

# 漢字字体と典籍の性格との関係 —「漢字字体規範データベース」が主張するもの—

高田 智和<sup>†</sup>

「漢字字体規範データベース」の成り立ちと現状を紹介する。また、「漢字字体規範データベース」の実現を通して構築者グループが主張する、漢字字体と典籍の性格との関係についても述べる。

## Correlation between the Hanzi Glyph Standard and the Codicological Features of Classical Materials

TOMOKAZU TAKADA<sup>†</sup>

In this paper, I describe the development of Hanzi Normative Glyphs Database since 1998. And I discuss the correlation between the Hanzi Glyph Standard and the codicological features of classical materials.

### 1. はじめに

「漢字字体規範データベース」(Hanzi Normative Glyphs, 略称 HNG, [http:// http://joao-roiz.jp/HNG11/](http://http://joao-roiz.jp/HNG11/)) は、各時代・各地域(国)には漢字字体の標準が存在し、そして、漢字字体の標準が各時代・各地域(国)により変遷することを明らかにし、それを明示することを目的に構築された。

2004年度から、東京外国語大学アジア・アフリカ言語文化研究所の豊島正之教授の管理のもとで公開を始めた。公開当初は、漢籍・仏典・国書等の典籍16点の用例を収録していたが、その後公開典籍数を増やし、2009年度末の時点では、典籍64点の用例を公開するに至っている。

「漢字字体規範データベース」の前身は、北海道大学の石塚晴通名誉教授が、およそ30年前から作成を続けてきた字体資料(「石塚漢字字体資料」と呼ぶ)である。「石塚漢字字体資料」は、紙カードで整理され、2005年9月の時点で、典籍67点、総用例数40万字が保存されていた。紙カードの電子化に着手したのは10年ほど前からである。

一方、比較的近年に「漢字字体規範データベース」に収録された典籍は、「石塚漢字字体資料」由来ではなく、典籍の原本(書写本や印刷本)、あるいは、複製本の撮影画像から直接用例を収集したものである。2004年度から現在までの構築期間において、コンテンツの作り方も変わってきている。

本発表では、「漢字字体規範データベース」の成り立ちと現状を紹介するとともに、「漢字字体規範データベース」の実現を通して構築者グループが主張する、典籍の書誌的な性格と、典籍に現れる漢字字体には関連があるという点についても述べる。

### 2. 「漢字字体規範データベース」の目的と方法

「漢字字体規範データベース」の目的は、冒頭に述べたように、下記の2点を明らかにすることであり、単純明快である。

1. 各時代・各地域(国)には、漢字字体の標準が存在する。
2. 漢字字体の標準は、各時代・各地域(国)により変遷する。

また、これらを明らかにするために用いた方法は、石塚晴通ほか(2005)から抜粋すると、次のとおりである。

- 各時代・各地域の漢籍・仏典・国書の典籍として標準的な文献(楷書体)を選定し、
- 文献の選定に当たっては、漢字字体の標準の存在と性格を明らかにする為に、公的文献の外に私的文献も混じえてある。

ここでは、文献に現れる漢字字体が標準的なものであるか否かは、その文献の性格と大きく関わりがあると述べている。字体と文献の性格に関連があることは、書誌学研究者にとっては自明のことなのであるが、「漢字字体規範データベース」の目的と方法を説明する上で、最も理解されにくい点の一つである。

「漢字字体規範データベース」は、漢字字体史研究のためのデータベース(資料)である。構築者グループにとっては、漢字字体史を記述するための、いわば資料編にあたるものが、「漢字字体規範データベース」なのである。名筆から字形を収集する書道辞典や、各種資料から「異体字」を収集する字体辞典とは、制作目的が異なっている。「電子

<sup>†</sup> 国立国語研究所  
National Institute for Japanese Language and Linguistics

くずし字字典データベース」「木簡字典」「拓本文字データベース」とは、集字対象が異なるだけでなく、制作目的も異なる。

漢字字体史研究のためのデータベース（資料）として、「漢字字体規範データベース」が実現すべきことは、次の3点であると考えられる。

1. 時代ごと、地域ごとに、漢字字体の標準が存在することを明示する。
2. 時代ごと、地域ごとに存在する漢字字体の標準が変わることを明示する。
3. 漢字字体の標準が、文献の性格や属性に関係していることを明示する。

これらが「漢字字体規範データベース」でどのように実現されているか、現状と課題については後述する。

### 3. 紙カードからデータベースへ

#### 3.1 「石塚漢字字体資料」の電子化

2005年まで、北海道大学大学院文学研究科石塚晴通研究室では、過去20年以上にわたって、漢字文化圏の各時代の標準的な文献（漢籍・仏典・国書等の典籍）を選定して、漢字字体の調査を行い、紙カードの状態に蓄積していた（図1）。奈良時代から江戸時代初期にかけての日本資料、南北朝時代から南宋までの中国資料、および新羅・高麗時代の韓国資料から成り、全67点余、総用例数40万字をこえる、手作業による大規模な調査資料であった。

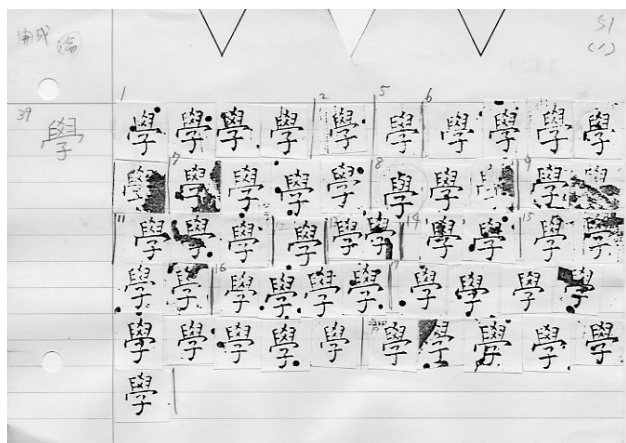


図1 紙カード

紙カードであるため、経年劣化が著しいものもあり、1998年ころから紙カードの保存を兼ねて、電子化を行うことになった。電子化作業の内容は、紙カードそのものの画像化と、紙カード情報のテキスト化の大きく2点である。

図2は、開成石経『論語』（開成2（837）年刻、東洋文

庫蔵拓本）の「學」の用例を整理した紙カードである。紙カードには、資料に出現した文字を貼りこみ、見出し字、部首番号、用例数を記述してある。紙カードの見出し字には、『大字典』（上田万年ほか編、1917年初版）の掲出字（康熙字典体）を書くのが原則である。

「石塚漢字字体資料」は、個々の資料内部での漢字字体のゆれを把握することを主眼として設計されている。したがって、図1のように、「學・學」と複数字体が出現している。一枚の紙カードで整理している。つまり、それぞれの紙カードは、複数字体をまとめた字種を表していると考えられる。

紙カード情報の電子化では、紙カードの整理番号、見出し字、部首番号、用例数を記述する（表1）。原本の用例の字体を再現して記述することは、当初から放棄している。見出し字の記述はUnicodeを用いるが、現在のUnicodeを用いたとしても、古典籍の書写本や版本の漢字字体を、すべて再現することは難しい。

表1 紙カードの見出し字の電子化

紙カード		電子化
原本の用例	見出し字	
學	學	學 (U+5B78)
學		

#### 3.2 紙カードの電子化終了以降

「石塚漢字字体資料」の電子化が終了してからは、原本や複製本の撮影画像から用例を収集し、データベースに登録している。カードを作成することはせず、原本の翻字テキストから字種を抽出している。

紙カードの見出し字が康熙字典体（『大字典』の見出し字字体）であったことから、これと整合性をとるために、原本の翻字テキストの作成においても、康熙字典体を用いるようにしている。以下は、図2の『金剛頂大教王経』（平安時代末期写、国立国語研究所蔵本）の翻字テキストである。「𠄎」は「就」、「未」は「來」、「𠄎」は「剛」のように翻字している。

如是我聞一時婆伽梵成就一切如來金剛  
 加持殊勝三昧耶智得一切如來寶冠三界  
 法王灌頂證一切如來一切智智瑜伽自在  
 能作一切如來一切印平等種種事業於無

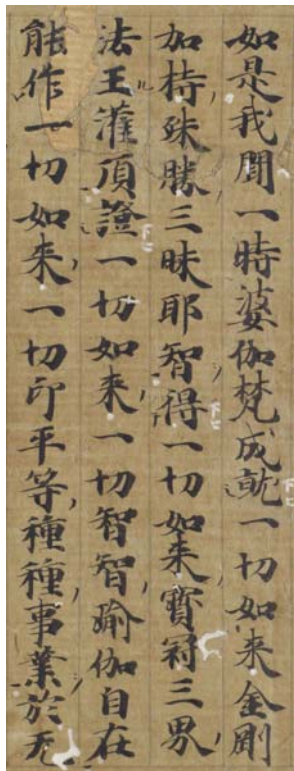


図 2 国語研本『金剛頂大教王経』

#### 4. 「漢字字体規範データベース」の現状と課題

漢字字体史研究のためのデータベース（資料）として、「漢字字体規範データベース」が実現すべきことは、3点であると前に述べた。再掲する。

1. 時代ごと、地域ごとに、漢字字体の標準が存在することを明示する。
2. 時代ごと、地域ごとに存在する漢字字体の標準が変わることを明示する。
3. 漢字字体の標準が、文献の性格や属性に関係していることを明示する。

1.と 2.については、検索結果の表示によって実現しようとしている（図3）。検索結果の表示画面では、第1段目に中国写本、第2段目に中国版本、第3段目に日本写本・版本、第4段目に韓国写本・版本、第5段目に中国周辺写本・版本を配置している。段の中は年代順である。なお、第6段目は比較のための字書用例である。

現在の公開文献は約100点であるため、図3の「所」のように多数の文献に共通して出現する字種であれば、一覧させた時の可読性が低下する。また、一つの文献に複数字体が見れる場合は、図3で示すように（四角囲みの部分）、字形を並べて配置している。

今後、年表表示や、字体を分類して表示するなどの工夫も必要であると考えている。

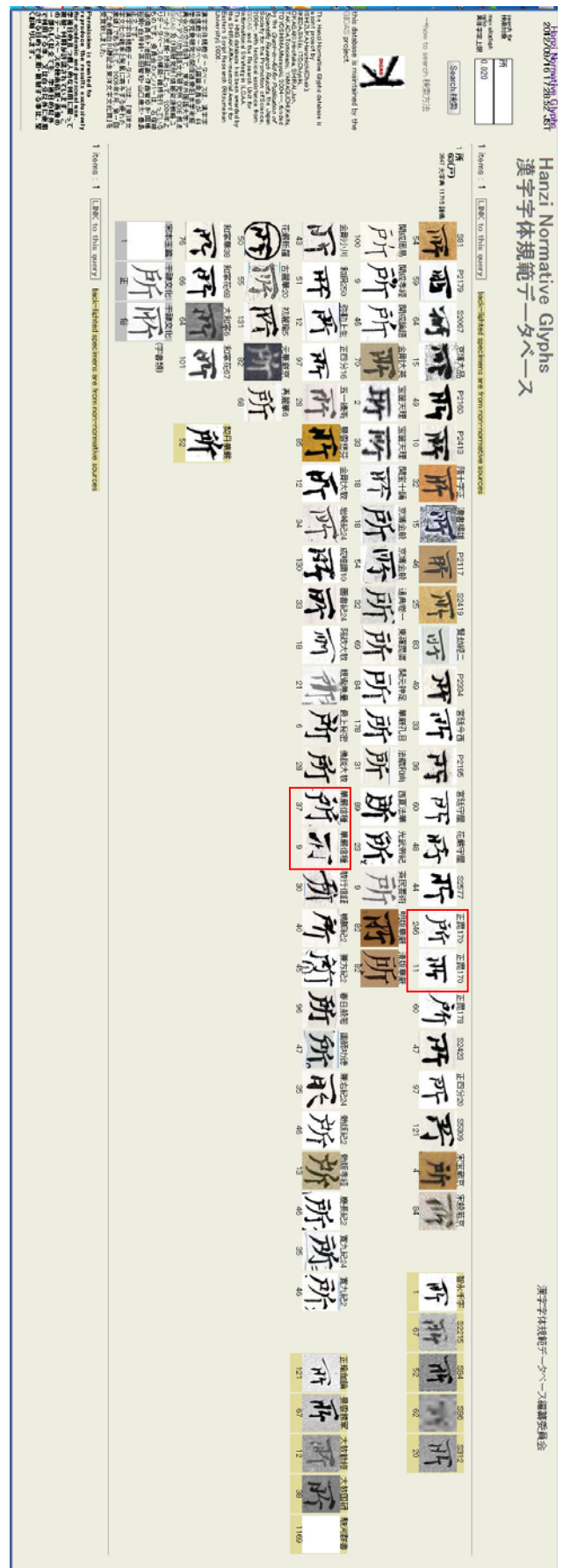


図 3 「漢字字体規範データベース」検索結果画面

3.については、異体率が高い非標準的文献と、異体率の低い標準文献との違いを、左右の配置分けで表現している。左側が標準文献の用例、右側が非標準文献の用例である。

ここで、異体率について説明しておく。異体率は、漢字を複数回書く時に字体がゆれる割合を示し、次の式で算定する。なお、ここで言う「異体」は、字体のゆれのある字種のうち、度数が少ない方の字体を指しての、便宜上の呼称である。字体規範の正俗などとは異なる概念であるので、注意されたい。

$$\text{異体率} = \frac{\text{異体の総字数}}{\text{文献の総字数} - \text{孤例の総字数}} \times 100$$

例えば、開成石経『論語』の場合、異体率は0.03%と算定され、字体の統制とれている文献であることがわかる。異体率の低さは、規範の規範に対する強い意識を持って文献が作られたことを表し、開成石経『論語』が唐代の『論語』の標準テキストであり、科挙のテキストとして用いられたことを考え合わせれば、文献の性格と漢字字体とに相関があるとする言にも、合理性が見いだせよう。

文献が標準的あるか、非標準的であるかは、第一に、料紙、寸法など本の形態的な特徴が、時代・地域の典型的なものであるか否か (typical/untypical) によって区分される。このほかに、文献が公的なものであるか、私的なものであるか (public/private) も、漢字字体に関係していると考えられる。

例えば、同じ初唐の法華経であっても、宮廷写経は異体率が低く (1.00%以下)、S.2577『妙法蓮華経』巻第八 (7世紀末写、大英図書館蔵本) は2.03%である。宮廷写経は公的写本であり、一方、S.2577は私的な学習本である。両写本の性格の違いが、異体字率の違いに反映していると考えられる。

現在の「漢字字体規範データベース」は、図4のように、典籍名 (所蔵先)、年代、地域、写刊の別を、極めて簡潔に示すのみである。

略称	公開年度	地域	種別	分類	年紀	文献
1 S81	2005	C	M	敦煌南北朝	天監五年 (506)	S.81大般涅槃經卷十一 (BL)
2 P2179	2005	C	M	敦煌南北朝	延昌三年 (514)	P.2179說實論卷八 (BNF)
3 S2067	2005	C	M	敦煌南北朝	延昌三年 (514)	S.2067華嚴經卷十六 (BL)
4 S767	2005	C	M	敦煌南北朝	S.767大般涅槃經卷二十三 (BL)	
5 P2160	2005	C	M	敦煌南北朝	至德四年 (586)	P.2160摩訶摩訶經卷上 (BNF)
6 P2413	2005	C	M	隋	開皇九年 (589)	P.2413大般涅槃經卷三 (隋經) (BNF)
7 京博大品	2005	C	M	高昌期		大品經卷二十八 (高昌) (京都国立博物館)
8 智永千字	2011	C	M	隋		千字文 (伝智永筆) (個人蔵)
9 隋十字正	2010	C	M	隋		十字經論卷一 (隋經) (正倉院)
10 P2117	2010	C	M	隋	大業四年 (608)	P.2117大般涅槃經卷三十三 (隋經) (BNF)
11 S2419	2010	C	M	隋	608)	S.2419妙法蓮華經卷三 (隋經) (BL)
12 寶劫經一五	2012	C	M	隋	610)	寶劫經卷一、卷五 (正倉院・隋經) (正倉院)
13 寶劫經二	2005	C	M	隋	610)	寶劫經卷二 (正倉院・隋經) (正倉院)
14 P2334	2005	C	M	隋	大業十三年 (617)	P.2334妙法蓮華經卷五 (隋經) (BNF)
15 S2215	2011	C	M	初唐	咸亨元年 (670)	S.2215法華經卷二 (BL)
16 S84	2011	C	M	初唐宮廷写経	咸亨二年 (671)	S.84法華經卷五 (BL)
17 宮廷今西	2004	C	M	初唐宮廷写経	咸亨二年 (671)	妙法蓮華經卷五 (今西本) (京都国立博物館)
18 S86	2011	C	M	初唐宮廷写経	咸亨三年 (672)	S.86金剛經 (BL)

図4 「漢字字体規範データベース」書誌情報画面

3.を実現するためには、より詳細な書誌情報が提示されなければならない。現状に加えて、下記のような項目を記

述する必要がある。現状の書誌記述では、典籍からの単なる集字データベースと大差ない。書誌情報の充実は、直近の最重要課題である。

- 装丁
- 料紙
- 寸法
- 紙数
- 1紙中の行数
- 界の種類
- 1行中の文字数
- 訓点・書込の有無
- 奥書・識語・刊記
- 公的文献／私的文献の別

## 5. おわりに

「漢字字体規範データベース」は、構築者グループが漢字字体史を記述するために作成した資料である。構築者グループと同様に、「漢字字体規範データベース」を漢字字体研究に利用する事例がある一方で、文献解読のために、字を調べるために「漢字字体規範データベース」を利用することも多いようである。

文献を解読するために漢字の字体を調べるといったニーズに対しては、典籍を対象とした「漢字字体規範データベース」だけでなく、良質な一次資料を収録した「電子くずし字字典データベース」「木簡字典」「拓本文字データベース」と、横断検索によって相互利用ができるようになると、利用者に益することもあるだろう。今後は、漢字字体史研究のための「漢字字体規範データベース」の整備を進めていくとともに、一次資料を対象とした文字画像データベース間の連携についても考えていきたい。

**付記** 本研究は、平成22年度～平成25年度日本学術振興会科学研究費補助金基盤研究(B)「漢字字体変容の原理—敦煌文献から現代日本戸籍漢字まで—」(研究代表者:高田智和、課題番号:22320087)による成果の一部である。

## 参考文献

- 1) 石塚晴通: 漢字字体の日本的標準, 国語と国文学, 76(6), pp.88-96 (1999).
- 2) 石塚晴通・豊島正之・池田証寿・白井純・高田智和・山口慶太: 漢字字体規範データベース, 日本語の研究, 1(4), pp.94-104 (2005).
- 3) 高田智和: 漢字処理と『大字典』, 訓点語と訓点資料, 109, pp.99-107, (2002).
- 4) 武内義雄: 論語之研究, 岩波書店, (1939).