

# 前近代の漢字字形に対する 字体の包摂モデルの適用に関する諸問題

守岡 知彦<sup>1</sup>

概要：字体の包摂規準に基づく漢字字形の整理は現代一般的に使われている漢字を符号化する上では有効な手法であるといえるが、前近代のさまざまなバリエーションに富んだ漢字字形を扱うためには問題があると言える。ここでは拡張包摂規準の利用を始めとする字体の包摂モデルに基づく手法の可能性と限界について議論し、問題点について整理したい。

## 1. はじめに

JIS X 0208/0213 や UCS (ユニコード) 等の現在一般的に使われている文字符号において、ある漢字の符号位置 (あるいは、それに対応するものとしての抽象文字) はある範囲の漢字字形を包摂したものとなっている。即ち、ある1つの抽象文字というインスタンスであると同時に字形の集合を指す概念であるといえる。ある漢字の符号位置にどういう漢字字形が包摂され得るかを判断するためのものとして JIS X 0208:1997 や JIS X0213 では字体の包摂規準が定義されており、UCS 統合漢字についても IRG Working Document Series (IWDS) 1: List of UCV (Unifiable Component Variations) of Ideographs [1] という同様なリストが存在する。こうした包摂規準は例示字形のある部分を別の形に置き換えても包摂可能であることを示したものであり、これによって字体の揺れの問題に対処したものである。なお、字形デザイン差に属するような細かい字形差に関しては字体の包摂規準で規定するまでもなく包摂されるものと見なされるが、字形デザイン差に属する差異に関して陽に包摂規準が設けられる場合もある。ある差異が字形デザイン差なのか字体差なのかを判断するのは字体というものの定義が文字符号の規定の外にあることもあって潜在的に微妙な問題をはらんでいるといえるが、現代の一般的なテキストで扱われる漢字においては文字符号で提示される例示字形との差異が少ないものが大部分といえ、差異がある場合においても包摂規準は上手く機能しているといえる。

しかしながら、前近代のテキストに現れる漢字の場合、

現代の漢字程簡単ではないといえる。[2] で指摘されているように、近代の活字においても JIS X 0213 の包摂規準では判断できない例が存在すること (明治初期に発行された『明六雑誌』の場合、1.5%程度) が指摘されており、デザイン差や包摂規準の拡張を試みているが、前近代のテキストの場合、字形デザインの揺れや字体差がより大きくなるものが少なくなく、より大規模に包摂規準の拡張を行うことが必要となる。例えば、長安宮廷写経は手書きの文字としては驚く程字体の揺れが少ないテキストであるが、著者が今西本妙法蓮華經卷五と守屋本妙法蓮華經卷三に現れる合計のべ 1237 字体の代表字形に対して IWDS-1 を適用して UCS 統合漢字に包摂可能かどうかを判定した所、24% に相当するのべ 291 字体の代表字形が包摂できなかった。<sup>\*1</sup>これに対し、包摂規準を拡張することで、15% に相当するのべ 182 字体の代表字形が包摂可能となり、IWDS-1 で包摂可能なものを含めると 91% がカバーできた。[3] このように前近代のテキストに現れる漢字をなるべくそのままの字体で符号化しようとする場合、包摂規準の拡張が不可避であるといえる。

ここではこうした包摂規準の拡張を含めた字体の包摂規準に基づく漢字字形の整理の適用可能性と限界について議論する。

## 2. 字体の包摂規準の適用範囲

字体の包摂規準は楷書 (ないしは、明朝体) における字体 (標準) の存在を暗黙の前提としたものであるといえ、この書体における筆づかいをその基礎に置いたものといえ

<sup>1</sup> 京都大学人文科学研究所  
Institute for Research in Humanities, Kyoto University

<sup>\*1</sup> IWDS-1 はおおむね JIS の包摂規準を包含するものと考えられ、レパートリとしても UCS 統合漢字は JIS X 0213 を包含するので、もし仮に JIS の包摂規準で実験したとしたら包摂できない文字は更に増えるものと考えられる。

る。故に、草書や行書、篆書といった楷書以外の書体で書かれた漢字は原則的には字体の包摂規準で扱えないといえるが、実務的には隸変以降の書体（広義の隸書）を無理矢理、楷書（ないしは、明朝体）とみなして楷書ベースの漢字符号化の枠組に載せることが行われてきたといえる。

UCS 統合漢字の中にも草書や先秦時代の漢字（古漢字）の隸定字をベースにした漢字が収録されており、例えば、「喜」の草書を楷書化した「𠄎」（U+3402, JIS X 0213:1-14-03）や金文の隸定字である「𠄎」（U+2C2DD）はその一例である。あるいは、ベトナムの漢字系文字である𠄎<sup>チヌム</sup>は草書で書かれることが多かったようだが、楷書化したものが符号化されている。

このように、楷書以外の書体の字形も楷書化することで形式的に字体の包摂規準を適用することができる訳であるが、実際には楷書ではない字形に楷書を想定して作られた包摂規準を適用することには原理的な問題があるといえる。一つには楷書における字体標準や筆づかいの埒外にあり、その書体における字形の揺れを適切に扱うことができず、また、楷書の場合と骨格（漢字構造）が異なるケースにおいて、適切な対応関係を表現できないという問題がある。

一方、甲骨文字や戦国楚簡文字のように、一見、現代の楷書体の漢字と見掛けが大きく異なっているにもかかわらず、漢字を構成する部品を見た場合、対応する現代漢字の部品を見出すことができ、各部品を現代のものに置き換えることで形式的に楷書や明朝体の字体を構成可能なことが少なくない（たとえ、その文字がどのようなものであったかが判ってなくても）。これが隸定の原理であるが、このことは書体をまたがるような抽象部品のようなものが想定可能であることを示している。しかしながら、ここで想定可能な抽象部品は意符や音符のような意味のある部品（それはしばしば部品単独で1つの文字として使われ得るようなもの）であるが、字体の包摂規準ではしばしばそれ単独では意味のある部品とは看做せないような筆画のパターンも対象となっている。実際の所、ある2つの抽象部品の形状がある書体でははっきりと違った形をしているのに別の書体では似た形になってしまっているケースがあるため、書体をまたがるような抽象部品を想定するとある字形の見掛けだけでは判定不能なケースが生じてしまう訳である。これは結局の所、ある字形をどういう文字として解釈するかという翻刻の問題に他ならないといえるが、現実的には、なるべく解釈を保留にしたまま記述したい場合も少なくなく、結局の所、どこかでこうした曖昧性を引き受ける必要があるといえる。

こうした抽象部品の決定に関する曖昧性がある場合でも、楷書では見掛け上の部品を導入することでとりあえず見掛け上の漢字構造記述を行うことができ、また、こうした見掛け上の部品に対して字体の包摂規準を適用すること

ができる（あるいは、拡張包摂規準を想定することもできる）。逆にいえば、こうしたことができないということ（すなわち、翻刻において抽象部品の選択を迫られるということ）が包摂規準を楷書ではない字形に適用することの原理的困難さであり、言い替えれば、包摂規準の楷書依存性ということの結果でもある。

### 3. 包摂規準の適用可能性

石塚晴通氏は各時代・地域における漢字時代の標準の存在とその変遷の実態を明らかにするために、「石塚漢字字体資料」およびそれを元にした「漢字字体規範史データベース」（HNG）を実現した。[4] 石塚晴通氏はそれ以前から楷書の字体に関して、初唐頃における標準的な字体（初唐標準字体）が大量の文物の受容とともに日本でも受容された一方、中国では、その後、開成石經を字体の基準とする規範の字体が宋版によって実践・普及したという見通しを持っていたが、[5] HNG のデータはこの仮説を支持するものといえる。いづれにしても、楷書の字体に関して初唐標準と開成石經規範という2つの大きな系列があり、今日の康熙字典を規範とする字体は開成石經規範の流れを汲んでおり、漢字符号もその影響を強く受けているといえる。\*2

当用漢字・(旧) 常用漢字によって普及した今日の日本のいわゆる『新字体』は日本で通行していた字体をベースにしているが、康熙字典規範の受容後に俗字とされたこうした字体は概ね初唐標準の流れにある字体ととらえることができる。初唐標準字体のうち、現代日本の標準字体として受容されたものは字体の包摂規準のカバーする範囲に含まれることとなったが、受容されなかった字体も少なくない。しかし、その部分の幾つかも「大字典」～「新大字典」を経由し、戸籍統一文字・文字情報基盤に取り込まれ、UCS 統合漢字や IVD に追加提案された。また、仏典を典拠として SAT や 韓国等からも提案され、拡張漢字 F には少なからぬ数の初唐標準系の字体を例示字形とする漢字が収録されるに至った。しかしながら、JIS 包摂規準や IWDS-1 は基本的に初唐標準字体を十分にカバーできるものとなっていないため、結果的に重複の疑いのある符号位置も存在するが、包摂規準が十分に整備されていない以上、形式的に仕方がないと判断せざるを得ない面もある。とはいえ、実際のテキストコーパスの運用上、難しい問題が生じてしまったことも事実ではある。

初唐標準字体をカバーする拡張包摂規準の集合としては前述のように著者が試作したものがあがるが [3]、長安宮廷

\*2 [6] では宋版の漢字字体に関して JIS 包摂規準で処理できる例が多いことを指摘している。また、宋版の漢字字体は「いわゆる康熙字典体」に近いものであるが、『新字体』に近い例もあることを指摘している。『新字体』は初唐標準の流れにあるものといえるから、これは開成石經規範受容後も残った楷書的な字体という風にとらえることができるかも知れない。いづれにしても、『新字体』もまた JIS 包摂規準でカバーされているため、そのカバーを高める結果になっているといえる。



て記述するという仕組みを導入すれば良いと考えられる。

## 5. おわりに

JIS X 0208:1997 / JIS X 0213 の包摂規準 (JIS 包摂規準) や IWDS-1: List of UCV (Unifiable Component Variations) of Ideographs はそれぞれ JIS 漢字・UCS 統合漢字という現代の漢文字符を対象とした字体の包摂規準であるが、宋版以降の版本や楷書の字形に対してもある程度有効であるといえるが、楷書以外の書体で書かれたテキストの字形を扱うことは基本的にできないものと考えた方がよい。字体の包摂規準は社会的に共有され普及した概念としての字体標準や明確に定められ強制力を持った字体規範の存在を暗黙の内に想定したものといえ、そうした標準や規範意識があつてはじめて字形のゆれが揺れ (あるいは、逸脱) として認識でき、それらを包摂する抽象的な形状としての字体や複数の異字体を包摂するものとしての抽象文字を定義することが可能になるといえる。

そういう観点に立つならば、草書や行書、隸書、あるいは、小篆といった楷書以外の書体に関しても、「石塚漢字字体資料」が行ったように、対象となる資料をその性格に基づいて選定し、字形用例を採取したグリフコーパスを開発し、その書体における字体の標準 (規範) 意識を浮かび上がらせる作業が必要であるといえるが、これは簡単なことではない。また、今日の漢文字符が暗黙のうちに想定している (字体ないしは抽象文字の) 書体独立性に反する立場であり、現実的には容易に実践可能な道ではないといえよう。とはいえ、テキストコーパスを安定的に記述する上でこうした観点を頭に入れておき、固有の包摂ポリシーを設計することも重要であると思われる。

一方、楷書においては、原理的に字体の包摂規準の考え方が適用しやすいといえるが、戸籍統一文字や仏典に見られるような『俗字』の幾つかを含む初唐標準字体の流れを組む字形に関しては JIS 包摂規準や IWDS-1 では必ずしも十分ではなく、初唐標準字体を扱うための拡張包摂規準の追加が必要だと思われる。

## 参考文献

- [1] : IRG Working Document Series, <http://appsrv.cse.cuhk.edu.hk/~irg/irgwds.html>.
- [2] 須永哲矢, 堤 智昭, 高田智和: 明治前期雑誌の異体漢字と文字コード—『明六雑誌』を事例として—, じんもんこん 2011 論文集, 情報処理学会シンポジウムシリーズ, Vol. 2011, No. 8, 情報処理学会, 情報処理学会, pp. 381-388 (2011).
- [3] 守岡知彦: CHISE による HNG データ収録の試み, 漢字字体史研究 二 — 字体と漢字情報 (石塚晴通監修, 高田智和, 馬場 基, 横山詔一, 編), 勉誠出版, pp. 185-203 (2016).
- [4] 石塚晴通, 高田智和: 漢字字体と文献の性格との関係—「漢字字体規範史データベース (石塚漢字字体資料)」の文献選定, 漢字字体史研究 二 — 字体と漢字情報 (石塚晴通監修, 高田智和, 馬場 基, 横山詔一, 編), 勉誠出版, pp.

- 349-359 (2016).
- [5] 石塚晴通: 漢字字体の日本的標準, 国語と国文学, Vol. 76, No. 5 (1999).
- [6] 石塚晴通, 池田証寿, 岡崎裕剛: 漢字字体規範データベースとその応用, 東洋学へのコンピューター利用第 17 回研究セミナー, pp. 53-63 (2006).
- [7] 守岡知彦: 漢字構造記述再考, 東洋学へのコンピューター利用第 28 回研究セミナー, pp. 337-346 (2017).
- [8] 李 媛: 篆隸万象名義の掲出字の文字同定について, 東洋学へのコンピューター利用第 28 回研究セミナー, pp. 347-366 (2017).