

## 部分字形組合せを用いた手書き風フォントの自動生成

安本 護<sup>\*\*\*</sup> 池田 尚志<sup>\*\*</sup> 豊倉 完治<sup>\*</sup>

<sup>\*</sup>三洋電機株式会社ハイパーメディア研究所  
<sup>\*\*</sup>岐阜大学工学部

〒503-01 岐阜県安八郡安八町大森180  
E-mail:yasu@gf.hm.rd.sanyo.co.jp

あらし マルチメディア環境の進展に伴い、従来のような画一的なフォントだけではなく、多様なフォントデザインが求められるようになってきている。その中に、ユーザの個性を表現できる手書き風文字を望むものがある。しかしながら、日本語文書の表現には非常に多くの漢字が必要であり、これをユーザごとにデザインすることは事実上不可能である。そこで、我々は漢字の構造的特徴に着目し、漢字をへんやつくりなどの部分字形に分解し、その組合せによって手書き風フォントを自動生成することを考えた。本稿では、常用漢字1945文字を対象に入力された手書き文字データから基本部分字形及び漢字生成規則を抽出して個性的なフォントの生成を試みた。

和文キーワード フォント, 漢字, 手書き, 部分字形, パターン生成, 個性

## A Generation of Handwritten Style Font by Combining Parts of Kanji

Mamoru Yasumoto<sup>\*\*\*</sup> Takashi Ikeda<sup>\*\*</sup> Kanji Toyokura<sup>\*</sup>

<sup>\*</sup>Hypermedia Research Center, Sanyo Electric Co.,Ltd.  
<sup>\*\*</sup>Faculty of Engineering, Gifu University

180 Omori, Anpachi-cho, Anpachi-gun, gifu, 503-01, Japan

### Abstract

The progress of the multimedia environment seems to give a boost to the need for various font designs. The individual handwritten style font is one of them. But because thousands of kanji characters are necessary for Japanese, it is practically impossible to design them all by oneself. We propose in this paper to generate kanji script by combining its parts. First we extract the combining rules of the parts of kanji and the basic kanji parts from handwritten kanji data inputed by the user. And then a trial of generating 1945 characters of individual handwritten style kanji font is shown here.

英文 key words Font, Kanji, Handwrite, Radical, Pattern Generation, Individuality

## 1 はじめに

日本語ワードプロセッサの開発に代表される文書処理の機械化によって、手書きによる文書作成は、フォントを用いた印刷へと変化した。手書き文字には、人それぞれの個性が現れ、それが味わいとなっている。しかしながら、手書き文字には、書き手による上手下手があり、極端なクセ字では判読することが不可能な場合さえある。また、手書きでは、十分な速度で文書を作成することが出来ず、生産性が低い。一方、明朝体やゴシック体などのフォントは、1文字1文字が美しく、読みやすいようにデザインされており、生産性も手書きとは較べものにならないほど高くなっている。たしかに、フォントを用いた文書は美しく読みやすいが、手書き文字とは全く異質なものである。また、既存のフォントの中にも筆文字やペン字など手書き風のものも存在するが画一的であることに変わりはなく、書き手の個性を伝えることは難しい。多様な表現力要求するマルチメディア時代の到来によって、フォントにもより多彩なデザインが求められるようになってきている。この中に書き手の個性を表現できる手書き風文字の要望がある。しかしながら、日本語文書の表現には数千もの漢字が必要であり、これをユーザごとにデザインすることは事実上不可能である。

日本語フォント作成における作業量を軽減するために漢字を部分字形などの構成要素から生成する試みとして、明朝体 [1] やゴシック体 [2] に関するものが報告されている。特に [1] では構成要素を組合せる際に大きさや位置の自動調整が行われており、非常に興味深い。しかしながら、明朝体やゴシック体などの印刷系書体のデザインにユーザの個性を反映させることは考えられない。手書き系文字の生成では、毛筆楷書体を扱った [3] があり、良好な品質の毛筆文字が生成されているが、この方式はパラメータを指定して文字を生成するものであるためエンドユーザが扱うのは難しいと思われる。このほかにも内臓フォントパターンに幾何学的ひずみを印加して手書き風文字を得る方式 [4] があるが、筆記者の手書き文字に似た文字パターンを生成するための係数を得る方法は課題とされている。

そこで、我々は漢字の構造的特徴に着目し、エンドユーザの筆記した漢字からヘンヤツクリなど

の部分字形を抽出し、それらの組合せによって、ボールペンで書いたような手書き風フォントを自動生成することを提案した。[5] さらに、本稿では入力された2組の手書き文字データから基本部分字形及び漢字生成規則を抽出して常用漢字1945文字の個性的なフォント生成を試みている。

## 2 生成対象文字

生成対象をボールペン風文字としたことには、2つの理由がある。第1の理由はボールペンが日常多くの場面で使われる筆記具だということである。2番目の理由は、文字データの入力および再現性に関するものである。手書き文字の形状データを部分字形ごとに入力するには、スキャナのようなオフラインの入力装置は向かないため、ディジタイザを使用することとした。文字の形状は骨格及び筆触形状(筆記具と記録素材の接触によって形成されるもの)によって作られる。しかしながら、文字データの入力にディジタイザを用いる場合、筆触形状を入力することは不可能である。そこで、筆触の太さが一定でカスレやトギレのないボールペン文字を生成対象とすれば、筆触を円で近似することができるため骨格のみの入力でも文字形状を再現可能であると考えた。

また、筆記具の特性は文字の骨格にも影響を及ぼすことが知られている。例えば、古代中国の甲骨文字が直線的なのは亀の甲らや骨のような硬い素材にキリのようなもので文字を刻みつけたからであり、行書体や草書体の流れるような形状は、柔らかい筆を用いることによって始めて生まれたものである。ペン入力タイプのディジタイザは、ボールペンによる筆記とほとんど同じ筆記特性(書き味)を持つと考えられるので文字骨格形状の再現性からもボールペン文字を生成対象に選ぶことが妥当である。

## 3 システムの構成

図1に本手法による手書き風文字生成システムのプロック図を示す。各ブロックは、次のようなはたらきを持つ。

- 手書き文字データ入力部  
手書き文字データをディジタイザから入力

する。

- 生成規則抽出部  
手書き文字データから生成規則を抽出する。
- 部分字形抽出部  
手書き文字データから基本部分字形を抽出する。
- 文字コード入力部  
生成する文字のコードを受け付ける。
- 文字生成部  
入力された文字コードに対応する生成規則を取得し、それに基づいて基本部分字形を組合せて手書き風文字を生成する。
- 生成規則格納部  
文字生成に用いる部分字形のコード及び外接四角形情報を格納する。
- 基本部分字形格納部  
文字生成に使用する部分字形の形状データを格納する。

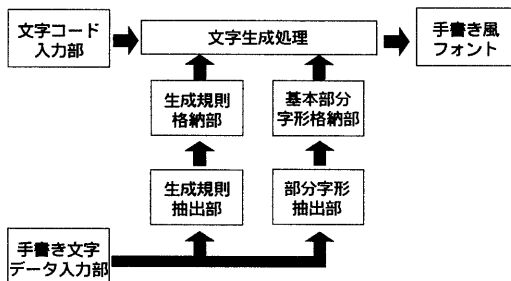


図 1: システムの構成

## 4 字形データベース

### 4.1 漢字の構造

漢字の成立には次に示す 4 通りがあるといわれている。

- 象形 物体の形状を描いたもの  
例:日、月、山、川

- 指事 線や点で事物の性質等を表したもの  
例:一、二、上、下
- 会意 2 個の文字を結び合わせて新しい意味をつくったもの  
例:日+月=明
- 形声 形状を示す文字と音声を示す文字を合わせてつくったもの  
例:水+可=河

多くの漢字はこの中の会意または形声によるものであるため、ヘンやツクリなどの構成要素を一定の規則に従って組合せたものとして表すことができる。また、構成要素をさらに小さく分解すれば、会意や形声以外の漢字にも構成要素の組合せで表現できるものがある。本稿では漢字の構成要素を部分字形、部分字形の組合せ方を構成パターンと呼ぶことにする。

### 4.2 部分字形

漢字の構成要素である部分字形には、漢和辞典で用いられる部首の他に対象となる漢字集合の中で 2 回以上出現した構成要素を加えている。例えば、「葉」のクサカムリから下の構成要素は漢和辞典に定められた部首ではないが、JIS X0208 に収録された葉、喋、牒、蝶、課、櫟、滌、縹、鏢という漢字に合計 9 回出現する。そこで、この「葉」のクサカムリから下の部分を新たに部分字形とする。

### 4.3 構成パターン

構成パターンは、文字枠を水平または垂直の線で分割し、部分字形の入る領域を示したものである。基本となる構成パターンには、次の 4 種類を定義した。

- Unique 分割不可タイプ
- Parallel 左右分割タイプ
- Cascade 上下分割タイプ
- Fense カマエやニョウに分かれる周辺分割タイプ  
(分割の形によりさらに 9 通りに分かれる)

構成パターンの表現は図によるものが直観的で分かりやすいが、図的表現では計算機による取り扱いが難しいため、分割タイプを表す記号  $u, p, c, f_1 \sim f_9$  と部分字形を表す記号  $x$  を用いる記号表現を考案した。これによって、計算機上で容易に構成パターンを扱えるようになった。

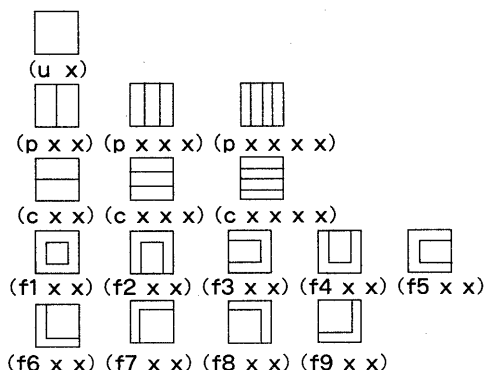


図 2: 構成パターン

図 2 に基本的な構成パターンの図的表現と記号表現を示す。記号表現において部分字形を示す記号  $x$  の出現順が左から右へ並ぶ順と図的表現における部分字形領域の並びは一意に対応づけられている。すなわち、記号  $x$  が左から右へ並ぶ順は左右分割タイプ  $p$  では左から右の部分字形領域、上下分割タイプ  $c$  では上から下の部分字形領域、周辺分割タイプ  $f$  では周辺から内部の部分字形領域へそれぞれ対応する。また、基本構成パターンでは表せない複雑な構造を持つ漢字は、図 3 に示すように複数個の分割タイプを組合せた複合構成パターンによって表現できる。複合構成パターンの記号表現は基本構成パターンの入れ子表現となる。

#### 4.4 字形データベースの構築

字形データベースは、字形テーブル、部分字形テーブルならびに構成パターンテーブルの 3 つから構成されている。字形テーブルとは、漢字をその構成パターンと 1 段下位の部分字形で記述したものである。同様に、部分字形テーブルは、部分字形をその構成パターンとさらに 1 段下位の

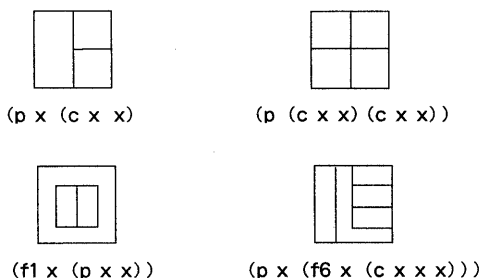


図 3: 複合構成パターン

部分字形で記述したものである。字形テーブル及び部分字形テーブルのレコードを表 1、表 2 に示す。ここで、字形テーブル及び部分字形テーブルに記述される下位部分字形の出現順と構成パターンの記号表現において部分字形を示す記号  $x$  とは 1 対 1 に対応している。また、構成パターンテーブルは、表 3 に示すように構成パターンのコードと記号表現を記述したものである。

表 1: 字形テーブルのレコード

フィールド	内容
1	JIS 区点番号
2	部首番号
3	画数
4	構成パターンコード
5~10	下位部分字形のコード

表 2: 部分字形テーブルのレコード

フィールド	内容
1	部分字形コード
2	画数
3	構成パターンコード
4~8	下位部分字形のコード

表 3: 構成パターンテーブルのレコード

フィールド	内容
1	構成パターンコード
2	記号表現

「葉」を例に字形データベースへの登録手順を説明する。「葉」はクサカムリ(部分字形コード 243)とその下の部分字形(部分字形コード 248)

に上下2分割(構成パターンコード2)され、字形テーブルへ登録される。また、部分字形248は「世」(部分字形コード943)と「木」(部分字形コード75)に上下2分割(構成パターンコード2)され、部分字形テーブルに登録される。「葉」は最終的には3つの部分字形243、943、75に分解されるが、1度に3つの部分字形には分解せず、クサカムリと他の漢字にも使用される部分字形248への分解にとどめる。このような階層的な登録方法を用いた結果、字形データベース作成にかかる作業量を削減するとともに漢字によって部分字形への分解状態が異なるという分析結果の揺らぎをなくすことができた。我々は、JIS X0208に収録されている6355文字の漢字を分析し、字形データベースを構築した。この結果、部分字形は1012個、構成パターンは105個となった。

この字形データベースは本稿で扱う文字生成だけでなく、通常の漢和辞典では不可能な部分字形からの漢字検索などに応用することも可能と考えられる。

#### 4.5 出現位置属性の導入

字形データベースでは、同一概念を持つ部分字形には原則として1つの部分字形コードを与えるようにしている。すなわち、図4に示す「禾」のようにヘンとなる時とカムリとなる時で形状の変化するものであっても部分字形コードは1つである。部分字形から文字を検索する場合などは、これで問題ないが、文字生成では出現位置による部分字形の形状変化を正しく表現する必要がある。そこで、図5に示すように部分字形が大きく形状変化する位置を経験的に分類し、部分字形がヘンの位置に来る場合のP属性、カムリの位置に来る場合のC属性、カマエなどの位置に来る場合のF属性、その他位置に来る場合のE属性という4つ属性を付加した。ここで、例外的なのは構成パターンF8の内部に来る場合をE属性とせず、P属性としたことである。このような文字には、「武」や「式」がある。

### 5 手書き文字データ収集

文字生成規則並びに基本部分字形を抽出するには、手書き文字データの収集が必要である。そ

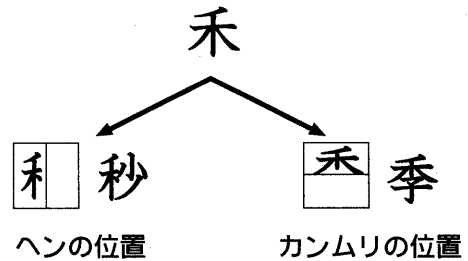


図4: 出現位置による変形

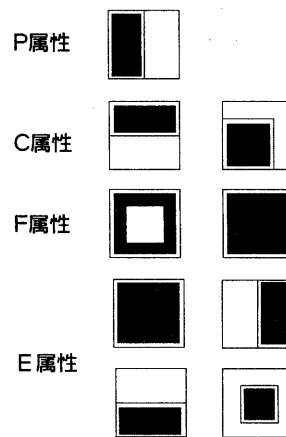


図5: 出現位置属性

こで、字形データベースに登録された漢字6355文字の約31%に相当する常用漢字1945文字の手書き文字データを収集し、後の実験に使用した。入力システムの構成は、ペン入力タイプのデジタルタイザをUNIXワークステーション<S4-330>に接続したものである。従って、手書き文字データは時系列の座標データとして入力される。文字は25mm角の矩形枠に記入するものとし、デジタルタイザのサンプリング速度は20点/秒とした。座標データの解像度は記入枠の一辺を512の点に分解しているので20.48点/mmである。

手書き文字座標データと字形データベースの対応づけを手書き文字入力後に行うことは非常に困難である。そこで、このシステムでは入力文字の

構成パターンを階層的に図示し、システムが入力要求している部分字形領域を反転表示する方法を使っている。図6は、X-Window上に作成したデータ収集プログラムの画面である。

入力された手書き文字には大きさのばらつきを始めとする様々な変動が含まれるが、これらも手書き文字の個性を表すものであると考え、文字中心を文字枠中心に揃える位置の正規化のみを行った。

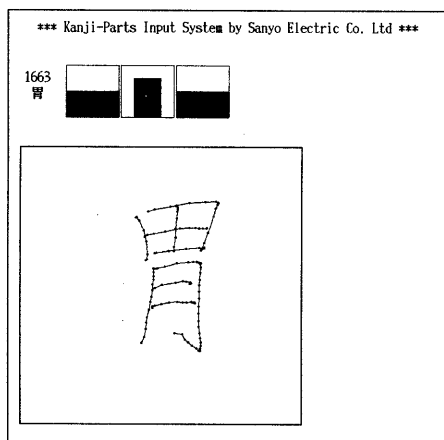


図 6: 手書き文字データ収集システム

ここで、基本部分字形を抽出するための入力文字は、できるだけ少ないほうが文字入力の負担を軽くできて好ましい。そこで、最も部分字形を多く含む漢字から候補に入れる以下のアルゴリズムによって基本部分字形抽出に用いる漢字を絞り込んだ。

#### <入力文字絞り込みアルゴリズム>

- (i) 文字ごとに含まれる相異なる部分字形を数えあげ、その数を  $N$  とする
- (ii)  $N$  が最大の文字  $C$  をサーチする
- (iii) *if*  $N \neq 0$   
*then* 文字  $C$  を基本部分字形抽出の対象に加える  
*else* 終了
- (iv)  $N > 0$  である全ての文字について、文字  $C$  に含まれる部分字形を取り除く
- (v) *goto* (i)

このアルゴリズムにしたがって、基本部分字形抽出に必要な漢字を求めた結果を表4に示す。

表 4: 部分字形抽出に必要な文字数

生成対象文字	出現属性なし	出現属性あり
常用漢字	257(13.2)	388(19.9)
JIS 第1	282(9.5)	454(15.3)
JIS 第1第2	345(5.4)	650(10.2)

カッコ内は、生成対象文字集合全体に対する割合 (%)

## 6 手書き風フォントの生成

### 6.1 基本部分字形の抽出

#### 6.1.1 手書き入力文字の絞り込み

基本部分字形データを作成するには、部分字形を単独に入力するという方法も考えられるが、漢字として入力された中から部分字形を抽出し、基本部分字形を作成するほうがより自然な形状のデータを得られると考えた。

#### 6.1.2 基本部分字形決定方法

以上のアルゴリズムによって決定された文字には文字生成に必要な部分字形が、部分字形の各カテゴリにつき少なくとも1つ以上含まれる。そこで、同一カテゴリの部分字形が2つ以上存在する場合には、それら全ての外接四角形の面積を求め、面積最大すなわち最も大きい部分字形を基本部分字形とする。これは、文字生成する際に基本部分字形を配置情報にしたがって変形するので基本部分字形をできるだけ拡大しない方が、線のブレなど手書きデータに含まれるノイズの目立たない品質の良い生成結果が得られるからである。

## 6.2 文字生成規則

文字生成規則は、生成に用いる部分字形の選択方法と選択された部分字形の配置方法を記述したものである。

### 6.2.1 部分字形の選択

字形データベースでは漢字を階層的に部分字形に分解し、登録している。従って、漢字をツリーに展開すると、ツリーのルートに漢字があり、リーフにはそれ以上分解不可能な部分字形が来る。さらに、ルートからリーフへ至る経路の中間ノードには、リーフの部分字形から構成される部分字形あるいは漢字が存在する。リーフの部分字形は、入力された手書き文字中にならざる含まれている。

今回用いた生成規則では、ルートからリーフに向かってたどり、中間ノードの部分字形が手書き入力された文字の中に含まれていれば、それより下のレベルへは降りずに、その部分字形を使用する。これは、入力文字中に既に適切な組合せの部分字形があれば、それを用いて不要な分解を避けるほうが、良い生成品質を得られると考えたからである。図7は、「花」のツリーをたどったところ、クサクナムリと「化」から生成できたことを示す例である。

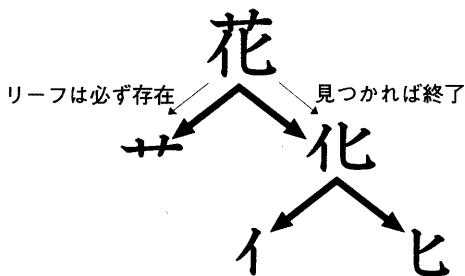


図7: 部分字形選択の例「花」

### 6.2.2 部分字形の配置

部分字形の配置には部分字形に外接する四角形の左上頂点を基準とし、幅と高さが一致するように変換する方法を用いる。図8に示すように基本

部分形を構成する座標を  $P_b(x_b, y_b)$ 、基本部分字形に外接する四角形の左上頂点座標を  $(X_b, Y_b)$ 、幅を  $W_b$ 、高さを  $H_b$  とし、生成文字の部分字形に外接する四角形の左上頂点座標を  $(X_g, Y_g)$ 、幅を  $W_g$ 、高さを  $H_g$  とすれば、生成される部分字形の座標  $P_g(x, y_g)$  は、次の式で表される。

$$x_g = (x_b - X_b) \cdot \frac{W_g}{W_b} + X_g \quad (1)$$

$$y_g = (y_b - Y_b) \cdot \frac{H_g}{H_b} + Y_g \quad (2)$$

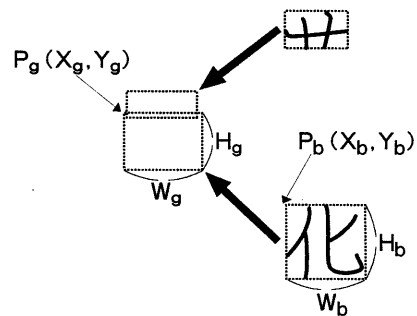


図8: 部分字形の配置

## 6.3 かなの扱い

本稿で生成の対象としているのは漢字のみであるが、日本語文書にかな文字は欠かせないものである。しかも、かなには漢字以上に書き手の個性が現れると考えられる。かな文字は、漢字よりもはるかに数が少ないので、全文字を登録するにしてもユーザへの負担はそれほど大きくない。しかし、少しでも負担を軽減するために濁音及び半濁音の付く文字については1回の入力ですむようにした。すなわち、か行、さ行、た行の文字は濁音付き文字のみを入力し、清音の文字は濁点を取り除いて生成した。また、は行は半濁音付き文字のみを入力し、清音は同様に半濁点の除去で生成し、濁音については、半濁点の位置に他の文字の濁点を配置することで生成した。この結果、ひらかなとカタカナで合計50文字の入力を省略することが可能となった。

## 6.4 生成実験と考察

### 6.4.1 手書き文字データの収集

生成実験には、次の2組の手書き文字データを用いた。

- セット A: ペン習字の手本から入力
- セット B: 研究室の女性が筆記

この2組の手書き文字データから生成規則と基本部分字形を抽出して手書き風フォントの生成実験を行った。

### 6.4.2 出現位置属性の効果

手書き入力されたオリジナルデータ並びに出現位置属性の有無による手書き風フォントの生成結果の違いを図9に示す。生成規則及び基本部分字形は、ともにデータセット A から抽出したものをを用いている。出現属性を用いない場合、「境」と「社」、「特」と「件」ではヘンとツクリに、「秋」と「季」ではヘンとカンムリに、また「処」と「夏」ではニョウとアシに同じ部分字形が使用されてしまい、文字の生成品質を著しく低下させていることが分かる。一方、出現属性を用いた場合には、正しい基本部分字形が選択され、より良い生成品質の得られていることが分かる。

### 6.4.3 基本部分字形の交換による個性化

データセット A と B それぞれから抽出した生成規則と基本部分字形を組合せて得られた4通りの手書き風フォント生成結果及びデータセット A と B それぞれのオリジナル手書きデータを図10に示す。この結果から同じ生成規則を用いても、基本部分字形を交換することにより全く異なる文字の生成されることが分かる。すなわち、ユーザは生成対象文字全てを入力するより、はるかに少ない文字の入力で、自分だけの個性を含む文字を生成し、使用できるようになる。

## 7 おわりに

漢字を部分字形と構成パターンで表現する字形データベースの構築を行い、これを用いて手書き文字データから抽出した文字生成規則と基本部分

電子情報通信学会  
境社特件秒季処夏

(a) オリジナル

電子情報通信学会  
境社特件秒季処夏

(b) 出現属性なし

電子情報通信学会  
境社特件秒季処夏

(c) 出現属性あり

枠付き文字はオリジナルであることを示す

図9: 出現属性の効果

字形によって手書き風文字を自動生成する手法を提案した。この結果、常用漢字1945文字の生成が、全文字数の19.9%に相当する388文字の手書き文字入力によって行えることを示した。また、異なるデータセットから取り出した生成規則と基本部分字形の組合せによっても文字生成の行えることを確かめ、ユーザ独自の個性的な文字を生成することが可能な見通しを得た。

今後の課題としては、外接四角形のサイズと位置を用いた現状の生成規則では、文字の生成品質があまり良くないのでより精密に部分字形相互の位置合わせができる手法を開発し、生成品質の向上を行うこと、基本部分字形抽出に必要となる文字数を削減するとともに手書き文字データ入力システムのユーザインタフェースを改善して文字入力の負担を軽くすることがあげられる。



美 しい 日 本 語

(a) セット A のオリジナルデータ

美 しい 日 本 語

(b) 生成規則:A、基本部分字形:A

美 しい 日 本 語

(c) 生成規則:A、基本部分字形:B

美 しい 日 本 語

(d) 生成規則:B、基本部分字形:A

美 しい 日 本 語

(e) 生成規則:B、基本部分字形:B

美 しい 日 本 語

(f) セット B のオリジナルデータ

図 10: 手書き風文字の個性化

書き風文字パターン生成の一手法, 信学論 D-II, Vol. J74-D-II, No.2, pp.209-219(1991).

- [5] 安本護, 池田尚志, 豊倉完治: 手書き風フォントの自動生成について, 情報処理学会第 49 回全国大会, 5S-1(1994).

## 参考文献

- [1] 田中哲朗, 石井裕一郎, 長橋賢児, 竹内幹雄, 岩崎英哉, 和田英一: 漢字スケルトンフォントの生成支援システム, 第 32 回プログラミング・シンポジウム報告集, pp.1-8(1991).
- [2] 上原徹三, 国西元英, 下位憲司, 鍵政秀子, 菊池純男: ストローク種別に基づく漢字形状生成方式, 情報処理学会論文誌, Vol. 31, No.2, pp.209-218(1991).
- [3] 張憲栄, 真田英彦, 手塚慶一: 漢字楷書毛筆字体の計算機による生成, 信学論 D, Vol. J67-D, No.5, pp.599-606(1984).
- [4] 塩野充: 非線形な幾何学的ひずみを用いた手