

包摂と分離

—多漢字文献翻刻の問題—

池田証寿

北海道大学

(発表要旨) コンピュータで扱える漢字は拡大の方向にあり、原文の字体を「見たまま」に翻刻することが可能となっている。院政初期、法相宗の学僧による編纂とされる図書寮本類聚名義抄を例にして、新字体、旧字体、原文の字体のいずれに比重を置くのがよいか、方法論の吟味を行う。

Unification and Separation in Chinese Character Information Processing: A Case Study of the Transcription of *Ruiju-Myōgi-Shō* from Archives and Mausolea
Department of the Imperial Household Agency

IKEDA, Shōju
Hokkaidō University

In the past several years, we can process many Chinese characters now by computer. This paper discusses the following issues related to the transcription of *Ruiju-Myōgi-Shō*(the Chinese-Japanese dictionary in the Heian period).

1. How to transcribe from old manuscript to printing.
2. How to encode Chinese characters in old manuscript.
3. What is the advantage of using new Chinese character standard, old Chinese character standard, and the shape of character in manuscript?

1 包摂の問題

1.1 JIS X 0208:1997 における包摂規準とその意義

JIS X 0208:1997 における包摂 (unification) とは、「複数の字体を区別せずに、それらに同一の区点位置を与えること」である。それまでの規格において、「同値」「字形の違いがわずかであると認めるもの」として漢字の字体・字形のゆれを記述していたが、その範囲が必ずしも明確でなかった。「包摂」は、その曖昧さの解消を意図したものであり、これによって一般に存在する漢字の字体をどの符号位置の図形文字に対応させるかが明確になった。符号化できない漢字字体も明確になり、漢字の追加拡張に道を開くことになったのである。

1.2 JIS X 0213:2000 における包摂規準の変更

JIS X 0208:1997 の包摂規準の精度の粗さに多くの批判があるが、その最たるものは、常用漢字に対応する旧字体 (康熙字典体、正字体) が独立に符号化できないという点である。しかし、旧字体とは何かという問題は、その範囲が明確でなく、突き詰めれば、新字体 (常用漢字表の字体) 以外の字体すべてということになる。

JIS X 0208:1997 では、いわば妥協の産物として、人名用漢字許容字体・常用漢字表康熙字典体別掲字の追加として、都合 105 字の面区点位置を新たに設けて、包摂規準を一切適用しないこととした。現実社会の要求に押し切られた格好だが、包摂規準の運用の難易度を上げるものである。たとえば、「勉」の新字体と旧字体との違いは、非常に微妙で、画数こそ違え、瞬時

に区別することは困難である。ここに分離 (separation) の問題が生じる。

1.3 Unicode の統合漢字

当初の統合漢字 20,902 字に、拡張漢字 A の 6,582 字、拡張漢字 B の 42,711 字の追加があり、総計 70,195 字となっている。これだけの漢字が使えるようになると、古文書の翻刻においても、異体字をいかに使いこなすかが課題となる。

Unicode は当初から、source code separation という矛盾を抱えていた。非常に類似した字体でも原規格で分離するものは、統合漢字でも分離するという例外の規則である。拡張漢字 A・B により問題が複雑化することが予想される。

2 活字翻刻と電子テキスト

2.1 原文から活字翻刻まで

翻刻とは、「原本通りの内容を一字一句の増減・改変も無く伝えるように出版すること」(『新明解国語辞典』第 5 版、東京：三省堂、1997 年) と説明される。しかし、原文が版本・写本である場合には、変体仮名や草体の漢字までそのまま翻刻することはせず、通行の活字に置き換えられ、ここに翻刻者の解釈が介在する。活字の書体は、例外なく明朝体が用いられる。また、従来の活版印刷では、足りない文字は造字して対応した。原文を生かそうと思えば、通行の活字に存在しない異体字の作成が必要となり、出版経費に跳ね返ることになった。

また、訓点資料の釈文・訓読文の作成では、原文のヲコト点、返り点、傍訓を、加點時期を区別して、なるべく忠実に翻刻しようとする。かなり複雑な組版となる。

2.2 原文から電子テキストまで

JIS X 0208:1997 に代表される文字コードは、漢字字体そのものを規定しない。あくまで一般に存在する漢字字体を符号化し、情報交換するのが目的である。したがって、原文を活字翻刻したものを、JIS の包摂規準 (またはそれに代わる字体認定規準) によって符号化し、電子テキストを作成する。

ややもすれば、JIS の例示する字体が規範化し、それまでの活字字体とのずれが解消できないという事態となるが、これは JIS に対する誤解である。また、JIS の包摂規準は、翻刻の字体認定基準としても使えるが、両者の区別は必要である。

2.3 見たままの翻刻 (「原文に忠実な翻刻」問題)

JIS X 0208 は、現代日本語の表記を目的とするもので、古典文献の翻刻を目的としていない。そのため、古典文献をなるべく忠実に翻刻しようとしたとき、原文の異体字が存在しないケースがある。たとえば、経典名の「大毘盧遮那成仏神變加持經」の「毘」は、古典文献では「田」と「比」を上下でなく左右に配置して書くことが多い。「毗」は Unicode で使えるが、JIS X 0208 では使えない。「貌」の異体字の「兕」は JIS X 0208 にあるが、「良」はない。(「良」は Unicode 拡張漢字 A に追加された。) しかし、「翻刻」は解釈であるから、「毘」「毗」、「貌」「兕」「良」のいずれを選択するかは、翻刻者の判断に任される¹⁾。

3 翻刻字体の選択

3.1 新字体と旧字体

新字体は、「当用漢字字体表」(1949 年) 及びそれを継承する「常用漢字表」(1981 年) の字体であり、いわゆる略字体や筆写体・俗字体などを採用する。それまでの正字体は、旧字体となった。旧字体は康熙字体とも呼ばれる。

活版印刷の時代には、学術的な出版物は、旧字体で印刷されることが多かった。たとえば、『高山寺資料叢書』(1971 年～2003 年、東京：東京大学出版会) は第 17 冊までが活版印刷、

¹⁾この問題に関しては、豊島正之の JIS 漢字に関する一連の論考、特に「「原文に忠実な翻刻」をめぐる」(<http://jcs.aa.tufs.ac.jp/mtoyo/on-JCS/index.html>) を参照。

それ以後、23冊までがLaTeXによる組版となった。

3.2 手書き筆写体

写本には、旧字体(康熙字典体)でなく、新字体かそれに近い字体の漢字が用いられており、翻刻でそれが生かされるケースもある。たとえば『明恵上人資料第一(高山寺資料叢書第1冊)』(1971年、東京:東京大学出版会)の「通巻凡例」では「原本に存する漢字の行書體、草書體、異體字及び略體字は、何れも、康熙字典所掲の楷書體に翻字することを原則とする。但し原本で次の諸字體の使用されてゐる場合には、例外として、原本通りに略體字又は異體字(常用漢字新字體にあるものはその字體)によつて翻字することとする」とある。「哥・歌」「円・圓」「塩・鹽」「无・無」「帀・紙」等は翻字で併用され(Aとする)、「亜・亞」「婦・歸」「權・權」「恵・惠」等は康熙字典所掲の字体に改められている(Bとする)。

3.3 文字コードの制約と可能性

上のAに例示したのは、JIS X 0208:1997で別区点位置に符号化される例であり、JIS X 0208:1997で符号化できない字体は例示を略した。Unicodeを使うと、「季・年」「良・貌」「災・災」等、今昔文字鏡を使うと「冨・海」「慥・慥」等が区別できる。(前者がJIS X 0208にない。)

また、『明恵上人資料第一』が康熙字典所掲の字体に改めた異体字も、「兂・光」「獸・厭」「高・高」等はUnicodeで、「吳・異」「吉・吉」等は今昔文字鏡で区別できる。(後者が康熙字典所掲の字体。)つまり、Unicodeや今昔文字鏡を使えば「見たまま」の翻刻もある程度まで可能になっている。こういう状況では、翻刻方針の設定がより重要となる。

4 正字と俗字

4.1 中国及び日本における漢字の標準字体の推移

ここで、石塚晴通「漢字字体の日本的標準」²により中国および日本における漢字の標準字体の流れを見ておく。

統一王朝が成立した初唐期(618-712)に異体字を強く意識するようになり「正俗通訛」の基準が形成される。初唐の標準字体は、『開成石経』(837年)の基準と大きな差があるが、『開成石経』及び『干祿字書』(大暦9[774]年頃顔真卿書写)の漢字字体の規範は、宋版において実践・普及していく。一方、日本では、初唐の標準字体がおおむね日本的標準となり、定着していく。その後の大きな変化は、『康熙字典』(1716年)以後であり、近代の活字印刷はこれをよりどころとした。さらに第2次世界大戦後の当用漢字(1946年)、JIS漢字(1978年)が大きな変化をもたらす。以上、漢字の標準字体の流れを概観した。

4.2 字体意識の変容—旧字体と俗字・異体字

「旧字体」とは、康熙字典体のことで、新字体を容認しない立場から「正字体」とも呼ばれる。しかし、この「旧字体」なる概念は実に曖昧で、突き詰めてゆくと、新字体(常用漢字表の字体)でない漢字字体すべてを包含する概念となる³。

「俗字」の「俗」の概念は、『干祿字書』の「俗通正三体」の区別が参考になる。「俗」は浅近で戸籍・文章の草案・契約書・処方箋に用いて雅言でないもの、「通」は上奏文・書簡文・

²石塚晴通「漢字字体の日本的標準」(『国語と国文学』76・5、東京大学国語国文学会編、東京:至文堂、1999年)。

³この問題の詳細は、小池和夫「『東京セブンローズ』の文字について」(小池和夫ほか『漢字問題と文字コード』、東京:太田出版、1999年)、小池和夫「出版と文字コード」(『bit別冊 インターネット時代の文字コード』、東京:共立出版、2001年4月)、豊島正之「文字の符号化—新JIS漢字第3・第4水準の開発から見た—」(『第64回研究セミナー報告』、京都:京都大学大型計算機センター、2000年3月24日)、及び池田証寿「旧字体への欲望が映し出す日本語の現在—『東京セブンローズ』論 漢字というものの現在[連載第2回]」(『d/SIGN』no.2、筑波:筑波出版会、2002年3月)を参照。

判決文に用いて久しく用いられてきたもの、「正」は著述・文章・科挙の答案・碑文に用いて証拠のある正しいものと規定される。「俗字」を誤字に近いニュアンスで用いる向きもあるようだが、「俗字」は前代の正字体であることが多い。

「異体字」の概念は、更に多様である。本来は、類形異字（類似した字形だが異なる字）、異形同字（異なる字形だが同じ字）という過誤に陥りやすい字体を区別する必要から出てきた考え方である。『干禄字書』の「俗通正」の別は、その典型である。しかし、最近の傾向として、非常に微妙な差異のある字形を異体字と呼ぶことが多い。「辺」「邊」「邊」の異体字は数十存在する、といった説明がその典型である。

5 古字書の漢字字体規範記述—図書寮本『類聚名義抄』の場合

5.1 古字書の掲出字体と注文の字体注記

古字書の体裁は、見出し字（掲出字）を大きな文字で掲げ、その発音と意味・用法を小字で説明する。漢字字体の規範は、見出し字そのものが示すことになるが、「俗」「通」「正」「或」「古」の字体規範が注文に示されることも多い。すなわち、古字書の字体規範は、見出し字と注文の字体注記によって表現される。ときには、非常に類似した字形を見出し字で区別して表示しようとする場合もあり、その場合には、原本の字形そのものを示すことが必要になる。

5.2 図書寮本類聚名義抄とは

図書寮本『類聚名義抄』は、零本ながら院政期書写にかかる漢字字書であり、原撰本系『類聚名義抄』唯一の伝本である。院政初期、法相宗の学僧によって編纂されたとされる。出典を明記するのが特徴で、その数は、仏典 67、外典 36、点本 27 に及ぶ⁴。仏典と外典からは音注釈義を、点本からは片仮名和訓を、原典に忠実に引用転載する。辞書史、漢文訓読史、音韻史、語彙史の面から平安時代の国語史料としてもっとも重要度の高い文献の一つである⁵。

5.3 図書寮本『類聚名義抄』の翻刻方針

その翻刻方針はおおよそ次のとおりである。原文縦書き。

- 1 これは図書寮本『類聚名義抄』の本文及び訓点を可能な限り判読した翻字である。
- 2 原本では、一頁あたり七行四段で、掲出字を大字、注文を小字割注で書写しているが、印刷の困難を避けるために、掲出字をやや大きな文字で表して、注文と区別するにとどめる。原本通りの配行も印刷が困難であるので、特に示さない。
- 3 掲出項の所在は、複製本の頁数を漢数字で、行数を洋数字で、括弧内に示す。
例、水（四二）
- 4 漢字は、原本に近い字体を採用することを原則とする。
- 5 旧字体、新字体いずれかに統一することをせず、JIS 漢字の範囲内で表現可能なものはそれを用いることとする。
仏 → 仏、佛 → 佛（両者とも JIS X 0208 にあり）
良 → 兒、毗 → 毘（前者は JIS にないが、後者が JIS にあり）
- 6 虫損等により判読できない文字は「□」で表す。字画の一部により本来の文字が推定できるものは「□^(?)」のように推定される文字を右横の括弧内に入れて示す。
- 7 明らかな誤字も原本のとおり翻字し、正しい文字を「^(正)誤」のように右横の括弧内に入れて示す。

⁴吉田金彦「類聚名義抄」（『国語学大辞典』、東京：東京堂、1980年）。

⁵築島裕「国語史料としての図書寮本類聚名義抄」（『図書寮本類聚名義抄』、東京：勉誠社、1976年）参照。出典研究の詳細は、池田証寿「図書寮本類聚名義抄出典略注」（『古辞書とJIS漢字』3号、札幌：北海道大学文学部池田研究室、2000年3月）を参照。

- 8 略字は、次の基準により翻字する。
- (1) 頻用される略字のうち、次の例は通行の字体に改める
例、メ → 反、ㄆ → 音、ㄨ → 也
 - (2) 頻用される略字のうち、出典名に用いられ、JIS X 0208 で表現できるものはなるべく原典のままとする。
例、广 (玄応)、茲 (慈恩)
 - (3) それ以外は右横に括弧に入れて本来の漢字を示す。
例、罍 (經)、类 (類)
- 9 同じ漢字が出現する場合に「一」を用いているが、これは右側に括弧に入れて本来の漢字を示すこととする。

例、一^(罍)淨 (一二三)

- 10 踊り字「」は「々」を用いて表す。
- 11 原本の片仮名は、現行の片仮名に改める。
例、ㄨ → キ、爪 → ス、ㄆ → シテ
- 12 原本のヲコト点は、平仮名で表す。
- 13 漢字の左下に施された返点は、「一」と表す。
- 14 原本の声点は、被注字の右下にやや小さな文字で(平)、(平軽)のように示す。
- 15 片仮名和訓に付された小字の出典注記は、煩瑣になるのを避けるため片仮名和訓と同じ大きさを翻字する。
- 16 掲出字の周囲に、朱書による片仮名の字音注が見えるが、「朱書」である旨は特に注記しない。
- 17 出典名の頭に付された点発は「・」(中黒)で表す。

以上の方針で作成した本文を、最後につけたので参照されたい。一つは LaTeX で組んだもの⁶、もう一つは、向井裕一氏に DTP で組んでいただいたものである⁷。

5.4 包摂と分離の実態

図書寮本『類聚名義抄』の冒頭の 13 項目を例として、包摂と分離の実態を説明しよう。この 13 項目には、21 字の掲出字と 246 字の注文、計 267 字の文字がある。

まず、JIS X 0208 でなるべく表現しようとした場合、対象とした 267 字は次の処理結果となる。数字は延べ数。括弧内の例は出現順。

JIS で表示可	261 字
JIS で表示不可	6 字
そのうち Unicode で表示可	2 字 (濃金)
今昔文字鏡で表示可	6 字 (濃濃湊湊金湊)

次に、旧字体を原則とする翻字方針をとった場合には、対象とした 267 字は次の処理結果となった。

JIS で表示可	261 字
そのうち旧字体 JIS 変換	5 字 (眞爲爲乗爲)
旧字体 Unicode 変換	1 字 (清)
旧字体今昔文字鏡変換	9 字 (益者清延類類類者者)
JIS で表示不可	6 字

⁶金水敏氏作成の訓点資料用マクロを使用した。

⁷池田証寿「なぜ日本語の書記の歴史なのか—小松英雄と藤枝晃とを手がかりに」(第 7 回句読点研究会、小石川後楽園「涵徳亭」、2002 年 2 月 17 日)の資料として作成。発表に当たっては前田年昭氏にお世話いただいた。

凶書寮本類聚名義抄の解読―水部

池田証寿

水(四二)

弘云戸癸^(七)反。・中云所^(一)以潤万物^(一)也。・玉云五行一曰^(水)北方行也。・魴曰江海淮濟河泉皆曰^(水)一也。井泉曰^(水)一也。・月^(水)俗云佐波利^(上上上)類聚抄莖垂類

二(四三)

類云正水今

大水(四三)

玉云海也

洪^(水)一(四三)

上音鴻^(平)・真云大也 オホキナリ^(平平平聲平上)書

麗^(水)一(四三)

千字文註云在益州永昌郡中有金^(以下補入)往々浮出

一^(水)手(四四)

順云日本紀私記云賀古^(平平聲)今案書注賀古者鹿子義。俗或呼^(水手)一^(水)為賀古^(上)。

岩清^(水)一(四四)

延喜式云イハシミツ^(上上上平平濁)

法(四五)

中云方乏^(入聲)反。常也令也。玉云則也數也。慈曰^(送)一者^(送)一則、放學為義。又云軌則也又楷式也。又^(送)一者^(送)道理義・唯識論云^(送)一謂軌持ノリ^(上平)律ノトル^(上平上)記コトハリ^(平平平平)智論^(以下補入)古點

浩灑(四六)

類云二正

沍(四六)

金類云二古

沍(四六)

類云俗法今

法界(四六)

慈云三乘妙法所依相故名為法界々々者體義但妙法依。

16 掲出字の周圍に、朱書による片仮名の字音注が見えるが、「朱書」である旨は特に注記しない。

17 出典名の頭に付された点発は「・」(中患)で表す。

◎ 翻字本文

水 (四二)

弘云 戸癸反・中云所以潤万物也・玉云五行曰北方行也・鮎曰江淮濟河泉皆曰一也・井泉曰一也・月一俗云佐波利(上) 類聚抄選垂類

〔注〕

- (1) 弘法大師空海(七七四〜八三三)撰『篆隸万象名義』に「戸癸反」(第帖九オ、第五帖八〇ウ)とある。
『篆隸万象名義』六帖は、高山寺本が現存唯一本。『類聚名義抄』の「名義」は『篆隸万象名義』に基つくとも考えられ、最も重要な出典の一つ。吉田(一九五四)をはじめとして数多くの研究がある。名義抄はこの書を全載の方針ではなく常用性のある漢字を採録(巻二九八七、池田一九三三)、採録の順位は慈恩に次いで第二位であること(池田一九九二)などが明らかになった。引用箇所の一覧は宮澤(一九七三)にある。高山寺本の影印は各種あるが、高山寺資料叢書所収のものが索引・掲出字一覧表を備えており至便。出典名(類度)は、「弘」(2)、「弘」(弘)(1)、「弘」(中掲)(1)、「弘回」(1)、「又弘」(2)、「又弘」(1)。『篆隸万象名義』の直接引用は五二七条。
- (2) 中算撰『妙法蓮華經釈文』に「水 式軌反 郭知玄云、所以潤万物也、今案々々香者、本草云、沈香堅而沈水者也、兼名苑云、々々一名陸里也」(卷三三オ3)とある。
「中」として引く書は、興福寺の学僧松室仲算(九三五〜九七六)撰の『法華經釈文』三巻。醍醐寺本の影印が古辞書意義集成に所収。『法華經釈文』については、吉田(一九五三)、馬淵(一九七二)、宮澤(一九七五)、西原(一九九二)、西原他(一九九二)、西原・河野(一九九三)など参照。図書寮本との関係については、吉田(一九五四b)、宮澤(一九七七)を参照。中算『法華經釈文』はほぼ全載、主要出典中の採録の順位は慈恩・弘法・玄心に次いで第四位であることが明らかになっている。出典名(類度)は、「中」(2)、「中引」(1)、「中引茲」(1)。「中」の直接引用は三六〇条。吉田(一九五四b)は「約四百五十條」とするが教え違いだらう。
- (3) 顧野王撰『玉篇』三十巻。残巻、逸文なし。『大広益玄玉篇』に「水 戸癸切、流津也、禮記曰、水曰清澗、儀禮所謂玄酒也」(中七〇ウ)とある。
- (4) 『釋勅切韻』五巻。釋勅は『東宮切韻』に見えないため直接引用。吉田(一九五)を参照。出典名(類度)は、「勅」(3)、「又勅」(2)、「五」(19)、「見勅」(1)、「將一切上麻韵」(1)、「五將勅武玄之」(1)。直接引用一六一条。