で使われていないコードを選んで与えた。このような例外は６３字あった。

例） 請 ＝ＢＦ乙

cf. 靖 ＝ＢＦＭ

なお３打目には主にＸ～Ｚを与えておいた。表５に基本字形３つ以上で構成される漢字とそのコードを一部分あげておく。

7. おわりに

本コードによって実際に日本語テキスト[10]が入力できるシステムを作成した。使用機器はＨＩＴＡＣ Ｍ－１８０（広島大学情報処理センター）とグラフィック端末テクトロニクス４０１０である。

試行の結果、入力速度としては１分間に１７～１８文字程度であった。これは製作者自身の場合で、コード作成期間約２週間、練習期間約１ヶ月の場合である。まだ漢字を見て反射的にコードが出るという段階までいたってなかったので、訓練しだいではもっと速くなると思われる。なお全くの初心者には試していないが、英文タイプやカナタイプに熟練した人であればコードの規則さえ覚えればすぐ利用できる。

終わりに、字形データとして、電子技術総合研究所のＦＯＮＴ４０００を使わせて頂いた。ここに謝意を表する。

〈参考文献〉

1) 田中二郎，山田尚勇：タッチ打鍵による日本文入力法の研究，125 pp.，東京大学理学部情報科学科 TECNICAL REPORT 78-01

2) 山田尚勇：日本語テキスト入力法の人間工学的比較，32 pp.，東京大学理学部情報科学科 TECNICAL REPORT 78-06

3) 木澤誠：漢字の入力，電気学会雑誌，９７巻２号，pp.90-92（昭52-2）

4) 渡辺定久：漢字入力装置，電気通信学会誌，vol.63，No.7，pp.707-712（1980年7月）

5) 増田功：日本語文字読み取り装置，電気通信学会誌，vol.63，No.7，pp.719-723（1980年7月）

6) 楊維楨：コンピュータによる漢字処理をめぐる問題，システムと制御，vol.24，No.6，pp.372-378（1980）

＊ ＊ ＊

7) 渡辺茂：漢字と図形，245 pp.，ＮＨＫブックス 264（昭51）

8) 読売新聞社：日本語の現場 第一集，210 pp.（昭51）

9) 文部省：外国人のための漢字辞典，大蔵省印刷局（昭41）

10) Donald E. Knuth：基本算法／情報構造，pp. 87-148，米田信夫・筧捷彦共訳，サイエンス社（昭53）

＊　＊　＊

[表5] 基本字形3つ以上で構成される漢字とそのコード（1部分）

A 1st stroke

3rd stroke →, 2nd stroke ↓

	AB	CD	EF	GH	IJ	KL	MN	OP	QR	ST	UV	WX	YZ
AB			滴	渡	濠	濃済	瀧涼	滝	溶	液	流	湾淳	泳
CD		淡	淀		演			潜		湊			滋
EF	装						潮湖 清	渚瀆 瀞	濆厳	鴻	凌法	濤	潔
GH				淋	満		滞漢 湘	漠瀆	洪濫 漆	灘		港	
IJ	漣	漂 減	涯	渠	濡 滅		瀦	漕	瀬	漸 源			
KL	渋裂				涙 瀕	渉	消		凝		淑		凄
MN			渇	深	洞	測	潰		営	灌	覚 況		
OP			潤	戦	湯 獣濃	混湿	澗 滑渦	温			湯漫 濁		婆
QR		渓	塗将		奨	冷	沿 浩	溢		淫	洗	浮	海
ST	激		準			添	活	溜	浃		湯		
UV	漁			染	澄		資洛 涌治	盗 湧	次淘	濯	没	泡 潑	姿 (浄) (治)
WX			渥		漏	泥	浸				澱	涵	
YZ				梁			沼						

B 1st stroke

3rd stroke →, 2nd stroke ↓

	AB	CD	EF	GH	IJ	KL	MN	OP	QR	ST	UV	WX	YZ
AB			痔症	譲	病	諒	諮	瘢 識諦	疹疾	痩	疫 諺	痕 廓	該
CD	謎	談詫				記 試	誼説 讃	譜		議	麟	謙廊	
EF		誌	詩 試			庇	廟詰 請	諸	敵		読 諏		(適) (諳)
GH	庶	遮		諜 麻	諾 磨		謹 親麿	席		護 摩新	度 魔		(諜)(講)
IJ		痘 誇				療	語	庵		鷲			
KL	襲	斎	齎		就	誹	店	詣	廠				(剤)
MN			調望		豪	龍 剖	膏高 亭			離 哀	亮 競	誤部	
OP		意	章		端	謁	韻 謂		裏	衰	彰 啟		
QR	於	癒謡 遊	座詮	旗			諭商 痴	論	卒 族	旅	診訟 旋	郊 施	効
ST		底	府			痢	話 腐褒			鷹夜	誘庭 誕		(謝) (毅)
UV	熟	廃	塾	棄	顔 詞		痛		竣		説 充謬	庖 序	
WX	瘁 烹		塵鹿				庸唐		敦		慶	郭	
YZ	率	認				刻	詔	畜	亥				劾