HOKKAIDO UNIVERSITY

Title	『篆隷万象名義』の本文データベースの構築 ： 『玉篇』残巻対応部分を中心にして
Author(s)	李, 媛
Citation	国語国文研究, 146, 80(1)-65(16)
Issue Date	2015-06-30
Doc URL	https://hdl.handle.net/2115/93160
Type	journal article
File Information	kokugokokubunkenkyu_146_80-65.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

『篆隷万象名義』の本文データベースの構築
――『玉篇』残巻対応部分を中心にして――

李　　媛

1　はじめに

弘法大師空海撰述の『篆隷万象名義』は平安初期（830年以降）の成立とされるが、その内容は梁の顧野王の『玉篇』から音注や義注を簡約した字書であり、完本であることも含め、玉篇系字書の中で重要な位置を占めている。従来、掲出字の翻刻、電子化、音韻、訓詁などの面で諸先学に数多くの優れた研究成果がある。また、近年日本と中国それぞれで『篆隷万象名義』に関しての研究が進展し、中でも異体字や本文についての研究成果が注目される。それらを踏まえ、本稿は『篆隷万象名義』本文データベースの構築について考察したものである。篆隷万象名義データベースは、『篆隷万象名義』、原本『玉篇』、宋本『玉篇』を対象として、掲出字画像、掲出字テキスト、注文の三つの部分から構成される。筆者は、公開済みの掲出字テキストデータを利用し、注文データの構築に取り組んだので、データベース構築の過程で得られた知見を述べる。まず、『篆隷万象名義』、原本『玉篇』、宋本『玉篇』三者の関係と先行研究を概観する。次にデータベース構築の実際とその問題点をまとめる。最後に漢字字体史研究の観点からデータベースを利用して、『篆隷万象名義』及び原本『玉篇』諸本の字体差を観察する。漢字字体から『篆隷万象名義』は初唐標準字体を基本とするものであるが、原本『玉篇』諸本には一部に開成標準を用いるテキストがあることを指摘する。

2　玉篇系字書の概観と先行研究

『篆隷万象名義』は六帖の構成で、約一万六千字が収録される。第一帖から第四帖までは空海撰述であるが、第五・六帖は別人の手による撰述である。1114年に書写された高山寺本が唯一の古写本である。

『玉篇』は中国南北朝時代（543年）に成立した典拠や用例を充実させた部首別の漢字字書であり、後世の字書に大きな影響を及ぼした。原本『玉篇』は中国ではすでに散逸し、約八分の一の残巻が日本に残されている。原本『玉篇』を受け継ぐものとして今に伝わっているものに日本側の『篆隷万象名義』と中国側の宋本『玉篇』がある。

両書の編纂の方針は類似しているが、相違点も少なくない。

『篆隷万象名義』、原本『玉篇』、宋本『玉篇』に関連する主な研究成果を、本文研究、音韻、文字、電子化などの方面から、次にまとめてみよう。

『玉篇』に関しての研究は、岡井（1933）から始まるが、その逸文蒐集をさらに進めた研究成果が馬淵（1952）である。貞苅（1958）は『篆隷万象名義』が原本『玉篇』を簡略化したものであることを明確に論証したものである。胡（1989）は、宋本『玉篇』の掲出字の配列を規準にし、原本『玉篇』残巻を取り込み、さらに原本『玉篇』の出典にあたる典籍との校勘を行って、原本『玉篇』の逸文も考証したものである。

『篆隷万象名義』掲出字の翻刻について、白藤（1977）と宮澤（1977）は『高山寺古辞書資料第一』に収録され、写真版テキストと照らし合わせ、利用に至便であるが、後者は原本『玉篇』残巻及び逸文、宋本『玉篇』の情報をまとめている。近年、中国側も『篆隷万象名義』への関心が高まり、研究成果としては呂（2007）が挙げられる。その『篆隷萬象名義校釋』は中華書局写真版のテキストを基礎とし、『高山寺古辞書資料第一』に収録されたテキストも参考に、『説文解字』、宋本『玉篇』、『干禄字書』などの文献と照合し、校勘を加え、全文を翻刻したものである。

周（1966）と河野（1979）はほぼ同時期に出され、『篆隷万象名義』の反切を利用し、原本『玉篇』の音韻体系を研究する優れた研究成果である。上田（1986）は、上記の二つの研究成果の上に、『玉篇』の反切に残巻、逸文及び『篆隷万象名義』の所在を記して、声類・韻類に分別した著作である。

孔（2000）は、玉篇俗字の類型、変遷に着目し、唐宋字書の俗字と比較研究を行い、玉篇俗字の現象とその形成した源流を総合的に論じ、朱（2004）は、玉篇における文字を類型化し、特に合併・新増などの現象を取り上げ、反切用字で六朝時代の常用字の状況を調査している。

大柴（2008・2009・2011）は、後漢から魏晋まで、魏晋から隋唐までの書体・字体変化の視点から、多くの実例を集め、『篆隷万象名義』の俗字の分析を行い、所用の俗字を共通する文字ごとに250種類に分類した。

池田（1994）は、篆隷万象名義データベースの詳細について符号化の問題を中心に論じていたが、近時の池田（2014）では、篆隷万象名義データベースUCS対応版は、掲出字・所在情報などの部分が完成し、公開済みとしている。なお、王（2005）と大柴（2008・2009・2011）によれば、それぞれ『篆隷万象名義』のデータベース化が進行中の模様である。

3　用語の定義

本稿で用いる漢字の書体・字体・字形・字種の定義は、石塚（1984）および石塚（2012）により、次の通りとする。

書体：漢字の形に於て存在する社会共通の様式。多くは其の漢字資料の目的により決まる。楷書・草書等。

字体：書体に於て存在する一々の漢字の社会共通の規準。

字形：字体に於て認識する一々の漢字の書写（印字）された形そのもの。
字種：社会通念上同一のものとして認識された、一般的に音訓と意味が共通する漢字字体の総合。

4 『篆隷万象名義』と原本『玉篇』残巻との対応

4.1 篆隷万象名義データベース

北海道大学大学院文学研究科言語情報学講座池田研究室では、平安時代の日本の古辞書（『篆隷万象名義』、『新撰字鏡』、『類聚名義抄』）の総合データベースを構築するプロジェクトを進めている[1]。篆隷万象名義データベースは2011年にUnicodeに対応した掲出字、所在情報などの部分が完成、公開されていたが、これをさらに発展させたものである。筆者は『篆隷万象名義』の部分を担当し、篆隷万象名義本文データベースの構築に取り組んでいる。現在、このプロジェクトでは『篆隷万象名義』と原本『玉篇』残巻の2,087字に対応する部分の本文データベース化をほぼ終えたところである。データベース構築中に問題となった点やこのデータベースを利用して得られた知見を述べる。

4.2 原本『玉篇』残巻

原本『玉篇』は中国ではすでに散逸したが、日本では巻8、巻9、巻18、巻19、巻22、巻24、巻27の7巻の残巻が存する。『篆隷万象名義』の編集方針を考察するため、原本『玉篇』残巻にある部首を対象としてまず本文データベースの整備をはじめた。原本『玉篇』残巻の7巻は、書写年代や底本の相違から、それぞれに異なる性格を有する。上田（1970）は原本『玉篇』残巻の反切用字の特色から、各巻の性格を述べている（詳細後述）。

これらの残巻は、首尾完存、首尾欠、あるいは中間部が欠損するなど現存状況が異なる。このため、残された各部首の収字状態も異なっている。例えば、「心部」の場合、『篆隷万象名義』では、450字の掲出字があるが、原本『玉篇』は巻8の断片17行のみであり、次に翻刻で示すものしか残っていない。全存する字は4字であり、欠損する字は3字である。

（悍）（前欠）与怨憾同為忿恨之恨也。
戇　眉隕反。『説文』、古文、閔字也。閔、病也。存門部。
悬　古顔反。『説文』、古文、姦字也。姦、私也、邪也、為也、賊、在外也。在女部。
悍　公翻反。『字書』、亦諱字。諱、更也、變也。飾也、謹也、戒也。在言部。
愆　所革反。『周易』愆々終吉。王弼曰處多懼之地故曰愆々也。『公羊傳』愆而再拝。何休曰愆、驚皃也。『説文』、亦訞字。野王案訞、告也、譖也。又音蘇故反。在言部。

---

[1] 池田（2014）

惸　仇營反。『尚書』、無害惸獨、孔安国曰、惸單也、謂無兄弟也、無子曰獨。『周礼』、凡遠近惸獨老幼之欲有復於上者、鄭玄曰、無兄弟曰惸。『字書』亦煢字、煢々、無所依也。在丮部。或為嬛字。（以下欠）

惷　力高反。『説文』、古文勞（以下欠）

原本『玉篇』の残存する7巻、63部首について、その存欠の状況をまとめれば、次の通りである。

巻8　心部（首尾欠）（1部首）

巻9　言部（首欠）、語部、曰部、乃部、丐部、可部、兮部、号部（首欠）、亏部、云部、音部、告部、凵部、皿部、品部、桑部、龠部、册部、眀部、只部、卤部、欠部（中欠）、食部、甘部（中欠損）、旨部（中欠）、次部（中尾欠損）、率部（中尾欠損）（27部首）

巻18　放部、丌部、左部、工部（尾欠）、卜部（首欠）、兆部、用部、爻部、爨部、車部（中欠）、舟部（中欠）、方部（12部首）

巻19　水部（首尾欠）（部首）

巻22　山部、屾部、嵬部、屵部、广部（ほぼ全存、虫喰欠損）、厂部、高部、危部、石部（中欠）、磬部、皀部、阜部、鼾部、厽部（14部首）

巻24　魚部（首尾欠）（1部首）

巻27　糸部（首欠）、系部、素部、絲部、衞部、率部、索部（7部首）

4.3 『篆隷万象名義』と原本『玉篇』との比較

次に、『篆隷万象名義』、原本『玉篇』残巻、宋本『玉篇』の対応部首の字数を表1に示す[2]。

（原本『玉篇』残巻に欠損のある字を統計から除く）

表1　原本『玉篇』残巻・『篆隷万象名義』・『宋本玉篇』対応部分の字数

	掲出字数 (A)	注文字数 (B)	平均注文字数 (B/A)
原本玉篇	2,049	80,614	39.3
篆隷万象名義	3,279	21,928	6.7
宋本玉篇	4,864	35,079	7.2

表1の通り、『篆隷万象名義』の掲出字3,279字（延べ字数）の注文の字数は全21,928字である。平均一つの掲出字に約7字の注文がある。実際の注文は最少2字で、最も多いのは31字である。注文平均約7字の内訳は、反切（音注）3文字で、意

[2] 本文の点検・校正により数値は変動すると見込まれるので、一応の目安である。

味注記（義注）4字が基本となる。『篆隷万象名義』は、反切が一つ、意味注記に単字注の二つ付く注文様式が最も一般的であると推測できる。一方、原本『玉篇』残巻には、集計対象掲出字2,049字に、注文字数80,614字を存する。一つの掲出字に平均約40字の注文がある。これらのデータにより、『篆隷万象名義』が顧野王の原本『玉篇』の長文の注文を2割程度にまで簡略化した字書であるという性格が認定できる。

下記の例でも簡略化の程度が確認できる。（対応する箇所に下線を施した）

託　a他各反、b寄也、c依也、d累也。（『篆隷万象名義』3帖・11丁・裏）注文9字

託　a他各反、『公羊傳』、託不得已、何休曰因託以也、『論語』可以託六尺之孤、野王案『方言』託b寄也、凡寄爲託、『廣雅』託c依也、託d累也、或爲佗字在人部。（原本『玉篇』巻9）注文54字

託　a他各切、b寄也、c依憑也。（宋本『玉篇』上篇・83丁・裏）注文8字

5　本文データベース化における情報処理上の問題

5.1　翻刻について

5.1.1　包摂

包摂（unification）とは、複数の字体を区別せず一つの区点位置を与えることをいう[3]。

この概念を翻刻するときの文字同定に適用し、同じ抽象化の字体について、具象化の字形は違っても、同じ字と認める字体の範囲が存在する。一つの漢字が視覚的に実現する際、特に手書き文献において、それぞれの実現形は、筆画や角度に揺れが現れやすい。しかしながら、コンピュータで情報処理をする上で、そのような揺れは整理せざるを得ない。そのため、写本の字形から離れた字形をコンピュータで扱うことになる。そのようなときの規準として包摂規準が必要となる。例えば、「しんにょう」は1点「辶」に書かれても、2点「辶」に書かれても、文字の同定上区別しない。このようなものを文字同定の「包摂規準」という。

漢字の書写は、筆画の角度や長さなどの手書きの影響で、個人差が出るものである。古辞書の写本も例外でなく、字形の揺れが生じる。そして漢字を処理し、符号化するときに、手書きの揺れの要素をなるべく排除する。故に、古典文献を翻刻する際に、文字コードと写本の字形のずれが生じる。このずれを情報処理の障りにならないように、上記の「包摂規準」を運用する必要がある。

豊島（1999）は文字計量研究における「明示的で操作可能な文字の同定規準」の重要性を強調し、文字調査では「暗黙の文字包摂を排除出来ない」と指摘した。さらに、池田（2003）は「明示的で操作可能な文字の同定規準」について、「字形・字体を異にしても、同じ文字（字種）と認める文字の一覧表を作成し、それによって文字同定を

[3] JIS X 0208：1997（p.4）包摂規準を用いている。

行うものである」とし、「暗黙の文字包摂」に関して、「文字調査者が漢字の常識を働かせて、異なる字形・字体の文字を同じ文字として処理してしまうことである」と定義する。

大柴（2011）は、『篆隷万象名義』の俗字の分析を行い、所用の俗字を共通する文字ごとに250種類に分類している。この研究成果は『篆隷万象名義』文字同定の一種の明示的で操作可能な規準となり、可能な限り暗黙の文字包摂を排除する有力な手段としても利用できるものである。

5.1.2 翻刻の方針

古写本を電子テキスト化するときに、異体字の処理は避けられない問題である。例えば、「於」と「扵」、「因」と「囙」のような通行字体と古い字体は、両方ともUnicodeによって区別できる。翻刻方針により、どちらかの字体で文字コード化することとなる。古い字形情報を保つ方針だと、古写本に近い字体「扵」・「囙」で表現すべきであるが、古典文献を解釈した結果を広く提供する方針では通行字体「於」・「因」に改める方がよい。

また、データ検索を考慮すれば、異体字の翻刻には一貫した方法をとることが求められる。実際の古写本をデータベース化する作業を進めながら、上記の翻刻方針の問題の深刻性を感じる。古典文献を如何に分かりやすく紹介できるかを特に考慮して、古写本に忠実翻刻する方針（古い字形情報を保つ観点）より現行字体に直す方針（古典文献を解釈する角度）を採ることとした。

掲出字の場合は字体を示す役割を担うため、注文とは異なり、字形が大きく異なる場合、現行字体と写本に近い字体の二通りで対応する必要もあるが、これは当面の例外的処理としておく。

通行字体を基本とする翻刻と、古写本に近い字体を基本とする翻刻の二つの方法がある。検索の便を考慮して、当面は前者とするが、将来後者を併用することも視野に入れている。

5.2 Unicodeにおける漢字の扱い

5.2.1 CJK統合漢字

古辞書はその時代の漢字体系を示す書物であり、現代の視点からみると常用性の低い漢字の比率が高く、翻刻や文字コード化の作業にさまざまな難点をもたらす。しかしながら、近年Unicodeの拡張により、扱える漢字が数量的に増加し、古典文献の文字コード化についてもより対応しやすくなった。

Unicode 6.2では、漢字を扱う文字のブロックは基本多言語面[4]と追加漢字面[5]に配置される。詳細は次の通りである。

---

[4] Basic Multilingual Plane、BMP、Unicodeの第0面。最初の65536の符号位置である0000_{16}～$FFFF_{16}$からなる。

[5] Supplementary Ideographic Plane, SIP、基本多言語面に入りきらなかった漢字を収録する補足面の一つ。補助漢字面とも呼ばれる。

CJK 統合漢字

○基本多言語面

CJK 統合漢字	CJK Unified Ideographs	U+4E00～U+9FFF	20,941 字
CJK 統合漢字拡張 A	CJK Unified Ideographs Extension A	U+3400～U+4DBF	6,582 字

○追加漢字面

CJK 統合漢字拡張 B	CJK Unified Ideographs Extension B	U+20000～U+2A6DF	42,711 字
CJK 統合漢字拡張 C	CJK Unified Ideographs Extension C	U+2A700～U+2B73F	4,149 字
CJK 統合漢字拡張 D	CJK Unified Ideographs Extension D	U+2B740～U+2B81F	222 字

上記を合わせると 74,605 個の漢字が利用できる。後から追加されたものほど使用頻度は低い。一方、Unicode では次のように符号化されていない漢字を表現する手段が用意されている。

5.2.2 IDS 漢字構成記述文字列

Unicode 6.2 では次の 12 個の漢字構成記述文字 IDC（Ideographic Description Character）が定義されている。

IDC	内容	Unicode 番号	IDC	内容	Unicode 番号
⿰	左→右	U+2FF0	⿱	下→囲む	U+2FF6
⿱	上→下	U+2FF1	⿵	左→囲む	U+2FF7
⿲	左→中央→右	U+2FF2	⿶	左上→囲む	U+2FF8
⿳	上→中央→下	U+2FF3	⿷	右上→囲む	U+2FF9
⿴	四方→囲む	U+2FF4	⿸	左下→囲む	U+2FFA
⿵	上→囲む	U+2FF5	⿹	重ねる	U+2FFB

漢字構成記述文字は、任意の漢字の構成要素を表す複数の漢字に使用され、漢字を構成する各部分の要素の位置関係を示す。この方法により、符号化されていない漢字の構造を示すことが可能になる。IDS を使って漢字を表現した例について、『篆隷万象名義』における実例は次の通りである。

遂⇒⿺辶家　（2 帖 83 丁表）

脩⇒⿰亻⿰左月　（2 帖 86 丁表）

蘗⇒⿳艹辟米　（3 帖 33 丁表）

これらの IDC「漢字構成記述文字」を利用し、漢字の「部品」の合成により符号化されていない漢字を表現する手段が IDS 方式であり、そのコード列は IDS「漢字構成記述文字列」(Ideographic Description Sequence) という。IDS で漢字を表現する場合、その漢字の構造を考慮しなければならない。一般的な規則としては、形声構造の形の旁と音の旁を分けて記述すべきである。記述の順序は左から右へ、上から下にするのがよく、文字列は短いのが好ましい。古典文献の場合、符号化できない漢字が出現する比率が比較的高く、IDS 方式によって、符号化されていない漢字の図形的な構造を示すことが可能になる。

5.3 翻刻方針と包摂の具体例

Unicode で扱える漢字数は膨大な数にのぼり、古典文献を翻刻する際に、利用でき

る字体は豊富である。現在、日本では新字体と旧字体（康熙字典体）があり、中国では簡体字と新体字が共存し、韓国や台湾などもそれぞれ違う字体が使われている。これらの諸国・地域の文字集合はUnicodeのCJK統合漢字に統合され、多くの漢字の種類が利用できる。ただ、すでに述べたように、古写本の字体の細部まで精確忠実に表現するのは、無理がある。翻刻方針として、古写本の漢字を符号化する規準を立てないと、混乱をもたらす。そこで、まず康熙字典体を規準とすることを翻刻方針として、入力作業を行う。次に、実例を挙げ、写本の字形と符号化された文字のずれについて検討する。実例は、写本字形_「符号化字形」／翻刻用例（該当字に下線）の順に示す。

① 写本の字形と康熙字典体に極く僅かな違いがある場合、Unicodeの包摂の範囲とみなして、康熙字典体で入力する。
詰_「詰」／**詰**　去質反。治也、謹也、責也、無也。（3帖・16丁・裏）

② 写本の字形が康熙字典体と明らかに異なり、ほかの新字体などでも写本の字形に近い字体がない場合、康熙字典体の字体に改める。
對_「對」／潰　胡**對**反。乱也、漏也、遂也。（5帖・91丁・表）

③ 写本の字形が簡略字体（日本の新字体や中国の現行簡体字など）に近い場合、これらの簡略字体で対応する。
乱_「乱」／潰　胡對反。**乱**也、漏也、遂也。（5帖・91丁・表）

④ 写本の字形と康熙字典体の漢字の構成要素が一致するが、構成の位置が違う場合、康熙字典体の字体で翻刻する。
蘇_「蘇」／量　**蘇**庭反、辰也。（3帖・26丁・裏）

⑤ 符号化できない漢字の場合、IDS方式で表現する。
虌_「⿰嵬⿱台丰」／**⿰嵬⿱台丰**　子内反、繒也。（6帖・143丁・裏）

現在のデータベースでは、この方針を取るが、それ以前は入力者が複数であり原本通りの字体を表現しようとした部分もある[6]。

6　篆隷万象名義データベース化における文字コード化の処理状況

6.1　『篆隷万象名義』・原本『玉篇』・宋本『玉篇』の項目の構造

Unicodeで『篆隷万象名義』を文字コード化し、項目構造の各部分のコード化された状況を確認したい。そのため、まず、符号化対象の『篆隷万象名義』と対照する原本『玉篇』、宋本『玉篇』の項目の構造を確認する。

漢字の三要素は、形、音、義であり、字書はこれらの三要素について説明を行う。字書の項目の各構成部分は形、音、義によって分類すると、次のようになる。

---

[6] そのため、「於」と「扵」、「因」と「囙」のような同字種異体字について、原文に合わせて、通行字体「於」・「因」と「扵」・「囙」とを混在させて翻刻してきた。以前の方針の問題点としては、検索の用として一字種のものを二度検索しなければならない効率上の問題が挙げられる。

形…掲出字・字体注記

音…反切・直音注

義…意味注記

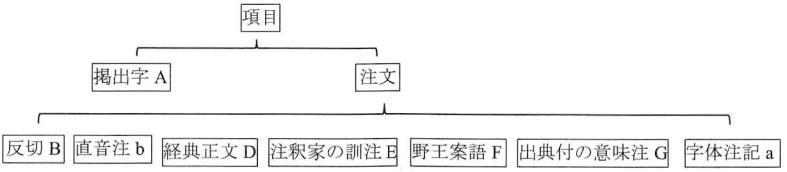
掲出字、反切、意味注記は、それぞれ形、音、義を解釈する役割を担う。字体注記は形、直音注は音の説明を補助する。

原本『玉篇』の構造は、まず反切を示し、次いで多くの儒家経典や訓詁書を網羅引用するのが基本である。さらに必要に応じて、著者顧野王の案語、異体注記、部首の違う異体字を参照する注記を加える。「託」を例にその項目の構造と内容を示せば、次の通りである。

○原本『玉篇』の項目の構造

A 託　B 他各反。D『公羊傳』託不得已、E 何休曰因託以也。D『論語』可以託六尺之孤。F 野王案『方言』託寄也、凡寄爲託。G『廣雅』託依也、託累也、a 或爲佗字在人部。（原本『玉篇』巻９）

構造：



内容：

形　A 掲出字　a 字体注記

音　B 反切　b 直音注

義　D 経典正文　E 注釈家の訓注　F 野王案語　G 出典付の意味注

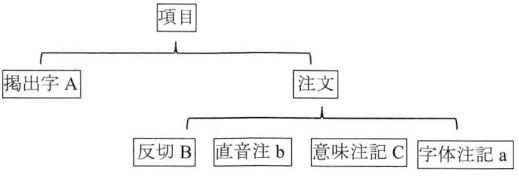
原本『玉篇』はこのような特徴を有するため、項目構造も『篆隷万象名義』と宋本『玉篇』より複雑である。

『篆隷万象名義』と宋本『玉篇』は、いずれも、原本『玉篇』の注文を縮約した字書であり、それぞれ「怨」、「言」を例にその項目の構造と内容を示す。

○『篆隷万象名義』の項目構造

A 怨　B 古多反。C 法也、範也。a 柯字。（『篆隷万象名義』３帖５丁表）

構造：



内容：

形　A 掲出字　a 字体注記

音　B 反切　b 直音注

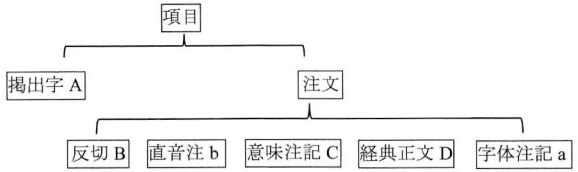
義　C 意味注記

○宋本『玉篇』の項目の構造

A 言　B 魚韈切。C 言辭也、我也、問也。D『説文』云直言曰言、論難曰語。

（宋本『玉篇』上篇 82 丁裏）

構造：



内容：

形　A 掲出字　a 字体注記

音　B 反切　b 直音注

義　D 経典正文　C 意味注記

意味注記の主体は両書とも単字注によるが、『篆隷万象名義』の方が網羅的で、単字注の数が多い。一方、原本『玉篇』の反切が二つか三つある場合、『篆隷万象名義』は一つしか採らないことが多いが、宋本『玉篇』は全部を採ることが多い。

『篆隷万象名義』の原本『玉篇』を簡略化した様相を見ると、形、音、義それぞれの抄出のしかたは大略次のようになる。

形　字体注記の部分には異体字だけを挙げる。「古文」「同上」のような字体注記だけのものもある。部首の違う異体字を参照する注記は省略する。

音　原本『玉篇』で反切が二つ以上の場合、原則一つしか採らない。

義　意味注記のうち、経典の正文、注釈家の名前は省略し、訓注と野王案語では単字注を網羅的に載せる。

6.2　各構成部分の Unicode による符号化率

『篆隷万象名義』の各構成部分は担う機能が異なり、それらの用字の特徴も異なることから、符号化の状況が異なると推測される、この点を確かめるために、次に、『篆隷万象名義』の原本『玉篇』残巻に対応する部分の各構成部分の Unicode による符号化率を掲出字、注文、反切、意味注記、字体注記に分けて表 2 から表 6 に示す。

表 2　掲出字の符号化率

文字コード	字数	CJK 統合漢字	拡張 A	拡張 B	IDS
延べ字数	3,279	2,406 (73.4%)	373 (11.4%)	495 (15.1%)	5 (0.2%)
異なり字数	3,256	2,391 (73.4%)	372 (11.4%)	493 (15.1%)	5 (0.2%)

表 3　注文の符号化率

文字コード	字数	CJK 統合漢字	拡張 A	拡張 B	IDS
延べ字数	21,928	21,644 (98.6%)	90 (0.4%)	123 (0.6%)	71 (0.3%)
異なり字数	3,932	3,674 (93.4%)	80 (2.0%)	112 (2.8%)	66 (1.7%)

表 4　反切の符号化率

文字コード	字数	CJK 統合漢字	拡張 A	拡張 B	IDS
延べ字数	9,430	9,376 (99.4%)	11 (0.1%)	28 (0.2%)	15 (0.2%)
異なり字数	1,413	1,365 (96.6%)	8 (0.6%)	27 (1.9%)	13 (0.9%)

表 5　意味注記の符号化率

文字コード	字数	CJK 統合漢字	拡張 A	拡張 B	IDS
延べ字数	11,596	11,409 (98.4%)	66 (0.6%)	76 (0.6%)	45 (0.4%)
異なり字数	2,221	2,054 (92.5%)	59 (2.7%)	66 (3.0%)	42 (1.9%)

表 6　字体注記の符号化率

文字コード	字数	CJK 統合漢字	拡張 A	拡張 B	IDS
延べ字数	922	879 (95.2%)	13 (1.9%)	19 (2.8%)	11 (1.6%)
異なり字数	298	255 (85.6%)	13 (4.8%)	19 (7.0%)	11 (4.1%)

以上の『篆隷万象名義』の項目の各構成部分を、CJK 統合漢字で処理できる割合の小さいものから大きいものの順にならべると次の通りである[7]。

掲出字（73.4%）＜字体注記（85.6%）＜意味注記（92.5%）＜反切（96.6%）＜直音注（100%）

掲出字の延べ字数と異なり字数からは、重複字が少ないことが窺える。これは、掲出字は字書の骨組みであり、説明対象であるため、ユニークな存在である性格を反映する。また、注文部分の符号化率と比較すると、掲出字は CJK 統合漢字で処理できる

[7] また字音の説明に直音注を含む項目は、今回の調査範囲で 11 項目があり、直音注に使われる文字コードはすべて CJK 統合漢字に含まれるものだった。

割合が低く、拡張Aと拡張Bが高い。掲出字は字形を示す役割を担うので、認識度の低い難字や異体字が多いことが反映されたものである。

注文の、CJK統合漢字、拡張A、拡張B、IDS方式で処理できる割合について、掲出字データと見比べると、違う傾向が見える。また、延べ字数は21,928字があるに対し、異なり字数は3,918字であり、重複字が多いことが分かる。字書の注文部分は比較的理解しやすい文字を利用して掲出字を解釈するという性格が窺える。

IDS方式によるものは、いずれの場合も僅少である。康熙字典体を基本とする翻刻方針によったため、僅少となったのであり、古写本の字形に忠実に翻刻しようとすれば、この数値は大きく変わると考えられる。

7 『篆隷万象名義』と原本『玉篇』の漢字字体史研究へのアプローチ

7.1 HNG[8]を利用した『篆隷万象名義』の字体研究

字体に関する研究は通時的な視点による研究と共時的な視点による研究の二つがある。通時的な研究は、各時代・各地域の標準的な文献の中の相違を明らかにしようとする。それに対して、共時的な視点からの研究は、同時代の文献の字体を見比べることになる。一つの文献の同一字種を検討すると、その文献の字体の体系的な特徴を把握できる。古典文献をデータベース化すると、大量のデータの集約や検索などの作業ができるため、同じ文献の中の同一字種の使用字体を把握しやすくなり、字体研究に進展をもたらすものとなる。

石塚字体変遷モデルに基づき、斎木・池田（2011）は「開成標準」と「初唐標準」が一致しない60字種を指摘した。本稿では、その中の12字種を取り上げ、『篆隷万象名義』（原本『玉篇』残巻との対応部分）、原本『玉篇』残巻諸本での使用実態の調査を試みる。

調査対象は「事・流・来・定・所・礼・悪・土・説・能・於・為」の12字種とする。篆隷万象名義本文データベースで所在情報を確定したうえで、そして高山寺本の画像で確認する。12字種について、開成標準と初唐標準のそれぞれの画像を対照して、調査した結果は次の通りである。字種、「開成標準」—「初唐標準」の順に示す。

---

[8] 漢字字体規範データベース（HNG）は、漢字の字体には時代・地域（国）による標準が存し、その標準は時代・地域（国）により変遷することを示すものである。このデータベースは、漢字字体の変遷について次のような考え方によっている。中国では、初唐期（618-712）に、異体字が強く意識されるようになり、「正俗通訛」の規準が生まれ、「初唐標準」と呼ぶことのできる標準字体が形成される。「開成標準」は開成石経（837）において示された字体を指し、宋版において実践され、普及していく。一方、日本では「初唐標準」を受容、そして定着する。宋版の影響は部分的であり、近世初期以降の版本から「開成標準」の影響が大きくなる。この漢字字体変遷に関する見解は「石塚字体変遷モデル」と呼ばれる。

・事　「事」-「事」　・所　「所」-「所」　・説　「說」-「説」
・流　「流」-「流」　・礼　「禮」-「礼」　・能　「能」-「能」
・来　「來」-「来」　・悪　「惡」-「悪」　・於　「於」-「扵」
・定　「定」-「定」　・土　「土」-「土」　・為　「爲」-「為」

12字種について『篆隷万象名義』とを対照した結果をまとめれば表7となる。例えば、「事」は第2帖に2例あり、その字体はA初唐標準であることを示す。表7から『篆隷万象名義』はすべて初唐標準であることが分かる。

表7　『篆隷万象名義』の字体調査

万象名義	事		流		来		定		所		礼		悪		土		説		能		於		為	
第2帖	2	A	1	A	2	A	0	-	0	-	0	-	0	-	0	-	0	-	0	-	20	A	1	A
第3帖	2	A	1	A	6	A	0	-	11	A	2	A	4	A	1	A	3	A	2	A	54	A	9	A
第5帖	2	A	39	A	4	A	5	A	19	A	8	A	12	A	3	A	0	-	1	A	34	A	10	A
第6帖	6	A	2	A	3	A	1	A	19	A	3	A	3	A	6	A	0	-	1	A	50	A	10	A

7.2　HNGを利用した原本『玉篇』残巻諸本の字体研究

次に、同じ手順で原本『玉篇』残巻の字体の状況を確認する。詳細は表8の示した通りである。Aは初唐標準、Bは開成標準を示す。

表8　原本『玉篇』残巻の字体調査[9]

原玉	事		流		来		定		所		礼		悪		土		説		能		於		為	
巻8	0	-	0	-	0	-	0	-	0	-	1	A	0	-	0	-	4	A	0	-	1	A	3	A
巻9	10	A	9	A	11	A	5	A	29	A	116	A	22	A	5	A	74	A	18	A	74	A	105	A
巻18	19	A	6	A	3	A	3	A	20	A	27	A	1	A	2	A	9	A	7	A	26	A	43	A
巻19	14	A	13	A	1	A	2	A	15	A	37	A	0	-	4	A	4	A	6	A	19	A	39	A
巻22	14	A	4	A	7	A	7	A	54	A	71	A	4	A	52	A	74	A	7	A 2 B 5	58	A	20	A
巻24	0	-	0	-	0	-	0	-	0	-	0	-	0	-	0	-	6	A	0	-	2	A	0	-
巻27	28	A	5	A	7	A	1	A	61	A	102	A	4	A	5	A	8	A	5	A	33	A17 B16	158	A

[9] 原本『玉篇』残巻の字体調査用例を資料として付した。

調査対象の12字を原本『玉篇』残巻の巻ごとに確認すると、「事・流・来・定・所・礼・悪・土・説・為」の10字種の字形は「初唐標準」に類似する。字体の揺れの見られる字種は「能」と「於」である。「能」の場合、巻22の7箇所に確認され、「初唐標準」の字体は2例、「開成標準」の字体は5例ある。「於」の場合、巻27の33箇所に揺れが確認され、「初唐標準」の字体は17例、「開成標準」の字体は16例ある。

ところで、原本『玉篇』残巻は、書写年代や底本の相違から、それぞれ異なる性格を有する。これについて上田（1970）は原本『玉篇』残巻の反切用字の特色から、各巻の性格について、「総合して玉篇の残巻を大別すると、A巻十八・十九が古い姿を示すもの、B巻九・二二・二七を新しい姿を示すものと二つに分ち、さらに後者を、b万象名義に最も近いものとして巻二七、c名義にやや近いものとして巻二二、d名義に最も差のあるものとして巻九と、三分することができる」ことを指摘した。

調査で確認した「能」と「於」の字体の揺れからは、巻22と巻27が開成標準字体を用いる新しいものであることが窺え、上記の上田（1970）の指摘と一致する。『篆隷万象名義』は、調査範囲では、字体の揺れが観察されず、用いられた字体は「初唐標準」に近いと言える。

8　おわりに

本稿では、『篆隷万象名義』の原本『玉篇』と対応する部分の本文をデータベース化し、その諸問題を検討したほか、原本『玉篇』、宋本『玉篇』と対照データを確認して、『篆隷万象名義』の構造特徴を分析した。『篆隷万象名義』と宋本『玉篇』はともに原本『玉篇』の注文を簡略化した字書であり、『篆隷万象名義』と宋本『玉篇』と同じく注文は7字程度であることを明らかにした。

現在、Unicodeによる符号化では、掲出字（異なり）の99.8%、注文（異なり）の99.7%の符号化が可能である。符号化できない文字はIDS方式により処理が可能である。Unicodeによる古辞書のデータベース化は実用可能の段階と言える。

HNGと篆隷万象名義データベースを利用し、12字種の漢字字体について『篆隷万象名義』と原本『玉篇』残巻での使用状況を確認した。『篆隷万象名義』と原本『玉篇』残巻の字体研究の一つの試みにすぎないが、先行研究で指摘された原本『玉篇』残巻の特色を字体研究の面で確認できたのは一つの成果である。『篆隷万象名義』と原本『玉篇』残巻の字体状況をより明確に把握できるように、今後も様々な観点から分析を行いたいと考えている。

参考文献

池田証壽（1994）「篆隷万象名義データベースについて」『国語学』178集
池田証壽（2003）「コーパスによる文字の研究」『日本語学』22巻5号　明治書院
池田証壽（2014）「平安時代漢字字書総合データベースの構築」『北海道大学文学研究科紀要』（142）

石塚晴通（1984）『圖書寮本日本書紀　研究篇』汲古書院
石塚晴通（2012）「漢字字体研究史―序に代えて―」『漢字字體史研究』勉誠出版
上田正（1970）「玉篇残巻論考」『神戸女学院大学論集』第 17 巻 1 号
上田正（1986）『玉篇反切総覧』私家版
大柴清圓（2008）「『篆隷万象名義』における俗字の研究(1)―後漢の隷変字から魏晋の草書の楷書化まで―」『高野山大学密教文化研究所紀要』第 21 号
大柴清圓（2009）「『篆隷万象名義』における俗字の研究(2)―魏晋から隋唐までの楷書の俗字―」『高野山大学密教文化研究所紀要』第 22 号
大柴清圓（2011）「『篆隷万象名義』における俗字の研究(3)―附録・『篆隷万象名義』俗字表―」『高野山大学密教文化研究所紀要』第 24 号
岡井慎吾（1933）『玉篇の研究』東洋文庫論叢第 19　東洋文庫
河野六郎（1979）「玉篇に現れたる反切の音韻的研究」『河野六郎著作集』二「中國音韻學論文集」平凡社（1937 年東京大学卒業論文の再録）
斎木正直・池田証壽（2011）「漢字字体の変遷―HNG に見る変わる字体と変わらない字体―」『計量国語学』27 巻 8 号
貞苅伊徳（1958）「玉篇と篆隷万象名義について」『国語学』31 集
白藤禮幸（1977）「掲出字索引」『高山寺古辞書資料第一』東京大学出版会
豊島正之（1999）「書評　横山詔一・笹原宏之・野崎浩成・エリク＝ロング編『新聞電子　メディアの漢字―朝日新聞 CD-ROM による漢字頻度表―』」『日本語科学』6　国書刊行会
馬淵和夫（1952）「玉篇佚文補正」『東京文理科大学国語国文学会紀要』3 号　東京教育大学東京文理科大学国語国文学会
宮澤俊雅（1977）「掲出字一覧表」『高山寺古辞書資料第一』東京大学出版会
胡吉宣（1989）『玉篇校釈』上海古籍出版社
呂浩（2007）『篆隷萬象名義校釋』学林出版社
孔仲温（2000）『玉篇俗字研究』台湾学生書局
王平（2005）「《说文》《玉篇》《万象名义》联合检索系统的开发：从原本《玉篇》到宋本《玉篇》」『中国文字研究』第 6 輯　华东师范大学中国文字研究与应用中心
周祖謨（1966）「萬象名義中之原本玉篇音系」『問学集』　中華書局（1936 年北京大学卒業論文の再録）
朱葆华（2004）『原本玉篇文字研究』齊魯書社

使用テキスト

『高山寺古辞書資料第一』東京大学出版会 1977
『原本『玉篇』残巻』中華書局 1985
『東方文化叢書第六』東方文化学院 1932-35
『大広益会玉篇』中華書局 1987

規格

日本工業標準調査会「JIS X 0208:1997 7ビット及び8ビットの2バイト情報交換用符号化漢字集合」日本規格協会　2000

The Unicode Consortium, *The Unicode Standard, Version 6.2*, Mountain View, CA, Unicode Consortium, 2012

資料　原本『玉篇』残巻の字体調査用例

【能】

A 初唐標準

隑　渠鎧午哀二反。司馬如賦臨回江之隑州、漢書音義曰隑長也、方言隑佮立也、東齊海伐燕之郊謂委□广委而跪謂之隑佮、郭璞曰脚躄不**能**行者也、又曰隑倚也、郭璞曰江南人呼梯爲隑、所以倚物而攻者。（原本『玉篇』巻22）

B 開成標準

破　普餓反。毛詩矢舍而破、叓云發矢即中、如推破物也、礼記孔子曰君子語小天下莫**能**破焉、説文破碎也、廣雅破壞也。（原本『玉篇』巻22）

【於】

A 初唐標準

縛　扶矍反。左氏傳襄公縛秦囚、野王案説文縛束也、又曰許男面縛、杜預曰縛手**於**後見其面也。（原本『玉篇』巻27）

B 開成標準

縈　**扵**(ママ)宮反。毛詩葛藟縈之、傳曰縈旋也、説文縈收韋也。（原本『玉篇』巻27）

附記

本稿は平成25年7月13日に開催された北海道大学国語国文学会大会において研究発表した内容に加筆・訂正を加えたものである。北海道大学国語国文学会大会での発表、草稿に対する諸先生のご指導に感謝申し上げる。

（り　えん・北海道大学大学院博士後期課程）