

天治本新撰字鏡全文テキスト構築の方法と課題

池田 証壽 (北海道大学 文学研究科)

李 媛 (京都大学人文科学研究所・日本学術振興会)

天治本新撰字鏡 (略称 TSJ) は平安時代漢字字書総合データベース (略称 HDIC) の一部である。昨年度高山寺本篆隸万象名義 (略称 KTB) を対象にして Unicode 対応全文テキストの公開を実現した。次の段階として、天治本新撰字鏡を取り上げて、次の 3 点を踏まえた全文テキスト構築の方法と課題を述べる。
a) 古写本古辞書を対象、b) Unicode 対応、c) 典拠字書データの作成と利用。

Methods and Issues Regarding the Construction of Tenjibon Shinsen Jikyo Full Text

Shoju Ikeda (Graduate School of Letters, Hokkaido University)

Yuan Li (Institute for Research in Humanities, Kyoto University, JSPS)

Tenjibon Shinsen Jikyo Database (TSJ) is part of Integrated Database of Hanzi Dictionaries in Early Japan, abbreviated as HDIC. The full text and the publication system of Kosanjibon Tenrei Bansho Meigi Database (KTB) were released to the public in the previous year. In this paper, as the next stage, we will take up TSJ, base on the following three points, state the methods and issues regarding the construction of Tenjibon Shinsen Jikyo full text : a) case of ancient manuscript dictionaries, b) corresponding to Unicode, c) construction and utilization of source dictionaries data.

1. まえがき

本稿では、天治本新撰字鏡データベース構築の方法と課題を中心に述べる。天治本新撰字鏡データベース (Tenjibon Shinsen Jikyo, 略称 TSJ) は平安時代漢字字書総合データベース (HDIC) の一部である。昨年度、高山寺本篆隸万象名義を対象にして Unicode 対応全文テキストの公開を実現した。次の段階として、天治本新撰字鏡を取り上げる。

天治本新撰字鏡は、2 万字以上の掲出字と約 3,700 に及ぶ和訓を収録し、貴重な言語資料となっているが、古写本であることに加えて、異体字と難読字が多いことも手伝って、全文の翻刻本文が存しない。

以下では、昨年度公開した高山寺本篆隸万象名義データベース (Kosanjibon Tenrei Bansho Meigi, 略称 KTB) の全文テキストと比較しながら、天治本新撰字鏡全文テキスト構築の方法と実際を述べる。具体的には、古写本古辞書を対象、Unicode 対応、典拠字書データの作成と利用の三点に即して報告する。

2. 新撰字鏡の資料解説と先行研究

2.1 新撰字鏡の資料解説

新撰字鏡 12 巻は、その序文によれば、撰者である昌住が、寛平 4 年(892)夏、唐・玄応一切経音義 25 巻を元に「三軸」からなる部首分類の字書を編纂したことに始まる。その後、昌泰年中(898-901)に切韻、玉篇及び私記類を得て増補して 12 巻とする。現存本は完本と抄録本との 2 系統がある。

完本は天治元年(1124)書写の宮内庁書陵部蔵本(天治本)が唯一である。抄録本は 12 巻系統の本から和訓のある項目を抜き出した本であり、享和本、群書類従本、大東急記念文庫蔵本などがある。

天治本は、160 の部を分かち、単字をその字形により部首に分類した部(天・日・月・肉・雨など)と語句を意義分類した部(親族・小学篇字・本草・重点・連字・臨時雑要字)とから構成される。部首分類は、天・人・衣・食・住など意義分類が加味されている。引用は一切経音義(玄応)、切韻、玉篇の三書が中心をなす。それぞれ原典での本文の体裁を反映し、一群となって引用される(この一群を出典群という)。特に切韻引用部分(切韻群ともいう)と玉篇引用部分(玉篇群ともいう)はその本文・配列などの点でそれらの原態をうかがうことができる。切韻引用部分は玉篇の注文を併合する項目が存する。一切経音義を含む部分は、和訓・和音・正借音など一切経音義以外の出典に依拠した注文を多く含み、出所不詳とあわせて出典の究明が待たれている。意義分類部分では、「臨時雑要字」が重要で資料的価値も高い。その用字・和訓は源順の撰述した倭名類聚抄に引かれた弁色立成・漢語抄と酷似することから、この部は漢語抄と同類の上代の用字字書に依拠したとされる。その他、爾雅・文選・遊仙窟・正名要録・干祿字書・日本霊異記・大般若経音義などの引用が指摘されている。

2.2 先行研究

2.2.1 先行研究の概観

本書の出典研究は貞苺(1998)が基本文献であり、新撰字鏡の全巻にわたって、一切経音義を含む部分、切韻引用部分、玉篇引用部分、出所不詳部分等を明示しており有益である。一切経音義との関係は池田(1982)(1984)、切韻との関係は上田(1981)(1984)、玉篇との関係は井野口(1978)と上田(1986)が詳しい。

新撰字鏡収録の和訓を含む項目の校注は、山田(1916)と京都大学文学部国語学国文学研究室(1958)があり、阪倉(1967)は享和本の価値を再評価する。近年では、内田(2005)と大槻(2002)が和訓の研究を更に深めている。掲出字の字体認定の問題に関しては、字様である正名要録との関係を論じた西原(2015)の成果が重要である。音注等を総合した文字同定は、井野口(1978)、上田(1981)(1984)(1986)が基本である。近年、中国学者も新撰字鏡に注目しており、張(2012)は中国字書・敦煌写本と比較し、日本資料を利用した中国語史への研究である。

また、HDIC における天治本新撰字鏡データベースの進捗に関する段階的な報告は Ikeda (2017)、池田(2017)、李・池田(2016)及び Li and Ikeda(2017)が参考となる。

2.2.2 複製本の刊行

1916年、山田孝雄の攷異并索引1冊を付して刊行された天治本新撰字鏡(六合館)が最初の複製本(1933年再刊)であり、その後、京都大学文学部国語学国文学研究室編新撰字鏡(全国書房、1944年)が刊行され、さらに1967年、天治本と抄本二種(享和本・群書類従本)と収めた同研究室編天治本新撰字鏡増訂版(臨川書店)が出た。天治本の複製本はいずれも1916年の複製本によるものである。宮内庁書陵部蔵の天治本原本は、破損がひどく、長く閲覧できない状態であった。修理の完了とその複製が待たれる。大東急記念文庫蔵本は1976年に古辞書叢刊行会から原裝影印版が出ている。

2.2.3 全文テキストの作成

前述したように新撰字鏡の内容は、出典群によって、大きく一切経音義を含む部分、切韻引用部分と玉篇引用部分との三つに分かれる。切韻引用部分は主に切韻の出現順になり、玉篇も併合している。玉篇引用部分は主に玉篇の配列順になっているため、篆隸万象名義を参照しながら入力作業を進めた。切韻引用部分に関しては、漢字データベースプロジェクトで公開中の広韻電子テキストを参照した。一切経音義を含む部分は、公開され

た一切経音義の徐時儀校注の翻刻テキストが参考になった。

和訓のある項目の判読は、山田(1916)の攷異及び新撰字鏡和訓索引(京都大学文学部国語学国文学研究室編1958)を参照した。また、疑問箇所の確認は切韻佚文研究(上田1984)と玉篇反切総覧(上田1986)を参照した。

3. 所蔵者の許諾

天治本新撰字鏡を所蔵する宮内庁書陵部には、平安時代漢字字書総合データベース編纂委員会(代表、池田証壽)として掲出字の写真掲載と全文翻刻をインターネットで公開することについて、2014年2月17日付けで申請を行い、同年3月10日に許可を得た(宮内書発第343号)。インターネット公開の条件としては、①原本が宮内庁書陵部の所蔵であることを明記すること、②複写物を出版掲載等に再使用する場合は改めて宮内庁書陵部に申請すること、③インターネット公開時期を事前に宮内庁書陵部に通知することが求められている。

また、インターネット公開に使用する原本画像は、京都大学国語学国文学研究室編『新撰字鏡増訂版』(臨川書店、1967年12月15日初版、1999年11月30日再版第7刷)である。これは1916年に六合館から刊行された複製本を再複製し、抄録本との校異を頭注として加えており、利用に便利なテキストである。京都大学国語学国文学研究室には2016年8月23日付け、臨川書店には同年9月5日付けで許可を得た。

4. 古辞書データベース構築の方法論

2016年9月、高山寺本篆隸万象名義の全文テキストをTSVのデータ・フォーマットによってインターネットで公開した。この全文テキストの特徴は次の3点であった。

- a) 古写本古辞書を対象
- b) Unicode対応
- c) 典拠字書データの作成と利用

古写本古辞書の全文テキストの公開としては、篆隸万象名義が他に類のない初めての試みであり、UnicodeはそのVersion 7.0に対応するものである。篆隸万象名義全文テキスト構築に際しては、典拠字書データとして宋本玉篇(1013年、略称SYP)、原本玉篇(543年、残巻、略称YYP)を独自に構築し、それを土台として篆隸万象名義の全文テキストを完成させた点に方法上の特色がある。天治本新撰字鏡の全文テキスト構築に際しても、同様の方法を用いるが、高山寺本篆

隸万象名義と天治本新撰字鏡とを比べると、体系と構造の点で相違が認められ、篆隸万象名義全文テキスト構築で採用した方法を、新撰字鏡全文テキスト構築に適用する場合にどのようなことが課題となるかを検討することが必要である。そこでまず、天治本新撰字鏡の体系と構造を概観する。その後、上記 a), b), c) の順に篆隸万象名義と新撰字鏡とを比較しながら、新撰字鏡全文テキスト構築の課題を述べる。

5. 天治本新撰字鏡の体系と構造

5.1 部首体系

漢字字書の部首体系としては、古く説文解字の540部があり、これを改編した玉篇の542部がある。日本古辞書は玉篇542部首を土台に部首の統合と新設を行っている。新撰字鏡は160部からなるが、部首分類部分と意義分類部分とが混在して構成されている。大半は部首分類部分であるが、データ処理上、意義分類部分を区別することが求められる。親族部・本草・臨時雜要字は明確に意義分類部分であり、小学篇字・重点・連字は部首分類として扱わず意義分類部分に準じて扱う。

部首の統合は、例えば新撰字鏡火部に玉篇の火部・炎部・炙部の3部を含めるものである。部首の新設は、文尻八点・品字様などである。こうした部首統合の状況は、貞苺(1998)によりその全貌を知ることができる。部首の統合と新設は、部首体系の変遷を考察する際の重要なポイントであるので、この点を踏まえたデータ作成が求められる。

次に分巻と項目構造をみよう。部首体系を含む分巻一覧を表1に示す。説文解字・康熙字典のような著名な字書を除くと、古辞書の部首体系と分巻に関してインターネットで公開されている情報は少ない。

表1 天治本新撰字鏡分巻一覧

篇立	巻1	巻2	巻3	巻4
総 160 部	序・天～イ 11部	父～皮 15部	毛～力 10部	糸～門 8部
	巻5 馬～石 15部	巻6 山～金 8部	巻7 木～小学篇字 及本草異名4部	巻8 禾～虫 12部
	巻9	巻10	巻11	巻12
	龜～イ 15部	↑～十 20部	文～数字 38部	雜字～ 臨時雜用字 4部・跋

5.2 項目構造

天治本新撰字鏡の項目の基本構造は図1左に示す通りである。項目の各構成部分で掲出字の形・音・義・訓について説明する。高山寺本篆隸万象名義で

はほとんど単字掲出字であるのに対し、天治本新撰字鏡は、部首分類部分における①単字掲出項目、②異体併記掲出項目、及び意義分類部分における③熟語掲出項目が共存する。

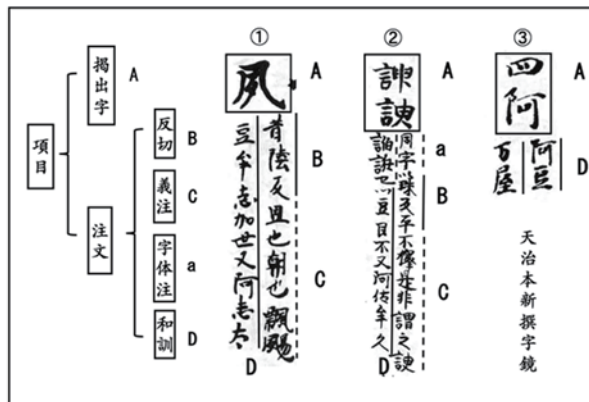


図1 天治本新撰字鏡項目構造

次に例を挙げる。図1右の①～③に対応する。

①単字掲出項目

例1 夙 昔陸反。旦也，朝也。「飄颻。豆牟志加世」。又阿志太。(巻1・18丁裏 夙部第七 弁別不能)

「夙」の項目は、一見、基本な単字掲出項目であるが、仔細に観察すると「飄颻。豆牟志加世」部分が「夙」の注文に繰り込まれた埋字項目である。「豆牟志加世」は万葉仮名で和訓「ツムジカゼ」を示す。

②異体併記掲出項目

例2 諛諛 同字。以珠反。平：不擇是非謂之諛，諛誑也。部豆良不，又阿佐牟久。(巻3・12丁裏 言部第卅 一切経音義引用部分)

「諛諛」の項目は、異形同字(異体字)を併記し「同字」と注記する。ここではUnicodeの「諛」と「諛」で翻刻したが、原本字形とのずれ(包摂規準の抽象度の高低)の扱いが課題である。弁別字体を明示するために、原本画像を併用して対応するのが望ましい。

③熟語掲出項目

例3 四阿 阿豆万屋。(巻12・27丁表 臨時雜要字第百六十・舍宅章)

「四阿」の項目は、意義分類部分に属し、万葉仮名で意味を説明している。「四」の原本字形は通行字体と異なり、②と同様に包摂規準の抽象度の問題が生じる。意義分類部分であるため、ここでは、包摂規準の抽象度を高く設定し、通行字体で翻刻することとする。

6. 新撰字鏡全文テキスト構築上の課題

6.1 三つの課題

古写本古辞書は、現代辞書に見られるような、詳細かつ明瞭な凡例が明示されない。辞書の体系

と構造,あるいはシステムの設計は暗黙の内に理解されるからである.そして,異体字・難読字が多く,欠損や誤写,誤脱の影響のため本文解説も難点がある.前述した a)古写本古辞書を対象, b)Unicode 対応, c)典拠字書データの作成と利用の3点の課題から,篆隸万象名義データベースと新撰字鏡データベースとを比較しよう.

6.2 古写本古辞書を対象

6.2.1 写本資料

高山寺本篆隸万象名義は同一人による全巻書写のため,書写者による影響は均一と仮定できる.一方,天治本新撰字鏡は複数人による寄り合い書きであり,書写担当者が丁寧に楷書で書写するかどうかという書写態度によって,判読の難易度に相違が生じる.具体例を図2・図3で示す.

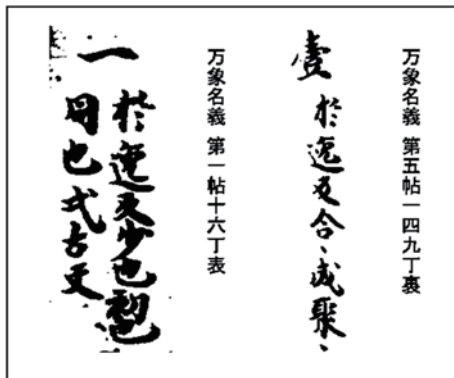


図2 篆隸万象名義における書写差異 (同一人書写)

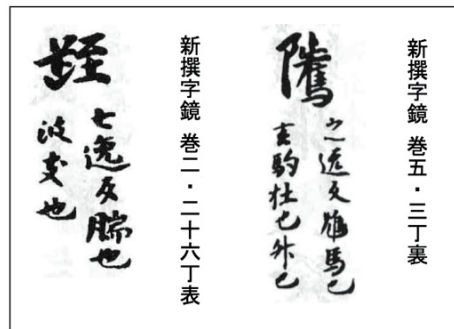


図3 新撰字鏡における書写差異 (複数人書写)

図2の左右の篆隸万象名義の「一」と「壹」の例で見比べれば分かるように,細かい箇所では違いがあるが,同一人書写であることは明瞭である.

図3の左右は,異なる筆跡であることが明らかである.左は新撰字鏡巻二の例であるが,この巻の書写担当者は「林幸」という名の学僧で,楷書で丁寧に書写している.右は巻五の例であるが,この巻の書写担当者は「覚嚴」という名の学僧で,行書体で書写しており,全体的に判読しにくい.天治本新撰字鏡は,都合五人の僧侶が手分けして書写している.各巻の巻末に書かれている奥書で書写担当者が確認できる.このような判読の難易度が均一でない古写本古辞書の本文入力・点検が課題となる.

現在,利用できる複製本は旧複製と新複製の二種類である.旧複製は1916年刊行の複製本とその再複製であり,広く利用されている.新複製は宮内庁書陵部で閲覧・入手が可能な最新の写真版である.この新複製によって旧複製で不鮮明な箇所が明瞭に判読できるようになった.また逆に,欠損箇所が修理状況によって,字画が判読しにくくなる場合もある.したがって,旧複製と新複製とを見比べて本文内容を確認する必要がある.

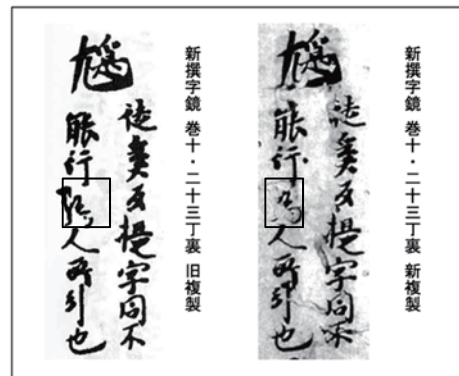


図4「熿」新複製による判読可能な例

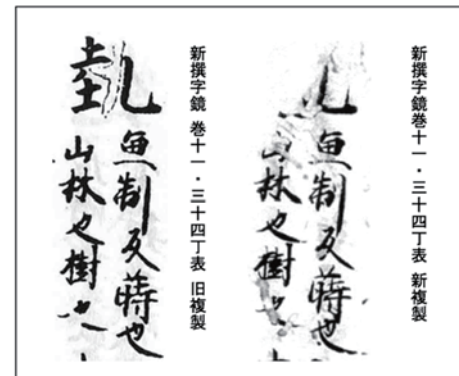


図5「𪗇」旧複製による判読可能な例

例4 熿 徒奚反。提字同。不能行爲人所引也。

例5 𪗇 魚制反。蒔也，山林也，樹也。

図4「熿」の例4は,義注字「為」が新複製で視認できるが,旧複製の方では判読不可である.逆に図5「𪗇」の例5は,掲出字「𪗇」が旧複製ではまだ字形の確認ができるが,新複製は虫損が進んだため,判読できない状態である.古写本の資料状態は,虫損と紙質の変容,修理の有無にもなって変化することには注意が必要である.

6.2.2 部首体系

篆隸万象名義の構成は部首分類であり,玉篇の部首配列を踏襲している.新撰字鏡の構成は部首分類を主に意義分類を付加するが,前者に部首の統合と新設が行われ,構成の多様性が顕著である.

新撰字鏡160部と玉篇542部とに共通する部首は134部である.新撰字鏡にのみ存在する部首は21部である.玉篇から分離した部首(準玉篇部首)は5部である.玉篇にのみ存在する部

首は 408 部である。部首をめぐる問題について、山田(1995)はつぎの四点を挙げる。

- (1) 部首立て：どのような部首を立てるか
 - (2) 部首配列：部首の配列原理は何によるか
 - (3) 部内配列：部首内の字の配列原理は
 - (4) 部首配属：字をどの部首に配属させるか
- この四つの視点から部首体系の変遷を考察するのは今後の重要な課題となる。

6.3 Unicode 対応

6.3.1 包摂規準

篆隸万象名義はほとんど単字掲出である。新撰字鏡は部首分類部分と意義分類とで項目構造に相違がある。部首分類部分は異体字関係の類似した字体・字形の掲出字が並列また連続掲出するので、文字同定の抽象度を下げる必要がある(例2「諛諛」)。これに対して、意義分類部分は熟語が大半で、文字同定の抽象度を高く設定することが可能である(例3「四阿」)。

入力済の新撰字鏡の Unicode 対応状況は、表2・表3の通りであり、篆隸万象名義と合わせて示す。新撰字鏡では柔軟な包摂規準をどのように設定するかが課題である。

表2 篆隸万象名義 (KTB)・新撰字鏡 (TSJ)

Unicode 処理状況				
DB	Unicode 漢字	IDS	その他	合計
KTB	15,872 (99.2%)	80 (0.5%)	48 (0.3%)	16,000 (100%)
TSJ	21,654 (89.2%)	14,68 (6.0%)	1,156 (4.8%)	24,278 (100%)

表3 表2の Unicode 漢字の内訳

DB	統合漢字	拡張 A	拡張 B	拡張 C
KTB	10,160 (63.5%)	2,000 (12.5%)	3,712 (23.2%)	0 (0.0%)
TSJ	15,286 (63.0%)	2,444 (10.1%)	4,010 (16.5%)	6 (0.0%)

表4 新撰字鏡 (TSJ) 符号化不能の内訳

巻	IDS	その他 ¹⁾	字数	その他の割合
1	108	58	2,046	2.8%
2	69	27	2,044	1.3%
3	73	42	1,577	2.7%
4	119	27	1,564	1.7%
5	150	41	1,848	2.2%
6	160	85	2,086	4.1%
7	156	97	2,377	4.1%
8	158	88	2,082	4.2%

1) 「その他」には記号類の「踊り字」が含まれており、これは符号化できない漢字と性質が異なるため、表4の新撰字鏡(TSJ)符号化不能の内訳から除いた。踊り字の内訳は、巻2が2字、巻12が371字ある。

9	128	68	1,552	4.4%
10	116	100	2,025	4.9%
11	148	84	2,010	4.2%
12	83	64	3,067	2.1%
計	1,468	783	24,278	3.2%

Unicode 漢字で処理できない項目は、「IDS」と「その他」で対応した。IDSは漢字構成記述文字を用い漢字の部品とその配列を指定するものである。「その他」は主に次の四種類のものを含む。

- 虫損で判読不能のもの 48 字
- 符号化できないもの 89 字
- と 踊り字 373 字
- β²⁾ 異体字 646 字

「□」は主に欠損状況によって、字画が視認不能なものである。関連文献と照合し、元来の内容を推測するしかない。

「■」は字画視認可能であるが、符号化できず、IDSの部品構成でも記述しがたいものである。

「と」は前字の重複との意味で(踊り字)使われており、記号類として、符号化できない文字と性質が異なる。

「β」の具体例は後述する。

巻七は李が担当し、それ以外は池田が担当した。担当者によって、入力方針は全同でないが、上記の詳細ではそれほど揺れが観察されない。

Unicode 漢字で処理できない異体字は、次の三つの方法で対応した。

- (1) 包摂規準の抽象度を上げる
- (2) IDS 方式
- (3) β 方式

字体記述を目的としない部分は、包摂規準を緩やかに設定し、代表字種で翻刻できるように処理した(例3「四阿」)。

字体弁別記述を目的とする部分は、「IDS」と「β方式」で対応するように処理した。例えば、次の「菌」の例(例6)は、同一項目に複数異体字が並列するので、二字目はIDSで表現する。

例6 「菌」の例

菌 菌

(新撰字鏡巻7・23 丁裏)

菌 □⁺⁺ 二同。一羊反。平。重薦皮類也。蓐也。席薦皮類曰菌。又作鞞，虎皮。

これとは別に、同一項目に複数異体字が並列する点は例6と同じだが、字画の差異が細微であるため、Unicode 漢字やIDSで表現しにくい場合もある(例7)。この場合、複数異体字を包摂してしまうと、意図的に字形を区別しようとする点を表現

2) HDIC における符号化できない異体字を表現する方法の一種であり、「β方式」と呼ぶ。「字β」の形式になり、「字」は字種代表字であり、「β」はその異体字であることを示す。実際に、異体字の種類に応じ、「β・γ・λ・δ」が使われている。「β方式」は豊島(1992)にいう diacritics(補助符号)を用いる方法である。

できない。このような異体字併記の例は、代表字とギリシア文字の組合せで示す「β方式」を用いることとした。

例7 「訓」の例

訓訓 (新撰字鏡 卷3・12丁表)

訓β訓 決運反。去: 導也, 教也, 善也, 道也, 従也, 誠也。

6.3.2 異なる辞書の連携

中国字書に対し、日本古辞書は表記・体例の面で多様性と複雑性を備えている(例8)。異なる古辞書の内容を連携するとき、複数の古辞書の間に、注文内容が一致するが、掲出字の原本字形が異なる項目の扱いが問題となる。

例8 「綌」の例

綌 (篆隸万象名義 第6帖 137丁表)

綌 祛逆反。麤也。

綌 (玉篇残卷 卷27糸部)

綌 祛逆反。毛詩為絺為綌。傳曰精曰絺, 麤曰綌。韓詩結曰絺, 辟曰綌。說文粗葛也。

綌 (宋本玉篇 下篇 58丁表)

綌 去逆切。粗葛也。

綌綌綌 (新撰字鏡 卷4・4丁裏)

綌・綌・綌 三同。去逆反。入: 麤葛布。

篆隸万象名義は掲出字を「綌」で示し、玉篇残巻も同様である。宋本玉篇は「綌」で示し、通行字体と一致する。一方、新撰字鏡では、同一項目に「綌・綌・綌」を並列する。注文の内容から同一項目であることが分かる。

このように古写本の原本字形と翻刻字体の包摂関係は古辞書のデジタル化において解決しなければならない課題であり、異なる古辞書の連携をはかる際に大きな課題となる。

6.4 典拠字書データの作成と利用

6.4.1 複数典拠

古辞書自体における出典文献の示し方について、池田(1998)は次の通りに指摘した。

先行文献の取り込み方には、出典の書名・人名を逐一明示するやり方と、明示しないやり方とがある。出典明示は、より学問的な編纂態度といえるが、これは、本文至上主義、典拠主義に立脚する。一方、出典を明示しない編纂態度は、より実用的、一般的な辞書を目指すといえる。出典明示で典拠主義の本邦古辞書は倭名類聚抄と類聚名義抄(図書寮本)であり、出典不明示で実用主義のそれは篆隸万象名義、新撰字鏡、類聚名義抄(観智院本)、字鏡(世尊寺本)、字鏡鈔である。

篆隸万象名義と新撰字鏡はいずれも出典を明示しない実用主義の古辞書であるが、篆隸万象名義は原本玉篇の節略本で単一出典に依拠したため、典拠特定に問題がなく、宋本玉篇もそのまま

利用できる。一方、新撰字鏡は複数出典に依拠しており、出典群の特定・参照の課題がある。一切経音義・広韻・玉篇系字書データを利用し、配列順序、掲出字体・音注・義注・字体注を総合した文字同定が課題となる。

6.4.2 別字併合

天治本新撰字鏡では特殊項目として、「別字併合」が存在する。Ikeda(2016)は「類似した字形の掲出字を同一の項目としているものを「併合」とする」と指摘した。

例9 「湊・湊」の例

湊 側洗反。水名。又倉候反。去: 會也, 聚也, 釐也, 最也, 進也, 盛也, 従矢也。

(新撰字鏡 卷6・15丁裏)

湊 水會也, 聚也。(広韻 轉・倉奏切 去声候韻)

湊 水名, 在河南。(広韻 臻・側洗切 平声臻韻)

点線は「湊」、直線は「湊」の説明注記である。

「湊」と「湊」とは別字であるが、類似字形のため併合した例である。このような「併合項目」の検出は課題となる。

7. 課題に対応する解決策

7.1 解決策とその進捗

前節では、篆隸万象名義と比較しながら、新撰字鏡全文テキストの構築上の課題を検討した。この節では、前述の課題に対応する現段階の解決策とその進捗を述べる。進捗状況は、A ほぼ解決済み・B 問題点を整理済みで解決策検討中・C 問題点を整理中という三段階に分けてみた。

7.2 資料特徴の解明

7.2.1 本文入力点検校正

解決策 天治本新撰字鏡は法隆寺における一切経書写事業の一環として、五名の僧侶により書写され、巻によって判読難易度の差異が出てくる。書写差異の視点からみた新撰字鏡における巻の判読難易度は次の三つのグループに分けられる。

I 巻1, 巻2, 巻6, 巻7, 巻9, 巻12

II 巻3, 巻4, 巻10, 巻11

III 巻5, 巻8

Iグループは明瞭な楷書で、虫損も少なく、判読に困難が少ない。IIグループはIよりやや困難の度合いが大きい。IIIグループは丁寧な楷書ではなく、行書体で書写され、判読に困難を感じる。

現在全文テキスト入力済みであり、公開に向けてデータの点検校正中である。解決策として次の四点がある。

- (1) 旧複製に依拠して入力した全文テキストを、新複製で確認・点検するのを基本方針とする。
- (2) 点検・校正の第一段階としてI・IIグループを対象に実施する。
- (3) 点検・校正の第二段階として、IIIグループを対象に実施する。

(4)旧複製の不鮮明箇所と新複製の不鮮明箇所を整理し、可能であれば原本調査を実施する。
進捗 A 活字翻刻のない古文献のテキスト化のモデルケースとして、電子テキスト化の工程を明示化できている、この課題はほぼ解決済みである。

7.2.2 玉篇部首体系照合

解決策 玉篇部首体系から新撰字鏡部首体系への変遷に関する考察は、部首分類体漢字字書の部首体系の日本的受容に関する重要な課題である。

HDIC では、玉篇系字書の原本玉篇残巻、篆隸万象名義、宋本玉篇はすでにデータ化済みである。新撰字鏡と照合すると、新撰字鏡部首と玉篇部首と共通するものは134部である。新撰字鏡にのみ存在する部首は21部である。玉篇分離部首は5部である。その詳細は次の通りである。

新設部首

天部、鼈部、冫部、尢部、文尻部、文下一点、止部、丘部、𠂔部、文下木点、首部、爻部、贏部、辨部（部首分類部分・14部）
 親族部、小学篇及本草木異名第六十九、小学篇字及本草木異名第七十一、雑字、重點、連字、臨時雑要字（意義分類部分・7部）

玉篇分離部首

連火部(火)、亻部(人)、犴部(犬)、扌部(手)、刂部(刀) (5部)
 *玉篇部首の表記を()で示す。

そして、部首内の掲出字レベルで確認すると、状況はさらに複雑になっている。「人部・亻部」を例にみてみよう。

玉篇系字書における「人部」の収録字数は、篆隸万象名義は457字あり、宋本玉篇は522字ある。新撰字鏡は「人部・亻部」があり、それぞれ61字と539字が収録される。Unihanは「人部」が1,025字ある。新撰字鏡「人部・亻部」の収録字と篆隸万象名義、Unihanとの字数対応の詳細を表5に示す。

表5 TSJ・KTB・Unihan「人部・亻部」

TSJ		KTB		Unihan	
人部	61	人部	4	人部	17
		人部以外	32	人部以外	32
		無	25	無	12
亻部	539	人部	387	人部	459
		人部以外	40	人部以外	35
		無	112	無	45

上記の詳細で示したように、データ照合だけでは新撰字鏡の「人部・亻部」は、部首の新設と分離が行われていると考えられる。そして、部首内の状況は、収録字の部首再配属が観察され、新撰字鏡の部首配属に独自性があると言える。

進捗 C 6.2.2 で挙げた(1)部首立て(2)部首配列(3)部内配列(4)部首配属の四点から見ると、(1)と(4)はHDICの古辞書テキストデータによって、分析可能である。(3)は出典群に関わる課題である。(2)は部首字の意味属性や字形属性等の分析に必要な情報付与が未了であり、さらなる課題になる。

7.3 柔軟な文字同定包摂規準の設定

7.3.1 異体字併記の処理

解決策 Unicode 漢字で処理できない異体字は、基本的にIDSで記述する。また、例7「訓」の例で示すように、元来字書編纂者は異体字として区別しようとする類似字体であるが、その点画の相違が分かりにくい。包摂規準を緩やかにし、字種レベルで翻字すると、異体字としての区別は表現できなくなる。そして区別して翻刻しようとしたら、対応するUnicode漢字はないため、代表字種とギリシア文字の組み合わせである「β方式」で異体字を表現せざるえない。利点としては、同じ字種であることは明確であり、翻字者の同一字種グループ分けの判読がはっきり反映できる。

進捗 A IDSとβ方式は、便宜的なものであるが、第一段階の入力方法としては有用性が認められる。もちろん、古写本字形の細部の表現に限界がある。そこで、画像システムで原本画像を併用するという方法も検討している。現在HDICにおける掲出字画像データベースは「切り出し画像」による。これはHNG（漢字字体規範史データベース）方式であり、ユーザ利用に利便性が高い。そして、許諾の問題をクリアして公開は難しくないかもしれない。しかし、技術的観点では拓本文字データベースやIIIFのようなシステムが望ましい。

7.3.2 古辞書翻刻階層モデル

解決策 異なる古辞書の連携については、翻刻階層モデルを設定・提案する。古辞書翻刻階層モデルとは、従来の研究が単一の正しい字体で翻刻してきたのに対し、原本字形を翻刻するに当って、抽象度を細かく分類し、古写本古辞書における原本字形を階層的に翻刻するモデルである(表6)。

[* HはHeadword(掲出字)の頭文字]

表6 古辞書翻刻階層モデル

H0	代表字・字種	代表字種に定めた判読結果
H1	原本字形近似字体	写本字形に近似する字体
H2	解釈字体	関連資料による判読字体
H3	原本再現字体	写本の字形を目視での結果
H4	原本字形	写本の字形

前述した例8の「絡」であれば、H0代表字・字種の「絡」の翻刻により、複数の古辞書間の異体項目の連携をとることが可能になる(表7)。

表7 翻刻階層モデルによる翻字

	KTB	YYP	SYP	TSJ
5	H0	絡	絡	絡
4	H1	絡	絡	綷・絡・𦉰系缶
3	H2	絡	絡	綷・絡・𦉰系缶
2	H3	絡	絡	綷・絡・𦉰系缶
1	H4	絡	絡	綷・絡・𦉰

また、H0~H4 のどのレベルで連携を取れるかということを経量的に記述する。上述の「俗」は、KTB と YYP は H3 のレベルで連携をとれ、数量的に 2 で記述できる。KTB と SYP は H2 のレベルで連携をとれ、数量的に 3 で記述できる。KTB と TSJ は H0 のレベルで連携をとれ、数量的に 5 で記述できる。こうすることで異なる古辞書の間の距離を計量的に考察することが可能になり、さらなる課題となる。

進捗 B この翻刻階層モデルは現在個別字例について進行中の段階であり、これから網羅的にリスト化する一覧を作成する予定である。また、現在は異体字の記述は、IDS 方式で処理しているが、β方式で入力した項目と対応がさらなる課題となる。

7.4 内容解読支援

7.4.1 出典明記対応

解決策 新撰字鏡は複数出典に依拠し出典明記がないため、出典群を特定し、それを明記することが課題となる。貞苺(1998)は、主要な出典がまとめて連続した項目として引用されていることから、新撰字鏡の全巻にわたって、一切経音義を含む部分、切韻引用部分、玉篇引用部分、出所不詳部分等を明らかにしている。現在、貞苺(1998)の成果は対応済みである。

進捗 A 新撰字鏡出典研究の基本文献である貞苺(1998)の成果をカバーし対応済みである。貞苺(1998)を部分的に修正した研究成果の上田(1984)を取り込むことは課題となる。

7.4.2 弁似字形確認

解決策 前述の例 9 の「湊・湊」のように、元来別字なのに、字形が類似するため同一項目にする例が多い。類似字形セットを作成して点検する。古写本文献で区別なしのセットは例示すると次のようなものがある。

[專-專] [崇-崇] [孟-孟] [至-至] [巾-巾]
[月-肉] [丹-舟] [角-甬] [易-易] [王-玉]

進捗 C 問題点指摘の段階であり、網羅的に類似字形セットのリストの作成と確認は今後の課題となる。

8. あとがき

天治本新撰字鏡は平安時代における漢字字書の日本化を典型的に示す日本古辞書であり、体系と構造に多様性を呈する。本稿は、その体系と構造の基本を踏まえ、全文テキスト構築に際して解決すべき課題を示した。現在、来年 3 月の公開に向けて全文テキスト・凡例・校記を準備しているが、安定的かつ長期的な公開方法の検討も必要になっている。

参考文献

- 1) 池田証壽：玄応音義と新撰字鏡，国語学 130 集，国語学会，pp.1-18 (1982).
- 2) 池田証壽：一切経音義音注索引，古辞書音義集成一切経音義索引，汲古書院 (1984).
- 3) 池田証壽：解説，新撰字鏡の研究，汲古書院，pp. 323-341 (1998).
- 4) 池田証壽：平安時代漢字字書総合データベースの構築及其公開，第四屆東亞文獻研究國際學術研討會，國際東亞文獻研究會・揚州大學尚書學研究中心，2017.4.15-16，中国・揚州 (2017).
- 5) 井野口孝：新撰字鏡「玉篇群」の反切用字，文学史研究 17-18，pp.49-62 (1978).
- 6) 上田正：新撰字鏡の切韻部分について，国語学 127，pp.13-20 (1981).
- 7) 上田正：切韻逸文の研究，汲古書院 (1984).
- 8) 上田正：玉篇反切総覧，私家版 (1986).
- 9) 内田賢徳：上代日本語表現と訓詁，塙書房(2005).
- 10) 大槻信：古辞書と和訓-新撰字鏡<臨時雜要字>-，訓点語と訓点資料 108，訓点語学会，pp.16-34 (2002).
- 11) 京都大学文学部国語学国文学研究室：新撰字鏡国語索引，京都大学文学部国語学国文学研究室 (増訂再版 1975 臨川書店) (1958).
- 12) 阪倉篤義：新撰字鏡の再検討-享和本を中心に-，本邦辞書史論叢，三省堂，pp.359-392 (1967).
- 13) 貞苺伊徳：新撰字鏡の研究，汲古書院 (1998).
- 14) 豊島正之：「JIS に無い字」をめぐる，月刊しにか 3(2)，pp.16-22 (1992).
- 15) 西原一幸：字様の研究 唐代楷書字体規範の成立と展開，勉誠出版 (2015).
- 16) 山田健三：奈良・平安時代の辞書，日本古辞書を学ぶ人のために，世界思想社，pp.68-118 (1995).
- 17) 山田孝雄：新撰字鏡攷異并索引，六合館(西東書房 1933 年改版) (1916).
- 18) 李媛・池田証壽：篆隸万象名義の全文テキストと公開システムについて，じんもんこん 2016 論文集，情報処理学会，pp.95-102 (2016).
- 19) 張磊：新撰字鏡研究，中国社会科学出版社 (2012).
- 20) Ikeda Shoji, Problems in the decipherment of the text of Sinsenjikyō: Introducing HDIC and its uses, International conference for the 20th Anniversary of the Society for Kugyol Studies, Gwanjeong Library Yangduseok Hall, Seoul National University, Seoul, Korea, October 7, (2016).
- 21) Li Yuan and Ikeda Shoji, Building a General Database System of Chinese Characters Dictionaries in Early Japan: Tenreibanshōmeigi in HDIC Project, EAJS2017, Lisbon, August 30 – September 2, (2017).

〔附記〕 本稿は、JSPS 科研費 (課題番号 16H03422 及び 課題番号 17F17301) による成果の一部である。