代表読み辞書を用いた交ぜ書き漢字変換

金子 宏† 建石由佳† 鳥原信一†

手書きによる日本語入力システムの補助として，任意のかな漢字交じりの入力を漢字表記に変換する交ぜ書き漢字変換を開発した．これは，かな漢字変換の辞書をかな漢字交じりの検索キーを持てるように拡張して実現した．単純な設計では，辞書の大きさがかな漢字変換の辞書の5倍以上に膨張するところを，漢字に単一の読みを振る「代表読み」と呼ぶ方式によって，辞書検索の対象となる表記に制限を設けることなしに約1.8倍の膨張に抑えた．この辞書を用いた交ぜ書き漢字変換システムを作成し，手書き入力システムに組み込んだ．

A Kana-Kanji Mixture Conversion with a Representative Reading Dictionary

Hiroshi Kaneko,† Yuka Tateisi† and Shin-ichi Torihara†

An auxiliary input method for handwriting system, called Kanji-Kana mixture conversion system, is developed. Kanji-Kana mixture conversion is an enhancement of kana-to-kanji conversion in which input strings may be partially represented in kanji. The system is implemented by keeping those dictionary entries that have search-keys represented in kana and kanji mixture. A new method of dictionary compaction called "representative reading method" is developed. In representative reading dictionary, a part of a key that corresponds to a kanji is converted uniquely to a kana string. The size of the representative reading dictionary is 1.8 times of that of kana-to-kanji conversion.

1. はじめに

1.1 「交ぜ書き漢字変換」

最近，ペンを用いて日本語文を手書きで入力するシステムが現れてきている．手書き入力システムでは漢字を直接入力できるが，画数が多い，文字認識システムが誤りやすい，などの理由で，漢字が直接入力しにくいことがある．この場合に，かな漢字変換を用いてかなから入力することが考えられる．しかし，かな漢字変換を使うと，直接入力しやすい字もわざわざかなで入力しなければならない．かな漢字変換を拡張して，入力に**図1**のようなかな漢字交じりの表記を許せば，入力しやすい漢字は漢字のまま，入力しにくい漢字はかなで入力することができ，手書きの入力の効率を上げることが期待できる[1]．

1.2 これまでの研究

従来，交ぜ書き漢字変換の研究はタッチタイプ入力に対して行われてきた．コード未習得の漢字やコードが定義されていない文字（外字）があるときに，打鍵できる漢字は直接，できない漢字は読みから変換して入力するためである．タッチタイプ入力に対する研究では，辞書に漢字交じりの検索キーを許す拡張により交ぜ書き漢字変換を実現している．

交ぜ書きに対応する辞書の単純な構成法は，おのおのの単語について可能な表記法（単語中のある漢字をそのまま書くかかなにするかの組み合わせ）を検索キーとしたエントリーをすべて辞書に持つことである．しかし，この方法では辞書の大きさがかな漢字変換の辞書の5倍以上になってしまい，実用的でない．辞書が大きくなる理由は，(1)エントリー数が，表記法の数だけ必要になり，おのおのの単語に含まれる漢字の数に対して指数的に増加する．(2)かな漢字変換では1バイトコードであった検索キーが，漢字を含めるために2バイトコードになる，の2つである．辞書圧縮のための工夫として，小野[2]は入力の表記に制限を設ける方針を取り，合成語およびコードの習得段階に応じて入力されないであろう表記法のキーを持つエントリーを辞書から除いて，かな漢字変換の辞書の約0.34倍～約2.3倍の大きさの辞書を作成している．塩見ら[3]は，漢字をインデックスとした辞書を別に作

† 日本アイ・ビー・エム(株)東京基礎研究所
Tokyo Research Laboratory, IBM Research

り，入力に漢字が含まれるかどうかで検索する辞書を分ける（「漢和辞書」）方式で，辞書の大きさを，かな漢字変換の辞書の約2.8倍に抑えている．

以上に述べたように，交ぜ書き漢字変換の中心課題は辞書圧縮であるといえる．

1.3 われわれの研究

われわれは，かな漢字変換の辞書の2倍以下の大きさの辞書を，検索対象となる入力表記を制限せずに実現したいと考えた．手書き入力システムは，多くの場合小型のコンピューターの上で動く．さらに，文字認識システムのためのデータと交ぜ書き漢字変換の辞書を同時に持つ必要がある．手書き入力のためには，タッチタイプ入力よりもさらに，辞書圧縮が重要である．また，手書き入力における漢字とかなの使い分けには前述のようにさまざまな要因があるのですべての表記法を検索キーとして辞書に持つことにした．

本論文は，手書き入力と併用するために新たに考案した辞書の圧縮法と，それをもとに作成した交ぜ書き漢字変換の概要について述べるものである．第2章では，われわれの考案した「代表読み」による辞書の圧縮法と，それを用いた交ぜ書き漢字変換について述べる．第3章では実用化したシステムとその評価について述べる．

2. 代表読みを用いた交ぜ書き漢字変換

2.1 かな漢字変換

われわれは，かな漢字変換プログラムを拡張して交ぜ書き漢字変換を実現した．まず，拡張のもととなったかな漢字変換について簡単に説明する．単文節かな漢字変換は

1. 自立語辞書参照
2. 付属語解析・接続検定
3. 最尤候補決定

の3つのステップからなる．「へんかんする」をかな漢字変換した例を**図2**に示す．

かな漢字変換を交ぜ書き漢字変換に拡張するためには，かな漢字交じりの入力から自立語辞書の検索ができるようにステップ1を拡張しなければならない．「へん換する」を交ぜ書き漢字変換した例を**図3**に示す．

図1 交ぜ書き漢字変換
Fig. 1 Kanji-kana mixture conversion.

図2と図3を比較して明らかなように，かな漢字変換と交ぜ書き漢字変換との違いは，入力に漢字が含まれることにより自立語辞書検索結果が異なることのみである．すなわち，かな漢字変換の自立語辞書参照部を，交ぜ書きに対応するように拡張すれば，交ぜ書き漢字変換を実現することができる．図2と図3では太線で囲った自立語辞書検索結果のみが異なる．

連文節かな漢字変換の場合には，まず上記のステップ1および2をループして入力文字列の一部分からなる文節を必要なだけ生成する．その後にステップ3を実行して変換結果を得る．この場合にもステップ1のみを拡張して交ぜ書き漢字変換を実現することができる．

2.2 かな漢字変換辞書の利用についての考察

1.2節に紹介した従来の研究では，可能な表記法をすべて検索キーにしたエントリーを辞書に持っていた．そして，辞書を実用的な大きさに圧縮するための方式を工夫していた．

辞書の大きさを重視するアプローチとしては，かな漢字変換の辞書をそのまま使う方式も考えられる．この方法では，各種の表記法に対応するためにエント

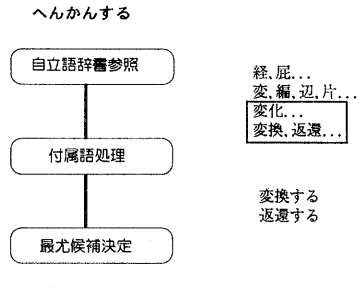


図2 かな漢字変換処理の流れ
Fig. 2 Process of kana-to-kanji conversion.

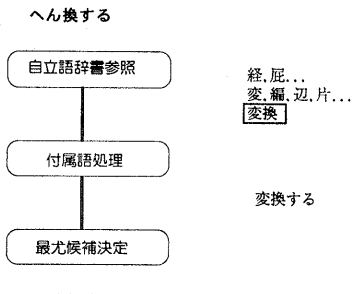


図3 交ぜ書き漢字変換処理の流れ
Fig. 3 Process of kanji-kana mixture conversion.

リーを増加させる必要がない．この場合，入力表記法から「かな表記」を推定することになる．例えば，(1)漢字かな変換を行う，(2)入力中の漢字の可能なかな表記のすべての組み合わせを生成してそのおのおのについて辞書検索を試みる，などの方法が考えられる．これらの方法(1)，(2)に共通の欠点は，かな表記の誤推定による誤変換が避けられないことである．さらに，(1)は漢字かな変換のための辞書の大きさが問題になる，(2)は無駄なかな表記を生成することで処理量が増大する，熟字訓（「紅葉（もみじ）」など）を漢字の組み合わせから生成しにくい，などの欠点がある．

上記の欠点について考えてみると，入力された漢字を辞書検索のためにかな表記に変換する際の不確定性に起因していることがわかる．そこでわれわれは，漢字かやかなへの変換を各漢字について一意に定めること（「代表読み方式」）にした．代表読み方式では，一意に定めたかながその単語のかな表記と一致している場合には，かな漢字変換辞書が利用できる．一致していない場合は，エントリーを追加して対応する．この方法により，エントリーの増加を実現的な範囲に抑えることができる．また，検索キーは「かな」なので1バイトコードである．以下に代表読み方式について詳述する．

2.3 代表読み方式の原理

「代表読み」とは，各漢字に対して一意に定めたかな列である．たとえば，

「変」には「へん」(変更)，
　　「か」(変わる)，
「換」には「かん」(換気)，
　　「か」(変わる)，
「書」には「しょ」(書物)，
　　「か」(書く)

などの読みがあるが，

「変」の代表読みは「へん」，
「換」の代表読みは「か」，
「書」の代表読みは「か」

などと定める．漢字から代表読みへの写像を「代表読みテーブル」と呼ぶ．ある入力に対してその漢字部分を対応する代表読みに変換したかな列は一意に定まる．これを「検索文字列」と呼ぶ．たとえば，「変換」「書き換え」の2語について，8通りの表記法に対応して次の検索文字列が生成される．

変換　→へんか
変かん　→へんかん
へん換　→へんか
へんかん→へんかん
書き換え→かきかえ
書きかえ→かきかえ
かき換え→かきかえ
かきかえ→かきかえ

生成された検索文字列は「へんか」「へんかん」「かきかえ」の3種類である．これを検索キーとした辞書を「代表読み辞書」と呼ぶ．すなわち，代表読み辞書には，「変換」，「書き換え」の2語に対して，

へんか　→変換
へんかん→変換
かきかえ→書き換え

の3エントリーを登録する．8とおりの表記法が3種類の検索文字列で表現されており，単純な拡張よりもエントリー数が削減されている．

以上に述べた代表読み方式の原理を**図4**に示す．「紅葉狩り」のように熟字訓を含む語については，熟字訓の部分は全体をかな書きするか全体を漢字書きするかどちらかである．すなわち，「紅葉狩り」，「もみじ狩り」，「紅葉がり」，「もみじがり」の表記法のみがあるとして，これに対応するエントリーを登録する．

2.4 代表読みの選び方とエントリー数

図4からわかるように，ある漢字の代表読みがその単語中での読み（以後，「自然読み」と呼ぶ）と一致する場合，その一致した部分を漢字書きしてもかな書きしても，検索文字列は同じである．このとき，一つのエントリーによって両方の表記法に対応することができる．代表読みと自然読みが異なるもの（たとえば図4における「変換」の「換」）についてのみエントリーを増大させればよい．代表読みを自然読みと一致するように選んで，エントリーの増大を抑制できる．

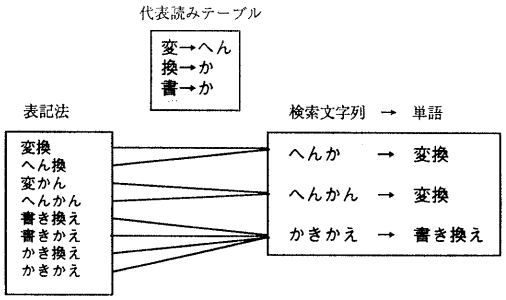


図4　代表読み方式の原理
Fig. 4　Representative reading method.

われわれの目的は，代表読みを上手に選んで辞書を小さくすることである．厳密に辞書サイズを最小化することは困難である．われわれは，3.2 節に示すように「各漢字が辞書中でもっとも多く読まれる読み方」を用いた．

2.5 代表読み辞書の参照

代表読み辞書参照の手順を**図5**に示す．

まず入力の漢字部分を一字ずつその「代表読み」に変換し，検索文字列とする．次に検索文字列をキーとして辞書を検索する．その後，過剰検索結果の削除を行う．過剰検索とは，入力中の漢字を代表読みに写像したときに検索文字列が偶然一致し，不要なエントリーが検索されてしまうことを指す．たとえば，図5で「へん換」の入力に対し，「へんか」を検索文字列として「変化」を検索してしまう．過剰検索結果の例を**表1**に示す．

2.6 過剰検索結果削除とチェック用フラグ

過剰検索結果は検索文字列の偶然の一致によるものである．したがって，各エントリーから対応する表記法（図4の右側から左側への写像）が構成できれば完全に除去することができる．代表読み方式では，代表読みと自然読みとが一致するときに漢字表記とかな表記とを1つのエントリーで処理している．一致しない場合には漢字表記，かな表記の一方のみに対応してエントリーが存在する．したがって，単語内の各漢字が検索文字列の中でどのように読まれているかがわかれば，エントリーに対応する（入力として許容される）表記法がわかる．このために必要なフラグを各エントリーに付加する．

（1） 検索文字列＝代表読み≠自然読み の場合

例：「へん換」と入力し，「へんか」を検索文字列として，「変換」を検索したときの「換」の部分

検索文字列＝代表読み＝「か」

自然読み＝「かん」

この場合，漢字入力のみが許容される（かな「へんか」を「変換」に変換することはない）．

（2） 検索文字列＝自然読み≠代表読み の場合

例：「へんかん」と入力し，「へんかん」を検索文字列として，「変換」を検索したときの「換」の部分

検索文字列＝自然読み＝「かん」

代表読み＝「か」

この場合，かな入力のみが許容される（「へん換」（漢字入力）は上記（1）のエントリーにより処理されている）．

（3） 検索文字列＝代表読み＝自然読み の場合

例：「へんか」「へん化」と入力し，「へんか」を検索文字列として，「変化」を検索したときの「化」の部分

検索文字列＝代表読み＝自然読み＝「か」

この場合，漢字入力，かな入力の両方が許容される．

上記（1）～（3）に対応して以下のようにフラグを検索文字列および単語中の文字列に付加する．付録に示すように，これらのフラグを用いて過剰検索結果を削除することができる．

（1） 単語内の漢字に1ビットフラグ（漢字入力必須フラグ）を付加する．

（2） 単語内の漢字に1ビットフラグ（かな入力必須フラグ）を付加する．さらに，その漢字の自然読み（検索文字列に含まれている）を求めるためにエントリーにフィールドを付加する．このフィールドを「必須入力かなフラグフィールド」と呼ぶ．このフィールドは検索文字列が8文字までなら1バイト，9文字以上16文字までなら2バイトとし，第Nビットは検索文字列のN文字目がかな入力を必須とするときは ON，そう

表1 過剰検索結果
Table 1 Excess candidates.

入力表記	検索文字	単 語	現 象
へ ん 換	へ ん か	変 化	漢字不整合
へ ん か	へ ん か	変 換	読み不整合
書き換え	かきかえ	書換え	かな喪失

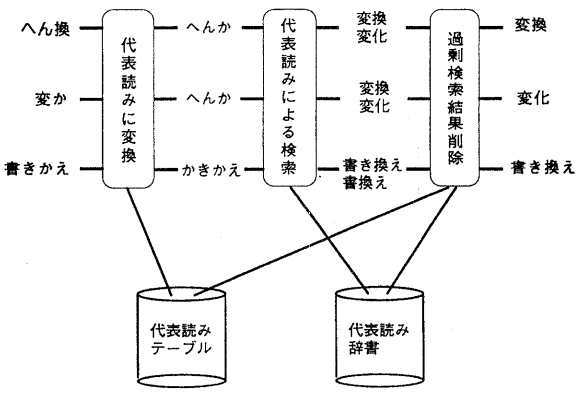


図5 代表読み辞書参照
Fig. 5 Consultation of representative reading dictionary.

でないときは OFF とする.

(3) フラグを付加しない.

ここで付加するフラグは,単語内の漢字(14ビットで表現可能)について2ビットであり*,フラグを付加しても検索文字列は1バイト,単語内の漢字は2バイトで表現される.したがって,フラグによる辞書の増大は必須入力かなフィールドのみである.チェック用フラグの例を**表2**に示す.

過剰検索結果削除のフローは**図6**に示すとおりである.この詳細な説明を付録に示す.

表2 過剰検索結果削除のフラグ
Table 2 Flags for excess candidate reduction.

	単語内の漢字	必須入力かなフラグフィールド	対応する表記
a	△△ 百発百中	▲▲▲▲▲ ひゃっぱつひゃくちゅう	ひゃっぱつ百中 ひゃっぱつ百ちゅう ひゃっぱつひゃく中 ひゃっぱつひゃくちゅう
b	△○ 百発百中	▲▲▲ ひゃっはつひゃくちゅう	ひゃっ発百中 ひゃっ発百ちゅう ひゃっ発ひゃく中 ひゃっ発ひゃくちゅう
c	○△ 百発百中	▲▲ ひゃくぱつひゃくちゅう	百ぱつ百中 百ぱつ百ちゅう 百ぱつひゃく中 百ぱつひゃくちゅう
d	○○ 百発百中	ひゃくはつひゃくちゅう	百発百中 百発百ちゅう 百発ひゃく中 百発ひゃくちゅう

ただし,

代表読みは		フラグは	
百	ひゃく	○	必須漢字入力
発	はつ	△	必須かな入力
中	ちゅう	▲	必須入力かな

3. 交ぜ書き漢字変換の実現と評価

3.1 交ぜ書き漢字変換の実現

われわれは,かな漢字変換プログラムをもとに,交ぜ書き変換を実現した.以下の2点のみがかな漢字変換からの変更点であり,ほかは変更していない.

(1) 辞書を代表読みを用いて作成した.

(2) 辞書参照に当たって,漢字を代表読みに変換するプロセスおよび 2.6 節に述べた過剰検索結果削除プロセスを加えた.

3.2 辞書の大きさの評価

代表読み辞書による交ぜ書き漢字変換を連文節かな漢字変換プログラムをもとに作成した.辞書は約7万7千語からなり,約 39 万4千とおり(語数の約5.1倍)の表記法がある.代表読みとしては「各漢字が辞書中でもっとも多く読まれる読み方」を用いた.この結果を**表3**に示す.登録エントリーの数は約 12 万2千(語数の約1.6倍)となった.辞書の大きさは,もとのかな漢字変換の辞書では約 0.9M バイト,代表読み辞書では約 1.5M バイト(代表読みテーブルを含む)で,もとの辞書の大きさの約1.8倍である.これは当初の目標であった2倍より小さい.

表3 辞書の大きさ
Table 3 Size of dictionaries.

	単語	表記法	エントリー	辞書サイズ(KB)
かな漢字変換(K)	76,538	76,538	76,538	864
交ぜ書き漢字変換(M)	76,538	394,102	121,994	1,522
M/K	1.0	5.1	1.6	1.8

3.3 処理速度の評価

代表読み辞書を用いた交ぜ書き漢字変換は,かな漢字変換と比べ3.1節に述べた分だけ処理量が増えている.また,辞書が大きくなることで,ハードウェア的に所要時間が増える可能性がある.そこで,速度の低下が小さいことを確認するため,所要時間を測定した.

測定には新聞記事などから取った3,122文(平均23字/文)を用いた.文中の漢字がそのまま書かれる割合(「漢字書き率」)を p とする.$p=1$(原文通り),$p=0$(かな書き)のテキストに加え,

$p=0.25$ (3/4 の確率でかな書き),

$p=0.5$ (1/2 の確率でかな書き),

$p=0.75$ (1/4 の確率でかな書き)

のテキストを用意した.$p=0.25$, 0.5, 0.75 については,確率的変動を考慮してそれぞれ5組のテキストを用意した.各テキストを文単位で連文節一括変換し,1文当たりの所要時間 t を測定した.使用環境は,PS/55* モデル 5570-V, DOS-J 5.0/V である.この結果を**図7**に示す.

* たとえばシフト JIS コードでは漢字を16進4桁表記と見て最上位桁が 8, 9, E, F のものが使われている.すなわち,最上位桁に2ビット分の余裕がある.

* PS/55 は IBM Corp.(米国)社の登録商標.

入力がすべてかなの場合（$p=0$）交ぜ書き漢字変換はかな漢字変換と等価になるが，交ぜ書き漢字変換の所要時間（1文当たり約0.96秒）はかな漢字変換（1文当たり約0.83秒）の約1.15倍であった．これは代表読みを用いることにより約15%処理時間が増加することを示している．また，漢字書き率pが高いほど所要時間が減少しているが，これは，同音異義語などの選択候補が減ることによる処理時間の減少が代表読みを用いることによる処理時間の増加を上回っているためと考えられる．

3.4 変換率の評価

交ぜ書き入力においては，入力に漢字を用いることにより同音異義語などの選択候補が減少し変換率が高くなると考えられる．そこで，処理時間測定に用いたテキストについて変換率を測定した．変換率は句読点単位の正解率として測定した[4]．

結果を**図8**に示す．変換率はpに対してほぼ直線的に上昇していることがわかる．

代表読み辞書を用いた交ぜ書き漢字変換では，かな漢字変換と異なる部分は前述のように自立語辞書参照部分のみである．第2章で述べたことから，$p=0$に対しては変換結果はかな漢字変換と同じであるから，漢字表記によって変換率が上昇することがわかる．

$p=1$のときの変換率は，95.8%であった．すなわち，全く変換が不要な入力に対して4%程度の誤変換（かなのままが正解のものを漢字に変換してしまう）があった．これらの大部分は，かな漢字変換においても漢字に変換されてしまうものであった．しかし下のタイプaのように交ぜ書き漢字変換に特有な誤変換が発見された．また，交ぜ書き漢字変換とかな漢字変換に共通の誤りの中に，タイプbのように，交ぜ書き入力であることをうまく利用すれば減少させられるものが発見された．

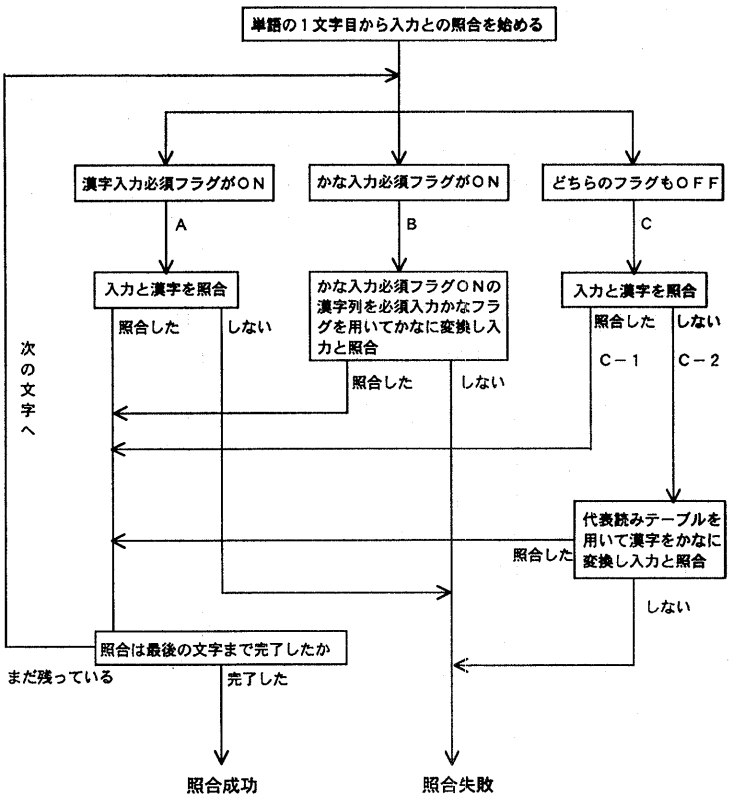


図6 過剰検索結果削除のフロー
Fig. 6 Flow of excess candidate reduction.

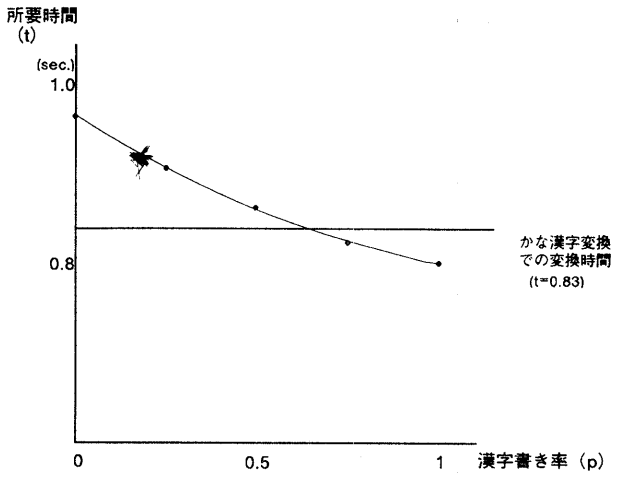


図7 変換時間測定結果
Fig. 7 Conversion time for several representations of texts.

タイプa．漢字書きとかな書きを合わせて，誤った語に変換してしまったケース

例：苦しんでいる点を考慮
→苦心でいる点を考慮
私だけが家で毎日
→私だけ画家で毎日

タイプb．わざわざかな書きされた語を過変換したケース

例：灯油をまいた跡があった
→灯油を巻いた跡があった
資本主義の巨魁（きょかい）
→資本主義の巨魁（巨魁）

タイプaの誤変換が起こる原因は，かな入力のときには音が異なるので対立しなかった単語同士が，漢字書きによって読みの情報が失われ，対立変換候補になるためである．タイプbの誤変換はユーザーが（漢字が入力できるにもかかわらず）かな書きを選択している，という意図が候補選択に反映されないことから起こる．

これらの現象は，交ぜ書き漢字変換においてはかな漢字変換とは異なる候補選択基準が有効であることを示している．

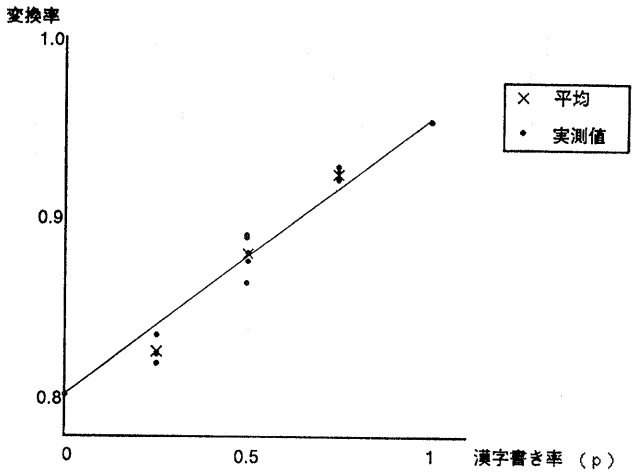


図 8 変換率測定結果
Fig. 8 Conversion results.

表 4 誤変換の分類
Table 4 Conversion errors.

p	0	0.25	0.5	0.75	1
タイプaの誤変換 (%)	0	0.34	0.53	0.51	0.49
タイプbの誤変換 (%)	2.1	2.1	2.1	2.1	2.1
その他の誤変換 (%)	17.7	14.9	9.27	4.7	1.6
合　計 (%)	19.8	17.3	11.9	7.3	4.2

4. おわりに

部分的に漢字を含む入力を許す交ぜ書き漢字変換を実現するために，代表読みによる辞書の構成法を示した．この方式では，すべての表記法に対応することができる．辞書の膨張は，かな漢字変換の辞書に対して約1.8倍（単純にすべての漢字とかなの組み合わせを持った場合の 1/3）に抑えることができた．

この辞書に基づいた交ぜ書き漢字変換は，かな漢字変換の約 1.15 倍の時間でかな入力を変換することができ，実用上問題のない速度となった．また，入力に漢字が混在する場合には処理時間が減少した．われわれは，この交ぜ書き漢字変換を実際に手書き入力システムに組み込んだ[5]．

交ぜ書き漢字変換の変換率は漢字書き率に対してほぼ直線的に向上した．また，(1)出力どおりの表記を入力としてもひらがな部分の過変換などによりもとどおりの変換結果にならないことがあること，(2)漢字を入力に交ぜることによってひらがなのみの入力からは起こりえない誤変換が起こること，がわかった．これは，交ぜ書き漢字変換では，入力表記を考慮した新しい候補選択法が有効になることを示している．

参考文献

1) 福永，井上，鈴木：ペン入力インターフェースとしての漢字混じり仮名漢字変換評価，第47回情報処理学会全国大会論文集5，pp. 189-190 (1993).
2) 小野：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，情報処理学会論文誌，Vol. 31, No. 3, pp. 405-413 (1990).
3) 塩見，喜多，河合，大岩：2ストローク入力のための仮名漢字変換，情報処理学会論文誌，Vol. 33, No. 7, pp. 920-928 (1992).
4) 建石，金子，鳥原：かん字漢字変換の変換率について，第46回情報処理学会全国大会論文集 3, pp. 277-278 (1993).
5) Toyokawa, K., Kitamura, K., Katoh, S., Kaneko, H., Itoh, N. and Fujita, M.: An On-Line Character Recognition System for Effective Japanese Input, *Proceedings of the Second International Conference on Document Analysis and Recognition*, pp. 208-213 (1993).

付録　過剰検索結果削除についての詳細な説明

過剰検索結果削除は，エントリーの持つチェック用

フラグを用いて，各エントリーの入力として許容される表記法を作成することにより実現する．入力どおりの表記法が生成されない場合には，異なる入力の検索文字列が偶然一致したものと判定され，過剰検索であることがわかる．

入力として漢字，かな，その両方のいずれが許容されるかは，漢字必須フラグ，かな必須フラグによりわかる．

(A) 検索文字列＝代表読み≠自然読み
　漢字必須フラグが ON（漢字表記のみ許容）
(B) 検索文字列＝自然読み≠代表読み
　かな必須フラグが ON（かな表記のみ許容）
(C) 検索文字列＝代表読み＝自然読み
　どちら のフラグも OFF（両標記を許容）

(A) 漢字表記のみが許容される場合には，単語中の文字どおりの表記が許容される表記法である．

(B) かな表記のみが許容される場合には，エントリーの漢字に対応する「かな」を求めることにより，許容される表記法がわかる．これは，必須入力かなフラグがONの部分として得られる．ここで，単語中にかな必須フラグがONの漢字が複数含まれる場合に各漢字のかな表記が得られるかどうかが問題になる．

(B-1) かな必須フラグ ON の漢字が続いている場合，例えば表1の(a)のエントリーの「百発」の部分には漢字ごとのかな表記は得られないが，このような連鎖全体に対応するかな表記が得られるので，問題はない．上例では，全体を「ひゃっぱつ」と読むことがわかれば許容される表記法は生成できる．

(B-2) かな必須フラグONの漢字が続いていない場合，検索文字列の必須入力かなフラグがONの部分は1漢字分の読みである．(かな必須フラグ OFF の漢字に対応する読みによって区切られている)．「かな必須フラグONの漢字の連鎖の数」と「必須入力かなフラグONの読みの連鎖の数」は同じであるから，左から順に対応させればそれぞれの連鎖に対応する表記を得ることができる．

(C) 漢字表記，かな表記（自然読み）の両方が許容される．この場合，自然読み＝代表読みなので代表読みテーブルを用いてかな表記を得ることができる．

以上により(A)～(C)のいずれの場合にも表記法が生成できるのでエントリーの情報から対応する表記法が生成できることがわかる．

フラグを用いて生成される表記と入力との照合は，図6のように単語内の漢字について左から逐次照合することにより，すべての組み合わせを生成せずに1漢字ずつ処理することができる．図6中のパスA～Cは上のA～Cに対応している．

以下，表2のCにある「百ぱつ百ちゅう」を入力した場合について説明する．入力を代表読みに変換すると「ひゃくぱつひゃくちゅう」となり，「百発百中」が検索される．1文字目「百」は漢字入力必須フラグがONなので，図6のAのパスを通り，漢字「百」が入力と照合される．2文字目「発」はかな字入力必須フラグが ON なので，図6のBのパスを通り，「発」に対応するかな「ぱつ」が入力と照合される．3文字目「百」はどちらのフラグも OFF なので，図6のCのパスを通り，漢字「百」が入力と照合され，照合するのでC-1のパスを通り，次の文字の照合に進む．4文字目「中」はどちらのフラグも OFF なので，図6のCのパスを通り，漢字「中」が入力と照合されるが，入力はかなであるので照合せず，C-2のパスを通り，「中」の代表読み「ちゅう」が入力と照合される．以上の結果，照合に成功する．

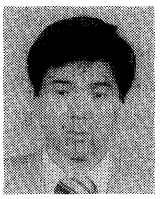
(平成5年2月 5 日受付)
(平成6年2月17日採録)

金子　宏（正会員）

1957年生．1980年東京大学工学部計数工学科卒業．1982年工学修士．1984年日本アイ・ビー・エム(株)入社．現在同社東京基礎研究所文書・音声システム担当．日本音響学会，日本オペレーションズ・リサーチ学会，日本応用数理学会，計量国語学会各会員．

建石　由佳（正会員）

1961年生．1989年3月東京大学理学系研究科情報科学専門課程単位取得退学．同年4月日本アイ・ビー・エム(株)入社．同社東京基礎研究所において，日本語校正システム，形態素解析などの研究に従事．理学博士．計量国語学会会員．



鳥原　信一（正会員）

1956 年生．1980 年獨協大学外国語学部英語学科卒業．1982年日本アイ・ビー・エム(株)入社．現在同社東京基礎研究所文書・音声システム所属．