電子化テキストの国際的共有

篆隷万象名義データベースについて

池田証寿

キーワード：篆隷万象名義、JIS漢字、ISO/IEC 10646-1、Unicode、統合漢字

要　旨

このデータベースは、篆隷万象名義の本文研究を主要な目的とするものであり、篆隷万象名義に掲出される一万六千余字の所在、諸橋「大漢和辞典」番号、玉篇巻数・部首番号、JIS区点番号等の情報を得ることができる。

作成上問題となるのは、第一にJISに無い字の扱い方。JISに無い字は篆隷万象名義の掲出字の七割、ISO/IEC 10646-1でさえ三割に及ぶ。第二に掲出漢字の同定。これは、篆隷万象名義の本文研究に役立つようなデータの扱い方が必要である。

漢字字書データベースとして見た篆隷万象名義データベースの特色は、漢字字書の採録字やその配列の規準を解明するのに有効であることと、JISやISO/IEC 10646-1等の漢字コード表に採用された漢字の典拠を示すことである。

1　はじめに

篆隷万象名義[注1]は、原本玉篇[注2]を抄録した漢字字書であり、弘法大師空海の撰述に係る。ただし、全六帖の前半四帖が空海自撰、後半二帖は別人の続撰である。高山寺本が唯一の古写本で、国宝に指定されている。本書の価値は、大部分散逸した原本玉篇（梁・顧野王撰）の内容を伝えている点にある。言うまでもなく、原本玉篇は、説文解字（後漢・許慎撰）、康熙字典（清・張玉書外編）等と並んで中国を代表する漢字字書の一つであり、日本の古代中世の文芸・学問に甚大な影響を与えた字書である。[注3]

現在、筆者は、この篆隷万象名義の掲出漢字のデータベースを作成している。これは、篆隷万象名義の本文研究や、玉篇ないし篆隷万象名義を土台として編纂された日本の古字書の研究に応用することを主な目的とするものである。本稿では、その概要、作成上の問題点等を述べる。

2　作成の目的

篆隷万象名義の本文研究の成果は、次の三つの業績に集約される。

白藤禮幸「掲出字索引」（「高山寺古辞書資料第一」東京大学出版会、1977年）

宮澤俊雅「掲出字一覧表」(同上)

上田　正「玉篇反切総覧」(著者刊、1986年)

上田正「新撰字鏡の切韻部分について」(国語学127集、1981年)は、白藤「索引」と宮澤「一覧表」について「これは用意の行き届いた誤のほとんどない秀れた労作である」と評している。宮澤「一覧表」は、篆隷万象名義、原本玉篇、宋本玉篇の対校作業を篆隷万象名義の側から調査したものである。上田「総覧」は、原本玉篇と篆隷万象名義に収集した逸文を加えて、玉篇のすべての反切をまとめた書である。

三氏の研究は、信頼度が極めて高い。これらの成果を踏まえて篆隷万象名義のデータベースを作成すれば、篆隷万象名義の研究に役立つばかりでなく、篆隷万象名義や玉篇を利用して成立した日本の古字書の研究に活用することができるであろう。このような考えから、白藤「索引」と宮澤「一覧表」とに基づき、篆隷万象名義をデータベース化する作業に取り掛かった。しかし、実際にその作業を行ってみると、両者の解読に違いがあることに気が付いた。むろん例はさほど多くない。違いがあるといっても異体字の関係にある場合が大半である。白藤「索引」の掲出漢字の同定は篆隷万象名義の原本に近く、宮澤「一覧表」のそれは宋本玉篇に近い、という傾向。ただ両者に異同があるということは、篆隷万象名義の本文の解読に未解決の点があるという見方ができる。両者ともデータとして採用しておけば、篆隷万象名義の本文を研究する材料がかなり得られる。また、玉篇ないし篆隷万象名義を利用した字書や玉篇の逸文と比較すれば、依拠した玉篇の系統についても何等かの手がかりが得られるであろう。

篆隷万象名義や玉篇の研究と、それを利用することで成立した古字書の研究とは、密接な関係にあり、これを切り離すことはできない。この二点を目的として篆隷万象名義データベースを作成することにした。

3　作成の経過

3.1　図書寮本類聚名義抄対応版の作成

このデータベースは元来、図書寮本類聚名義抄と篆隷万象名義との関係を調査するために作成を開始したものである。作成の範囲は、図書寮本に対応する篆隷万象名義の十三の部首(玉土邑足心言水山石阜糸巾衣)である。その作成の手順は、図書寮本名義抄掲出字データから篆隷万象名義に対応する字を抜き出し、図書寮本に採録されていない字を篆隷万象名義から追加するというものである。そのため、データの形式に不統一の部分を多く含んでおり、その統一に時間をとられた。新規に篆隷万象名義のデータを作り、これを図書寮本のデータと照合した方がよかったと反省している。

3.2　篆隷万象名義データベースの作成

図書寮本類聚名義抄対応版作成の反省を踏まえて、次の手順で作成中である。

(1) 諸橋編「大漢和辞典」の検字表のコピーに白藤「索引」を転記。

(2) JIS漢字(第一水準・第二水準・補助漢字)分を入力。

(3) 非 JIS 漢字分を入力。

(4) 宮澤「一覧表」と照合。

(5) 白藤「索引」と照合。

(1)の方法を採ったのは、類聚名義抄などの所在データの追加を考慮したためである。JIS 第一水準と第二水準（JIS X 0208-1990）[注4]は、豊島正之・金水敏・古田啓作成の計算機用漢字字書（Ydic）[注5]を利用し、JIS 補助漢字（JIS X 0212-1990）[注6]は、筆者作成のファイルを利用した。現在は、(4)の段階。覆製本との照合は適宜行っているが、最終的に全体を通しての照合が必要なのは勿論である。

4 データの形式

図書寮本類聚名義抄対応版は、次の六つのフィールド（欄）からなる。

(1) 大漢和辞典番号

(2) 玉篇巻数と部首番号

(3) 掲出字

(4) 図書寮本類聚名義抄との対応

(5) 篆隷万象名義の所在

(6) JIS 区点番号

内容は、次のようになっている。

M2722100, 27425, 糸,g, 6/126-51, 0-2769

M2722100, 27425, 糸,z, 6 = 126-61, 0-2769

M2722101, 27425, 糸 λ,z, 6/126-61, 4-0000

M2744800, 27425, 絲,z, 6@126-61, 0-6915

M2794400, 27425, 繭,z, 6/126-62, 0-4390

M2746600, 27425, 見 χ,z, 6/127＋11, 1-5178

M2784700, 27425, 惨 ω,z, 6 = 127＋12, 1-5246

(以下略)

4.1 大漢和辞典番号

大漢和辞典番号は、上 5 桁に諸橋轍次編「大漢和辞典」(大修館書店、1955-60 年初版）の番号を、下 2 桁に付加情報を示す。大漢和番号に ’ を付すものは下 2 桁を 10 に、大漢和に無い字は、上 5 桁に該当する部首・画数の最後の字の番号を採り、下 2 桁を 90、91、…とする（豊島・金水・古田作成の計算機用漢字辞書 KLEX の仕様。豊島が 1982 年に決定した方法の由)。ただし、所属部首や画数が不明だが、注文から某字の古文などであることが明白な例は下 2 桁を 01、02、…とする。

4.2 玉篇巻数と部首番号

玉篇巻数と部首番号は、巻第二十七糸部第四百二十五を 27425 のように示す。玉篇の巻

数と部首番号は宮澤「一覧表」による。目録部分の巻数は 00 とする。

4.3 掲出字

掲出字の字体は、白藤「索引」および宮澤「一覧表」の同定に従う。両者異なる場合は、原本に近い字体を採ることにする。上田「総覧」も適宜参照する。

掲出字のうち JIS 漢字に無いものは以下の方法で示す。

M2746600, 見 χ　糸偏に見
M3545200, 洲 ω　言偏に州
M3542500, 察 λ　察の古字（通用・譌字等）
M0487300, 経 μ　経の旁の部分の字
M2273500, 百 2 百を二つ並べた字
M1082600, 心 3 蕊の草冠の無い字
M2806300, ▼　以上で駄目な時

この方法は、かつて岡田希雄が「類聚名義抄の研究」（一條書房、1944 年）で難字を×（糸偏に見）のように示したことや、豊島正之「「JIS に無い字」をめぐって」（しにか 3-2、大修館書店、1992 年）に示された考え方を参考にしたものである。

4.4 図書寮本類聚名義抄との対応

篆隷万象名義データベース完成版ではこれを削除する予定なので、説明を省く。

4.5 篆隷万象名義の所在

篆隷万象名義の所在は、三帖七丁表六行一字目を 3/07＋61 のように、三帖七丁裏六行二字目を 3/07-62 のように示す。

また、/を !、#、@、｜、= に替えて、以下の情報を示す。

! 万象名義に見えないが本来玉篇に存したと考えられる字
万象名義に見えないが宋本に存する字
@ 白藤「索引」における掲出漢字の同定を参考として記す場合
｜ 宮澤「一覧表」における掲出漢字の同定を参考として記す場合
= 万象名義の注文中に見える字体注記（白藤「索引」にある例のみ）

4.6 JIS 区点番号

JIS の区点番号を以下の要領で示す。

0-xxxx　JIS 第一水準・第二水準（JIS X 0208-1990）
1-xxxx　JIS 補助漢字（JIS X 0212-1990）
4-0000　JIS に無い漢字

JIS 第一水準・第二水準は、豊島・金水・古田作成の計算機用漢字字書による。

3-xxxx は、JIS 漢字（JIS X 0208-1990 と JIS X 0212-1990）に無く、ISO/IEC 10646-1 ないし Unicode の統合漢字にある漢字のためとする。統合漢字は以下の規格書による。

The Unicode Consortium. The Unicode Standard: Worldwide Character Encoding Version 1.0, Volume 1-2. Addison-Wesley.1991-92.

ISO/IEC 10646-1:1993 Information technology - Universal Multiple Octet Coded Character Set (UCS) Part 1: Architechture and Basic Multilingual Plane. 1993.

ISO/IEC 10646-1 の BMP (Basic Multilingaual Plane) は Unicode を採用している。

5 作成上の諸問題

5.1 JIS 漢字の問題

特に問題になるのは、やはり JIS に無い漢字の処理。これについては、いろいろな考えがあるが(豊島正之「「JIS に無い字」をめぐって」前掲。當山日出夫「コンピュータであつかえない漢字——「和漢朗詠集」の場合——」汲古 12、汲古書院、1987 年等)、前述の方法によった。万全ではないが、大漢和辞典をコードブックとして利用する方法はよく行われているし(田嶋一夫「漢字シソーラスの構想と課題」日本語学 3-3、明治書院、1984 年。齋藤秀紀「漢字コードの拡張法に対する試案」国立国語研究所報告 83 研究報告 6、1985 年等)、部首情報を付記すればある程度の実用性は確保できよう。

篆隷万象名義第一〜四帖(目録部分を含む)の範囲の掲出漢字(/を付した字のみ)について、JIS と ISO/IEC 10646-1[注7] の漢字でどれだけ処理できるかを調査した結果を以下に示そう。また、参考として、史記(李暁光・李光編「史記索引」北京、中国広播電視出版社、1989 年)について調査したデータも示す[注8]。

規格	篆隷万象名義	史記
JIS X 0208-1990	2899/8861 (32.7%)	3543/4883 (72.5%)
JIS X 0212-1990	1881/8861 (21.2%)	957/4883 (19.6%)
ISO/IEC 10646-1	6121/8861 (69.1%)	4750/4883 (97.3%)

例えば、規格 JIS X 0208-1990、篆隷万象名義の欄の 2899/8861 は篆隷万象名義第一〜四帖の掲出字 8,861 字のうち、2,899 字が JIS X 0208-1990 で処理できることを表す。32.7% はその比率である。これによって、JIS の第一・第二水準では篆隷万象名義データベースの漢字の三割程度しか処理できないことが分る。

ISO/IEC 10646-1 ではどうであろうか。これは、上に示したように篆隷万象名義第一〜四帖の掲出字 8,861 字中、6,121 字が ISO/IEC 10646-1 で処理できるという結果であった。ISO/IEC 10646-1 の漢字約 20,000 字に、JIS 漢字 (JIS X 0208/0212) はすべて含まれるから、ISO/IEC 10646-1 で篆隷万象名義の掲出字約 16,000 字の七割近くが処理できる。つまり、JIS 漢字 (JIS X 0208) に無い字は、篆隷万象名義の掲出字の七割、ISO/IEC 10646-1 でさえその三割に及ぶのである。

史記の例に見るように、通常の文献であれば、外字での対応も現実的である。しかし、万を越える所収文字数の漢字字書のデータベースの場合、外字での対応は困難である。

古字書のデータベースの試みに田島毓堂「類聚名義抄の注釈的研究－電算機利用による－」(平成元年度科学研究費報告書、1990 年) がある。この研究で JIS 漢字 (JIS X 0208) に無い漢字は「現段階は検討段階にあるので、とりあえず「○」をいれておく」とする。処

理の困難さがうかがえる。データの形式の詳細は不明であるが、大漢和辞典番号が利用できる形式で公開されれば、より有用なデータとなろう。

5.2 掲出漢字の同定

篆隷万象名義第一～四帖の範囲について調べたところ、白藤「索引」と宮澤「一覧表」とで掲出漢字の同定が異なる例は、400 箇所以上あった。篆隷万象名義全体ではおそらく 800 箇所以上になるであろう。次に一例を挙げる（難字は ABCD の符号で示す)。

M0069300, 03023, ▼, z, 1/055＋31, 4-0000　(B)
M0088500, 03023, ▼, z, 1@055＋31, 4-0000　(D)
篆隷万象名義　A　(I 55 オ 3)
宋本玉篇　B　(上 26 オ 10)
新撰字鏡　C　(82 頁 3 行)　玉篇引用部分。
白藤「索引」[B]　A 1-55.3 (375 頁 4 段目)
　D　1-55.3　(376-1)
白藤「校勘篇」A－B　F　(487-5)　F は「憶説（白藤）の考へ」の意。
宮澤「一覧表」B　(507-1)
上田「総覧」D

(注) A＝人偏に巻の己を、二つとする字。B＝人偏に火の下に廾を書く字。
C＝人偏に巻の己が無い字。D＝人偏に巻の己を女に作る字。

篆隷万象名義の字体は A と見たが、影印が不鮮明で C のようにも見える。崇文叢書は明らかに C である。C が本来の字体であれば、上記のデータは次のようになる。

M0061900, 03023, ▼, z, 1/055＋31, 4-0000　(C)
M0069300, 03023, ▼, z, 1@055＋31, 4-0000　(B) 白藤「索引」
M0088500, 03023, ▼, z, 1@055＋31, 4-0000　(D) 白藤「索引」
M0069300, 03023, ▼, z, 1｜055＋31, 4-0000　(B) 宮澤「一覧表」

こうした作業を積み重ねることで、より正確な解読を目指しているのである。

6 古字書研究への応用――観智院本類聚名義抄の「玉篇字順群」を例として――

観智院本と玉篇との関係を実証した研究に、貞苅伊徳「観智院本類聚名義抄の形成に関する考察　その 1　字順をめぐる問題」(第 48 回訓点語学会口頭発表、1983 年 5 月 20 日)があり、玉篇ないし篆隷万象名義と同じ字順の「玉篇字順群」の存在が指摘されている。これを以下確認して見よう。例は、貞苅のあげた巾部とする。観智院本類聚名義抄(風間書房)、図書寮本類聚名義抄 (宮内庁書陵部、1950 年) の所在は、覆製本のページ数等を示す (104.13 は 104 頁 1 行 3 段目の意。さらに 1 桁付けて順序を調整したところがある)。

大漢和	漢字	玉篇	万象名義	観本（法中）	図本	JIS	備考
			(前略)				
M0905300	票 χ	28432	6/145-61	106.54	281.74	1-2822	

M0899700 刺χ	28432	6/146+22	106.64		4-0000
M0908100 無χ	28432	6/146+52	106.71		3-5E60
M0877700 匕χ	28432	6/145-32	106.74		4-0000
M0903800 祭χ	28432	6/145-41	106.82		4-0000
M0896900 兪χ	28432	6/145-42	106.83		4-0000
M0914900 孅ω	28432	6/146+31	107.11		4-0000
M0902000 蒙ω	28432	6/146+41	107.14		4-0000
M0902600 家χ	28432	6/146-61	107.22		3-5E4F
M0899300 幃	28432	6/146+62	107.23	282.73	0-5475
M0905400 婁χ	28432	6/147-31	107.31		4-0000
M0888100 券ω	28432	6/146-11	107.33		3-5E23
M0914100 奮χ	28432	6/146-31	107.34		4-0000
M0898500 盾χ	28432	6/146-32	107.41		4-0000
M0880100 及χ	28432	6/146-41	107.43		4-0000
M0910900 賁χ	28432	6/146-42	107.51		3-5E69
M0915100 ▼	28432	6!146-421	107.52		4-0000 !
M0894200 弦χ	28432	6/146-62	107.53		4-0000
M0910000 壁ω	28432	6/147+12	107.54		3-5E66
M0899400 幄	28432	6/147+21	107.61	280.71	0-5474
M0886700 百χ	28432	6/147+31	107.62	284.41	3-5E1E
M0883700 粲ω	28432	6#147+311	107.63		4-0000 #
M0910200 操ω	28432	6｜147+32	107.71		3-5E67
M0897800 愁ω	28432	6/147+41	107.72		4-0000
M0892400 峽ω	28432	6/147+52	107.731	285.73	4-0000
M0888000 合χ	28432	6/147+61	107.732		3-5E22
M0894500 諂ω	28432	6/147+51	107.74		4-0000
M0893100 卑χ	28432	6/147-11	107.81		4-0000
M0905700 幗	28432	6/147-12	107.82	283.23	0-5478
M0907700 喬χ	28432	6/147-32	107.84		4-0000
M0891300 貌ω	28432	6/147-21	108.12		4-0000
M0888700 多χ	28432	6/147-22	108.14		4-0000
M0911000 著χ	28432	6/147-41	108.21		4-0000
M0891500 佇ω	28432	6#147-401	108.22		4-0000 #

(以下省略)

紙幅の都合で引用は巾部の一部を示すにとどめたが、この部分をとってみても、篆隷万象名義に無く宋本玉篇にある字（!や#を付す）の存在が理解できよう。観智院本における宋本系の玉篇の影響は確実である。[注9]

日本撰述の古字書のデータベース化は種々試みられなければならない。新撰字鏡や類聚名義抄を直接データベース化するのも一つの方法であるが、その土台となった玉篇や切韻(広韻)のデータベースを作成した上で、日本の古字書のデータベースを作成するのが、より堅実な方法と考える。漢字音の研究では、切韻(広韻)を規準にして、それとのズレを見ていくのが常道である。古字書(特に部首分類体の漢字字書)の研究では、玉篇を規準にして、その採録範囲や配列の順序のズレを検討していくのが有効であろう。

7 おわりに

この研究は、篆隷万象名義の本文研究を主要な目的とするものであるが、電子化テキストの国際的共有という点を勘案して、その特色を挙げれば次の二点になろう。

(1) 漢字字書の採録字やその配列の規準を解明するのに有効である点

(2) JIS や ISO/IEC 10646-1 等の漢字コード表に採用された漢字の典拠を示す点

特色の(1)漢字字書の採録字やその配列の規準を解明するのに有効である点は、従来の漢字字書データベースに無かった視点であると思う。

特色の(2)漢字コード表に採用された漢字の典拠を示す点は、ある漢字が正しいかどうかを判断する際の規準に関係してくる。一般に、ことばの正しさを判断する規準として、伝統性、一般性、合理性等が指摘されている。[注10] 篆隷万象名義データベースは、JIS や ISO/IEC 10646-1 等の文字表に含まれる漢字について、特に、その「伝統性」を検証する上で大きな力となるのである。

注1 高山寺典籍文書綜合調査団編「高山寺古辞書資料第一」(東京大学出版会、1977年)所収本による。

注2 「東方文化叢書第六」(東方文化学院、1932-35年)所収。

注3 小島憲之「上代に於ける学問の一面——原本系『玉篇』の周辺——」(文学 39-12、岩波書店、1971年)等、関連論考は多い。

注4 「JIS X 0208-1990 情報交換用漢字符号」は、第一水準漢字集合 2,965 字、第二水準漢字集合 3,390 字、合せて 6,355 字の漢字を含む

注5 Ydic は JIS X 0208-1983 に基づく。JIS X 0208-1990 より 2 字少ない。

注6 「JIS X 0212-1990 情報交換用漢字符号－補助漢字」は 5,801 字の漢字を含む。

注7 Unicode, Inc.が公開している Unicode と他漢字コードとの対照マップを利用した。anonymous FTP で入手可能。unicode.org(IP Address 192.195.185.2)。

注8 王蜀豫「漢字の常用性についての基礎的研究——日中古典籍を資料として——」(信州大学大学院人文科学研究科修士論文、平成五年度)による。延べ 570,681 字。

注9 玉篇字順群は図書寮本類聚名義抄にも見出せる。拙稿「図書寮本類聚名義抄の単字字書的性格」国語国文研究 94 号、北海道大学国語国文学会、1993 年) 参照。

注10 岩淵悦太郎「現代日本語——ことばの正しさとは何か——」(筑摩書房、1970年)。

〔付記〕

1 この研究で作成した篆隷万象名義データベース試作版[図書寮本類聚名義抄対応]は、情報処理語学文学研究会(JALLC)のテキストアーカイブに登録されている。

2　豊島正之氏・金水敏氏・古田啓氏作の計算機用漢字字書(Ydic)に多大の恩恵を受けた。記して感謝の意を表す。

3　この研究は、平成四年度新村出記念財団研究助成金による成果の一部である。

――信州大学助教授――