

電子計算機および電子印刷用の標準漢字字形の開発

勝村哲也 丹羽正之
京都大学人文科学研究所

電子計算機を人文科学に利用する際に発生する漢字処理に関する問題を、
抜本的に解決するための基礎研究を進め、標準漢字字形を開発中である。

当面の緊要事と考えているのは、JIS第1・2水準漢字，JIS補助漢
字、さらにISO10646などの漢字コード系では解決できない異体字の問題
や文字種の不足，漢字属性情報のあいまいさの克服であるが、国際的な標準
字形の作成を最も重視している。

Making a Standard Kanji Glyph for Computers and Printers

Tetsuya Katsumura Masayuki Niwa

Institute for Research In Humanities Kyoto University
47 Higashioguracho, Kitashirakawa Sakyo-ku, Kyoto 606 JAPAN

Katsumura and Niwa are making a Standard Kanji Glyph System in
order to solve the Kanji problems in text processing and printing
for human science.

There are some problems in JIS 1st and 2nd Set, JIS Supplementary
and ISO10646, which are known as; variants unification, no ability
of Kanji addition, ambiguity of Kanji Information Data.

Standard Kanji Glyph System will clear up these problems for
worldwide usage.

【“形としての漢字”に関する定義】

中国学の伝統に従えば、漢字は3つの要素を持つと理解できる。1は、形としての漢字、2は、音としての漢字、3は、意味としての漢字である。

ここでは主として第1の“形としての漢字”について考えることにする。

字種あるいは字体 … 意味に基づく概念的な形を指す。

字種はより抽象的な概念である。ここではなるべく使わないこととする。

書体あるいは書風 … 明朝体・仿宋体・草書体・ゴシック体など、ひとつの字体に対してさまざまな書体がある。同じ字体、同じ書体であっても、書き手の個性や印刷メーカーによってデザインが変化する。これを書風という。書風はデザインの特徴を指すことが多いので、ここでは使わないこととする。

字形 … グリフのこと。フォント・セットの中の各文字の実際のパターンを指す。本発表で論ずるのは、おもに字形についてである。

0. 問題の所在

コンピュータで扱う漢字の問題：

- (a) 漢字が違う、あるいは漢字が足りない。
- (b) 漢字の字形が一定しない。(規格の違い、機種の違いで字形が変わる)
- (c) 外字、拡張漢字の規格・作成手法が一定していない。

印刷される漢字の問題：

- (d) 漢字が足りない。→旧活字文献の印刷がきわめて困難になってきた。
(東京では大蔵省印刷局以外では満足な旧活字印刷ができない)
- (e) 書体ばかり増えて、字種は増えない。→電算写植機の問題。

- (f) ワープロ入力，パソコン入力が増えて，使用がJ I S漢字のみに絞られてきた。→入力，コードの問題。
- (g) 字形違い，外字，拡張漢字のコード・作成手法が一定していない。

1. 電子漢字(電字)とは

コンピュータによって合理的に生成された漢字を電子漢字(電字)と呼ぶ。電子漢字(電字)は、J I S 1・2水準漢字・J I S補助漢字・I S O 10646等の既存の漢字コード系にとらわれず、全ての漢字字形をカバーするのが理想である。

既存の漢字コードに依らないのは、次のような問題点があるためである。

- (a) 漢字数が少なすぎる。

J I S 1・2水準漢字 (6355字)

J I S補助漢字 (5801字)

I S O 10646 (20902字，ただしJ I S 1・2・補助を含む)

- (b) コードは、字形を定義したものではなく、字体もしくは字種，より厳密に言えば“意味としての漢字”を定義したものであるため、字形に乱れが生じる。あるいは同じコードに複数の字形が対応し得る。(例えば、天か天，吉か吉，京か京，辻か辻，祇か祇 など)

- (c) 新たに漢字を追加するための規約が定義されていない。

2. 電字明朝体の定義と生成

電子漢字(電字)を開発するための第1段階として、明朝体書体のドットフォントを試作中である。これを便宜的に“電字明朝体”と呼んでいる。明朝体を選択したことや、ドットフォントを選択したことに深い理由はなく、ゴシック体でもアウトライン・フォントでも、もちろん可能である。

電子漢字(電字)は次のような手順で定義・生成される。

- (a) サポートする漢字集合を決定する。
- (b) 各漢字を部分品に分解し、漢字を部分品の組み合わせで定義する。
- (c) 部分品の字形データを作成する。
- (d) 部分品の字形データと漢字の定義データとによって、漢字を動的に生成する。

3. 静的な字形から動的な字形へ

従来のコンピュータや印刷機の文字の中にも使用時にその都度生成されるものもあった。例えばアウトライン・フォントは計算によって字形が生成されるし、写植機はレンズによって変倍、変形が可能であった。

しかし、いずれの場合も、元になる字形(「字母」と呼ばれる)は一字一字、文字デザイナーによって作画されたものであり、字形そのものは固定的であった。いわば、静的な字形といえる。

しかし、電字明朝体は、一部の漢字を除いては漢字ごとの字母は存在しない。存在するのは部分品ごとの字母と漢字の定義式のみである。したがって、ほとんどの漢字は必要に応じてその都度、動的に生成される。

このような手法をとれば、

- (a) (部分品は共用するので)部分品ごとに字形を統一できる。
- (b) 部分品をとりかえれば容易に異体字を生成できる。
- (c) (部分品があれば)組み合わせの定義のみで新字(外字)を生成できる。
- (d) 静的な字形にくらべて、必要なデータ量が少なくてすむ。
または、同じだけのデータ量でより多くの漢字をサポートできる。

一方、克服されるべき欠点としては、

(e) 静的な字形(一字一字作画されたもの)は一字ごとのバランスが良いが、部分品の合成字形では一字ごとのバランスを調整することに限界がある。

(f) データ量が少なくて済むかわりに、動的な生成を行うための時間がかかる。高速なCPUが求められる。

4. 技術的な工夫

漢字字形の動的な生成には、さまざまな技術的工夫が必要である。漢字はその成り立ちからして、部分品(部首)の組み合わせに基づいていることは自明である。しかし単純に部分品を合成しただけでは漢字にならない。

例えば「金」という部分品を含む漢字には **錦** や **淦** や **鑿** や **銜** が存在するが、一体どのように組み合わせれば漢字になるのか、どのようなバランスが“美しい”のか、の検討が常になされねばならないし、そのために高度な技術的工夫が必要である。

もし、漢字の美しさが数量的に表現できるならば、計算機でも美しい漢字が作れるはずである。書道ならぬ電子漢字(電字)道が生まれるかもしれない。

5. その他の作業と近い将来の問題

電子漢字(電字)の研究は、

(a) 標準漢字字形の定義

が中心であるが、次のような付随するデータの研究と整備も同時に行っている。

(b) 標準画数の定義

(c) 標準漢字音の定義

(d) 漢字コードと漢字属性字書の作成

(e) 歴史的に存在するさまざまな漢字もしくは従来コード化されなかった漢字の収集と研究

また、近い将来、

(f) 事実上印刷・出版が不可能な旧漢字を大量に含む文献の電子漢字(電字)印刷による作成(これを電字印本と呼ぶ)

(g) 漢字OCRへの応用

(h) 電子漢字(電字)を日本のみならず国際的に標準化していくための、普及推進組織

などが日程にのぼってくるであろう。

【合成例】

碇	05/112 = 100140	石	08/167 + 100242	金	
綵	06/120 = 100160	糸	08/167 + 100242	金	
鉤	08/167 = 100240	金	06/137 + 100362	舟	
衿	05/145 = 100760	衤	08/167 + 100242	金	
釵	08/167 = 100240	金	01/005 + 118892	し	
釵	08/167 = 100240	金	01/005 + 100372	乙	
釵	08/167 = 100240	金	02/021 + 119112	七	
釘	08/167 = 100240	金	02/006 + 119062	了	
釘	08/167 = 100240	金	02/018 + 101212	刀	
釘	08/167 = 100240	金	02/018 + 119222	リ	
釘	08/167 = 100240	金	02/001 + 102562	丁	
釘	08/167 = 100240	金	02/025 + 119002	卜	
釘	08/167 = 100240	金	02/005 + 104152	九	

【部分品例】

03/038 100590	女子	03/038 100599	女子	03/038 100592	女子
03/039 100230	木	03/039 100039	木	03/039 100032	木
04/075 100070	火	04/075 100079	火	04/075 100072	火
04/086 100100	牛	04/086 100109	牛	04/086 100102	牛
04/093 100500	王	04/093 100509	王	04/093 100502	王
04/096 100110	羊	04/096 100119	羊	04/096 100112	羊
06/123 101030	虫	06/123 101039	虫	06/123 101032	虫
06/142 100200	金	06/142 100209	金	06/142 100202	金
08/167 100240	言	08/167 100249	言	08/167 100242	言
07/149 100400	足	07/149 100409	足	07/149 100402	足
07/157 100220	馬	07/157 100229	馬	07/157 100222	馬
10/187 100660	鳥	10/187 100669	鳥	10/187 100662	鳥
11/196 100420	豆	11/196 100429	豆	11/196 100422	豆
07/151 101100		07/151 101109		07/151 101102	