

## 漢字四角号碼の代数構造と‘声形法’による 中国語漢字入力について†

川口 喜三男<sup>††</sup> 王 思 鴻<sup>†††</sup>

漢字四角号碼が中国語漢字入力の有力な手段として利用し得ることに着目し、はじめに、四角号碼の数学的構造を明確にするため、四角号碼を伴う漢字の抽象的モデル—字形モデルという一を導入する。この字形モデルは一種の代数系として取り扱うことができ、そこでは、あらゆる漢字の集合は14の字形類に類別され、2種の並置演算と6種の包摂演算に関して閉じた系をつくる。このため、すべてのより複雑な漢字はより単純な字または部首からこれらの演算を施すことによって得られると同時に、その四角号碼もこれに伴い代数演算によって機械的に求められる。さらに、四角号碼と拼音を組み合わせた中国語漢字入力法として‘声形法’を提案する。この声形法によれば、最大4文字(4打鍵)からなる漢字コードの入力によって目的の漢字は90%以上の確率で正しく選択されることが実際例をもって示される。

### 1. はじめに

漢字辞典を引くとき、目的の字に到達するには、わが国では、部首索引によるか音訓(中国語のときは拼音 pin yin)索引によるのが普通のことである。しかし、これは、特に読めない漢字の場合、時間がかかっていかにもわずらわしい。ところが、中国では、特に年輩者の間で四角号碼(シカクゴウマ)による索引が愛好されているという事実がある。実は、わが国の漢字辞典でも一部のものは四角号碼索引を別冊付録としてではあるが持っている<sup>1)</sup>。この方法による索引は、他の索引法によるよりも迅速である。

四角号碼とは、漢字の四すみ(角)の筆形・筆画をみて、それぞれを0から9までの番号に置き換え、一字の漢字を四桁の数で表現するコードである。たとえば、‘漢’は、左上(さんずいの上)の点を3、右上(くさかんむりの右)の‘+’を4、左下(さんずいの下)の左下からのハネを1、右下の捺(ヒッパリ)を3で番号付けし、左上・右上・左下・右下の順に3413と書いて、この四桁の数3413を‘漢’の四角号碼とするのである。

しかしながら、この四角号碼もある程度慣れなければ必ずしも正しく号碼表通りに作り出せないことがあ

る。たとえば、‘左’は4001であって、‘工’の下部から得られる番号1は右下の角位置に置かれる。これに対し‘右’は正しくは4060であって、‘口’に対応する番号6は左下の角位置に置かれる。また、[新方式]<sup>2)</sup>では、‘弓’は1702であって、右上からのハネ‘丿’の2は右下に位置するのに、‘彡’は1020であって、同じハネが左下に位置する。

さらに、たとえば、‘戈’は5300であるが、‘文’に対応する番号5は左上角の番号として現れている。より複雑な文字の中に存在するとき、たとえば、‘刈’(5200)、『戎’(5340)では番号5はやはり左上角番号として現れるが、『戔’(5350)の下部、『劊’(5250)の左下部では左下角番号として現れ、『戕’(2325)、『戕’(4345)、『戕’(4355)、『戕’(2265)、『戕’(4455)では右下角番号として現れる。このような例は、ほかにも‘九’、『寸’、『白’、『弋’、『头’(頭の中国簡体字)、等々たくさんある。このように、基本的な造形単位、部首、字を上下左右に並置したり、一方が他方を包摂する(たとえば、‘虎’が‘彡’を包摂すると‘彪’となる)と、基本字のある角番号は消滅したり、また他の角番号は別の角位置に移動したりして合成字の号碼が新しく生成される。上述のような例によって四角号碼はややつかみどころがないという印象を与えがちでもある。

そこで、号碼生成の一般法則を抽出し、厳密規則体系として確立したいという自然な欲求が生じるのは避けがたい。そのみならず、工学上の要求、すなわち、計算機による漢字処理システムの入力時における視覚にたよる号碼作成を迅速かつ確実ならしめること、将

† The Algebraic Structure of the Four-Corner Codes and a Method for the Input of Chinese Characters by KIMIO KAWAGUCHI-IZAWA (Department of Computer Science and Communication Engineering, Nagoya Institute of Technology) and SIHONG WANG (Beijing Research Institute of Chemical Industry).

†† 名古屋工業大学情報工学科

††† 北京化工研究院

来の機械による漢字自動認識が効率化をねらって基本的な造形単位に分解して行われるとすれば、基本単位の号碼から元の合成漢字の号碼が一種の代数演算を施すことによって‘機械的’に求めることを可能ならしめるなどの要求に応じるためには、どうしても個々の基本単位、部首、字の造形機能を抽象化し、号碼生成の代数的構造を明確にしておかなければならない。

基本造形単位から出発して、できる限り‘少数’の演算規則を用いてより複雑な字の号碼を生成できるようにすることが望ましい。また、この演算が定義される字の抽象的構造—字形という—の種類も‘少数’であるほどよい。ところが、現在ではいく通りかの号碼表が存在し、なおも一部改変が続けられているようである。そうであるならばなおのこと四角号碼の背後に隠れている法則性を明確にし、可能な限り少数の規則によってすべての号碼が生成できるように理論体系を整備しておき、背後に存在する数学的構造を複雑化するような個々の号碼の変更は避けるという原則を設けることが賢明であると思われる。

2. 四角号碼の作成法

ここでは、個々の漢字が与えられたとき、その四角号碼をどのようにして定めるのかをまず解説する。ここでの作成法は、基本的に文献 1) と一致するが、一部に相違がある。なお、以下の記述では、多くの部分が文献 1) における解説の引用から成っている。

I. 筆形と番号

表 1 筆形番号  
Table 1 Stroke numbers.

番号	筆名	筆形	字例	説明
0	頭 (クベツア)	一	言主ノ	独立の点と独立の横線とが結合したものと、点と一と合したものと、いわゆるナベツア形。
1	横 (ヨコ)	一	天 土 地 江 元 風	横線。左下からのハネ(地、江の左下すみ)や、カギハネ(元、風の右下すみ)を含む。
2	垂 (タラシ)	丨	山 月 千 則	垂直のタテ線(丨)と右上からのハネ(ノ)を含む。
3	点 (チ)	・	ハネ 一 ム 之 衣	点と横(ヒツバリ)し、水車もこれに入れる。
4	又 (マタ)	十	草 杏 皮 封 刈 大	二筆が交差するもの。
5	挿 (ウツ)	ㇿ	打 戈 申 史	一筆が他の二筆以上を貫いている場合。中・羊・書などの中央タテ線もこれに入れる。
6	方 (シカタ)	口	國 鳴 目 四 甲 由	四すみがりついている四角形。(■)且などはこれに含まれない。
7	角 (ツノ)	厶	羽 門 灰 陰 雪 衣 印 兎	四すみのいずれにあるを問わず縦線の二筆が後して作るウツ。
8	八 (ヤチ)	八	分 頁 羊 余 火 全 足 午	八の字とその変形。
9	小 (シヨウ)	小	尖 糸 鼻 木 尗	小の字とその変形。

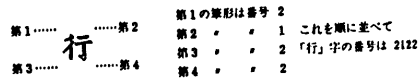
漢字の筆形を 10 種類に分け、表 1 のように 0 から 9 までの番号をつける。

この番号のうち、1, 2, 3 は、単筆であるが、その他の 0, 4, 5, 6, 7, 8, 9 は、すべて二筆以上で構成された複筆である。四すみとして番号を与える際には、できるだけ複筆のほうをとるようにする。たとえば、

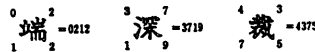
- 一は 0 として、3 とはしない。
- いは 8 として、32 とはしない。
- すは 4 として、2 とはしない。
- 小は 9 として、23 とはしない。
- 厶は 7 として、2 とはしない。

II. 漢字の四すみのとり方と順序

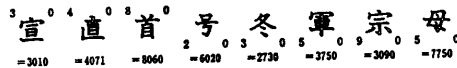
- 1) 漢字の四すみをとる順序は、1. 左上, 2. 右上, 3. 左下, 4. 右下の順とする。たとえば、



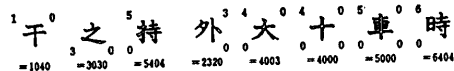
同様に、



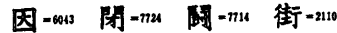
- 2) 上記 1) 項で定められた順序に従って一つのすみで一度用いられた筆形が、他のすみにもたがう場合は、再び該当する番号を与えるのではなく、0 (欠番) を与える。



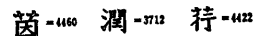
- 3) 字の上部か下部に、一個の単筆(つまり、1, 2, 3 のもの)か一個の複筆(つまり、0, 4, 5, 6, 7, 8, 9 のもの)しかない場合は、その番号を左すみとして与え、右すみには欠番として 0 を与える。



- 4) 文献 1) の号碼法においては、口, 門, 鬥, 行の「かまえ」で構成されている字では、第 3, 第 4 のすみは、その「かまえ」の内側の筆形からとる。



このような構成のとき、その「かまえ」は内部筆形を半包摂するという。ただし、上下または左右に、他の筆形がつけ加わっている場合は、この限りではない。



このような場合は、内部筆形は、それを包む筆形によって全包摂されるという。

5) 現代中国語の簡化体である門 (3722) も 4) 項に準じる。

6) 文献 1) とは違い、冂, 凵, 凵の「かまえ」も 4) 項に準じる。

### Ⅲ. 諸注意

1). 文献 1) では、字体は、通行の筆写体楷書によっている。たとえば、

正	宀 <sup>3</sup> 住 <sup>0</sup> 匕 <sup>1</sup> 反 <sup>0</sup> 言 <sup>3</sup> 尸 <sup>3</sup> 心 <sup>3</sup> ト <sup>3</sup> 斥 <sup>3</sup>
異	冂 <sup>7</sup> 住 <sup>0</sup> 匕 <sup>2</sup> 反 <sup>2</sup> 言 <sup>1</sup> 尸 <sup>1</sup> 心 <sup>1</sup> ト <sup>4</sup> 斥 <sup>4</sup>

正	业 <sup>3</sup> 亦 <sup>2</sup> 羽 <sup>1</sup> 衣 <sup>2</sup> 佳 <sup>2</sup> 禺 <sup>2</sup>
異	业 <sup>2</sup> 亦 <sup>3</sup> 羽 <sup>2</sup> 衣 <sup>1</sup> 佳 <sup>1</sup> 禺 <sup>1</sup>

このほか、下記の類に注意を要する。

- 敵→前 8022 倫→2822 (倫)
- 兼→兼 8033 謙→0863 (謙)
- 戸→戸→戸 3027 週→週→3330 (週)
- 井→井 8044 餅→8874 (餅)
- 发→发 4304 祇→1394 (祇)
- 令→令 8030 琴→1030 (琴)

また、字の下側にくる夕や攵の類も同じように通行の筆写体によっている。たとえば、

童→童 2724 夢→夢 4020

1) 資料 3) では、汉语拼音索引<sup>4)</sup> に使用されている字体に対して四角号碼表が作成された。

#### 2) 筆形をとる際の注意

(a) 冂, ネなど、点(丶)の下の横線が他の筆形と接続するものは、すべて3とし、0とはしない。

(b) 尸, 皿, 且, 門など、四角形を構成する筆形が、外側に延長している場合は、すべて7とし、6とはしない。たとえば、皿は、7710である。

(c) 角(カド, カギ)を構成する筆形の両端は7とはせず、'フ'のように、その両端の筆形にそれぞれ番号を与える。

(d) 筆形の「ハ」(ハチ)は、他の筆形と交錯する場合には8とはしない。たとえば、美。

(e) 业, 亦は、「小」字に類似してはいるものの、中央に二筆があるので、「小」(9)とはみなさない。水, 小も余分の筆形をもっているため、「小」(9)とはしない。

#### 3) 四すみを定める際の注意

(a) 独立した形や並行をなす形の筆形は、その高

低にかまわず、最も左側、または最も右側をとる。

非<sup>1</sup> 青<sup>1</sup> 疾<sup>3</sup> 浦<sup>3</sup> 帝<sup>2</sup>

(b) 最も左側、または最も右側の筆形があっても、もし他の筆形が上部にかぶさっていたり、下部でうけている場合には、上にかぶさっている筆形を上すみとし、下でうけている筆形を下すみとしてとる。

宗<sup>3</sup> 幸<sup>4</sup> 寧<sup>2</sup> 共<sup>2</sup>

(c) 複筆(つまり、0, 4, 5, 6, 7, 8, 9)のものが二つあると考えられる場合は、上すみとしてはより高いもの、下すみとしてはより低いものをとる。

功<sup>4</sup> 盛<sup>5</sup> 顔<sup>4</sup> 鳴<sup>5</sup> 奄<sup>7</sup>

(d) 右上からのハネが、他の筆形によってうけられている場合には、下でうけている筆形を下すみとする。

春<sup>2</sup> 奎<sup>1</sup> 石<sup>0</sup> 衣<sup>7</sup> 辟<sup>6</sup>

(e) 左上部にあるハネは左すみとなるが、左上に位置していない場合は、ハネよりもさらに右にある筆形を、右すみとしてとる。

句<sup>7</sup> 鈞<sup>7</sup> 鳥<sup>7</sup> 鳴<sup>7</sup> 牟<sup>3</sup> 倅<sup>3</sup>

#### 4) 方式の違い

四角号碼の方式は、中国でも何度か訂正変更され、時期によって異なっているが、その違いは、四すみの定め方にある。

以上が四角号碼の作成法である。

### 3. 字形類とその演算

すべての筆形、または現存の漢字の部分(例、‘六’は‘六’の部分である)に一致するような筆形どうしのあらゆる組合せ(例、‘リ’、‘ロ’、等)を部首と呼ぶ。組合せ方は、横並置(例、‘リ’), 縦並置(例、‘古’), 重量(例、‘彡’), 左上, 右上, 左下, 右下包摂(例、‘斗’、‘戈’、‘才’、‘尢’), 半包摂(例、‘冂’、‘区’、‘凶’、‘囟’), 全包摂(例、‘向’、‘用’)である。

すべての部首または字どうしの重量を除く可能なすべての前述の組合せ(たとえば‘三’に‘丨’を重ねて‘丰’を作ることは許さないし、また‘一’で他の部首をどのような意味でも包摂することは不可能である)を字という。この定義によれば、現存のあらゆる漢字は字である。以下、一般に字は、その四角号碼

$i j k l$  を四角に添字として付加した英文字  $ia$ , または号碼の一部を省略した  $ia', \dots, 'a', \dots, 'a, \dots, a$ , または  $a$  などで表す. 添字の値は筆形に応じて  $0, 1, \dots, 9$  のいずれかであるが, 筆形番号  $0$  と欠番の  $0$  を区別するため, 欠番は以後  $\_$  で表示する.

字と字の組合せ操作を次のように演算記号を用いて表現する. 2個の字をそれぞれ  $a, b$  と置くと,

1. 横並置  $a|b$  例, 井|多=形
2. 縦並置  $a/b$  または  $b/a$   
例, 幸/幸=幸\幸=筆
3. 左上(からの)包摂  $a \lceil b$  例, 厂 \lceil 万=疒
4. 右上包摂  $a \rceil b$  例, メ \rceil 气=氣
5. 左下包摂  $a \lfloor b$  例, 又 \lfloor 壬=廷
6. 右下包摂  $a \rfloor b$  例, ム \rfloor 十=斗
7. 半包摂  $a ] b$  例, 冂 ] 厶=冂
8. 全包摂  $a \supset b$  例, 宀 \supset 口=向

上記の演算はすべての字に対して定義されるわけではない. 半包摂は, 現存の部首または字に対しては2章IIの4), 5), および6) 項によって指定される「かまえ」に対して定義される. ただし, 4) 項の但し書の条件の下では全包摂となる\*. 一般に字  $a$  と字  $b$  が演算  $\cdot$  によって字  $c$  を合成するとき,  $a \cdot b = c$  と書く.

あらゆる字の集合を, 同じ類(クラス)に属する字は演算が定義されさえすればすべて同一の字形合成機能を持つように14種の類に類別することができる. このように類別された類を字形類という. 各字形類の字形合成機能は, 類ごとに下記のように定義される. 類の表現(代表元)には, 浮動点(白丸 $\circ$ )と固定点(黒丸 $\bullet$ )の2種の点の組合せ

(たとえば,  $\overset{i}{\bullet} \underset{j}{\circ}$ ,  $\overset{i}{\circ} \underset{j}{\bullet}$ ,  $\overset{i}{\circ} \underset{j}{\circ}$ ) が用いられる. 固定点は, 角位置が固定されていることを表す. たとえば,  $\overset{i}{\bullet} \underset{j}{\bullet}$  (例,  $\overset{i}{\bullet} \underset{j}{\bullet} \text{寸}$ ) の固定点  $j$  は,

つねに第3(左下)角に対応し, 演算によって角位置を移動させられることがなく, 結果の字形にそのまま現れる(例,  $\overset{i}{\bullet} \underset{j}{\bullet} \text{寸}$  の固定点  $2$  は,  $\overset{i}{\bullet} \underset{j}{\bullet} | \text{白} = \overset{i}{\bullet} \underset{j}{\bullet} \text{𠂇}$  のように $\overset{i}{\bullet} \underset{j}{\bullet}$ の第3角番号になるか), または演算の相手の第3角に対応する筆形によって取って代わられる(例,  $\overset{i}{\bullet} \underset{j}{\bullet} \text{寸}$  の固定点  $3$  は,  $\overset{i}{\bullet} \underset{j}{\bullet} \text{木} | \text{寸} = \overset{i}{\bullet} \underset{j}{\bullet} \text{村}$  のときは,  $\overset{i}{\bullet} \underset{j}{\bullet} \text{木}$  の9が残るので消える)かのいずれかである. 浮動点は,

\* 半包摂と全包摂の相違は, 第3(左下)と第4(右下)の角の番号が, 包摂される内部筆形からとられるか, 内部筆形を包む周囲の筆形から定められるかの違いである. 前者が半包摂, 後者が全包摂である. たとえば,  $\overset{i}{\bullet} \underset{j}{\circ} \lceil \overset{i}{\bullet} \underset{j}{\circ} = \overset{i}{\bullet} \underset{j}{\circ} \lceil \overset{i}{\bullet} \underset{j}{\circ}$  (半包摂),  $\overset{i}{\bullet} \underset{j}{\circ} \supset \overset{i}{\bullet} \underset{j}{\circ} = \overset{i}{\bullet} \underset{j}{\circ} \supset \overset{i}{\bullet} \underset{j}{\circ}$  (全包摂).

角位置が固定されていないことを表す. たとえば, 下記の単心類 $\overset{i}{\circ}$ の場合には, それ自身では, たとえば,  $\overset{i}{\circ} \lceil \text{口}$  のように第1(左上)角に対応するが, 演算を行うと, その結果, 第1~第4のどの角へも移動し得る. たとえば,  $\lceil \overset{i}{\circ} \text{人} = \text{仄}$  では,  $8$  は第1角から第4角へ移動する. 個々の字形類における浮動点が演算によってどのように角位置を移動するかは, 下記の各類の字形合成機能を定義する演算式によって明示される. なお, 全体を通じて, 浮動点は常に第1, 第2, 第3, 第4の順で欠番があれば可能な限り第1角またはそれに近い角位置を占めようとする性質を持っている.

1. 単心字形類(または単心類)

単心類は, たとえば  $\overset{i}{\circ}$  のように, 号碼  $i \_ \_ \_$  ( $i \in \{0, 1, \dots, 9\}$ ) を持つ字の類である. この類の代表元を記号 $\overset{i}{\circ}$  または $\circ$  で表す.  $\circ$  は浮動点である. 単心類の字形合成機能は下記の1), ..., 16) によって定義される.

1)  $\overset{i}{\circ} | a = \overset{i}{\circ} b \_$  が成立するならば,  $p = i, r = \_$  である. この命題は次の簡略式で表される.

- 1)  $\overset{i}{\circ} | a = \overset{i}{\circ} b$
- 2)  $a | \overset{i}{\circ} = b \overset{i}{\circ}$
- 3)  $\overset{i}{\circ} / a = \overset{i}{\circ} b \_$
- 4)  $a / \overset{i}{\circ} = \overset{i}{\circ} b \_$
- 5)  $\overset{i}{\circ} \lceil a = \overset{i}{\circ} b \_$
- 6)  $a \lceil \overset{i}{\circ} = b \overset{i}{\circ}$
- 7)  $\overset{i}{\circ} \rceil a = \overset{i}{\circ} b$
- 8)  $a \rceil \overset{i}{\circ} = \overset{i}{\circ} b \_$
- 9)  $\overset{i}{\circ} \lfloor a = \overset{i}{\circ} b \_$
- 10)  $a \lfloor \overset{i}{\circ} = b \overset{i}{\circ}$
- 11)  $\overset{i}{\circ} \rfloor a = \overset{i}{\circ} b$
- 12)  $a \rfloor \overset{i}{\circ} = \overset{i}{\circ} b \_$
- 13)  $\overset{i}{\circ} ] a = \overset{i}{\circ} b \_$
- 14)  $a ] \overset{i}{\circ} = \overset{i}{\circ} b \_$
- 15)  $\overset{i}{\circ} \supset a = \overset{i}{\circ} b \_$
- 16)  $\overset{i}{\circ} a \_ \_ \_ \supset \overset{i}{\circ} = \overset{i}{\circ} b \_ \_ \_ \_$

たとえば,  $\overset{i}{\circ} \lceil \text{口} = \overset{i}{\circ} \text{香}$ ,  $\lceil \overset{i}{\circ} \text{人} = \text{仄}$ .

2. 二心字形類(二心類)

二心類には, それぞれ記号  $\overset{i}{\bullet} \underset{j}{\circ}$ ,  $\overset{i}{\circ} \underset{j}{\bullet}$ ,  $\overset{i}{\circ} \underset{j}{\circ}$ ,  $\overset{i}{\bullet} \underset{j}{\bullet}$ ,  $\overset{i}{\circ} \underset{j}{\circ}$  で表される代表元を持つ6種のものがある. 黒丸 $\bullet$ は固定点である.

2.1  $\overset{i}{\bullet} \underset{j}{\circ}$  を代表元として持つ類

この類は, 演算が定義される限り下記の1), ..., 15) で定義される字形合成機能を持つ字(たとえば,  $\overset{i}{\bullet} \underset{j}{\circ} \text{寸}$ )の集合である.

- 1)  $\overset{i}{\bullet} \underset{j}{\circ} | a = \overset{i}{\bullet} \underset{j}{\circ} b$
- 2)  $a | \overset{i}{\bullet} \underset{j}{\circ} = b \overset{i}{\bullet} \underset{j}{\circ}$
- 3)  $\overset{i}{\bullet} \underset{j}{\circ} / a = \overset{i}{\bullet} \underset{j}{\circ} b$
- 4)  $a / \overset{i}{\bullet} \underset{j}{\circ} = \overset{i}{\bullet} \underset{j}{\circ} b$
- 5)  $\overset{i}{\bullet} \underset{j}{\circ} \lceil a = \overset{i}{\bullet} \underset{j}{\circ} b$
- 6)  $a \lceil \overset{i}{\bullet} \underset{j}{\circ} = b \overset{i}{\bullet} \underset{j}{\circ}$





3.5  $\begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix}$  を代表元として持つ類

たとえば,  $\begin{smallmatrix} 2 & 2 \\ \triangle & 2 \end{smallmatrix}$ ,  $\begin{smallmatrix} 1 & 4 \\ \triangle & 3 \end{smallmatrix}$ ,  $\begin{smallmatrix} 3 & 7 \\ \triangle & 2 \end{smallmatrix}$ ,  $\begin{smallmatrix} 2 & 1 \\ \triangle & 2 \end{smallmatrix}$ ,  $\begin{smallmatrix} 1 & 4 \\ \triangle & 3 \end{smallmatrix}$ ,  $\begin{smallmatrix} 1 & 4 \\ \triangle & 3 \end{smallmatrix}$  はこの類に含まれる.

- 1)  $\begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} | a = i b$     2)  $a | \begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} = b_i$
- 3)  $\begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} / a = i b$     4)  $a / \begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} = i b_k$
- 5)  $\begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} \lceil a = i b$     6)  $a \lceil \begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} = b_k$
- 7)  $\begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} \rceil a = i b$     8)  $a \rceil \begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} = i b_i$
- 9)  $\begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} \lfloor a = i b_k$     10)  $a \lfloor \begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} = b_j$
- 11)  $\begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} \lrcorner a = i b$     12)  $a \lrcorner \begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} = i b_i$
- 13)  $a \lrcorner \begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} = i b_k$     14)  $\begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} \supset a = i b_i$
- 15)  $? a ? \supset \begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} = ? b ?$

たとえば,  $\begin{smallmatrix} 凶 & 2 \\ \triangle & 2 \end{smallmatrix} = \begin{smallmatrix} 2 \\ \triangle & 2 \end{smallmatrix}$ ,  $\begin{smallmatrix} 2 & 1 \\ \triangle & 2 \end{smallmatrix} \supset \begin{smallmatrix} 2 \\ \triangle & 2 \end{smallmatrix}$

3.6  $\begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix}$  を代表元として持つ類

たとえば,  $\begin{smallmatrix} 1 & 2 \\ \triangle & 1 \end{smallmatrix}$ ,  $\begin{smallmatrix} 1 & 4 \\ \triangle & 2 \end{smallmatrix}$ ,  $\begin{smallmatrix} 2 & 1 \\ \triangle & 1 \end{smallmatrix}$ ,  $\begin{smallmatrix} 1 & 4 \\ \triangle & 2 \end{smallmatrix}$ ,  $\begin{smallmatrix} 2 & 2 \\ \triangle & 2 \end{smallmatrix}$ ,  $\begin{smallmatrix} 1 & 4 \\ \triangle & 1 \end{smallmatrix}$ ,  $\begin{smallmatrix} 1 & 4 \\ \triangle & 2 \end{smallmatrix}$  はこの類に含まれる.

- 1)  $\begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} | a = j b$     2)  $a | \begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} = b_i$
- 3)  $\begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} / a = i b$     4)  $a / \begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} = j b_k$
- 5)  $\begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} \lceil a = i b$     6)  $a \lceil \begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} = b_k$
- 7)  $\begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} \rceil a = j b$     8)  $a \rceil \begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} = i b_i$
- 9)  $\begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} \lfloor a = j b_k$     10)  $a \lfloor \begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} = b_i$
- 11)  $\begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} \lrcorner a = i b$     12)  $a \lrcorner \begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} = j b_i$
- 13)  $a \lrcorner \begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} = j b_k$     14)  $\begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} \supset a = j b_i$
- 15)  $? a ? \supset \begin{smallmatrix} i & j \\ \triangle & k \end{smallmatrix} = ? b ?$

たとえば,  $\begin{smallmatrix} 1 & 2 \\ \triangle & 1 \end{smallmatrix} / \begin{smallmatrix} 1 & 2 \\ \triangle & 1 \end{smallmatrix} = \begin{smallmatrix} 1 & 2 \\ \triangle & 1 \end{smallmatrix}$ ,  $\begin{smallmatrix} 1 & 4 \\ \triangle & 2 \end{smallmatrix} \lrcorner \begin{smallmatrix} 1 & 4 \\ \triangle & 2 \end{smallmatrix} = \begin{smallmatrix} 1 & 4 \\ \triangle & 2 \end{smallmatrix}$

4. 四心字形類 (四心類)

四心類は, 記号  $\begin{smallmatrix} i & j \\ \square & l \end{smallmatrix}$  で表される代表元を持つ類であって, たとえば,  $\begin{smallmatrix} 1 & 4 \\ \square & 2 \end{smallmatrix}$ ,  $\begin{smallmatrix} 2 & 1 \\ \square & 3 \end{smallmatrix}$ ,  $\begin{smallmatrix} 1 & 4 \\ \square & 2 \end{smallmatrix}$ ,  $\begin{smallmatrix} 2 & 1 \\ \square & 3 \end{smallmatrix}$ , 等の字がこの類に属する.

- 1)  $\begin{smallmatrix} i & j \\ \square & l \end{smallmatrix} | a = i b$     2)  $a | \begin{smallmatrix} i & j \\ \square & l \end{smallmatrix} = b_i$
- 3)  $\begin{smallmatrix} i & j \\ \square & l \end{smallmatrix} / a = i b$     4)  $a / \begin{smallmatrix} i & j \\ \square & l \end{smallmatrix} = i b_l$

- 5)  $\begin{smallmatrix} i & j \\ \square & l \end{smallmatrix} \lceil a = i b$     6)  $a \lceil \begin{smallmatrix} i & j \\ \square & l \end{smallmatrix} = b_l$
- 7)  $\begin{smallmatrix} i & j \\ \square & l \end{smallmatrix} \rceil a = i b$     8)  $a \rceil \begin{smallmatrix} i & j \\ \square & l \end{smallmatrix} = i b_i$
- 9)  $\begin{smallmatrix} i & j \\ \square & l \end{smallmatrix} \lfloor a = i b_l$     10)  $a \lfloor \begin{smallmatrix} i & j \\ \square & l \end{smallmatrix} = b_j$
- 11)  $\begin{smallmatrix} i & j \\ \square & l \end{smallmatrix} \lrcorner a = i b$     12)  $a \lrcorner \begin{smallmatrix} i & j \\ \square & l \end{smallmatrix} = i b_i$
- 13)  $\begin{smallmatrix} i & j \\ \square & l \end{smallmatrix} \lrcorner a = i b_l$     14)  $a \lrcorner \begin{smallmatrix} i & j \\ \square & l \end{smallmatrix} = i b_l$
- 15)  $\begin{smallmatrix} i & j \\ \square & l \end{smallmatrix} \supset a = i b_i$     16)  $? a ? \supset \begin{smallmatrix} i & j \\ \square & l \end{smallmatrix} = ? b ?$

たとえば,  $\begin{smallmatrix} 2 & 1 \\ \square & 2 \end{smallmatrix} \lrcorner \begin{smallmatrix} 2 & 1 \\ \square & 2 \end{smallmatrix} = \begin{smallmatrix} 2 & 1 \\ \square & 2 \end{smallmatrix}$ ,  $\begin{smallmatrix} 1 & 4 \\ \square & 2 \end{smallmatrix} \lrcorner \begin{smallmatrix} 1 & 4 \\ \square & 2 \end{smallmatrix} = \begin{smallmatrix} 1 & 4 \\ \square & 2 \end{smallmatrix}$

以上が, 四角号碼を伴う漢字の字形モデルである.

この字形モデルでは, 字形を14種類に分類し, すべての字はそのいずれかに当てはまるものとして, より複雑な字は8種の演算を用いてより簡単な字からすべて合成できるという考えに基づいている. それならば, このモデルは字の合成に関して果して閉じた体系を作っているであろうか.

代表元  $i, j, \dots, \begin{smallmatrix} i & j \\ \square & l \end{smallmatrix}$  の演算をもって字形類の演算であると定義する.  $C = \{i, j, \dots, \begin{smallmatrix} i & j \\ \square & l \end{smallmatrix}\}$  は, 代表元によって表した14種の類の集合であるとする.

〔命題1〕 字形類の集合  $C$  は演算  $|, /, \lceil, \rceil, \lfloor, \lrcorner, \supset$  に関して閉じている.

〔証明〕 演算によって2個以上の浮動点を持つ三心構造を生成しないことを示せばよい.

(1) 演算  $a | b = c$  によって  $c$  の左右両側にはそれぞれ a) 上角に浮動点, 下角に空白, または b) 上下角とも固定点を生じる. したがって,  $c$  は2個の浮動点と1個の固定点を持つことはない.

(2) 演算  $a / b = c$  によって  $c$  の上下にはそれぞれ a) 左角に浮動点, 右角に空白, または b) 左右角とも固定点を生じる. したがって, 2個の浮動点と1個の固定点を持つ三心構造を生成することはない.

(3) 演算  $a \lceil b = c$  によって  $c$  の右下角は固定点になる. 残りの3角は  $a$  によって定まるが,  $a$  の固定点が浮動点に変化することはない.  $a$  が2個の浮動点を持つ  $\begin{smallmatrix} i & j \\ \square & l \end{smallmatrix}$ ,  $\begin{smallmatrix} i & j \\ \square & l \end{smallmatrix}$  である場合,  $c$  は  $\begin{smallmatrix} i & j \\ \square & l \end{smallmatrix}$ ,  $\begin{smallmatrix} i & j \\ \square & l \end{smallmatrix}$  という構造になることが字形合成機能定義式から分る. したがって,  $c$  は2個の浮動点と1個の固定点を持つことはない.

(4) 演算  $\lrcorner, \lfloor, \lrcorner$  についても (3) と同様である.

(5) 演算  $a \lrcorner b = c$  によって  $c$  の下部には a) 左角に浮動点, 右角に空白, または b) 左右角とも固

定点を生じる。したがって、このとき2個の浮動点、1個の固定点を持つ三心構造を生成することはない。

(6) 演算  $a \supset b = c$  によって  $c$  は  $a$  と同じ字形になる。(終)

命題1は、演算によって  $C$  に属さない字形を生成する恐れがないことを保証する。

字形の合成に関しては下記の諸式が成り立つ。以下、 $\alpha, \beta, \dots$  は、集合  $C$  の上の変数であるものとする。

$$[1] (\alpha|\beta)|\gamma = \alpha|(\beta|\gamma)$$

$$[2] (\alpha/\beta)/\gamma = \alpha/(\beta/\gamma)$$

$$[3] (\alpha|\beta)\lceil\gamma = \alpha\lceil(\beta/\gamma)$$

$$[4] \gamma\lceil(\alpha|\beta) = (\gamma\alpha)|\beta$$

$$[5] (\alpha|\beta)\lceil\gamma = \alpha\lceil(\beta/\gamma)$$

$$[6] \gamma\lceil(\alpha|\beta) = (\gamma\alpha)|\beta$$

$$[7] (\alpha/\beta)\lceil\gamma = \alpha\lceil(\beta/\gamma)$$

$$[8] \gamma\lceil(\alpha\backslash\beta) = (\gamma\alpha)\backslash\beta$$

$$[9] (\alpha\backslash\beta)\lceil\gamma = \alpha\backslash(\beta|\gamma)$$

$$[10] \gamma\lceil(\alpha/\beta) = (\gamma\alpha)/\beta$$

たとえば、 $\lceil\text{夾}\lceil(\lceil\text{犬}\lceil/\lceil\text{心}\lceil) = (\lceil\text{夾}\lceil|\lceil\text{犬}\lceil)/\lceil\text{心}\lceil = \lceil\text{愁}\lceil$  は [10] の実例である。

$\bar{\varepsilon}$  は空でない式、 $\cdot$  は任意の演算を表すものとする、

$$[11] \bar{\varepsilon} \cdot (\alpha|\beta) = \bar{\varepsilon} \cdot \alpha$$

$$[12] (\alpha|\beta) \cdot \bar{\varepsilon} = \alpha \cdot \bar{\varepsilon}$$

一般に、式の中に  $(\alpha|\beta)$  が埋め込まれているときには、 $\alpha|\beta$  が単独に存在するときの式  $\alpha|\beta = \gamma$  によって得られる字形  $\gamma$  で  $(\alpha|\beta)$  を置き換えることができない。たとえば、 $\lceil\text{口}\lceil$  或  $\lceil\text{囧}\lceil$  の字形は  $\lceil\text{口}\lceil$  であるが、 $\lceil\text{口}\lceil$  の右半部にある  $\lceil\text{囧}\lceil$  の字形は  $\lceil\text{口}\lceil$  として扱わねばならない。

$$[13] \alpha \supset \beta = \alpha$$

(注意) 上記の式 [1], ..., [13] は字形類の演算に関して成立する式である。それぞれの左右両辺が字として全く同一の造形であるとは言っていない。たとえば、[8] の実例として、 $\lceil\text{虫}\lceil$  ( $\lceil\text{虫}\lceil \backslash \lceil\text{四}\lceil$ ) は  $\lceil\text{蜀}\lceil$  であると造形的に理解するべきであろうが、 $(\lceil\text{虫}\lceil \backslash \lceil\text{四}\lceil)$  は  $\lceil\text{蜀}\lceil$  である。しかし、字形としては両者とも  $\lceil\text{蜀}\lceil$  であって等しい。

次に、通行の四角号碼の取り方(たとえば、文献1))では、現存の漢字はすべて上述の字形モデルに当てはまるであろうかという疑問が生じる。

問題は皆無というわけではない。

1) 通行の四角号碼では、 $\Delta$  の号碼は 2370 ある

いは 2073 である<sup>1)</sup>。どちらであるべきであろうか。 $\Delta$  の号碼は 2223 であるから、 $\Delta$  の字形は  $\lceil\text{口}\lceil$  のようにも見えるが、一方、 $\lceil\text{允}\lceil$  の号碼は 2321 であるから、 $\lceil\text{允}\lceil$  のようにも見える。このようなあいまいさを退けて、どちらか一方に決め、 $\Delta$  を部首として用いる字の号碼もそれに従って整合させるべきである。 $\Delta$  は  $\lceil\text{允}\lceil$  であるとするのがよい。そうすると、 $\lceil\text{私}\lceil$ 、 $\lceil\text{私}\lceil$ 、等の号碼はそれぞれ 2320、2390、等となる。幸い通行の号碼表では私に対しては 2390、2293 のどちらも使用できるようになっている。

2)  $\lceil\text{史}\lceil$ 、 $\lceil\text{史}\lceil$ 、 $\lceil\text{史}\lceil$ 、 $\lceil\text{史}\lceil$  のような字を部首に用いる字は上述のモデルの字形合成規則に合わないことがある。 $\lceil\text{一}\lceil/\text{史}\lceil = \lceil\text{更}\lceil$  であるから、 $\lceil\text{更}\lceil$  の四角号碼は 1050 である。通行の号碼法では 1050 でも 1054 でもよいようにしてあって、更に他の部首を接続すると、 $\lceil\text{便}\lceil$ 、 $\lceil\text{便}\lceil$ 、... というようになっている。この例のように号碼5が合成の結果として別の番号(4)に変わるというのは、個々の字の形を捨象して抽象的な字形の代数演算によって複雑な字の四角号碼を機械的に算出しようとするときにはきわめて具合が悪い。このような例外は少なくとも機械によるパターン認識を行うような際には排除するべきである。この不整合は、 $\lceil\text{便}\lceil$ 、 $\lceil\text{便}\lceil$ 、等の右下角号碼を5にすることにより回避することができる。実用上は、人は従来の号碼法と変更後の号碼法の両者を用い、機械は変更された四角号碼のみを用いるということになる。

なお、これまでの調べでは、以上のほかに重大な不整合は発見されない。

#### 4. 声形法による中国漢字入力方式の提案

われわれは、1) 目的の漢字(現代中国簡体字)が、第一級および第二級漢字<sup>4)</sup> 6,763文字を含む数千字の中から最初のコード入力により90%以上の確率で正しく選択されること、2) コード作成法が簡単で学習容易であり、しかも、個々の漢字が与えられたとき、そのコードが字を見ただけで瞬時に作れること、3) 入力装置として特殊な鍵盤を用意する必要がなく、普通の英文タイプライタと同等の鍵盤を使用すること、等の設計方針を立て、これに沿う方式として‘声形法’を考案し、かつ実現した。

声形法は、漢字一字に最大4文字(4打鍵)、最小1文字(1打鍵)からなるコードを割り当て、最大の4文字の場合、最初の2文字(英字)は、拼音の声母





漢字への変換—漢字出力は、将来性を期待されている方式である<sup>6),7)</sup>。ところが、現代中国語簡体字に対する四角号碼表は、現在のところ、中国においてもまだ完全ではないのが実状である。声形法を実用化するためには、簡体字を基本とする中国語第一級、第二級漢字を含む6,763字について四角号碼表を整備することが前提として必要である。われわれは、本稿に述べた事柄に矛盾しないようにこれを整備した<sup>3)\*</sup>。その際、原典ともいうべき文献5)\*\*の方式を尊重した。なお、この表を作成するに当たっては、当研究室所属の学生井川智、宇野誠両君ほか数名の努力に負うところが大きであった。

### 参 考 文 献

- 1) 諸橋轍次, 渡辺末吾, 鎌田 正, 米山寅太郎: 新漢和辞典, 改訂版, 大修館, 東京 (1982).
- 2) 新華字典, 1971年修訂重排本, 商務印書館, 北京 (1971).
- 3) 川口喜三男, 王 思鴻: 四角号碼付汉语拼音索引, 名古屋工業大学情報工学科川口研究室, (1982).
- 4) 中华人民共和国 国家标准 信息交換用汉字編碼字符集 基本集 GB 2312-80, 汉语拼音索引, 国家标准总局发布, 中华人民共和国, 1981年5月1日实施 (1981).
- 5) 王 雲五: 四角號碼國語字典, 第二次增訂版, 商務印書館, 南京, 1945年 (初版, 1935年).
- 6) 村田 茂: 中国におけるコンピュータ技術と漢字処理, インターフェース, 1980年9月号, pp. 174-179.
- 7) 村田 茂: 中国語処理システム, *bit*, Vol. 13, No. 1, pp. 44-50 (1981).

(昭和57年11月17日受付)

(昭和58年1月17日採録)

\* 要望があれば、コピーの配布が可能である。

\*\* 四角号碼は文献5)の著者王雲五氏によって考案されたものであり、初めて発表されたのは1915年であるようである<sup>1)</sup>。