

篆隸万象名義の全文テキストと公開システムについて

李媛・池田証壽（北海道大学 文学研究科）

本稿は、先日公開された高山寺本篆隸万象名義データベース（平安時代漢字字書総合データベース HDIC の一部）について、そのデータベースの構築、全文テキストと公開システムを中心に述べる。内容は、篆隸万象名義の資料解説と先行研究、全文テキスト、公開システム、課題からなる。このたび、古辞書研究者での活用と文字情報処理での応用を期待してデータを公開し、本稿はそのための公開データの内容を解説するものである。

About the Full Text and Publication System of Tenrei Bansho Meigi

Yuan Li / Shoju Ikeda (Graduate School of Letters, Hokkaido University)

This paper presents an explanation of the construction, the full text and the publication system of KTB (Kosanjibon Tenrei Bansho Meigi database, which is part of Integrated Database of Hanzi Dictionaries in Early Japan, abbreviated as HDIC), of which the full text data has been recently released to the public. It consists of the summary, preceding studies, the full text, the publication system and problems in future. It is hoped that the data made public, could be made practical use by the researchers both in the domain of premodern Chinese characters dictionaries' studies, and information processing.

1. まえがき

本稿では、日本の平安時代に編纂された字書である篆隸万象名義を例にして、古写本の全文テキストの電子化とその公開システムについて、その現状と課題を報告する。

漢字古文献の研究は Unicode の普及と定着により大きな進展を見せている。漢籍・仏典の全文テキストをオンラインで検索するのは、ごく日常的なことであり、難字を多く含む康熙字典ですら簡単に検索できる。しかし、それらの依拠するテキストは、校訂された活字テキストあるいは版本であり、所用の漢字字体は康熙字典に準拠するものが用いられている。これは中国学や仏教学において宋版・元版・明版および清版、さらに高麗版や和刻本など、版本を基礎においた研究がなされてきたことによる。筆者等が研究対象とする日本古辞書は、中国撰述の広韻、玉篇、一切経音義（玄応）を参照することが多いが、広韻と一切経音義は中国版本に基づく電子テキストが利用できる。

一方、日本においては奈良時代以来の古写本が数多く伝存し、それらの影印も容易に利用可能となっている。我々が対象とする日本古辞書は、国語学と中国語学の重要資料であるが、ほとんどが写本として伝わったものである。写本の常として誤写が避けられない。体例（本文の体裁）の不統一、難字、異体字、誤字、脱字、衍字、顛倒、錯簡について、関連の典籍を参照

して校訂して行くことが不可欠である。

篆隸万象名義の本文研究の状況は次節で詳しく紹介するが、近時、中国の研究者により校訂テキストが初めて出版されるなど、関連研究が続出しているが、日本の研究成果を参照することは少ない。

このような状況の中、本年9月1日、高山寺本篆隸万象名義データベースの全文テキストを公開した。これは、日本古辞書では初めての全文電子テキスト化とその公開となる。難字の多い古写本を電子テキスト化した点と日中の本文研究成果を総合する点に特色がある。

この高山寺本篆隸万象名義データベース（Kosanjibon Tenrei Bansho Meigi, 略称 KTB）は平安時代漢字字書総合データベース（Integrated Database of Hanzi Dictionaries in Early Japan, 略称 HDIC）の一環をなすものであるが、HDIC は高山寺本篆隸万象名義のほか、天治本新撰字鏡、凶書寮本類聚名義抄、観智院本類聚名義抄を収録するデータベースであるが、それらについては池田(2014a, b)・池田ほか(2016)で詳しく報告済みである。

本稿では、Unicode による篆隸万象名義データベースの構築、全文テキストおよび公開システムを中心に述べる。特に古辞書研究と文字情報処理の上で扱いやすい TSV データフォーマットで全文テキストを提供し、利用する際に必要になるデータインフォメーションについて解説を行う。

2. 対象資料解説と先行研究

2.1 対象資料解説

篆隸万象名義は平安初期(830年頃)に編纂された日本現存最古の字書である。

内容は原本玉篇から抜粋し、六帖の構成で、約16,000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。

永久二年(1114)に書写された高山寺本が唯一の古写本である。原本玉篇は中国で早く散逸し、残巻と逸文を伝わるのみである。それゆえ、完本である篆隸万象名義は原本玉篇の元の姿を窺う資料として価値が高い。

ただ、字体の誤写や義注の誤りが多いと言われる。さらに、一見、誤写かと思われる例の中には、中国南北朝時代以来の字形の古態を残すものもあって、誤写か古い字体かを判別するのは容易ではない。高山寺本篆隸万象名義の本来の価値を見出すには、精密な校訂が要求される。

2.2 先行研究

2.2.1 先行研究の概観

これまでの原本玉篇、篆隸万象名義に関する先行研究は、岡井(1933)から始まるが、原本玉篇の逸文蒐集をさらに進めた研究成果が馬淵(1952)である。音韻研究の成果[周(1966)・河野(1979)・上田(1986)]は顕著であるが、訓詁と字体の研究は少ない。

掲出字の全体を解説した先行研究には白藤(1977)・宮澤(1977)・池田(2014a)・呂(2007)があげられる。

白藤(1977)は、高山寺古辞書資料第一に収録された高山寺本篆隸万象名義影印テキストの掲出字を検出するための索引である。索引は検字が主目的であるため、通行字体たる康熙字典体を用いながら、原本字形を重視する特徴がある。

宮澤(1977)は、篆隸万象名義の骨組みとなる掲出字を解説したうえ、高山寺本の配列形式で一覧にした研究成果である。説文解字、玉篇残巻等の先行文献とも対照し、埋字・脱字に相当するものも一覧表に反映させた。

呂(2007)は、説文解字、原本玉篇残巻、宋本玉篇等の関連文献と照合し、校勘を加え、篆隸万象名義全文を翻刻したものである。

解説積文の公刊である呂(2007)が注目を集めたが、紙媒体故に情報处理的な側面において、制約があった。それゆえ、全文テキストデータベース化が期待されていた。

池田(2014a)は、篆隸万象名義掲出字データベース UCS 対応版(2011年公開)を踏まえ、HDIC プロジェクトの構想と進捗について詳しく解説したものである。

なお、上田(1986)は、すべての反切に厳密な校訂を加えており、掲出字の同定にも参考になるところが大きい。

2.2.2 複製本刊行とその画像公開

篆隸万象名義は、空海の著作であるため、しばしば複製本が刊行されてきた。

最初の複製本である崇文叢書本(1926)は、国立国会図書館デジタルコレクションに登録されその画像が公開されている。

この他の複製本としては、『高山寺本古辞書資料第一』(1977)と『弘法大師空海全集』(1986)に収録された影印テキストがある。

2.2.3 全文テキスト公開に至る経緯

篆隸万象名義の電子テキストは、全掲出字を対象とするものが池田(1994)で公開した暫定版データベースがある。篆隸万象名義データベースは、本文研究を主要な目的とするものである。公開した暫定版を通して、篆隸万象名義に掲出される16,000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。

池田(2003)の調査により、篆隸万象名義の全掲出字のうち46字のみが今昔文字鏡(ver.3単漢字10万字版)で未登録であり、今昔文字鏡を使えばほとんどの掲出字が処理できるようになったのである、しかしながら、データ処理にはかなり制約があった。

池田(2011)では、この論文発表時点での情報処理環境では、UnicodeのCJK統合漢字は約7万の漢字が使用可能であり、篆隸万象名義の掲出字について調査したところ、ほぼすべてを扱うことが可能となったとした。

池田(2014a)は篆隸万象名義全文テキスト入力の完成を機にその概要と課題をまとめ、新撰字鏡と類聚名義抄に拡大する計画を述べた。

池田(1994)以来、古辞書のデータベース化を目指し、Unicode等の漢字情報処理に関する研究と、日本古辞書編纂に大きな影響を与えた篆隸万象名義のデータベース化に関する研究に取り組み、本年9月に全文テキストを公開したものである。

3. 全文テキスト

3.1 篆隸万象名義の体系と構造

データベース化作業を行う際、古辞書自体の体系と構造が、データベースの設計に大きな影響を与える。よって、この節では部首分類と配属・配列で同一の体系を持つ原本玉篇と宋本玉篇と比較しながら、篆隸万象名義の全体の体系と構造を説明する。

3.1.1 帖・巻から見た部首体系

高山寺本篆隸万象名義は六帖の構成で、各帖はさらに巻に分かれる。各巻は複数の部首を収める(「艸部」「木部」など、字数の多い部首は複数の巻にまたがる場合もある)。各半丁を上

下二段、左右六列に書写する体例が基本である。この枠内に篆書と隸書の大字で掲出字が掲げられ、下には注文が付されている。

11世紀に簡略した宋本玉篇も原本玉篇を受け継ぐものとして今に伝わっているものである。両書の編纂方針には類似点も多いが、形式面・内容面において相違点も少なくない。

篆隸万象名義の構成は前半の原撰部分(空海自撰による)と後半の続撰部分(他の人による)とに分かれるが、前半の原撰部分は構成において依拠した原本玉篇の三十巻と異なり、百巻に分巻する方針が確認できる。一方で、宋本玉篇は宋太宗祥符六年(1013)に、勅令で陳彭年らによって重修増補されたものである。体系上、篆隸万象名義と違って、原本玉篇の三十巻構成は保たれている¹⁾。

原本玉篇・篆隸万象名義・宋本玉篇の全体を見渡すと、次のようになる。

原本玉篇: 巻-部首-掲出字-注文
 篆隸万象名義: 帖-巻-部首-掲出字-注文
 宋本玉篇: 篇-巻-部首-掲出字-注文

原本玉篇は卷子本であり、篆隸万象名義と宋本玉篇は冊子本である。篆隸万象名義と宋本玉篇は原本玉篇の節略本となる。分量が節略された複数の巻を帖(宋本玉篇の場合は篇)になす。三書の対応関係は表1に示す。篆隸万象名義における巻の分け方は前後が異なって、前の原撰部分は細かく分巻する方針が見られる。一方で、続撰部分は、前半の編纂方針を踏襲せず、原本玉篇の構成をそのまま継承している。

表1 篆隸万象名義と宋本玉篇の帖・巻の対照表²⁾

篆隸万象名義			宋本玉篇
帖数	対応部首	巻数	巻数
一	一~目 49部	1-12	1-4
二	(目)~心 40部	13-26	(4)-7
三	(心)~中 70部	27-38	8-11
四	木~禾 38部	39-50	12-15
五	(禾)~氏 148部	15-21	(15)-21
六	山~亥 197部	22-30	22-30

1)この特徴から残巻となる原本玉篇の体系を考察する際に、代わりに宋本玉篇を用いる。

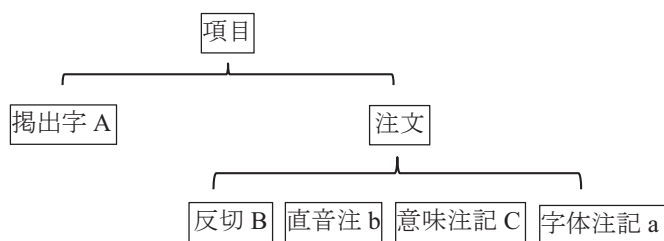
2)()を付したものは該当部首・巻の後半部分を示す。

3.1.2 部首内の項目の配列構造

各部首における掲出項目の配列の順序は、①最初に説文解字によるものを説文解字の出現順に掲出し、②その次に説文解字に見えないものを配列する。宋本玉篇では、①と②の後に、説文解字、原本玉篇(=篆隸万象名義)に掲載のないものが増補されるのが原則のようである。説文解字と玉篇との部首体系は異なるので、事情は簡単ではない。篆隸万象名義の掲出項目の配列の順序が確定しているわけではない。篆隸万象名義の本文研究から言えば、項目の配列順序を検討するためにデータベース化が必要という研究段階である。

3.1.3 項目構造

篆隸万象名義の項目構造は比較的単純である。



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

掲出字、反切³⁾、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。次の「誑」で解説しよう。

A 誑 B 俱放反。C 欺也、誤也。a 慙字。
 (篆隸万象名義 第3帖13丁裏)

掲出字「A 誑」そのものが字体を示し、反切「B 俱放反」は字音を説明するものである。「C 欺也、誤也」は単字による義注である。「a 慙字」は異体字の説明である。

関連して、説文解字・原本玉篇・宋本玉篇では「誑」の項目は次の通りである。

誑 欺也。从言狂聲。(説文解字巻3言部)
 誑 俱放反。《國語》: 天又誑之。賈逵曰: 誑猶惑也。《左氏傳》: 是我誑吾兄。杜預曰: 誑欺。《爾雅》: 佞張誑。郭璞曰: 《書》云無或佞張為眩, と欺誑人也。《聲類》: 或為慙字。在心部。(原本玉篇巻9言部)

3)二字の組み合わせで音を表す方法である。反切上字の声母と、反切下字の韻母および声調を組み合わせ、その漢字の音を表す。

誑 俱放切。惑也，欺也。
(宋本玉篇上篇 84 丁表)

義注の「欺也」は説文解字から，原本玉篇・篆隸万象名義・宋本玉篇まで継承するものであり，反切「俱放反」は原本玉篇から伝わった音注である。「誑」は説文解字に存し，原本玉篇では反切による音注，典拠による例文・意味注記を加え，篆隸万象名義・宋本玉篇まで伝わった掲出項目である。

篆隸万象名義をデータベース化し，本文を研究することは，大部分が逸文となっている原本玉篇の原貌を再構する一助になると考えられる。よって，原本玉篇と宋本玉篇との比較した情報もデータベースに加える（詳細後述）。

3.1.4 掲出字の分類

篆隸万象名義は書名の「篆隸」に示されるように，その掲出字は本来篆書と隸書により掲げられるべきものである。ただし，現存の高山寺本では篆書掲出字が約 6%しか存せず，隸書掲出字のみの項目が主となっている。

高山寺本篆隸万象名義が書写されるまでに，掲出字の脱落，誤写，重出，さらに掲出字が注文に繰り込まれるといった問題が生じた。したがって，篆隸万象名義の掲出字を扱う際に，大字で掲げられる掲出字を対象とすることだけではなく，脱落した掲出字（脱字）も，注文に繰り込まれる掲出字（埋字）もすべて視野に入れるべきである [李(2015)]。

この点を考慮し，篆隸万象名義の掲出字を四つに分類した。

隸書掲出字（約 88%）・埋字（約 4%）・脱字（約 3%）の三つを基本の分類とする。これらに篆書か隸書かの情報を加えて掲出字の下位分類を行った（後述の Group a, b, c）。

残りは補足的な分類で，宋本玉篇で増補された異体字（約 5%）を加えた（後述の Group d）。これは玉篇系字書である原本玉篇と宋本玉篇との関連を考えるための措置である。

これらの四分類を次のように説明した。

Entry_type

Group a -- Regular headword

- 1 Regular (= Regular_clerical)
- 2 Regular_seal

Group b -- Embedded headword

- 3 Embedded_clerical
- 4 Embedded_omitted
- 5 Embedded_seal

Group c -- Omitted headword

- 6 Omitted
- 7 Omitted-regular

Group d -- Variant of the Songben Yupian 8 Songben-Yupian

Group a には，基本的に篆隸万象名義では隸書大字で掲げられる掲出字を含めた。したがって，英語表記は Regular headword とした。篆書掲出字の有無で 2 種類に分類した。さらに，書体を付加情報として加えた。

- 1 Regular_clerical は隸書掲出字のみのものを指す。
- 2 Regular_seal は篆書掲出字も併存するものを指す。

隸書は，本来 clerical とするべきであるが，最も一般的な掲出字タイプであるので，書体情報を省略して Regular とした。

Group b には，基本的に篆隸万象名義では注文に繰り込まれる掲出字（埋字）を含めた。したがって，英語表記は Embedded headword にした。さらに，書体を付加情報として加えられ，篆書は seal，隸書は clerical とした。埋字は脱落した場合，omitted で注記した。

- 3 Embedded_clerical は隸書で示す埋字を指す。
- 4 Embedded_omitted は見出しが脱落した埋字を指す。
- 5 Embedded_seal は篆書で示す埋字を指す。

Group c には，基本的に篆隸万象名義では脱落した掲出字（脱字）を含めた。したがって，英語表記は Omitted headword にした。掲出字のみ脱落する場合（注文存），regular で注記した。

- 6 Omitted は諸文献に本文が存し，元来篆隸万象名義にもあったと考えられる掲出字を指す。
- 7 Omitted_regular は掲出字のみ脱落（注文存）した掲出字を指すのである。

Group d には宋本玉篇で増補された異体字(8 Songben-Yupian)を含めた。

3.2 翻刻

3.2.1 手書きによる従来成果公表

現在のパソコン環境では，Unicode が実装され，多漢字文献の処理に対応している。しかし，従来研究成果公表は手書きだった。

篆隸万象名義の掲出字についての認定である白藤(1977)・宮澤(1977)・上田(1986)はいずれも手書きで公開された。

3.2.2 Unicodeの普及と利用

2001年のUnicode3.1以降、CJK統合漢字が拡張Bを追加し約7万字を扱えるようになった。近年、Unicodeの基本多言語面のCJK統合漢字・拡張A~Eにより7万を超える漢字が処理可能となった[Lunde(2008)]。これを背景に、海外では説文解字・広韻等の古い辞書のオンライン版が続々公開され、呂(2007)は篆隸万象名義の活字翻刻として出版された。

3.2.3 KTBにおけるUnicode対応

KTBの入力作業もUnicodeで扱える漢字の範囲で進めた。翻刻方針は、康熙字典体を基準とした。一部、とくに原本において字体の差異を問題とする場合、原文の字体を採用する。Unicodeで符号化できない文字はIDS (Ideographic Description Sequence, 漢字構成記述文字列)で処理した。すなわち、IDC (漢字構成記述文字, □□□□□□□□□□□□□□□□, U+2FF0~U+2FFB)を利用し、漢字部品の合成により漢字を表現したのである。

実際にHDICでは、Unicodeで扱える漢字の全掲出字に占める割合は、篆隸万象名義99.2%・宋本玉篇99.8%・新撰字鏡89.2%となる。処理できない漢字は篆隸万象名義で128字、宋本玉篇で46字、新撰字鏡で2,592字となる(表2)。表3はUnicode漢字の内訳である。

表2 篆隸万象名義(KTB)・宋本玉篇(SYP)・新撰字鏡(TSJ)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%)
SYP	22,954 (99.8%)	23 (0.1%)	23 (0.1%)	23,000 (100%)
TSJ	21,408 (89.2%)	1,512 (6.3%)	1,080 (4.5%)	24,000 (100%)

表3 表2のUnicode漢字の内訳

DB	統合漢字	拡張A	拡張A	拡張C
KTB	10,160 (63.5%)	2,000 (12.5%)	3,712 (23.2%)	0 (0.0%)
SYP	13,386 (58.2%)	3,243 (14.1%)	6,325 (27.5%)	0 (0.0%)
TSJ	15,048 (62.7%)	2,400 (10.0%)	3,960 (16.5%)	5 (0.0%)

3.2.4 文字の解説

高山寺本篆隸万象名義の本文解説上で課題となるのは、上述の難字・僻字の処理だけではない。高山寺本の中では誤写・誤脱が多く存在するのである。同時に中国南北朝以来の字形の古態を残すものもあって、本文内容、関連文献を確認しながらの文字解説が必須となる。「隊」の例で詳細を見てみる。



隊の例

隊 階古諧也。道也，梯也，陞。
(篆隸万象名義第6帖24丁表)

この項目は、原本画像で確認したところ、形式的に見る限り、篆隸万象名義ではごく普通な項目である。だが、注文内容を読んでみると、不自然な点があることに気づいてくる。この点について、内容・字順・字形の三方面からみることにする。

内容：原本玉篇・宋本玉篇の内容と見比べ、注文は「階」に相応しい説明であり、注文の1字目「階」は掲出字に相当する。「古諧也」は反切の誤写である。従って、原本に掲出字の位置にある「隊」の字は異常とみられる。

字順：この箇所について、原本玉篇残巻、宋本玉篇、篆隸万象名義自身の前後の掲出字の配列順を比較すると、

原：…階・除・階・阡…
宋：…階・除・階・阡…
万：…階・除・隊(階)・阡…

ここでは、三書における字順は一致しているように見える。

字形：「隊」は現在「隊」の簡体字として使われており、字形が「隊」としていつ頃から使われていたのかは不明である。この字は説文解字・原本玉篇残巻・宋本玉篇・新撰字鏡等の古い時代の文献では出現していない。

よって、この位置に「隊」という掲出字が書かれた原因を考えてみると、その前後の掲出字「除」か「阡」かの傍の部分で、最初の筆画においていずれも「人」との類似により、誤写で「隊」になり、掲出字を修正するより(紙面の清潔さを考慮したか)、注文の1字目に書き添えた可能性は否定できない。すると「隊」は衍字となる。したがって、この項目本来の掲出字は「階」と認定すべきである。

4. 公開システム

4.1 公開の準備と方法

4.1.1 所蔵者の許諾

篆隸万象名義データベースの公開は、高山寺典籍文書総合調査団(代表者：石塚晴通北海道大学名誉教授)の配慮・指導の下に高山寺当局から許諾を得た。さらに、漢字字体規範データベース(HNG)との連携の許諾も得た。

- ①全文翻刻テキスト
- ②掲出字画像
- ③HNGとの連携

この報告は①に関するものである。②は、掲出字画像を公開し、IDS 検索を可能とするシステムを念頭においている。③は HNG のデータは、CHISE でも検索可能となっており、どのように連携をはかればよいか、検討中である。

4.1.2 公開計画

篆隸万象名義データベース (KTB) は平安時代漢字字書総合データベース (HDIC) の一環となすものである。

HDIC プロジェクトの全体のホームページは <http://hdic.jp> である。主に次の情報を示す。

- ・進捗・イベント等の最新情報
- ・プロジェクト全体の紹介
- ・編纂委員会の構成
- ・関連する論文、学会発表等の情報
- ・関連の漢籍・仏典のソース、ツール等へのリンク

篆隸万象名義のテキストを公開する形式は次の三種類がある。

- (1) テキストを一覧するウェブサイト

<http://hdic3.let.hokudai.ac.jp/ktb/>

- (2) 部品による掲出字画像の検索システム

<http://hdic2.let.hokudai.ac.jp/ktb/>

- (3) 最新版の全文テキスト

<https://github.com/shikeda/HDIC>

(1) の担当者は李媛で、TSV 形式のファイルを HTML 化して部首ごとに掲出字と注文を通覧できるようにしたものである。

(2) の担当者は劉冠偉で、IDS による部品検索と掲出字画像の表示を可能としたシステムである (詳細は未発表)。

(3) の担当者は池田で、TSV 形式のファイルを掲載し、順次構成をすすめているものである。最初に掲載したファイルは、原文に比較的忠実に翻刻する方針のもとに作成されているが、これに校訂を加え、明らかな誤字、脱字、衍字等を修正したものに作り変えている。

4.1.3 ライセンス

Github はソフトウェア開発プロジェクトのための共有ウェブサービスである。オープンソースのライセンスが要求される。HDIC のデータ公開は改善の共有に配慮する公開である。そのため概ね、公開されているデータを自由に複製・配布することが可能であり、メールで報告された意見・問題点等は修繕される。今後は、次のように GPLv2 の明記を考えている。

This data is distributed under GPLv2.

Send bug reports and feature requests via email.

4.2 データフォーマット

4.2.1 TSV

データフォーマットについては、TSV データフォーマットを採用した。

TSV (Tab-Separated Values) とは、複数の項目で構成されるデータを複数件列挙して表現することができる汎用的なデータ形式の一つで、項目間の区切り文字にタブ文字を用いるものである。

データを文字の連なりとして表現するテキストデータおよびテキストファイルの形式であり、複数の項目をタブ文字で区切って一件のレコードを構成し、複数件のレコードを改行文字で区切ってデータ全体を構成する。

TSV を採用したのは次の三点の理由である。

- (1) 篆隸万象名義は単字字書であり、比較的単純な項目構造をもつこと。
- (2) 古辞書研究および文字情報処理上では汎用性の高いデータフォーマットであること。
- (3) 古写本の古辞書の本文解読はマークアップ可能な精度に到達していないこと。

TSV はテキストファイルとして扱うことができるため、テキストエディタなどで手軽に編集することができ、大容量のファイルも比較的開きやすいといった利点がある。TSV は、すなわち、古辞書研究と文字情報処理の上で扱いやすいデータフォーマットと言える。

古辞書を研究資料とする人文系研究者は、エクセルで処理することが多く、本文解読の内容に関する意見を得やすいと考えたからである。

4.2.2 TSV インフォメーション

より多くの国・地域の研究者に参考に供するため、公開データの先頭には次のように、TSV Format Information を付した。01~10 の番号は、説明の便宜ため付けた。以下解説を行う。01 から 10 までの公開情報は、その内容によって次の三つのタイプに分けられる。

I. 基本情報: 所在及び体系・構造の情報 [01~03], 篆隸万象名義の掲出字と注文を翻刻したテキスト [04・07].

II. 関連情報: 玉篇系字書 (原本玉篇と宋本玉篇) の対応所在情報. [08・09]

III. 校勘情報: 作成者による掲出字の分類, 諸家による掲出字同定の相違, 校勘意見. [05・06・10]

I は篆隸万象名義に関する本文情報である。

II は関連字書の対応する対照情報である。

III は先行研究, 諸関連文献の内容を総合した上での作成者の校勘情報である。

TSV Format information:

- 01 TBID (v_www_xyz) : Book(v), leaf(www), recto-verso(x), line(y) and number(z)
- 02 TB_vol_radical (xx#yyy): Volume(xx) and radical number(yyy)
- 03 TB_radical : Radical of Chinese character
- 04 Entry: Headword
- 05 Entry_type: For details, refer to the following section
- 06 Entry_diff: Differences of transliteration with other scholars
- 07 TB_def: Definition of pronunciation, meaning and variant(s)
- 08 SYID (vwwwxyyyz): Book(v), leaf(www), recto-verso(x), line(yy) and number(z)
- 09 YYID (Ywwwxyy-z): Volume(ww), leaf(xxx), line(yy) and number(z)
- 10 TB_remarks: Editor's notes

解説 :

- 01 TBID は該当掲出字 ID を示す。篆隸万象名義における所在の帖数 v ・丁数 www ・表裏 x ・列数 y ・字順 z を v_www_xyz の形式で示す。TB は篆隸万象名義の略称。
- 02 TB_vol_radical は篆隸万象名義における該当掲出字の巻数 xx と部首番号 yyy を xx#yyy の形式で示す。
- 03 TB_radical は該当掲出字の篆隸万象名義に所属する部首の部首代表字である。
- 04 Entry は該当掲出字の翻刻された字体を示す。データベース作成者の判断が含まれる。本文は Unicode で扱える漢字の範囲で入力した。符号化されてない漢字は IDS 方式で表現する。
- 05 Entry_type は該当掲出字の分類である。詳細は 3.1.4 を参照。
- 06 Entry_diff は該当掲出字の判読について、諸家の認定意見と分かれるときに示すものである。主に白藤(1977b)・宮(1977)・

上田(1986)・呂(2007)を参照した。

- 07 TB_def は該当掲出字の注文内容を翻刻したものである。注文内容の音注・義注・字体注の解説を示すために、句読点を付し、書名に括弧を付す。
- 08 SYID は該当掲出字が宋本玉篇データベースに対応するものの ID を示す。宋本玉篇(SY)における所在を篇数 v ・丁数 www ・表裏 x ・列数 yy ・字順 z の順に示す。SY は宋本玉篇の略称。
- 09 YYID は該当掲出字が原本玉篇データベースに対応するものの ID を示す。原本玉篇残卷(YY)におけるの巻数 ww ・頁数 xxx ・列数 yy ・字順 z の順に示す。YY は原本玉篇の略称。
- 10 TB_remarks はデータベース作成者による総合的な校勘意見である。

4.2.3 データサンプル

例えば、掲出字「哥」(表 4) と「船」(表 5) の TSV データの詳細をみてみよう。

表 4 「哥」の TSV データの詳細

01	TBID	3_024_A31
02	TB_vol_radical	v9#96
03	TB_radical	可
04	Entry	哥
05	Entry_type	Regular
06	Entry_diff	
07	TB_def	古何反。詠言也。
08	SYID	a088a031
09	YYID	Y09a-21-12-2
10	TB_remarks	詠言：據尚書本文「詩言志，歌詠言」。詠字，尚書作永。

- 01 第 3 帖 24 丁表 3 列の 1 字目 (所在)
- 02 卷 9 ・ 部首 96 番目 (巻数・部首番号)
- 03 可部 (部首)
- 04 哥 (掲出字)
- 05 隸書掲出字 (掲出字タイプ)
- 06 諸家認定に異同なし (先行研究との照合)
- 07 古何反。詠言也。(注文内容)
- 08 対応する宋本玉篇の所在は上篇 88 丁表 3 列 1 字目 (関連字書所在)
- 09 対応する原本玉篇の所在は巻 9 の 21 紙 12 列 2 字目 (関連字書所在)
- 10 詠言：據尚書本文「詩言志，歌詠言」。詠字，尚書作永。(校勘意見)

表5「船」のTSVデータの詳細

01	TBID	5_077_B32
02	TB_vol_radical	v18#283
03	TB_radical	舟
04	Entry	船
05	Entry_type	Regular
06	Entry_diff	宮
07	TB_def	時專反。舟。
08	SYID	b069b072
09	YYID	Y18a-17-09-1
10	TB_remarks	

- 01 第5帖77丁裏3列の2字目(所在)
 02 卷18・部首283番目(巻数・部首番号)
 03 舟部(部首)
 04 船(掲出字)
 05 隸書掲出字(掲出字タイプ)
 06 宮澤(1977)の認定字「船」と異なることを示す。(先行研究との照合)
 07 時專反。舟。(注文内容)
 08 対応する宋本玉篇の所在は中篇69丁裏7列2字目(関連字書所在)
 09 対応する原本玉篇の所在は巻18の17紙9列1字目(関連字書所在)
 10 なし(校勘意見)

5. あとがき

古写本の古辞書の文字情報処理は、その前提として正確な翻刻本文が必要不可欠である。

日本の古辞書研究では、その土台となった中国字書、とくに玉篇系字書である原本玉篇残卷、篆隸万象名義、宋本玉篇との対照作業が不可欠となっている。これまでの古辞書データベース化では、玉篇系字書のデータベース化を疎かにしていた。

HDICは古辞書翻刻・入力作業の効率化をはかるため、まず古版本である宋本玉篇をデータ化し、これを土台に高山寺本篆隸万象名義の入力作業も完成した。本年4月に宋本玉篇、9月に篆隸万象名義の全文テキストデータを順次に公開した。また、研究の国際対応を考慮し、これまで国際学会での発表を行い [Li(2015)・池田(2015)・池田(2016)], 公開データには英文凡例と中文校記を付け加えている。公開したデータ、古辞書研究上の活用と文字情報処理での応用を期待する。

また、公開したテキストデータに基づき、IDS部品検索による字体研究・マークアップ言語のXMLによる本文の構造化・国会図書館(NDL)デジタルコレクションの画像データとの連携などは今後の課題にしたい。

参考文献

- 1) 池田証壽:篆隸万象名義データベースについて, 国語学 178, pp.60-68(1994).
- 2) 池田証壽:篆隸万象名義データベースの改訂, 漢字文献情報処理研究, 漢字文献情報処理研究会編 4, 好文出版, pp.4-11(2003).
- 3) 池田証壽:篆隸万象名義データベースの整備と問題点, 平成二十二高山寺典籍文書総合調査団研究報告論集, 高山寺典籍文書調査団, pp.46-60(2011).
- 4) 池田証壽: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, pp.79-90 (2014a).
- 5) 池田証壽:平安時代漢字字書総合データベース—現状と課題 2014 夏—, 漢デジ 2014: デジタル翻刻の未来, 京都大学人文科学研究所附属東アジア人文学情報学センター編, pp.3-43(2014b).
- 6) 池田証壽:佛經音義與日本古字書, 佛經音義研究: 第三屆佛經音義國際學術研討會論文集, 上海辭書出版社, pp.53-62(2015).
- 7) 池田証壽:新撰字鏡本文解説上の諸問題—HDICの紹介とその活用—, アジア諸民族の文字 2, 口訣学会, pp.3-18(2016).
- 8) 池田証壽・李媛・申雄哲・賈智・斎木正直:平安時代漢字字書のリレーションシップ, 日本語の研究 12(2), 日本語学会, pp.68-75 (2016).
- 9) 上田正:玉篇反切総覧, 私家版 (1986).
- 10) 岡井慎吾:玉篇の研究, 東洋文庫論叢第 19, 東洋文庫 (1933).
- 11) 河野六郎:玉篇に現れたる反切の音韻研究, 河野六郎著作集 2, 平凡社(1979).
- 12) 白藤禮幸:篆隸万象名義掲出字索引, 高山寺古辞書資料第一, 東京大学出版会, pp.369-474 (1977).
- 13) 馬淵和夫:玉篇佚文補正, 東京文理大学国語国文学会紀要 3号, 東京教育大学文理科大学国語国文学会, pp.1-153(1952)
- 14) 宮澤俊雅:篆隸万象名義掲出字一覧表, 高山寺古辞書資料第一, 東京大学出版会, pp.497-635(1977).
- 15) 李媛:埋字と脱字—篆隸万象名義の掲出字数をめぐる問題—, 訓点語と訓点資料 135, 訓点語学会, pp.37-56(2015).
- 16) 呂浩:篆隸萬象名義校釋, 学林出版社(2007).
- 17) 周祖謨:萬象名義中之原本玉篇音系, 問学集, 中華書局(1966).
- 18) Li Yuan: The Creation of a Tenrei Banshō Meigi Database and its Textual Study, 9th Conference of the European Association of Chinese, Stuttgart University (Germany), (2015).
- 19) Lunde, Ken: *CJKV Information Processing*, 2nd Edition. O'Reilly Media, Inc.(2008).

[附記] 本稿は、JSPS 科研費(課題番号 16H03422)による成果の一部である。