

文字分類方式の変更が字形に及ぼす影響

鈴木俊哉[†]

広く使われる漢字の分類方式として部首・画数分類方式がある。現在普及している部首体系は康熙字典の部首分類、および、康熙字典部首を常用漢字字体に基いて縮約したものである。しかし、康熙字典以前の字書の部首分類は必ずしも康熙字典の体系と一致しておらず、また、特定の部首体系が支配的だったわけでもない。そのため、ある漢字がどの部首に配置されるかは必ずしも一定しておらず、しかも配置された部首の字形が似せられてしまうため、字書ごとに字形が異なるという状況があった。

近年、漢字字典の収録字数を増やすために、それらの古字書から漢字を採集したため、実際には使い分けが不能であるにも関わらず、字書ごとに異なる部首に配置され、あたかも別字のように考えられて国際標準符号にとり入れられる問題が発生している。

本稿では説文解字に存在したが、その後消滅した部首について、そこに含まれていた漢字がどのように分散し、漢字分類方式によって字形がどのように変化したかの調査結果を報告する。あわせて、国際標準における漢字統合規則の改訂の必要性について考察する。

A Study of the Effect from Character Classification to Glyph Shape

suzuki toshiya[†]

Modern huge Hanzi dictionaries are indexed by KangXi radical and the number of strokes, and Hanzi blocks of UCS charset are same. But the history of KangXi radical systems is only the last 300 years, and there had been many dictionaries based on different radical systems. In addition, KangXi itself has a conflict that the ordering system is KangXi, but the description is based on Shuowen. Hanzi dictionaries are always designed to be larger than older one, and the ancient dictionaries are often used to inflate the size of new Hanzi dictionaries, without the consideration about their radical system. It causes some difficulties in the information interchange with the coded character sets.

In this report, the typical effect from character classification to glyph shapes in ancient Hanzi dictionaries are investigated, and the side effect to KangXi and UCS charset are also summarized.

1. はじめに

Unicode Consortium がメンテナンスする Ideographic Variation Database[1]は ISO/IEC 10646:2003/Amd.3:2008[2]から正式に参照された。これにより、既存の符号化済み漢字の異字形の標準文字符号による利用に関しては、ISO/IEC 10646 をメンテナンスする ISO/IEC JTC1/SC2/WG2 による長期間の審議を経ずに利用できるようになった。この異字形データベースの最初の登録は Adobe-Japan1 字形集合であったが、日本も汎用電子情報交換環境整備プログラム[3]で収集されたグリフ(以下、汎用電子外字と呼ぶ)のうち JIS X 0213 漢字の字形バリエーションと考えられるものを 2010年11月14日に、JIS X 0213 外漢字(JIS 漢字などに典拠をもたない、中国や台湾から提案された漢字)のバリエーションと考えられるものを 2012年3月2日に登録した[4]。



図 1: 現在の IVD のグリフ表

ただし、汎用電子外字でも、UCS 対応を検討した結果、ある UCS 符号位置に対応する字形が 1 つしか見当たらなかったものは登録していない(図 1)。これは、汎用電子グリフは行政用の標準の字形・デザインを定めようというのではないため、Adobe-Japan1 のように自社のアーキテクチャで利用できる漢字について全てのデザインを公布しようという意図とは若干異なる背景を持つことに注意しなければならない。また、最初に登録された JIS X 0213 漢字に対応づくと考えられる字形集合は、多くの JIS 漢字表の字形を含んでいる。特に、IVD に登録された漢字表は 1990 年以降の規格票印刷に用いられた平成明朝であるため、JIS 漢字表と全く同じものを少なくとも 1 個含んでいる。平成明朝体以降の日本のデジタルフォントは JIS 漢字部分においては平成明朝のデザインの影響が非常に強く、Adobe-Japan1 の字形例示フォントとして用いられた小塚明朝も例外ではない。その結果、Adobe-Japan1 字形集合と Hanyo-Denshi 字形集合には一見して例示字形に字形差を見出せないものが少なくなく、この 2 つを

1) ここでの「汎用」とは、グリフ収集・整理事業の対象となった既存の行政システムに存在するグリフ群、具体的には戸籍統一文字、住民基本台帳ネットワーク文字、登記統一文字を用いるシステムの間で共用できるという意味での汎用である。(行政外の)一般社会で広く用いられるという意味での汎用ではない。

混ぜて使うことは非常に難しい。

このことは、Hanyo-Denshi 字形集合によって、様々な場所で外字の需要の根拠とされるクサカムリの画数やシンニョウの点の数といった字形差を区別するための登録簿と期待した人々からは強い不満が示された[5]。例示字形の図形的な差を説明できないようなものに別々の字形 ID を与えるべきではないという考え方である。この考え方は、漢字圏に限らず広く通用しているものであるが、異なるデータベースの共役可能性に関する検討が不足している。たとえば、既存の Adobe-Japan1 字形集合では、例示字形は全て 3 画のクサカムリで示されている。これに対し、Hanyo-Denshi 字形集合では 3 画クサカムリと 4 画クサカムリは区別されている(図 2)。



図 2: Adobe-Japan1 と Hanyo-Denshi でのクサカムリに関する例示字形差

逆の例としては、Adobe-Japan1 は筆押さえの有無を区別しているが(図 3)、Hanyo-Denshi では常用漢字でデザイン差とされているため基本的には区別しない(常用漢字政策で区別しないとされるデザイン差を行政業務が区別することになるため)。



図 3: Adobe-Japan1 での筆押さえに関する例示字形差

一見すれば、「Adobe-Japan1 には 4 画クサカムリの漢字はない」と見えるが、実際には Adobe-Japan1 準拠のフォントデザインにおいてはクサカムリの 3 画と 4 画はフォントベンダの自由裁量と解釈されていた。Adobe-Japan1 は例示字形集合のデザイン基準を明示しているわけではなく、Adobe-Japan1 の字形集合が拡大されるにあたって拡大前は自由裁量と解釈されていた字形差が、自由裁量でなくなった事例もある。

また、Adobe-Japan1 も Hanyo-Denshi もこれらは需要があると認められれば作りわけているのであって、区別する字形差のポリシーをもとに、ありうる組み合わせを全て作りわける字形集合ではないことに注意しなければならない。Adobe-Japan1 の例示字形では筆押さえの有無のみで区別しているように見える文字があるが、全ての文字に対して筆押さえのある字形と筆押さえのない字形を用意しているわけではない。また、

Hanyo-Denshi は既存の JIS 規格票字形との整合性に配慮しているため、全ての字形で筆押さえの有無を揃えているわけではない。たとえば「芟」は Adobe-Japan1 は筆押さえのある字形 1 個のみであり、Hanyo-Denshi でも一見する筆押さえの有無を作り分けられているように見えるが、実際にはこれはクサカムリの差である(図 4)。



図 4: クサカムリと筆押さえの両方を含む場合の例示字形差

複数の漢字資源の共役可能性問題は、当用漢字政策に由来する日本特有の問題ではなく、編纂方針の異なる複数の字書をどのようにマージするかという、より一般的な問題として捉えることができる。本稿では、康熙字典の編纂における古字書参照の問題点と、その国際文字符号への影響を報告する。中国や台湾は現時点では IVD への登録を行っておらず、また、Adobe のこれらのマーケットでの字形集合(Adobe-GB1, Adobe-CNS1 など)も登録されていない。従って、これらの問題は IVS のようにデータの一部を機械的にフィルタして回避することができず、より深刻な問題である。本稿執筆直前にこれらの問題に関する寄書を SC2/WG2/IRG に提出し、その議論を行うことができたので、あわせて報告する。

2. 字書における部首分類と再排列

2.1 部首分類の発展の概要

漢字を人間が探すための排列方式として様々な方法が考えられてきた。字義によるもの、音韻によるもの、字形(見た目)によるもの、書き方によるものなどがあるが、検索対象となる漢字の分量、利用者に要求できる予備知識の違いにより、それぞれの方式の得失は異なる。

日本で出版されている漢和辞典の大半は本文を部首画数排列で編集し、総画数や音訓は補助的な索引として追加されている。音訓 50 音方式は同一漢字に複数の音訓がありうる日本では重複の検出が難しく、そもそも読みがわからない漢字を引くことができないという難点があり、一般に用いられない漢字を収めるような大部の字書に向かないためである。また、漢字の総画数は全ての画数(1 画から 64 画まで)均等に分布するものではなく、総画数で排列した場合に一部の総画数に大量の漢字が集まるため、その中から探すための方法が別途必要となるためと考えられる。

日本で用いられる部首画数排列はその基本を康熙字典を規範とし、常用漢字政策に基づく一部の部首の合併・簡略化などを行ったものである。中国、台湾、韓国で用いられる部首体系も基本的に同様である。しかし、康熙字典の部首体系は、部首による

漢字分類をはじめて行った説文解字の部首体系とは異なり、説文解字で形符とされているものがそのまま康熙部首に対応づけられているわけではない。説文部首は意味が体系的に整理できるものを収集したので、漢字の図形的要素を単純に整理したものではないため、「一」のようなものは説文には存在していなかった。

本節ではまず説文解字の部首がどのように縮約されたかを整理する。説文解字以降の字書と部首数についての概略は以下のようである。

字書(成立年)	部首数	収録字数	字数に関する備考
説文解字(100)	540	9353	序文による。
大広益会玉篇(1013)	542	約 22500 ²	原本玉篇では 16917 字とされる[6]。
龍龕手鏡(997)	242	26430	宋本による。
類篇(1067)	542	31319	序文による。
群籍玉篇(1188)	581	40086 ³ ~62236	
五音篇海(1208)	444	54595	
海篇心鏡(1596)	444	(未確認)	
字彙(1615)	214	33179	
康熙字典(1716)	214	47035	

表 1: 康熙字典以前の字書の部首数と収録字数

部首自体の画数排列を行ったのは明代の字彙が最初と考えられている。ここで、版本が広く流通せず、また、後続の字書に引用されないため、一般の字書史では資料とされないものであるが、部首数の変異の推移の傍証として群籍玉篇を含めた。

2.2 説文部首から康熙部首への縮約の流れ

中国字書史の研究では、一般に、玉篇・類篇は説文解字に由来する部首体系を踏襲しとされる。本稿ではその詳細に立ち入らないが、玉篇・類篇の部首体系も実際には説文の部首体系を部分的に変更していることには注意が必要である。説文解字の部首の順序は字義に何らかの規則性をあてはめて決まったものと考えられているが、その規則性の背景は現在でも研究が進められているもので、理解し易いものではない。特に大きな問題として、ある部首が説文の中で単体の部首として立てられているかどうかの一覧性に乏しい点あげられる。たとえば、説文の部首排列のうち、小→八→采→半→牛→犛→告などと示す順序は、ある部首字の一部に注目して次の部首字を選んでいることが想像できるが(図 5 を参照)、たとえば「小」「八」は見つかっても「大」や「七」「九」はどこに配置されるかは予想できない。

2) 玉篇は何度も改訂されているため、現在伝わる大広益会玉篇の序文がなく、明確な字数が宣言されていない。岡井による計数[6]では 28989 字とされており、この数値が広く参照されているが、ここでは大岩本による計数[8](p.189)を参照した。

3) 現在知られる群籍玉篇には欠落があるため、現在残っている字数である。大岩本による計数[8](p.189)を参照した。



図 5: 説文解字での部首の連続性の例

玉篇は 2 個の部首を新設するだけでなく、部首の排列を変更して不完全ながらも字義によるグループ化を図っている。たとえば、「四」～「十」などの漢字に関してはその(説文で説明される)背景を無視して、漢数字の一群として、「牛」「馬」「羊」「豚」なども動物の一群として、集められている。ただし、「一」～「三」は別の巻に配置されるなど、この方針が徹底されているわけではない⁴。

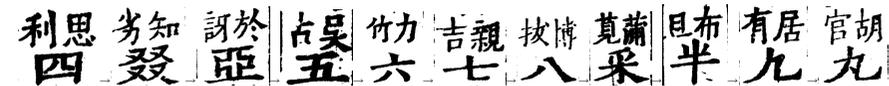


図 6: 玉篇での部首の連続性の例

宋代に編纂された類篇は玉篇の部首体系に対し数個の移動を行っているが、部首の縮約・追加はなく、玉篇で不徹底であった再配列を改善しようという強い意図は見られない。

さて、玉篇・類篇は部首の排列は変更したが、どの漢字をどの部首にわりあてるかという部分については説文を踏襲した。しかし、説文の部首排列は先秦時代の字義に対応づけたもので、漢代以降に通用している意味で説文部首を推定することは容易でない。龍龕手鏡ではこの問題に対して楷書字形による部首の再割り当てを行い、また、「部首が推定できない文字」を収めるための「雑部」を導入した(図 7)。雑部に排された漢字には「古文」などとされているものも多いが、難字を収めるための部首ではなく、「屯」「壺」「巫」など常用度の高いものも含まれている(つまり、龍龕編纂時にはこれらの文字の部首を推測することが既に難しかったと考えられる)ことには注意すべきである。楷書字形による再排列は必ずしも広く需要されなかったが、雑部という考え方は金代～明代の字書にはある程度需要された。群籍玉篇は(書名から示唆されるように)部首体系は玉篇により⁵、末尾に龍龕手鏡で新設された部首を 36 個追加して

4) 常用字に限定した小規模な字書における部首排列の縮約は唐代から活発であったようで、五経文字では 320 個(3235 字しか収録しない)、九経字様では 77 個(421 字しか収録しない)までに縮約されているが、この縮約を大規模な字書の編纂にも適用した例は見当たらない。

5) 群籍玉篇と五音篇海は類篇以降に編纂された字書だが、類篇は補助的にしか参照されていない。どちらも参照されている類篇は現存の類篇とは別の同名異書であったと考えられており、群籍玉篇が玉篇の部首体系

いる。追加された龍龕部首の排列は龍龕手鏡の順序とは異なっており、また、各部首に含まれている漢字も玉篇に含まれているものは玉篇の部首に移動した後に残ったものようであるが、雑部については玉篇部首との整理は行っておらず、先に述べた「屯」「壺」「匚」などが残る。



図 7: 龍龕手鏡(高麗本)の雑部

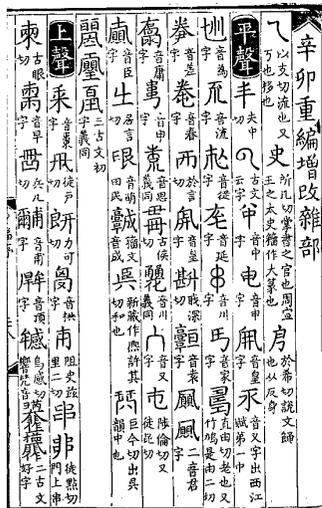


図 8: 五音篇海の雑部

五音篇海の部首体系は群籍玉篇と同様に、玉篇に龍龕手鏡の部首を加えたものであるが、さらに部首の合併を行って 444 個に縮約している。龍龕手鏡や群籍玉篇で雑部に排されていたものも、再度整理を行っているため、雑部に残る大半は難字だけになっている(図 8 参照)。

五音篇海が部首体系を縮約したものの、その縮約は 444 部首にとどめていた。この部首体系は明末の字彙で半分以下にさらに縮約されたが、明代に多数出版された海篇心鏡などのいわゆる「海篇類」ではどれも字釈は五音篇海の字釈をさらに削るなどして検索の便を図るにも関わらず、部首体系の変更は行っていない。しかし、海篇類のうち、部首体系を字義によって排列しなおした海篇心鏡の系統が他の系列より多く刊行されていたことを考えると、依然として五音篇海の部首体系には難があったと思われる。



図 9: 海篇心鏡の部首排列

海篇心鏡は五音篇海の部首を事典のように分類することで一覽性を向上させようとした⁶。海篇心鏡とは別系統のものとして、詳校篇海を海篇心鏡の排列方式で並べなおした篇海類篇がある程度流通していたこと(海篇心鏡は康熙字典では参照されていないが、収録には大差がないと思われる篇海類篇は参照されている)もこの分類方式の有効性を示している。しかし、龍龕手鏡が雑部を作ったのと同様に、海篇心鏡でも最終巻の通用門が 113 個の部首を含む一種の雑部として使われており、この分類方式でも一覽性の悪さが十分には解決されていない。この問題は字彙が部首を約半分の 214 個まで削減し、同時に部首に対する漢字としての要件を放棄することでようやく解決された。

2.3 部首体系の変更とその背景

以上に整理したように、字彙以前の字書の部首体系は、排列や部首体系の縮約の大きな変更を数度行っており、その際に編纂方針が大きく変更されている。新しい部首

に従う理由は現存する類篇に対応する字書が流通していなかったためと推測される。

6) 分類方式は華夷訳語のそれに似ているが、影響関係については明確でない。

の追加、もしくは、削減を行った字書として、玉篇、龍龕手鏡、五音篇海、字彙を選ぶと、その編纂方針は以下のような変更を与えている。

- 玉篇: 説文以来の編纂方針と異なり、掲出字から小篆を除外する。小篆の字形と関連している部首の関係性が見えなくなるため、部首の排列も変更する。
- 龍龕手鏡: 楷書において形が分化してしまったもの(折と崇など)は別部首としてもよく、逆に混同されるようになったものを同部首にしても良いとする。また、単体字として使われない図形的要素を部首としてもよいとする。部首が単体字としての固有の意味を持ちがなくなったため、部首排列を音韻で全面的に整理し直す。
- 五音篇海: 玉篇と龍龕の部首体系を合併し、444 個に整理し直す。楷書において混同されるようになったものを同部首にし、また、楷書において安定している図形要素が説文・玉篇の部首と異なっていれば、これを優先する。
- 字彙: 篇海類の部首体系に対して、部首字が他の部首を含んだ字であると見える場合にはこれら合併する。また、龍龕手鏡や五音篇海で収集された難字は削り、龍龕以来の部首である「雑部」も削除する。

3. 字書の再排列に由来する重複符号化

3.1 意図的な重複掲出のミスによる重複符号化

前項で整理したように、部首分類の変更は、先行する字書に対して明確な編纂方針の違いをたてた上で変更されている。一般に、部首分類の字書の中で部首の移動を行った場合、同部首に排列された漢字と類型化に拍車がかかる。図形要素がただか 2~3 個の部首字である場合は大きな問題とはならない(たとえば、「聞」は「門」「耳」のどちらに排されても字形が大きく変わるとは考えにくい)が、「次」の部首を削除すると「盜」「羨」などの同部首に置くことで字形の類型化を維持していたものが制御できなくなり、字書によって異なる字形で示される。宋代のいわゆる「篇韻」以降の大規模字書の編纂では、形音義が揃わないため文字の同定が困難な典拠(韻書は音しか示さない)であっても別項を立てて見かけの収録字数を増やす方針がとられているため、これらの字書を単体で評価して別符号とする必要があるのかどうかは判断できない。

極端な例として、字書内の誤記に由来する字形差がある。説文は小篆字形に加えて、同義と思われる古文文献に見られる字形を古文・籀文字形として示すが、その中には必ずしも排列された説文部首を図形要素として含まないものも多い。説文が依拠した古文文献が散逸したため、同音異義字による仮借なのか、そもそも同音ですらないのか現在ではわからないが、「排されている部首を図形要素としてまったく含まない」ことは検索の際に大きな難点となる。しかし、音韻分類を行う字書だけでなく、部首分類を行う字書でも、説文の掲出方法の影響は非常に大きく、楷書化された古文・籀文字形

を、図形要素としては関連性が見えなくても説文部首に排列するという方針が類篇の頃までには維持されていた。龍龕手鏡や五音篇海は意図的に部首体系を変更してこの問題に対応し、字彙は説文や玉篇に辿ることができない古文・籀文字形を掲出しないことによって対応した。康熙字典は字彙の部首体系を踏襲しながら、これらを説文・玉篇と同様に図形要素としては関連が見えない部首に排し、さらに楷書字形において部首のように見える項目にも同じものを排している。検索の利便性は改善しているが、常用されない漢字を同じ字形で大きく離れた掲出することを木版印刷で行うことは難しい。

康熙字典の本編で独立の項を立てられている漢字はほぼ全てが CJK 統合漢字拡張 B で符号化されたが、独立項を立てられていない見出し字および補遺から採集したものが CJK 統合漢字拡張 E には提案され、その中には翻刻ミスにより生じた字形差が疑われるものが少なくない。網羅的な確認は完了していないが、IRG#38 で報告された例を以下に示す。

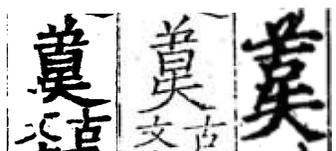
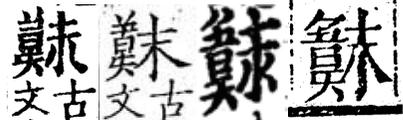
拡張 E 提案字	康熙字典拠	別箇所掲出	康熙字典が参照する古字書の字形
 G_K1393.07			 左より大広益会玉篇・宮内庁宋本、澤存堂翻刻本、四部叢刊元刊本
 G_K1393.26			 左より大広益会玉篇・宮内庁宋本、澤存堂翻刻本、四部叢刊元刊本、字彙補

図 10: 拡張 E に提案された康熙古文字形と別箇所掲出字形の比較

説文解字 (大徐本)	篆隸萬象 名義	大廣 益會 玉篇	類篇	群籍 玉篇	五音篇海	五音集韻	重訂直音篇	新校經史 海篇直音	海篇 朝宗	海篇 大全	海篇 心鏡	重刊 詳校 篇海	篇海類編	字彙・ 字彙補	正字通	康熙 字典

図 11: 康熙字典以前の「谷」「谷」の掲出字形

康熙字典	中華 大字典	中文 大辞典	漢語 大字典	中華字海	SuperCJK 10.2	SuperCJK 14.0
					0-8C37 52495 52495 1189.010 63902.010 G0-3948 T1-4947 J0-432B K0-4D5B V1-676B 谷 0	08C37 52495 52495 1189.010 63902.010 谷 0 G0-3948 T1-4947 J0-432B K0-4D5B V1-676B 谷 谷 KP0-D0A8
					2-7BAB 52496 52496 1189.020 63902.020 G_KX T5-2477 谷 谷 谷 0	27BAB 52496 52496 1189.020 63902.020 谷 0 G_KX T5-2477 谷 谷 谷 谷 KP1-77FC
					2-050C 02362 02362 0127.150 10244.090 G_KX T5-2339 谷 谷 八 5	2050C 02362 02362 0127.150 10244.090 谷 谷 八 5 G_KX T5-2339 谷 谷 谷 谷

図 12: 康熙字典以降の「谷」「谷」の掲出字形

3.2 編纂時の同定ミスによる重複符号化、および、字形標準化の誤り

大規模な字書の編纂中の同定ミスにより発生した字形差と、その文字符号への影響例としては「谷」(U+27BAB)に関するものがある(図 11)。この文字は説文では「谷」と区別されていたが、楷書では字形の区別がつけにくくなり、説文部首を踏襲している玉篇・類篇でも小篆の字形差を十分に表現できなくなっている。また、「谷」は形符

としてよりも声符として利用する漢字のほうが多かったため、これを字義の上で区別する動機づけが弱いという背景もあり、五音篇海では八部に統合された⁸⁾。明代の海篇類は基本的には複数の字書をマージしたのではなく、五音篇海の字形に従いハチャネをもつ「八」につくるものが多かったが、字彙・字彙補が編纂されるにあたり、玉篇・類篇の字形を参照しながら「谷」に配置したため、康熙字典ではこれを3箇所排することとなった。この段階ではまだ五音篇海の字形に従いハチャネを含めて作っており、図 12 に示すように初期の金属活字印刷字書である中華民国の中華大字典などでもそれを維持しているが、漢語大字典では中華人民共和国の印刷標準字体に従ってハチャネを削った。その結果、「谷」の小篆書体に近いものとなってしまい、中華字海などの字書では「谷の旧字」などと解説されている。事実、この漢字は康熙字典を典拠に拡張 B に追加されたが、web 検索などで用例を探すと「谷」として使われる例しか見当たらない。

もう一つの例として、「章」とその異字形群が挙げられる(図 14)。この漢字も本来は設問で独立の部首を立ててあり、海篇類でもこの部首は維持されていたが、龍龕は一部と雑部に、字彙は高部にこれを合併した。この漢字が「谷」と事情が異なるのは、説文はこれを部首として含む籀文字形を多数示していた点である。このため、「章」や

8) 従来、八部に排される「谷」様の字形は「併丁部頭」という逸書に見える漢字と解釈されてきた(たとえば漢語大字典など)が、五音篇海の掲出字について網羅的な調査を行った結果から、逸書に由来するものではなく、五音篇海が部首を縮約したために玉篇などに存在する文字を移動したことを明示したものであるという見解が出されている[8]。本稿でもこれに従った。

「土」を図形要素として含む漢字に対して「章」の字形で置換すれば古文になるという誤解が広まり、説文には存在しなかった字形が古文として作られたと推測される。玉篇にはそのように推測されるものが採録されているが、部首排列は説文に従っているため、擬古的に作られた字形群に対しては部首の類型化が十分に働かない。大広益会玉篇の宋版本ではある程度整理されているが、唐代の原本玉篇に基づくとされる篆隸万象名義ではこの混乱が明確に表れている。おそらく同様の混乱に基づいて発生した異字形群が龍龕にも採録されている。「谷」「谷」の事例と異なるのは、常用度の高く、かつ良く似た字形がなかったため、誤字の字形も安定化しなかったと思われる点である。

康熙字典	章	唐韻	   <small>王韻(郭)注釈 度 王韻裴務齊本 唐韻写本</small>
		集韻	  <small>張氏重刊宋本広韻 符山堂藏版重刻広韻</small>
		説文	   <small>世界書局(平津館本) 説文真本(朱筠本) 四部叢刊(岩崎本)</small>
		玉篇	  <small>四部叢刊(元刊本) 宮内庁宋本玉篇</small>
康熙字典	章	字彙	   <small>六書正詁 四庫全書本 早稲田明刊本 早稲田元刊本</small>
		字彙補	   <small>説文 世界書局(平津館本) 説文真本(朱筠本) 四部叢刊(岩崎本) 小徐本(華文書局)</small>
康熙字典	章	字彙補	  <small>白虎通 Related character is not found in this book.</small>
康熙字典	章	字彙補	  <small>説文長箋</small>

図 13: 康熙字典が掲出する「章」様字とその典拠字形

特に、唐韻・王韻の字形は部首排列字書と韻書の資料交換が活発となる前であり、部首分類の字書とはかなり異なる字形となっている(図 13)。また、説文の研究書は一般的な楷書体との差異を強調するために強引に変形した楷書体を示すことが少なくないが、往々にして既存の研究書で用いる字形を利用せずに新たに作る傾向が見られる。六書正詁は元代の説文研究書であるが、元刊本と明刊本、さらに清代の写本でも異なっている。多数の説文研究書・校訂書を参照すればそれらを個別に収集することに意味がないことは明らかになるが、字彙補は十分にこれを整理できなかったため、康熙字典にもこれらの異字形が採録された。

五音篇海は部首「章」を維持したが、これを含む漢字を全て同一部首に集約することにはしていない。後続の海篇類では字形差を十分に管理できなくなり、不均一が生じている。これらは康熙字典には採録されていないが、難字を収集することで見かけの分量を拡大しようという編纂方針は現在でも続いており、五音篇海や龍龕手鏡といったしばしば参照される難字資源が尽きた後には、これらから採集したものを別典拠の別字として統合漢字に提案される恐れが高い。統合漢字拡張 F には中華字庫典拠の提案は予告されていないが、古字書を典拠とした提案の取り扱ひの速やかな検討が必要であろう。

3.3 IRG#38 における提案とその結果

本稿執筆中に ISO/IEC 10646 のメンテナンスをする ISO/IEC JTC1/SC2/WG2 の漢字部会である Ideographic Rapporteur Group の第 38 回会議が韓国慶州で開催された。本稿で報告した問題点のうち、U+29ACF にかかわるものと、康熙字典の複数回掲出によるものを寄書として報告し、前者については統合、後者については一貫性のある対応をとるための審議と結論の延期を提案した。前者に関しては、U+29AC3 章、U+29ACF 章、U+29AD6 章、U+29AD7 章、U+20180 章、U+20177 章の 5 個、および、拡張 E に提案された U+20177 と G_ZJW00686 の統合(図 15)を提案した。後者の延期提案は合意されたが、前者については(拡張 B 漢字の統合可能性については審議未了となったものの)、G_ZJW00686 が既存の拡張 B 漢字と統合可能という主張は合意されなかった。

楷書化された結果、一点・一画を越える図形要素の増減が生じたものを同一視すべしという主張は、常用しない漢字の場合に直感的な判断がむずかしく、抵抗が強いと思われる。

	ISO/IEC 10646:2003 (高部)	康熙字典 (高部)	正字通 (高部)	字彙・字彙補 (高部)	説文長箋	音韻字海	海鏡大全	五音篇海	群籍玉篇	龍龕手鑑(高麗本) (十部・雜部)	類篇	大廣益 會玉篇	篆隸万象名義
Glyphs in same radical	章 章 章 章 294C3 294C6 294D6 294D7	章 章 章	章	章 章 章	章	章	章	章	章	章	章	章	章
Glyphs in different radical	章 章 20177 20180					章 章 (高部欠落)	章 章 (高部全欠)			章 章			
苙	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳
水	灇 灇 2415D 2417B	灇 灇	灇	灇	灇	灇	灇	灇	灇	灇 (本部全欠)	灇	灇	灇
雨	霽 霽 291D1 291CA	霽 霽	霽	霽	霽	霽	霽	霽	霽	霽 (雨部全欠)	霽	霽	霽
木	樟	樟 樟	樟	樟	樟	樟 樟 樟	樟	樟	樟	樟	樟	樟	樟
山/巛	嶂	嶂 嶂	嶂	嶂	嶂	嶂 嶂 嶂	嶂	嶂	嶂 (山部全欠)	嶂	嶂	嶂	嶂
鳥	鷓	鷓	鷓	鷓	鷓	鷓	鷓	鷓	鷓	鷓	鷓	鷓	鷓
𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳
𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳
𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳
𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳
𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳
𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳
𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳
𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳
𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳	𦉳

characters in the radical of 亭 etc (modernized character of 亭, 亭, 亭)
 characters in the radical that does not include 亭, 亭, 亭, or 亭 etc

図 14: 「亭」を含む漢字群とその典拠字形

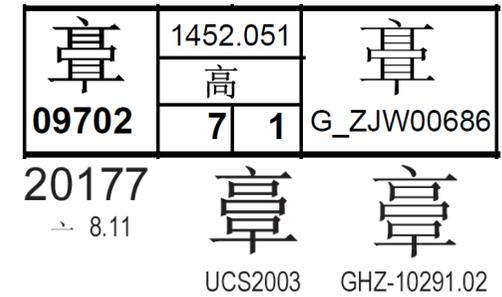


図 15: 拡張 E に提案された「亭」様字とその統合候補拡張 B 漢字

4. まとめ

本稿では、編纂方針が異なる字書を混合した場合に生じる問題について、康熙字典までを例にとって典型的なパターンを2つ(重複掲出で発生する字形差、同定ミスによる字形差)示した。また、部首デザインは各字書が個別に規範化しているものであり、これを無視した編纂および標準字体化を行うと本来の使い分け意図と異なる結果になることを示した。

本稿で整理した事例より、たとえば字書の見出し字形を単に符号化するだけでは、字書の参照関係すら正しく記述できず(たとえば字書Aは文字Bについて同書の別項、または字書Cを参照するが、文字Bがそこに見つからない場合、字書Aによって何も知見を得ることができない)、古文書以前に字書の電子化にも問題があると言える。字書の共役可能性には編纂方針について十分な配慮が必要である。

このことから、IVDのようなグリフ集において、「例示字形は単体で何らかの図形文字の範囲を暗黙で示すことができる」という前提に基づいて、様々な字形集合を混ぜ合わせて利用しようという考え方の困難さも示される。混ぜ合わせた字形集合がどのような編纂方針によって混ぜ合わされたのか、もとの字形集合を想定して作られたグリフ資源は常に混ぜ合わせた字形集合の資源の素材として適切なのか、といった問題を混ぜ合わせるユーザが考えなければならない。

今後の課題として、字書から字書への転記や翻刻の際に生じる字形差について、より広い範囲での調査が必要である。これまでの標準化動向は基本的に活字印刷以降の用例を想定しており、康熙字典字形と漢語大字典字形の差異に関しては積極的に統合を認めていた(中国が康熙字典を典拠に提案しながら、提出するフォントは漢語大字典をもとにしていたため)が、今後は康熙字典や漢語大字典が参照する原典から、これらの字書が掲出する字形と突合せをせず別途翻刻した字形が提案されることが予想される。これらが別文字として符号化された場合、字書間の参照関係の追跡を著しく困難にし、字書の画像としてのデジタル化は容易であっても電子文書としての利便性を損なう。転記・翻刻により生じる字形差をできる限り統合できるような統合規則が必要である。

謝辞

本研究は科学研究費補助金 若手研究B 課題番号21700113、および、基盤研究C 課題番号24500116の補助による。

参考文献

- [1] Unicode Consortium, “UTS#37: Ideographic Variation Database”, <http://www.unicode.org/reports/tr37/>
- [2] ISO/IEC JTC1/SC2/WG2, “ISO/IEC 10646: Information technology -- Universal

Multiple-Octet Coded Character Set (UCS) --”, edition 2003, amendment 3, 2008.

- [3] 日本規格協会、国立国語研究所、情報処理学会、“汎用電子情報交換環境整備プログラム成果報告書”, 2009年3月, <http://www.meti.go.jp/information/downloadfiles/c100806a04j.pdf>
- [4] 情報規格調査会, “国際規格 ISO/IEC 10646 の規定に基づく漢字字形データベースへの登録”, 情報規格調査会プレスリリース, 2010年4月1日付, 同9日修正. <http://www.itscj.ipsj.or.jp/pr/itscj/sc2100401.pdf>
- [5] CSSWG, “Minutes and Resolutions Kyoto F2F Fri: Transitions, IVS, CSS3 Text, Floats and Exclusions, Regions, Pagination”, <http://lists.w3.org/Archives/Public/www-style/2011Jun/0325.html>
- [6] 貝塚茂樹、小川環樹, “中国の漢字”, 中央公論社, 1981
- [7] 岡井慎吾, “玉篇の研究”, 東洋文庫論叢 19, 1933
- [8] 大岩本幸次, “金代字書の研究”, 東北大学出版会, 2007
- [9] 梁春勝, “《新修玉篇》《四声篇海》引書考”, 中国典籍与文化, 2008年04期
- [10] 鈴木俊哉, “Proposal for the Discussion How to Handle the Mistakenly Differentiated Glyphs in Huge Dictionaries”, ISO/IEC JTC1/SC2/WG2/IRG N1859, 2012.
- [11] 鈴木俊哉, “Feedback to IRG N1824 part 5; proposal to add new UCV for U+29ACF related characters”, ISO/IEC JTC1/SC2/WG2/IRG N1824 Feedback, 2012