Tコードの補助入力：字形組み合わせ法と交ぜ書き変換法†

小野 芳彦††

2ストロークコード入力であるTコードの使い勝手を良くし，また実務に就くまでの練習期間を短くするために，二つの補助入力方式を開発した．一つは字単位の合成を行うもので，Tコードで入力した2文字の字形を組み合わせてその字形をもつ漢字を入力フロントエンドが探索することによってコード化されていない文字でも入力できるようにしたものである．JIS X 0208 の全漢字について字形を二つの部品に分ける試みを行い，2文字から直接，あるいはその部品から間接に合成を行って目的の漢字を検索するアルゴリズムを実現した．これは，従来の字形入力がもつコードの重なりを極端に低くしている．もう一つはコード化入力方式に熟語のカナ漢字変換機能を融合したもので，被変換表記にTコードで入力できる漢字を交ぜるようにした方式である．これによってカナ漢字変換の欠点である同音語の選択の濃度を低くでき，変換結果を目視しないで打鍵を続ける可能性を高くした．通常の設計では辞書の大きさが数倍にふくれるのを，コードの習得グレード別の辞書を作るという方式で押さえている．

1. はじめに

現在，最も普及している日本文入力方式はカナ漢字変換方式である．この方式では，カナ文字の入力方法を習得するだけで一応の実務をこなせるという利点がある．これは，それほど頻繁に文章を入力しない利用者にとっては便利なものであり，この点がカナ漢字変換入力方式を普及させる大きな要因であった．

しかし，一方，比較的大量の文書作成においてもこのカナ漢字変換が利用されている．われわれは，このことがOA病といわれる精神的疲労の蓄積を招く大きな要因となっているのではないかとの疑いをもっている[1]．

われわれは，日本文入力のためにコード化入力方式「Tコード」を設計し[2)-4)]，数名のタイピストを養成するなどの実際の運用を行ってきた[1]．Tコードは通常の英文キーボードの2打鍵を日本文の1文字に対応させたコード入力法で，設計段階から疲労の原因となる要素をできるだけ排除しており，継続的な文書作成に向いた，いわゆる専任者向きの日本文入力方式である．

コード方式側の欠点として，その習得に時間がかかることが挙げられてきている．しかし，運動能力を身につけるには，本質的に練習を繰り返すことが必要であり，カナ漢字変換においてもその点に変わりはない．カナ漢字変換入力（詳しくいえばそのうちのカナ入力）も，繰り返し使ううちに習熟してくる．しかし，熟練した状態での入力速度や疲労度の点でカナ漢字変換には問題がある[5),6)]．コード入力方式の欠点として挙げるべき点は，実務に就くまでの練習期間が長いということである．

実務につくためにはタイピストが習得している文字セットによる文のカバー率がある程度以上の高さでなければならない．練習時間が長くなるのは，コード打鍵のみを入力手段としてカバー率を上げようとすることが原因である．ある程度の数しかコードを記憶・習得していない段階でも，初見テキストのコピータイプという実務に就けるようにするためわれわれはTコード入力の拡張手段となる補助入力手段を2通り開発した．もちろん，熟練したコードタイピストもこの方式を有効な補助入力とすることができる．

まず，2章でTコードの概要を示したのち，3章で第1の外字コード入力方式について述べる．これは，コードを割り当てていない文字（外字）をTコードを使って入力する手段である．これによって，適当な規模のコード化文字セットだけで非常に大きな文字セットを取り扱うことができるようになっている．4章では，第2の交ぜ書き変換方式について報告する．これは，コード化入力方式とカナ漢字変換入力方式とを融合させたものである．字単位から単語（あるいは文節）単位に拡張して入力する場合，通常のカナ漢字変換ならば宿命となる同音語の選択のわずらわしさを，この方式では著しく減らすことができる．

なお，本論文に述べる二つの補助入力方式は，MS-DOS のコンソール・ドライバの型式で実現しており，Tコードの実用化に大いに役立っている．

† Auxiliary Input Methods for T-Code System: a Kanzi-Form Combination and a Kanzi-Mixed Conversion by YOSHIHIKO ONO (International Research Center for Japanese Studies).

†† 国際日本文化研究センター

2. Tコードの概要

本論文の議論で必要なTコードに関する設計時の検討[2]の概要は以下のとおり．

(1) キー

—英文標準キーボードの4段中央10列のキーを使用する．その他のキーは，小指の異常な使い方を強いるので，Tコードの入力には使用しない．打鍵した場合は，キートップに刻まれた記号文字の入力となる．

—シフトキーは使用しない．ロック型のシフトによってシフトロックされていても，入力される文字には影響を与えない．

(2) 文字

—JIS X0208 (1983) に含まれる，ひらがな・カタカナ・算用数字・句読点・記号・漢字(第1水準) のうち使用頻度の高いものをコード化する．

—英字は，そのキーボード本来の配列によって一打で入力する．Tコードとはモードで切り換える．

(3) コード割り付け原則

キーボードの打鍵の解析から得られた下記の知見[7]に従って，日本文の打鍵列が最適になるようにコードを割り付ける．

—段・指の負荷分布（静的特性）

- 段の負荷は中段が小さく，再上段が最も大きい．
- 人差し指・中指・小指・薬指の順に負担能力が高い．ただし，人差し指には他の指の2倍以上の負担能力があるわけではない．

—打鍵組の特性（動的特性）

- 左右の手の交互打鍵が最も高速である．
- 同じ指の段越えなどの Awkward Sequence は低速で疲労が大きい．

これらの難易を数値化した式でキー組の打ちやすさの順位を計算し，出現頻度順に並べた(2)の文字の上位1200字に順に割り当て，文字コードとした．コードが割り当てられた文字を基本文字と呼んでいる．

コードの頭脳による記憶を助けるために2文字のニモニックを付加して，ニモニックの字をキーに割り当てるいわゆる連想コードや，かなの50音配列や子音・母音構造をコードの空間的配列に利用するいわゆるローマ字配列は，それぞれの文字の出現頻度に基づく最適化とは相いれないので採用できない．

3. 外字入力方式

3.1 コード化と外字の関係

Tコード・システムでは，2打鍵で入力できる基本文字に漢字が約1000字含まれている．原則として出現頻度の高いほうが基本文字となっているので，通常の文書ならば基本文字だけで延べ95%〜98%の文字をカバーすることができる[8]．しかし，この数字を裏返して読むならば，一行を40字として，一行当たり1ないし2文字の外字が出現することを意味する．これは，心理的にはかなり大きな出現率である．

この外字の出現率を小さくするために基本文字セットを大きくすることも考えられないではない．しかし，タッチタイプという観点からはキーの数を多くできず，タッチ数を増やすことになる．2打鍵のコード化の方針を同じように拡張して，3打鍵以上のコードを追加することは可能である．たとえば，前章のTコード割り当てで未使用の390組の2打鍵コードを3打鍵コードへの導入コードとすれば15600のコードが新たに割り当て可能になる．われわれと同様な方針のコード化入力である TUT コード[9]は，このような2打鍵と3打鍵の混在コード方式をとっている．ただし，そのコードの総数は2525で常用漢字と人名漢字をカバーするに留めている．

こうやってコードを大量に増やしてみても，これらのコードに割り当てられた文字の出現頻度は非常に低いので，コードの獲得にはますます困難が伴うようになる．しかし，タッチタイプに慣れると打鍵数の増加はほとんど苦にならない．コード空間は打鍵数を増やすことによって非常に大きくなるので，外字のためにはもっと冗長なコード化を行って，コードの獲得の困難さを救わなければならない．すなわち，外字には字の構造や意味などによる連想可能なコード化を行ったほうが得るところが多いと考えられる．

3.2 外字入力の概観

文字の入力の対象を JIS X 0208 の文字セットに絞る．これは，実用上大きすぎることはあっても，小さいということはない．その総計6877字個々を完全に指定できる方式には，あまりくふうの余地はない．つまり，JIS コード（あるいは等価なコード）を直接指定する，あるいは（一度には一部分だけの）一覧表を提示してそこから選択させるといった方法しかなさそ

うである．Tコード・システムも，生のJISコードを16進数字で指定する方式を最終的な外字入力手段として設けてある．しかし，これは，いちいちコード表を検索する手間が大きく，打鍵効率は極端に悪い．どうしても入力手段がない場合の最後の手段である．

コード入力の実用化においては，極端に使いづらいが完全であるJISコード入力と速度を重視した2ストローク入力との中間の段階の外字入力方式が望まれる．そして，こういう中間的文字入力の要件は以下の三つになろう．

（1） 入力キー列を組み立てあるいは思い出すために，単純で分かりやすい規則(性)か連想がある．

（2） 一つの文字に多くの入力列が対応しても，それをただ1通りだけに制限していない．

（3） 一つの入力列に多くの文字が（できれば）対応していない．

外字は出現頻度が小さいので，指（すなわち体得した記憶）ではもちろんのこと，頭（すなわち意識した記憶）ででも個々の入力列を覚えることは難しい．したがって，条件(1)は特に重要である．また，試行の回数を少なくする（できれば一度で入力できる）ために，考えられるいく通りかのどれを試してみても入力できるように，条件(2)の必要がある．条件(3)については条件(2)との両立が困難なことがあり，その意味で「できれば」という表現になっている．一つの外字入力列に二つ以上の文字が対応するような方法では，選択のための機構を用意する必要がある．しかし，入力作業にあまり多くの文字の選択や頻繁な選択が付随すると，作業の心理的な疲労が大きくなる．したがって，条件(3)は，ごく一部が二重または三重の対応をもつ場合に限定することを念頭においている．

3.3 字形組み合わせ方式の外字入力

Tコードの外字入力は6打鍵が1文字となる．まず，外字入力の始めを示す機能文字「◆」(基本文字の一つに定義済み）を打鍵する．続く4打鍵で基本文字2文字を入力する．入力システムがその2文字から対応する1文字を計算(検索)してプログラムに渡す．以下の説明で，英大文字はその字が基本文字であることを示し，英小文字は基本文字であってもなくてもよいことを示している．

（I） 漢字 k が二つの部品*に分かれていて，それぞれの部品が基本文字 V と W であるとき，その二つの基本文字を入力する．

$k=V+W$.

（II） 漢字 k が二つの部品 v と w に分かれていて，v を部品にもつ基本文字 L と w を部品にもつ M があるとき，その二つの基本文字 L と M を入力する．

$k=v+w$, $L=x+v$, $M=y+w$.

（III） 上記の(I)と(II)の混合．漢字 k が二つの部品 v と W に分かれていて，v を部品にもつ基本文字 L があるとき，二つの基本文字 L と W を入力する．

$k=v+W$, $L=x+v$.

（IV） 漢字 k 全体がある基本文字 L の部品として含まれていて，しかも，その残りの部分 X が基本文字であるとき，L と X の二つの基本文字を入力する．

$k=L-X$.

この4種類の入力例を**表1**に示す．打鍵する2基本文字はどちらを先に入力してもよい．

タイピストがこの方式で外字を入力する場合の難易は，基本文字打鍵とほぼ同じになっている．というのは，外字コードとして打鍵する基本文字（あるいはその部品）は，入力漢字 k の部品（あるいは k 自身）としてタイピストの目に見えているからである．(I)から(III)においては，タイピストは見えている V や W を打鍵できるかどうかを判断すればよいし，また，自分が打鍵できる L や M を一つ思い浮かべるのもそう難しくない．

同様に(IV)においても，L を思い浮かべるのは難しくないが，引算をした残りである X を自分が打鍵できるかどうかという条件判定が加わるため，L としてたくさんの候補を考慮しなければならなくなる．この点で，4通りのうちで(IV)は特異であり，実際の適用が難しくなっている．

多くの漢字に見られる部品がたまたまJISセットに

表1 外字の組み立ての例
Table 1 Examples of outside character code.

(I)	韻＝音＋員 翼＝羽＋異 褒＝衣＋保	(II)	悟＝性＋語 毒＝青＋毎 廷＝任＋建
(III)	竣＝立＋酸 芥＝花＋介 辿＝山＋道	(IV)	也＝地－土 貝＝貨－化 韋＝衛－行

* 漢字字形を分割してできる部品字形は部首と呼ばれることが多いが，正確には部品のうち辞書のインデックスとして採用されているほうだけを呼ぶ名前である．ここでは誤解を避けるために，部品という一般的な名詞を採用した．

存在しないがためにさらに細かく分かれるのは好ましくない．たとえば，「おおざと」は単独では JIS セットには存在しない．ここでおおざとに元の文字「邑」を当ててもよいのであるが，字形としてはあまりに掛け離れすぎている．そこで，おおざとを含む最もありふれた文字「部」をおおざとの代りに部品とする．実際の指定においては，タイピストは(Ⅱ)または(Ⅲ)のつもりで「部」を使うのであるが，システムは(Ⅲ)または(Ⅰ)の解釈で正しい外字を検索する．もし，おおざとの指定に「部」以外を使った場合は，その文字の部品として「部」が含まれているはずであるから，意図どおりの入力になるはずである．ただし，相手の文字によっては「部」の偏が活躍することがあるのは欠点である．

われわれの方式では，さらに，上記の基本文字の代りに(Ⅰ)から(Ⅳ)を適用して得られた外字を利用するような再帰的な適用も採用している．この方法を使って三つ以上の部品に分解して漢字を入力することもできる．

(1) 荻＝花＋(独＋火＝狄)，遜＝(子＋系＝孫)＋道，髭＝(長＋形＝髟)＋(止＋化＝此)

(2) 揖＝(打＋口＝扣)＋耳，葺＝(花＋耳＝茸)＋口，揖＝打＋(口＋耳＝ϕ)，葺＝花＋(口＋耳＝ϕ)

(3) 煙＝火＋(西＋土＝ϕ)，稲＝秋＋(受＋旧＝ϕ)

ただし，この方法が適用できるのは，二つの部品から組み立てられる中間の段階の漢字がJISの文字セットに存在する場合に限られる（例(1)）．この場合は，中間の漢字の形が大きく歪んでいても，とにかく存在しさえすればよい（例(2)）．また，組み合わせ方も，ただ1通りしか許されない．中間の漢字が存在しない場合（例(3)）は，外字入力法の再帰的な適用でも組み立てることはできない．

上記の例(3)のように，三つまたはそれ以上に分解して初めてその部品がJISの文字セットに該当するような漢字が1パーセント未満存在する．これらは，当初に示した枠組みでは決してコード化できない．しかし，その任意の組み合わせに該当する文字が存在しないのであるから，そのうちの適当な二つを部品としても差し支えはない．たとえば，煙＝火＋西，稲＝禾＋旧としている．タイピストがこれらを入力する場合，部品を組み合わせている途中の段階で目的とする漢字が得られてしまう，ということが起こるだけのことである．

3.4 漢字部品データベース

この方式を実現するために，基本部品となる漢字を除くJIS漢字の全部を二つの漢字部品に分解し，そのJIS コード三つ組をデータベースとした．

現在の漢字の字形は，元は同じ部品が見かけが変っていたり，元が異なっていても今は同じに書くものがあったりする．また，部品自身が入力の対象となる(Ⅳ)や再帰的適用を除いては，部品へのコード割り当ては，いくつかの漢字が同一の部品を共有していることを表現するために必要なのである．したがって，漢字部品のコード割り当てにおいては，漢字の字形に厳密な峻別を行わず，字形や配置の大胆な変形を許容している（たとえば**図1**）．また，部品を漢字に限る必然性はないので，漢字の部品に起源があるカタカナなどを援用している（たとえば**図2**）．歴史的経緯を経て，現在は部品が簡略化されている漢字が数多くある．しかし，部品単独を文字として使う場合にはその簡略化が及ばないものも多い．われわれは，このような場合に，簡略化された文字に対しても簡略化されていない漢字を部品とすることにした．

桜＝木＋嬰，験＝馬＋僉，残＝歹＋戔，譲＝言＋襄，棲＝木＋妻，

これに伴って，新しい字体と古い字体は同じ部品をもつことになってしまう．古い字体をわざわざ使用するのは限られた場合であるから，それらを入力する場合には，後に述べる外字検索の連想的な使用法で対処することにした．

このような部品コードの同一化は，上記以外にも存在する．一つは，同一の漢字が部品の配置を変えて表記される場合である．JIS X 0208 コードはこのような漢字にも独立な番号を与えている．たとえば，「峯」と「峰」は同じ意味で部品も同じ異字体である．第一水準内ではこの組しかないが，第二水準を含めると配置が異なるだけの異字体が 19 組存在する．これらの異字体も特別な場合にしか使われないであろうから，

奄＝大＋長，舎＝人＋吉，丘＝斤＋一，
道＝之＋首，育＝云＋月，第＝竹＋弔

図 1 部品を同一とみなした例
Fig. 1 Examples of equivalent treatment of slightly different parts.

阿＝ア＋可，休＝イ＋木，宙＝ウ＋由
礼＝ネ＋レ，軍＝ワ＋車，冷＝ン＋令
合＝△＋口，弔＝弓＋1，与＝5＋一，玉＝王＋、
エ＝工，カ＝力，タ＝夕，ニ＝二，ヘ＝人

図 2 カタカナなどの漢字部品への援用
Fig. 2 Examples of katakana, etc. as kanzi parts.

旧字体と同じ取り扱いとしてかまわない（後述）．

異なる漢字が同じ部品をもつ場合もある．

縦並び　栗某脅累含
横並び　栖柑脇細吟　杏（呆＝保ーイ）

上の例はいずれも第一水準内の重なりであるが，第二水準まで含めると総計 33 組の部品コードの重なりがある．これは，部品への分解を行った漢字6170字（第一水準 2965－137 と第二水準 3388－46）の実に 0.9% にしかすぎない．これらは，原理的にどちらか一方しか入力できない．

われわれは，1字について1通りの部品分解を前提とした高速の探索アルゴリズムを採用している．したがって，部品の分解方法が2通り以上ある場合は，そのうちの一つだけが正しく働く（たとえば表 2）．したがって，ユーザが部品の組み合わせを指定しても，うまく変換されないことがある．ただし，このような曖昧な例はあまりないので，データベース中に2通り以上の分解を登録するような救済の必要性は少ない．この問題は実現の方法にも関わる事項であり，データ空間の効率やアルゴリズムなどに影響を及ぼす．われわれは，データやアルゴリズムの単純性のほうを重視した．

3.5 外字検索の連想的利用法

漢字以外の文字についても，原理的にわれわれの外字検索法が利用できる．たとえば，各種の括弧は「開き」か「閉じ」かという属性とその固有の形からなっている．これを漢字の部品に見立てて表 3 のように分解する．ここに挙げた部品は字形として含まれているわけでは決してない．（出＝山＋山という構造も含めて）多少アドホックな名前付けであるが，これらをいったん記憶してしまえば，字形から部品を抽出するのと同じように字形から外字コードを連想することは容易になる．

表 2　漢字が2通りに分解できる例
Table 2　Examples of two ways to divide a character form into two.

漢字	採　用	不採用
克	十＋兄	古＋ル
章	立＋早	音＋十
灘	シ＋難	漢＋隹
彬	木＋杉	林＋彡
諸	サ＋諸	言＋著

表 3　外字括弧の部品
Table 3　Outside parentheses.

開き括弧	閉じ括弧
〈＝山＋（	〉＝山＋）
《＝出＋（	》＝出＋）
〔＝半＋（	〕＝半＋）
［＝大＋（	］＝大＋）
｛＝中＋（	｝＝中＋）
【＝黒＋（	】＝黒＋）
『＝「＋「	』＝」＋」

一般的な記号も，その連想の付け方の規則を簡明なものとすることで，字形とは関わりなくコードを付けることができる．たとえば，

＋＝た＋す　－＝ひ＋く
×＝か＋け　÷＝わ＋る

のようにしている．

この延長として，旧字体の文字は「(新字体)＋古」と部品化し，異字体の文字は「(本字体)＋異」と部品化している．これらは，真に「古」や「異」を部品としてもつ場合と区別がつかないが，衝突が起きる例はほとんどなかった．現在，これらの文字は使われなくなってしまって，タイピストが旧字体や異字体であるという知識をもたない可能性が強いことがむしろこれらの文字の入力の場合に欠点になっている．

3.6 字形組み合わせ入力方式の評価

われわれの外字コーディングは，本質的には1文字に順序のない二つの連想コード (associative code) を付加するものである．漢字の部品をその文字のコード構造として用いれば，前述の外字3条件をほとんど満足するものになることが，実際に作成して明らかとなった．

また，変換型の入力は音や読みの登録を前提とするため，頻繁には出現しない単独の字（たとえば人名の漢字など）の扱いに非常に苦労するが，本方式は字形を元にしているため，それらの取り扱いにも威力を発揮する．

日本では，読みを使った入力が一般的に好まれる傾向にあるが，中国の場合は単漢字を読みだけで区別することは難しい．中国語の文字の入力のための各種のコード化方式は，多く字形に基づいて考案されてきている[10),11)]が，コードの重なりが大きく，会話的な後選択が必須となっている．われわれと類似の漢字の部品を組み合わせて入力する方式も提案されている[12)]が，キーの数などの評価だけで実用に供しているのではなさそうである．われわれと異なるのは，部品にしか現れない文字のために効率の良い打鍵列を割り当てている点である．これは，記憶することの負荷を増す

とともに，全体の打鍵効率を落とすことになっている．

部品データベース自身は，原理的に基本文字セットとは独立である．Tコードの基本文字も外字として入力可能であるし，また，Tコードの個人向けのチューニングをしても，外字入力の性能には大きな差は現れない．さらに，Tコードでなくても，単漢字の入力が容易なシステムであればコード入力ですらなくても，本外字入力システムの有効な利用が期待できる．中国語入力に応用することも，字体やコード系の手当をすれば可能となろう．

実務についたタイピストが実際に使用した外字1文字の打鍵時間は以下のようになった．漢字の合成を意識的に考えた場合では，各文字の間で打鍵が滞るため平均的に1.8秒ほどかかり，すでに何度も打っているために打鍵が滞ることがない場合では 0.78 秒ほどかかる．Tコード1文字の打鍵の平均が0.2秒強なので，ほぼ3倍の時間で打鍵できている．本方式の外字入力は，打鍵速度に関しては熟練に支障がないことが明らかになった．

4. 交ぜ書き変換入力方式

4.1 カナ漢字変換の導入

コード化入力にカナ漢字変換を導入するのに，カナ漢字変換のカナ入力をコード入力を用いて行うだけのものがある．たとえば，前述の TUT コードは，カナのコードをローマ字型に構造化して初期獲得の時間を短縮することによって，積極的にこのレベルの導入を行って商品化の方向を進めている．

ここでは，もう一歩踏み込んで，被変換文字列を「読み」を表すためのカナ文字列に限定せず，漢字を含む「交ぜ書き」の文字列を対象として全漢字表記に変換するような導入を実現した．たとえば，変換の同音語のコンフリクトを示す例としてよく用いられる「貴社の記者は汽車で帰社した。」は，具体的には**図3**のような方法で変換される．ただし，ここで用いた「社」「者」「車」のコードは獲得されているとする．以下では，このような変換を「交ぜ書き変換」と呼ぶことにする．

き社の
↓変換
帰社の　き者は き車で き社する
↓次候補↓変換↓変換↓変換
貴社の　記者は 汽車で 帰社する。

図3 交ぜ書き変換の例
Fig. 3 An example of mixed-form conversion.

交ぜ書き変換の目的は，タイピストがコードを獲得している漢字を被変換文字に交ぜることによって，変換に伴う選択の幅（濃度）を小さくし，円滑な入力作業を進めようとするものである．同音異義語の数の最も多い「こうしん」について，いくつかの国語辞典にのっている単語をすべて集め，それらを第1字目および第2字目について出現頻度でソートし2次元に表示したものが**図4**である．熟語の一部を漢字で表記することによって，選択軸が2次元に増え，その度数が総数 26 から1から3に，最大でも8にまで減らせることが，この図から見てとれる．特に，交ぜ書きにした場合に候補が一つに限定される場合が全体の半数近くある．これは，変換結果を目視で確認せずにタイプ作業を続けても，かなりの場合支障がないことを示す．このことはタイプ作業にとっては重要なことであり，作業能率の向上・精神的疲労度の軽減に大きく貢献する．

4.2 交ぜ書き変換のデータ構造

交ぜ書き変換の実現法については，2通りの方式を検討した．一方を「読み方式」，もう一方を「交ぜ書き辞書方式」と呼ぶ．

交ぜ書き変換の実現法は，われわれとほぼ同時期に TUT コード入力でも検討され，「漢和辞典法」と名付けられた独自の方法を開発している[13]．この方法は，交ぜ書きされた漢字をインデックスとして別途用意された辞書を引くものである．なお，交ぜ書き辞書方式と同様に（後述），辞書の膨張が小規模システムでの障害となることが予想されるが，その対策は示されていない．

読み方式

被変換文字列に含まれた漢字について，その字に可能な読みすべてを組み合わせて，カナ文字列を作り，それを漢字変換して得られた文字列と元の入力の照合を行うという方式である．読み方式には以下のような技術上の問題点がある．

（1） 連音による読みの変更（連濁，連声，転韻）を手当しなければならない．

（2） 単語のつぎ合わせを行うことで未登録語を（見かけ上では正しく）変換するいわゆる複合語変換と呼ばれる方式を用いると，大量の候補が発生して交ぜ書き変換の利点を失ってしまう．

（3） 変換候補が辞書の複数の見出しに属するた

め，合理的な提示の順序を決められない．また，有効性の高い最尤法学習機能を用いることができない．

連濁，連声，転韻のルールとしていくつか知られているものはあるが，一般的に現代語で連濁，連声，転韻を正確に決定する法則は知られていない．対策として，

（a） 連音変化の可能性のあるすべての読みを登録する

（b） 連音変化を元に戻した読みで語を辞書に登録する

の二つが可能である．しかし，（a）では，本来存在しない読みを検索する無駄が辞書探索の時間を増やすことになり，（b）では「感謝」などの一連の正しい表記に「患者」などの候補が混ざることが新たな混乱を招く．

読み方式には，一般のカナ漢字変換の辞書とアルゴリズムをそのまま利用できるという利点があるが，上述の欠点が大きいため採用しなかった．

交ぜ書き辞書方式

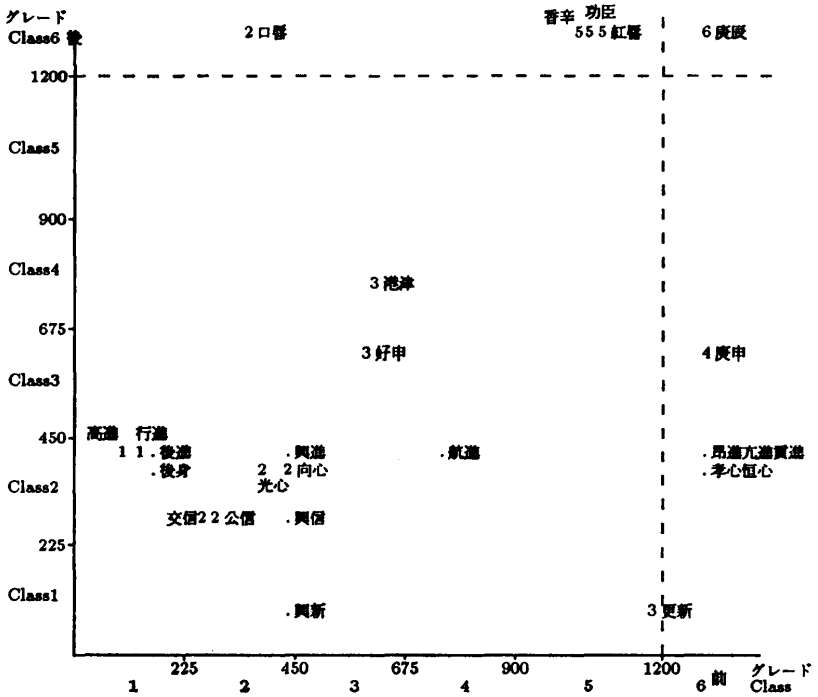
交ぜ書き辞書方式の着想は，交ぜ書き表記の見出しを辞書に登録するという単純なものである．たとえば，単語「汽車」に対して，「きしゃ」「汽しゃ」「き車」の三つの見出しを登録するということである．

ただし，単純に読みを展開しただけでは，辞書を数倍にふくらませてしまう．漢字入力システムは，小規模でも使えることが望ましく，辞書の肥大は深刻な障害となりかねない．そこで，展開に制限を設け，辞書を小さくする工夫をほどこす必要がある．

ひとつは登録語の見直しである．通常のカナ漢字変換では語長が長いほうが同音語の選択濃度が小さいため，積極的に合成語や接辞付きの語を登録している．交ぜ書き変換では短い基本単語でも同音語の選択濃度が十分小さく，また，接辞となる漢字の出現頻度は一般に高く，コードで直接打鍵できてしまう．このため，交ぜ書き辞書への登録は基本単語に限定できる．

われわれは一般のカナ漢字変換用の辞書から交ぜ書き辞書を作成した．元の辞書の合成語の分解等を行うことで，漢字文字数の分布は表4のようになり，平均漢字数は2.0文字になった．漢字の読みの展開数は漢字数のベキで増えるから，効果は大きい．

もうひとつは，利用者の漢字習得の程度に合わせて辞書を変えるというものである．利用者がコードを習得した漢字が少ない場合は交ぜ書きはあまり使えず，



単語の位置をピリオドまたは数字で示している．数字は，その単語が選択なしで確定する交ぜ書き変換のクラス2のグレードを示す．

図4 同音熟語の漢字頻度順2次元プロット
Fig. 4 Two dimensional plot of homophone words by the character rank.

表 4　カナ漢字辞書の単語の分解結果
Table 4　Change of the number of kanzi in a word vs. the number of words in a dictionary after dividing complex words.

漢字数	元の辞書	分解後
0 (記号など)	613	228
1 (単漢字)	3257	3164
1 (単語)	3759	3870
2	20448	20340
3	3965	3691
4	2670	205
5以上	568	0

習得数が多くなると交ぜ書きが増え，逆にカナのみの見出しは使われなくなる．そのときどきに使われる表記のみを辞書に登録することにすれば辞書を小さくすることができる．

交ぜ書き辞書の具体的な設計は，以下のとおり．

(A)　タイピストが到達している習得のグレードをいくつかのレベルに分ける．

(B)　各レベルごとに，以下のような漢字のクラス分けを行う．

1. 確実に獲得されていると仮定する漢字
2. 獲得されることが期待されている漢字
3. 獲得されているとは期待できない漢字

非基本文字は3とみなす．

(C)　1～3のどれに属するかによって，見出し表記を以下のように決定する．

(1)　クラス1に属する字は漢字表記とする．

(2)　クラス2に属する字は，カナ表記のものと，漢字表記のものを両方用意する．クラス2に属する字が複数存在する場合は，それぞれの組み合わせすべてを用意する．

(3)　クラス3に属する表記はカナ表記とする．

(D)　見出しが変換する必要のないものとなった単語は辞書には登録しない．

(C)の各項目および(D)で登録表記数の増減がある☆．

☆ これらのほかに，読み方の差異によって2通りに登録されていた語が交ぜ書き表記では1通りになるもの
例「あくしょう」/「あくせい」→あく性→悪性
読みが異なる語が同じ表記で入力されるため，選択の濃度が逆に高くなるもの
例「せいし」→せい子→精子/聖子/征子/清子/…
がそれぞれ若干ある．

以上の方針で，上述の約3万語の基本単語からなる交ぜ書き辞書をいくつか作成した．それに先立ち，利用者の漢字習得を出現頻度順位に従うとした仮定によって，交ぜ書き辞書の大きさを計算した．

まず，出現頻度順に漢字を6グレードに分ける（表5）．辞書の語をこのグレードの組み合わせで細分すると表6のようになる（表内の数字は全語数に対する百分率）．この表は，単語内の漢字のレベルの最大値と最小値を縦軸と横軸にとったものである．対角線上に＋で加算されている数字は，漢字が1字しか含まれない（したがって，両グレードが一致する）表記の割合を示している．

前述のように，辞書に登録される単語はほとんどが2文字である．その場合，漢字の習得のクラス 1,2,3 の組み合わせに従って図5のように見出しが作られ，

表 5　漢字のグレード分け
Table 5　Kanzi grades by usage rank.

グレード	頻度順位	漢字数
1	1-225	121
2	226-450	185
3	451-675	211
4	676-900	216
5	901-	296
6	外字	—

表 6　辞書の単語のグレード別分布
Table 6　Distribution of word grades.

1	5.7+2.0
2	9.4　3.2+2.1
3	7.4　4.5　1.5+2.1
4	5.3　3.4　2.0　0.7+2.0
5	4.7　3.1　2.0　1.4　0.6+2.6
6	8.4　5.1　3.0　2.2　2.1　1.8+11.7
最大グレード	1　2　3　4　5　6 最小グレード

第1文字クラス	単漢字単語	第2文字のグレード・クラス		
		1	2	3
1	字 0倍	文字 0倍	文字，文じ 1倍	文じ 1倍
2	字，じ 1倍	文字，も字 1倍	文字，も字，文じ，もじ 3倍	文じ，もじ 2倍
3	じ 1倍	も字 1倍	も字，もじ 2倍	もじ 1倍

図 5　見出しの型と見出し数の増え方
Fig. 5　Representation style and increment ratio of word indices.

辞書に登録される数が変化する．これを表6に適用して辞書の語数の増減を計算したものが**表7**である．最上段最右端は，全語を制限なしに展開した場合に辞書が2.3倍になることを示し，最下欄は全基本文字を習得した熟練者は単語数で 34% に縮小した辞書で十分であることを示している．利用者の上達に従って，クラス2のグレードを大きくしていく必要があるが，同時にクラス1のグレードを徐々に大きくしていけば，辞書の膨張を5割増くらいにおさえることが可能であることをこの表は示している．

4.3 交ぜ書き変換の入力形態

辞書方式を使った日本語入力フロントエンドでは，初心者と熟練者とでは異なる2通りの操作方法を用意している．

初心者は図3にあるように交ぜ書き変換を連続して利用するであろう．操作は次のようになる．

(0) 交ぜ書き変換モードに入る機能キーを押す．フロントエンドはスクリーン最下行に変換・表示バッファを確保する．

(1) Tコードで交ぜ書きの読みを打つ．読みはバッファに表示される．

(2) 変換機能キーを押す．フロントエンドは辞書探索を行い，変換結果をバッファに表示する．

(3) 変換結果が目的の語でなければ（次）変換キーを押す．フロントエンドは次候補をとってきて，バッファに表示する．

(4) 確定機能キーを押す．フロントエンドは変換結果をプログラムに渡す．
必要ならここで(1)に戻って変換を続ける．
(1)の直後に(4)を行えば，無変換操作となる．

(5) 交ぜ書き変換モードから出る機能キーを押す．フロントエンドは確保した最下行を解放する．

熟練者は頻繁に変換を用いることは少ないので，モードではなく単変換機能として実現する．(1)から(3)は同じである．

(0′) 単変換の始まりを示す機能文字「◇」(基本文字のひとつ) を打鍵する．フロントエンドはスクリーン最下行に変換・表示バッファを確保する．

(4′) 次の打鍵で変換が確定し，しかも交ぜ書き変換モードからぬける．この打鍵は通常のTコード打鍵となる．フロントエンドは変換結果をプログラムに渡し，確保した最下行を解放する．

Tコードの基本練習 (100〜200 時間) を終えた段階の上級者が実際に交ぜ書き変換を使うと，**図6**のようになる．「貴」と「汽」や「捨」はTコードの基本 1200 字には含まれない．

4.4 交ぜ書き変換の評価

辞書の中の真の交ぜ書きの割合，すなわち，漢字を見出しに含む割合は表6から求められる．これを**表8**に示す．これによれば，辞書の見出しの 60〜70% が交ぜ書きであり，ほぼそのくらいの交ぜ書き使用率になると考えられる．図4の「庚辰」の例が示すように，交ぜ書き辞書の場合はグレードが進むとカナだけの表記でも変換が確定するケースが増えてくる．したがって，コードを学習してそれに適応した辞書を使えば，交ぜ書き変換の利用はますます有利になってくる．

ここで問題となるのは，変換にたよってコードを学

◇き社　　◆シ気　　◇喜しゃ
↓変換　　⇓　　↓変換
貴社　の記者は汽　　車で　喜捨　に行く。

図 6 上級者の交ぜ書き変換の例
Fig. 6 An example of mixed-form conversion by advanced user.

表 7 グレード化辞書の大きさ
Table 7 Word size of graded dictionaries.

クラス1のグレード	クラス2のグレード					
	0	1	2	3	4	5
0	100	147	179	200	216	231
1		92	124	146	162	176
2			78	100	115	130
3				62	78	92
4					49	63
5						34

表 8 グレード化辞書の交ぜ書き率
Table 8 Kanzi-mixed rate of graded dictionaries.

クラス1のグレード	クラス2のグレード					
	0	1	2	3	4	5
0	0%	32%	44%	50%	54%	58%
1		44%	54%	61%	65%	68%
2			54%	56%	62%	66%
3				60%	68%	73%
4					61%	70%
5						61%

習しなくなる可能性があることであるが，こういう事情を説明してユーザを啓蒙することで対処している．

基本文字を交ぜ書き変換を使って入力した場合に，Tコードをフロントエンドが教えるという機能を企画したこともあるが，トレーニング機能を備えることでフロントエンドが重くなることを恐れて実現には至っていない．

交ぜ書き変換も中国語への応用の可能性がある．中国語にはカナ文字にあたる文字がなく，読みにはピンインが使われるのが通常である．しかし，漢字を発音を表すのに利用することも可能である．すなわち，同音漢字のなかで最も頻度の高い字をカナ表記にあたるものとして交ぜ書き変換を行うことができる．辞書を大きくしてもよければ，1音1文字に限る必要はない．

5. おわりに

本論文で述べた補助入力の方法は，一つは字を基準にしたものであり，もう一つは語を基準にしたものである．頻繁に行う日本文入力においては，大部分の基本となる文字や語を高速に疲れずに打鍵できることが最も重要であるが，残りの文字や語をうまく取り扱うこともおろそかにはできない．これらが使われる場合に応じて適切な方法を選ぶことによって，さらに快適な入力を行うことができるようにした．

タイピストの使用を目的としたシステムであり，今まではパーソナルコンピュータの入力フロントエンドとしてシステムを開発し，実用に供するようにしてきた．今後はワークステーションのためのデバイスドライバや日本語入力サーバとして開発を進めたい．

謝辞 本論文は，著者が東京大学に所属時のTコード開発プロジェクトの一部の成果の報告です．プロジェクトはいくつかの科学研究費の補助を受けています．プロジェクトを指導いただいた東京大学の山田尚勇教授（現 学術情報センター）と有益な助言をいただいた研究室の方々，Tコードのタイピストとしてご協力いただいた皆さん，および，NEC PC-9801 用のフロントエンドのプロトタイプを作成してくださった藤波順久氏（現 SONY-CSL）に感謝します．

参考文献

1) 山田尚勇：専任タイピスト向きタイプ入力法の研究経過，コンピュータソフトウェア，Vol. 2, pp. 344-354 (1985).
2) Hiraga, Y., Ono, Y. and Yamada, H.: An Assignment of Key-codes for a Japanese Character Keyboard, in *Proc. 8th Intern. Conference on Computational Linguistics* (*COLING-80*), pp. 249-256, Tokyo, Sep.-Oct. (1980).
3) 平賀 譲，小野芳彦，山田尚勇：タッチタイプによる日本文入力方式，情報処理学会日本文入力研究会資料，2-3, p. 8 (1981).
4) Yamada, H.: Certain Problems Associated with the Design of Input Keyboards for Japanese Writing, in Cooper, W. E. ed., *Cognitive Aspects of Skilled Typewriting*, chapter 13, pp. 305-407, Springer-Verlag (1983).
5) 岡留 剛，小野芳彦，山田尚勇：タイプ入力作業の構成要素間に起こる干渉，情報処理学会論文誌，Vol. 27, No. 3, pp. 304-311 (1986).
6) 岡留 剛，小野芳彦，山田尚勇：日本文タイプ作業時の認知処理過程の負荷，第3回ヒューマンインタフェースシンポジウム論文集，pp. 325-332, 大阪，計測自動制御学会ヒューマンインタフェース部会，Oct. (1987).
7) Hiraga, Y., Ono, Y. and Yamada, H.: An Analysis of the Standard English Keyboard, in *Proc. 8th Intern. Conference on Computational Linguistics* (*COLING-80*), pp. 242-248, Tokyo, Sep.-Oct. (1980).
8) 国立国語研究所（編）：現代新聞の漢字，秀英出版，東京 (1966).
9) 大岩 元，高島孝明：日本文タッチタイプ入力の一方式，情報処理学会論文誌，Vol. 24, No. 6, pp. 772-779 (1983).
10) Huang, J. K.: Three Corner Coding Method: Its Design and Application, in *Proceedings of 1983 Intern. Conference on Text Processing with a Large Character Set* (*ICTP '83*), pp. 228-230, Tokyo, Chinese Language Computer Society (USA) and Information Processing Society of Japan, Oct. (1983).
11) Jian, S.: A Coding System of Keyboard for Several Kinds of Characters, in *Proceedings of 1983 Intern. Conference on Text Processing with a Large Character Set* (*ICTP '83*), pp. 25-28, Tokyo, Chinese Language Computer Society (USA) and Information Processing Society of Japan, Oct. (1983).
12) Chen, T.: A New Design for Chinese Typewriting Systems, in *Proceedings of 1983 Intern. Conference on Text Processing with a Large Character Set* (*ICTP '83*), pp. 12-14, Tokyo, Chinese Language Computer Society (USA) and Information Processing Society of Japan, Oct. (1983).
13) 喜多辰臣，塩見彰睦，河合和久，大岩 元：2ストローク入力用仮名漢字変換システム，情報処理学会文書処理とヒューマンインタフェース研究会，16-4 (1988).

(平成元年 8 月 8 日受付)
(平成元年 12 月 12 日採録)

小野　芳彦（正会員）

1951 年生．1974 年東京大学理学部卒業．1976 年同大学院理学系研究科修士課程修了．1978 年同博士課程中退．同年東京大学理学部情報科学科助手．1989 年国際日本文化研究センター助教授．理学修士．日本研究における計算機支援についての研究を行う．ソフトウエアツール，日本語処理に興味を持つ．JUS，日本ソフトウエア科学会各会員．