

## 平安時代漢字字書総合データベースの拡張と和訓対応

劉 冠偉・李 媛・池田 証壽  
北海道大学

我々の研究室では、日本の平安時代（中古）に編纂された漢字字書である『篆隸万象名義』・『新撰字鏡』・『類聚名義抄』を総合したデータベースを構築している。中国撰述の『玉篇』系字書のデータベースを独自に構築し、それを土台として、公開済みの『説文解字』や『広韻』データをも参照して、対象字書の本文入力と校正を進めている。これらのデータは漢字データが中心であり、和訓（日本語）のデータが未整備である。平安時代漢字字書の和訓の同定の方法として、室町時代（中世）、江戸時代（近世）、明治時代（近代）の漢和字書のデータを追加すれば、より効率的かつ正確に行うことができると考えられる。中世は『和玉篇』、近世は『増続大広益会玉篇大全』（毛利貞斎）、近代は『大字典』（上田萬年）を拡張の候補としている。この発表では、まず、平安時代漢字字書総合データベースの現状を述べ、次に、試みに拡張した『大字典』データベースの概要を報告する。

### Expansion of Integrated Database of Hanzi Dictionaries in Early Japan and its Combination with Japanese Readings

Guanwei LIU    Yuan LI    Shoju IKEDA  
Hokkaido University

A General Database with the name of Integrated Database of Hanzi Dictionaries in Early Japan (abbreviated as HDIC) is being constructed by the members of the IKEDA Laboratory at Hokkaido University. The database includes the *Kōzanjibon Tenrei Banshō meigi* 高山寺本篆隸万象名義, *Tenjibon Shinsen Jikyō* 天治本新撰字鏡, *Zushoryōbon Ruiju Myōgi shō* 図書寮本類聚名義抄, *Kanchi'inbon Ruiju Myōgi shō* 観智院本類聚名義抄. A database of the dictionaries which are compiled in China, including *Yupian* 玉篇 and *Daguan yihui Yupian* 大広益会玉篇, is also being independently constructed in the laboratory. This Medieval China Dictionaries Database operates as a foundation for the HDIC. By comparing this data with the *Shuowen jiezi* 説文解字 and *Guanyun* 広韻, which have already been made public, we continue to proofread the database and input new data. The data discussed above is current focused on the Chinese characters, while the data for the Japanese readings is still under development. By supplying data from Chinese-Japanese character dictionaries from Muromachi 室町 (Medieval), Edo 江戸 (Early modern), and Meiji 明治 (Modern) periods, it is thought we can identify the Japanese readings for Hanzi Dictionaries in Early Japan with greater accuracy and efficiency than otherwise possible. The possible candidates this expansion are the *Wagokuhen* 和玉篇 for the Medieval era, *Zōzoku dai kōekikai gyokuhen taizen* 増続大広益会玉篇大全 (Teisai Mōri) for the Early Modern era, and the *Daijiten* 大字典 (Kazutoshi Ueda) for the Modern era. In this presentation we will report on the current situation of the HDIC, and the outline of the attempted *Daijiten* extension.

## 1 平安時代漢字字書総合データベースの現状と整備・拡張計画

### 1.1 現状と進捗状況

平安時代漢字字書総合データベース (Integrated Database of Hanzi Dictionaries in Early Japan, HDIC) とは、平安時代の日本で編纂された漢字字書を総合したデータベースである。その概要は池田 (2014) で述べたが、要点を繰り返せば次の通りである。

対象とする漢字字書は次の四点である。

- ・高山寺本『篆隸万象名義』空海撰、9世紀初、約 16,000 字
- ・天治本『新撰字鏡』昌住撰、10世紀初、約 21,000 字
- ・図書寮本『類聚名義抄』撰者未詳、11世紀初、約 3,600 項目
- ・観智院本『類聚名義抄』撰者未詳、12世紀初、約 32,000 項目

これらの漢字字書を解読するには、中国撰述の玉篇系字書を総合することが必要となるので、次の二つの字書のデータベースを独自に作成した。

- ・『玉篇』梁・顧野王撰、543年、現存 2,087 字 (原本『玉篇』)
- ・『大広益会玉篇』宋・陳彭年等撰、1013年、約 22,800 字 (宋本『玉篇』)

これらを土台として、公開済みの『説文解字』と『広韻』のデータ<sup>1</sup>を参照し、対象字書の本文入力と校正を進めている。原本『玉篇』、宋本『玉篇』、『篆隸万象名義』、『新撰字鏡』の順に本文入力作業は終了しており、現在、整備点検を進めている。これらの入力済みのデータは漢字データが中心であったが、二種の『類聚名義抄』は和訓 (日本語の訓読み) のデータが大量含まれており、それらの入力と整備が課題である。

### 1.2 整備と拡張計画

#### 1.2.1 漢字字書の項目構造

『篆隸万象名義』と宋本『玉篇』は比較的単純な項目構造なので、これらを例にしよう。参考として、原本『玉篇』も例示する。

『篆隸万象名義』と宋本『玉篇』は、いずれも梁の顧野王の原本『玉篇』の長文の注文を簡略したものである。原本『玉篇』は經典正文とそれらに対する訓詁注記を豊富に引用した字書であるが、中国では早く散逸し、その残巻が日本に伝存する。この三書は玉篇系字書と位置付けられる。

漢字字書の基本的な項目構造は、掲出字と注文に分けられる。注文は、音注・義注・字体注の三部分からなる。これらの関係を示すと図1になる。

『篆隸万象名義』と宋本『玉篇』の実例を挙げれば次の通りである。A 掲出字は「二重線」・B 音注は「一重線」・C 義注は「破線」・D 字体注は「波線」で示す。

- A 誑 B 俱放反。C 欺也、誤也。D 𠄎字。  
 (『篆隸万象名義』第3帖13丁目裏)
- A 誑 B 俱放切。C 惑也、欺也。  
 (宋本『玉篇』上84丁表)

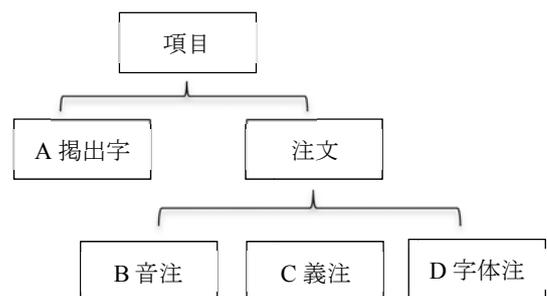


図1 漢字字書の基本的な項目構造

このように、『篆隸万象名義』と宋本『玉篇』は項目構造が単純であることが分かる。『新撰字鏡』もこれらに準じる項目構造である。一方、原本『玉篇』は、次のように經典正文・訓詁注記が豊富のため、より複雑な義注構造を持つ。

- A 誑 B 俱放反。C 《國語》：天又誑之。賈逵曰：誑猶惑也。《左氏傳》：是我誑吾兄。(中略) D 《聲類》：或為𠄎字。在心部。(原本『玉篇』卷9言部)

<sup>1</sup>漢字データベースプロジェクト <http://kanji-database.sourceforge.net/index.html>.

さらに、図書寮本『類聚名義抄』・観智院本『類聚名義抄』は漢字注記以外に、万葉仮名・片仮名による和訓をも記載されており、複雑な様相を呈する。

### 1.2.1 整備・拡張の目的

平安時代漢字字書総合データベースは、対象とする字書の全文テキストデータベースを構築することによって、それらに収録された字音・字訓・字体等を考察するための基礎資料としてすることを目指している。漢文注文部分の入力と整備は見通しが得られたので、今後は和訓と字音の整備拡張が課題となっている<sup>2</sup>。

### 1.2.2 整備と拡張の方法

まず、『新撰字鏡』の和訓を整備する。底本となる天治本には誤写・誤脱が多いので、異本及び従来の解説結果を参照しながら、入力と校正を行う。図書寮本『類聚名義抄』もこれに準じる。

次に、観智院本『類聚名義抄』は和訓の量が膨大であり、効率化がはかるため、参照データを準備して、データベースを拡張する必要がある。その拡張は、他字書の和訓データの取り込みによって行う。取り上げる他の字書としては、室町時代（中世）、江戸時代（近世）、明治時代（近代）からそれぞれ代表的な漢和字書の一つずつ選ぶこととした。中世は『和玉篇』、近世は『増続大広益会玉篇大全』（毛利貞斎）、近代は『大字典』（上田萬年）を候補としている。これらの日本字書を参照するのは、漢文注文の入力に際して、中国字書を参照したのと同じ考え方に基づく。

現在、上田萬年等によって編纂された『大字典』（初版 1917）について、その和訓データ（品詞情報を含む）を整備している。

## 2 『大字典』による拡張

### 2.1 『大字典』とは

『大字典』は収録の約 1 万 8 千の掲出字を、『康熙字典』と同じ部首画数順で配列する。明治以降刊行された他の漢和辞典と比べると、国語学者による編纂であること、親字に通し番号を付けること、重要視する和訓にその品詞を示すことなどの特徴がある。

『大字典』のすべての掲出字を通し番号に基づいてデータベースを作成し、これを『『大字典』データベース』と名付けている。高田（2002）はその概要を紹介し、高田（2003）は全体を一覧しているが、符号化はユニコードの拡張 A までにとどまっている（図 2 を参照）。

また、重要視した和訓の調査は管見に入らない。

字種	大字典	字形	部首	文字	x0208	UCS	大漢和	備考
00300	300	【倚】	9			4F9C	606	
00301	301	【依】	9	依	16-45	4F9D	607	
00301x	301x	【佹】	9			4F84	579	
00302	302	【御】	9			4F9E	610	
00303	303	【佻】	9				617	
00305	305	【佻】	9				609	
00312	312	【侮】	9	侮	41-78(U98)	4FAE	630	

図 2 『大字典』データベース 高田（2003）

### 2.2 掲出字の記載形式と項目構造

高田（2002）は掲出字の記載形式は次の四種類にわけて、その字数を算出している。

- ① 【】 で括られた番号付きの掲出字…14,860 字
- ② 【】 で括られた番号なし掲出字…550 字
- ③ □ で括られた番号なし掲出字…1,886 字
- ④ 項目内掲出字…820 字

<sup>2</sup>池田（2014）では、平安時代漢字字書総合データベースを言語史に活用するには、和訓と字音の研究に資する設計が必要であり、字音データの取り込みや和訓と義注（訓詁）との関連付けが、文字同定・本文研究に必要な基礎作業であると述べた。

①が原則であるが、これ以外の形式として②と③がある。④【】で括られた掲出字の項目内にある「俗字」「略字」と書いているものであり、掲出字に準じて扱う。

各項目は掲出字の字形・字音・字義を示し、次に熟語とその説明を示す。これが基本的な項目構造である。

字形には字源、異体字など、字音には漢音、呉音など、字義には訓義、名乗りなどが記載されている。訓義のところでは、編者が重要視する訓をゴシック体にし、その下に品詞を付ける。この重要視する訓を以下では「重要和訓」また単に「和訓」と呼ぶ<sup>3</sup>。

## 2.3 データベースの構成

我々の作成する『大字典』データベースは、次の五つのテーブルによって構成したい。

- 1 掲出字テキストテーブル
- 2 和訓付き掲出字テキストテーブル
- 3 和訓テキストテーブル
- 4 字音テキストテーブル
- 5 熟語テキストテーブル

掲出字テキストテーブルは、高田（2003）の『『大字典』データベース』を土台とし、符号化の範囲をユニコードの拡張 D まで拡大してすべての掲出字を符号化する。

和訓付き掲出字テキストテーブルは、作成した掲出字テキストテーブルに和訓を順次追加して作成する。

和訓テキストテーブルは、和訓付き掲出字テーブルから和訓を一つのレコードとして抽出して品詞と掲出字の対応を作成する。

字音テキストテーブルは、作成した掲出字テキストテーブルに字音を順次追加して作成する。

熟語テキストテーブルは、作成した掲出字テキストテーブルに熟語を順次追加して作成する。

これらのテキストテーブルは Excel で入力、それらのリレーションシップは MySQL を使う。

現在の進捗は2の和訓付き掲出字テキストテーブルと3の和訓テキストテーブルが作成済みである。次にその詳細を述べる。

## 2.4 データベース作成の詳細

### 2.4.1 作成前の準備

効率的に入力するために、対象とする底本をスキャンニングする。スキャンニングして作成した PDF ファイルをパソコンの画面で参照しながら必要な情報を入力する。

底本とする『大字典』は、高田（2003）の『『大字典』データベース』と同じく『大字典』普及版初版（1965）とする。

### 2.4.2 和訓付き掲出字テキストテーブルの作成

和訓付き掲出字テキストテーブルは Excel で次の四つのフィールドに入力する。

DJT_ID	『大字典』番号（擬似番号を含む）
DJT_radical	掲出字の部首
Entry	掲出字
UTF	掲出字のユニコード 16 進コード番号

高田（2003）『『大字典』データベース』の資料を OCR によりテキスト化する。テキスト化は『大字典』番号（DJT\_ID、図2の大字欄）と掲出字のユニコード 16 進コード番号（UTF、図2の UCS 欄）を対象にする。

OCR は ABBYY 社の FineReader 11 を使い、認識精度を考えて漢字のテキスト化を行わず、比較的正確に認識できるアルファベットと数字で構成された上の二つのフィールドを対象にして読み込む。「1」と「l」、「0」と「o」のような間違い多い組合を正規表現により、Perl スクリプトで修正した。

掲出字（Entry）のフィールドは読み込んで作成した CJK 統合漢字と拡張 A の UTF のデー

<sup>3</sup>『大字典』の「凡例」では「訓義の中最も重要視せらるゝものをば、特にゴシックを用ひて之を明らかならしめ、且つ其品詞を示したり」と定義している。

タを Excel 2013 の Unichar()関数により、一括で変換した。手打ちの入力より、ずいぶん効率的である。

CJK 統合漢字と拡張 A の掲出字はまず和訓付きの掲出字を中心に CHISE IDS 漢字検索を利用して拡張 B~D の範囲で入力した。念のため、Unihan データベースの部首画数索引<sup>4</sup> (Unihan Radical-stroke Index) も参考した。

### 2.4.3 和訓テキストテーブルの作成

和訓テキストテーブルは次の四つのフィールドを設定し、Excel で入力する。

- DJT\_wk\_ID 『大字典』 和訓 ID
- DJT\_ID 『大字典』 番号
- DJT\_wakun 重要和訓
- DJT\_hinshi 重要和訓の品詞

『大字典』 和訓 ID (DJT\_wk\_ID) は『大字典』 番号 (DJT\_ID) の後に「.1…n」を追加して作成する。

### 2.5 重要和訓の入力例

例えば、『大字典』の「訓」の項目は、次の図3のように記載されている。

『大字典』番号の 10909 と掲出字の「訓」は、和訓付き掲出字テキストテーブルに入力する。

ヨミ (名)	ヨム (動、 四)	サトシ	ヲシヘ (名)	ミチビク	サイト、	イマシム	ヲシフ (動、 下二)	説キ教フ	10909 【訓】
-----------	-----------------	-----	------------	------	------	------	-------------------	------	--------------

図3 『大字典』における「訓」の記述

重要和訓であるゴシック体の「ヲシフ」「ヲシヘ」「ヨム」「ヨミ」の4語は、品詞情報と合わせて和訓テキストテーブルに入力する。

MySQL を用いてこの二つのテーブルを結合する。

「訓」を検索した結果は次のように表示される。

DJT_wk_ID	Entry	DJT_hinshi	DJT_wakun	DJT_radical	UTF
10909.1	訓	動	ヲシフ	言	8A13
10909.2	訓	名	ヲシヘ	言	8A13
10909.3	訓	動	ヨム	言	8A13
10909.4	訓	名	ヨミ	言	8A13

### 2.6 和訓付き掲出字の符号化

『大字典』和訓は全掲出字約 18,000 字のうち 6,099 字、掲出字全体の約三分の一に付されている。この 6,099 字の和訓付き掲出字のユニコードによる符号化率は表1の通りである。

表1 和訓付き掲出字の符号化率

	総字数	CJK 統合漢字	拡張 A	拡張 B	その他
字数	6,099	5,917 (97.0%)	96 (1.6%)	73 (1.2%)	13 (0.2%)

### 2.7 重要和訓の詳細

重要和訓は延べ 10,514 語、異なり 4,278 語である。本発表ではこの 10,514 語の和訓を対象にする。これらのうち、品詞ごとに集計すると、主な品詞の比率の内訳は、名詞 5,194 (49.4%)、動詞 4,066 (38.7%)、形容詞 610 (5.8%)、副詞 335 (3.2%)、その他 309 (2.9%) である。

<sup>4</sup> <http://www.unicode.org/charts/unihanrsindex.html>

### 3 和訓対応の実際

#### 3.1 HDIC と大字典和訓データベースとの連携

HDIC の『篆隸万象名義』を例にして、『大字典』の和訓データベースと連携し、『篆隸万象名義』の掲出字に対応する和訓、そして部首ごとの特徴を調べてみよう。

『篆隸万象名義』データベースについて簡単に説明すれば、次の通りである。『篆隸万象名義』は 542 部首の約 16,000 字の掲出字を収録し、字音・字義・字体の記述を収める。『篆隸万象名義』データベースは、掲出字画像、掲出字テキスト、注文テキストの三つの部分から構成される。現在本文入力作業は終了して、点検校正中である。

参考として、『篆隸万象名義』の掲出字について Unicode による符号化率を、原本『玉篇』残巻に対応する部分と掲出字の全体と区別し、表 2 に示す<sup>5</sup>。

表 2 『篆隸万象名義』掲出字の符号化率

	字数	CJK 統合漢字	拡張 A	拡張 B	互換 漢字	IDS
原本玉篇 対応部分	3,279	2,406 (73.4%)	373 (11.4%)	495 (15.1%)	0 (0.0%)	5 (0.2%)
掲出字全体	15,797	10,588 (67.0%)	2,013 (12.7%)	3,136 (19.9)	35 (0.2%)	25 (0.2%)

#### 3.2 『大字典』の和訓データに対応する『篆隸万象名義』の掲出字

『篆隸万象名義』では、『大字典』の和訓が付く 6,099 の掲出字と対応のあるものは、542 部首中 459 部首にわたって、4,767 字ある。対応する和訓は 8,648 個である。

参考として、その品詞の内訳を記せば、名詞 4,136 語 (47.8%)、動詞 3,424 語 (35.6%)、形容詞 519 語 (6.0%)、副詞 297 語 (3.43%)、その他 272 語 (3.14%) である。

#### 3.3 和訓データからみる『篆隸万象名義』の部首の特徴

##### 3.3.1 和訓データに対応する『篆隸万象名義』の部首

『篆隸万象名義』では、和訓データに対応する 459 部首であるが、その 459 部首の総掲出字数は 15,566 字である。最も字数の多い部首は「艸部」であり、掲出字数が 732 字である。最も字数の少ない部首の掲出字数は 1 字であり、「𠂔・七・未・丙・三・戊・壬・六・𠂔・久・𠂔・燕・垚」の 13 部首がこれに該当する。

字数が 101 以上の部首は 42 個あり、これらを比較的字数の多い部首とする。

字数が 30 から 100 までの部首は 50 個あり、これらを比較的字数の少ない部首とする。

字数が 29 以下の部首は 367 個であり、これらは字数の僅少の部首とする。

字数が 30 以上にあたる 92 個の部首の掲出字数は 13,382 字であり、和訓データに対応する部首の総字数 15,566 字の 86.0% を占める。字数が 29 以下にあたる 367 個の部首の掲出字数は 2,184 字であり、和訓データに対応する部首の総字数 15,566 字の 14.0% しか占めない。

字数が 29 以下の 367 個の部首は、分析できるデータの分量が少なく、特徴の考察が容易ではない。また、これらの部首は、康熙字典の部首体系において、他の部首に併合したものが多数であり、今回の調査の対象外とする。

##### 3.3.2 和訓データによる『篆隸万象名義』の部首のグループ分け

字数が 30 以上の部首に所属する『篆隸万象名義』の掲出字を対象にし、『大字典』和訓データと照合し、掲出字数と和訓数の観点から検討してみる。

表 3 は『篆隸万象名義』に対応する部首ごとの和訓付掲出字の比率をまとめたものである。

<sup>5</sup> 李 (2013)

表3 『篆隸万象名義』に対応する部首ごとの和訓付掲出字の比率<sup>6</sup>

万象名義の部首	和訓対応掲出字数①	延べ和訓数	万象名義の掲出字数②	和訓付掲出字比率①/②
口	26	55	40	65.0%
子	17	39	30	56.7%
人	209	393	391	53.5%
口	171	262	320	53.4%
宀	48	153	91	52.7%
彳	31	59	60	51.7%
力	32	50	65	49.2%
土	97	155	202	48.0%
...	...	...	...	...
水	199	347	633	31.4%
足	60	86	198	30.3%
弓	14	23	47	29.8%
...	...	...	...	...
毛	4	4	43	9.3%
走	10	22	112	8.9%
鼠	4	5	45	8.9%
邑	15	29	194	7.7%
総計	4,767	8,648	15,566	30.6%

表3の総計の欄に示すように、和訓付掲出字全体の比率は30.6%であり、これを対応する和訓数の多寡の判断基準にする。さきに検討対象とした比較的字数の多い部首と字数の少ない部首について、和訓の多寡によって分類すれば、次の通り四つのグループに分けられる。

- a 字数多・和訓数多の部首：水・木・手・心・人・糸・言・口・女・土・彡・日・山・阜・巾・刀・貝 (17 部首)
- b 字数多・和訓数少部首：艸・虫・金・鳥・肉・竹・衣・魚・目・足・邑・火・馬・犮・玉・車・禾・石・頁・犬・革・支・食・走・酉 (25 部首)
- c 字数少・和訓数多部首：广・示・宀・門・穴・力・彳・口・戈・尸・彡・子・廾・匚 (14 部首)
- d 字数少・和訓数少部首：雨・欠・羽・米・牛・瓦・耳・髟・角・歹・隹・黑・舟・見・齒・骨・風・羊・网・大・弓・皿・田・豕・鼠・鹿・毛・厂・麥・虫・豸・韋・立・耒・缶・谷 (36 部首)

ここで得られた和訓による『篆隸万象名義』部首の特徴を参考にして、次に HDIC に中世・近世の漢和字書の拡張作業を進め行くことにする。

#### 4 和訓データと『篆隸万象名義』・宋本『玉篇』との連携による検索システム

##### 4.1 全文テキストの HTML 化

より多くの国・地域の学者に利用できる形の研究成果にするために、HDIC における篆隸万象名義データベースに入力済の本文データを HTML 化し、オンラインで篆隸万象名義テキスト一覧できるウェブサイト作成し、校正中である。図4は Google Site を利用した試作版である。

このように古字書をオンラインで閲覧可能とするシステムは、古字書の写本を読むときに役に立つと考えられる。実際、玄応の『一切経音義』はその全文が HTML して公開されており、



図4 『篆隸万象名義』HTML化した全文テキストサイト試作版

<sup>6</sup> 紙幅の関係で一部のデータを省略する。

参考になる<sup>7</sup>。

#### 4.2 HDIC Viewer の開発

HTML 化したデータは、検索や他の古字書データベースと連携に利用するには、有効ではない。そこで、HDIC の検索システムは WEB 開発にて広く使用されている PHP と MySQL を組み合わせて構築することとした。データは RDBMS の MySQL で管理されている。検索機能と表示機能は PHP と HTML で作られている。このシステムを HDIC Viewer と名付けた。

MySQL は UTF の文字コードをサポートしているが、デフォルト設定の utf8 はユニコードの BMP にある符号を扱えるが、それ以外の 4 バイト以上の符号を処理することができない。2010 年 3 月の 5.5.3 バージョンより、4 バイト符号を扱うことができたが、サーバ側でデフォルト設定の utf8 を utf8mb4 に変更する必要がある。

フロントエンド側にはツイッター社の Bootstrap というフレームワークを用い、CSS 知識を身につけずにほぼ全てのブラウザでレスポンス・ウェブデザインに対応できる。

現時点ではデータベースへの接続、基本の検索、ユーザーのログイン・ログアウト、レスポンス・ウェブデザインの機能が完成した。図 5 は検索画面、図 6 は検索結果画面を示す。検索画面の図 5 では、字書データベースにより、検索項目は掲出字・和訓・注文を一つ選択する。検索内容欄は検索項目に対応する文字を入れると検索できる。検索結果が複数ある場合、結果は図 6 の右のようにリストで表示される。項目をクリックすると、図 6 のような結果画面に入る。次の段階はデータベース導入の GUI と多条件検索を追加したいと考えている。



図 5 検索画面

図 6 結果画面

#### 5 おわりに

最後に今後の課題を述べる。『大字典』データベースについては、1 掲出字テキストテーブルを完成させ、4 字音テキストテーブルと 5 熟語テキストテーブルの設計と試作を行う。試作は、先に述べた和訓品詞情報を参考にして行う。同時に、試作段階の HDIC Viewer を実用レベルに向けて整備を進める。

『篆隸万象名義』データベースについては、点検校正を完了させ、HTML 化した全文テキストの年内の公開を目指す。

#### 参考文献

- 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
- 高田智和 (2002) 「漢字処理と『大字典』」、『訓点語と訓点資料』109、訓点語学会
- 高田智和 (2003) 『漢字情報処理のための基礎的研究』、平成 15 年度北海道大学大学院文学研究科博士論文
- 李媛 (2015) 『『篆隸万象名義』の本文データベースの構築 - 『玉篇』残巻対応部分を中心に』 - 『国語国文研究』(146)、北海道大学国語国文学会、(2015 年 6 月刊行予定)

<sup>7</sup> <http://tls.uni-hd.de/xuanying.html>