



Title	日本漢字辞書研究の資料と方法に関する基礎的研究
Author(s)	劉, 冠偉
Citation	北海道大学. 博士(文学) 甲第14572号
Issue Date	2021-03-25
DOI	10.14943/doctoral.k14572
Doc URL	<a href="http://hdl.handle.net/2115/81136">http://hdl.handle.net/2115/81136</a>
Type	theses (doctoral)
File Information	Guanwei_Liu.pdf



[Instructions for use](#)

令和2年度博士学位論文

日本漢字辞書研究の資料と方法に関する  
基礎的研究

劉 冠偉

北海道大学大学院文学研究科博士後期課程

# 目次

<b>第 1 章</b>	<b>序論</b>	<b>1</b>
1.1	研究の背景と目的	1
1.2	研究の方法	1
1.3	利用資料	1
1.3.1	資料の選定	1
1.3.2	観智院本『類聚名義抄』	2
1.3.3	夢梅本『和玉篇』	2
1.3.4	上田万年『大字典』	2
1.3.5	その他の資料	2
1.4	本論文の構成	3
<b>第 2 章</b>	<b>テキストデータに基づく日本漢字辞書の和訓に関する基礎統計</b>	<b>4</b>
2.1	観智院本『類聚名義抄』の和訓の基礎統計	4
2.1.1	HDIC 観智院本『類聚名義抄』全文テキストデータ	4
2.1.2	掲出字と和訓の分布	5
2.2	夢梅本『和玉篇』の和訓の基礎統計	7
2.2.1	夢梅本『和玉篇』全文データベース	7
2.2.2	漢字と和訓の分布	7
2.3	『大字典』の和訓の基礎統計	12
2.3.1	『大字典』和訓データベース	12
2.3.2	漢字と和訓の分布	12
2.4	三つのデータベースの掲出字と和訓の基礎統計	16
2.4.1	共通する掲出字	16
2.4.2	共通する漢字と和訓の組合せ	16
2.4.3	品詞から見た共通する漢字とその和訓	31
2.4.4	和訓数から見た共通する漢字とその和訓の対応	31
<b>第 3 章</b>	<b>日本漢字辞書の項目構造の多様性に対応するマークアップ</b>	<b>32</b>
3.1	はじめに	32
3.2	観智院本『類聚名義抄』全文データベースと注文項目の構造	32

---

3.3	段階的なタグセット拡張 . . . . .	33
3.3.1	辞書類を対象にしたマークアップの前例 . . . . .	33
3.3.2	マークアップの準備 . . . . .	33
3.3.3	段階的なタグセット拡張 . . . . .	35
3.4	マークアップツールの必要性と現状 . . . . .	36
3.4.1	文系研究資源の作成について . . . . .	36
3.4.2	データベースへのながれ . . . . .	37
3.4.3	tagzuke の必要性 . . . . .	37
3.4.4	「tagzuke」の開発と利用 . . . . .	37
3.5	ツールの有用性 . . . . .	38
3.6	ツールの改善 . . . . .	38
3.6.1	改善点 . . . . .	38
3.6.2	新しい tagzuke の全体像 . . . . .	39
3.6.3	操作性に関する改善 . . . . .	40
3.6.4	汎用性に関する改善 . . . . .	40
3.6.5	維持性に関する改善 . . . . .	41
3.6.6	改善後の効率 . . . . .	41
<b>第 4 章</b>	<b>構造化した注文に基づく分析</b> . . . . .	<b>42</b>
4.1	観智院本『類聚名義抄』の熟語項目と異体項目 . . . . .	42
4.1.1	はじめに . . . . .	42
4.1.2	先行研究 . . . . .	42
4.1.3	観智院本『類聚名義抄』の項目数 . . . . .	42
4.1.4	観智院本『類聚名義抄』の項目構造と要素 . . . . .	42
4.1.5	掲出字の形式と注文内容との対応 . . . . .	43
4.1.6	おわりに . . . . .	45
4.2	観智院本『類聚名義抄』の基本項目と拡張項目 . . . . .	45
4.2.1	はじめに . . . . .	45
4.2.2	判別方法 . . . . .	46
4.2.3	要素形式を用いて自動判別 . . . . .	46
4.2.4	基本項目と拡張項目の統計 . . . . .	46
4.2.5	おわりに . . . . .	47
<b>第 5 章</b>	<b>字形・字体に関する研究</b> . . . . .	<b>49</b>
5.1	GlyphWiki による観智院本『類聚名義抄』字形の再現 . . . . .	49
5.1.1	はじめに . . . . .	49
5.1.2	字形情報利用の制限 . . . . .	50
5.1.2.1	符号化による制限 . . . . .	50

5.1.2.2	著作権による制限 . . . . .	50
5.1.3	GlyphWiki によつての字形再現 . . . . .	50
5.1.3.1	GlyphWiki について . . . . .	50
5.1.3.2	字形データと Kage システム . . . . .	50
5.1.3.3	作字プラットフォームの実現 . . . . .	50
5.1.4	再現した字形とその利用 . . . . .	51
5.1.4.1	再現した字形のアクセス . . . . .	51
5.1.4.2	観智院本類聚名義抄データベースの画像代用 . . . . .	51
5.1.5	IDS を用いて作成した字形の検索 . . . . .	51
5.1.6	字形データ . . . . .	52
5.1.7	字形情報と漢字の包摂粒度 . . . . .	52
5.1.8	作成字形と Unicode の関係 . . . . .	53
5.1.9	おわりに . . . . .	54
5.2	漢字辞書から異体字情報の抽出：宋本『玉篇』を例に . . . . .	54
5.2.1	はじめに . . . . .	54
5.2.2	宋本玉篇テキストデータベース . . . . .	54
5.2.3	宋本『玉篇』各要素の特徴 . . . . .	54
5.2.4	マークアップした結果 . . . . .	55
5.2.5	異体字情報の抽出 . . . . .	55
<b>第 6 章</b>	<b>漢字辞書データベース公開用インターフェースの開発</b>	<b>58</b>
6.1	平安時代漢字字書総合データベース検索インターフェース：HDIC Viewer . . . . .	58
6.1.1	HDIC Viewer とは . . . . .	58
6.1.2	Web インターフェースの提供 . . . . .	58
6.1.3	Web API の提供 . . . . .	58
6.1.3.1	Web API の意義 . . . . .	58
6.1.3.2	HDIC Viewer Web API のスキーマ . . . . .	58
6.1.3.3	画像データの Web API . . . . .	59
6.1.4	サーバーへの実装 . . . . .	59
6.2	多漢字文献の検索効率向上とマルチデバイス対応の試み . . . . .	60
6.2.1	はじめに . . . . .	60
6.2.2	利用資料 . . . . .	60
6.2.3	漢字検索の効率化 . . . . .	60
6.2.3.1	多漢字文献における漢字の検索と入力の問題 . . . . .	60
6.2.3.2	IDS 漢字検索の必要性 . . . . .	61
6.2.3.3	IDS 画数漢字検索の必要性とその開発 . . . . .	62
6.2.4	マルチデバイスへの対応 . . . . .	62
6.2.4.1	マルチデバイスにおける問題点 . . . . .	62

---

6.2.4.2	画面サイズへの対応 . . . . .	62
6.2.4.3	フォントシステムの補助 . . . . .	63
6.2.4.4	ローカルストレージの利用 . . . . .	63
6.2.5	開発と実装 . . . . .	63
6.2.5.1	フロントエンドについて . . . . .	63
6.2.5.2	バックエンドについて . . . . .	64
6.2.6	IDS データの利用 . . . . .	64
6.2.7	対応するフォント . . . . .	65
6.2.8	検索速度の向上 . . . . .	65
6.2.9	おわりに . . . . .	66
6.3	古辞書における横断検索の実現 . . . . .	66
6.3.1	問題点 . . . . .	66
6.3.2	古字書の研究が求める横断検索 . . . . .	66
6.3.3	字体注によるの字体変換表 . . . . .	67
<b>第 7 章</b>	<b>結論</b>	<b>69</b>
7.1	本研究のまとめ . . . . .	69
7.2	本研究の意義 . . . . .	69
7.2.1	国語学において . . . . .	69
7.2.2	人文情報学において . . . . .	70
7.2.3	成果の発信において . . . . .	70
7.3	HDIC Viewer の利用者に関する統計 . . . . .	70
7.3.1	統計方法 . . . . .	70
7.3.2	統計結果 . . . . .	71
<b>参考文献</b>		<b>72</b>
<b>謝辞</b>		<b>74</b>



# 第1章

## 序論

### 1.1 研究の背景と目的

日本の漢字辞書は長い歴史があり、膨大な漢字情報と和訓を収録している。国語学、訓点語と訓点資料の研究には不可欠な材料である。

近年、日本の古辞書に関する発表・研究が増え、活発化する傾向が見える [1][2][3][4][5][6]。これは、日本の古辞書に関するデータベース・研究ツールが揃ったからであると指摘されている [7]。

また、コンピュータとインターネットの普及によって、研究資料としての文献画像・テキストの公開が徐々に増加しており、研究環境が大きく変化している。日本の漢字辞書の場合、平安時代漢字字書総合データベース (HDIC<sup>1</sup>) [8] や日本語史研究用テキストデータ集などのテキストデータが公開され、辞書に掲載の漢字と和訓を量的に研究することが可能になった。

一方、これらの辞書テキストデータ資料は以下の難点がある。

1. 注文の構造が複雑で、漢字の形・音・訓・漢文義注・引用出典の分析が十分ではない。
2. 写本資料では、それぞれに符号化方針が異なり、異体字・形近異字の検索が難しい。
3. 簡単に検索・表示できるインターフェースがない。応用のシーンが限られている。

本研究は、これらの新しい日本漢字辞書研究資

料を分析して、情報学と人文学を併用して、新たな研究手法を提案する。

### 1.2 研究の方法

本研究は、日本の漢字辞書と漢字研究資料を用いて、上記の三つの問題に対して、次のように解決していきたい。

1. 注文項目構造の多様性に対応したタグセットとマークアップ・ツールを開発する。
2. 漢字辞書テキストデータから和訓と異体字関係を抽出して、対応関係を分析する。
3. ウェブインターフェースを開発して、成果を利用性の高い研究資料として公開する。

### 1.3 利用資料

#### 1.3.1 資料の選定

ここに提案する方法を実際の資料によって応用するには以下の三つの条件を満たさなければならない。

#### 国語・日本語の研究に有益なもの

国語・日本語、特に語彙や漢字の研究では、漢字辞書は参照することを怠ることが許されない研究資源である。平安時代の類聚名義抄や色葉字類抄をはじめ、室町時代の和玉篇や節用集など、漢字がメインとして書かれた漢和辞書や類語辞書を資料の候補とする。

また、上記のような日本古辞書はほとんど写本

<sup>1</sup> <https://hdic.jp>



であるため、文字の同定や和訓の整形などにはより現代に近い資料を参照することが必要である。漢字の事態と仮名遣いの揺れが整えた近代の漢和辞書がこういう役割を担いやすいと考えられる。

### テキストデータが公開されたもの

本研究は主に新たな研究方法を開発・提案するものである。利用資料は研究方法を検証する材料となるため、一からテキストデータを作成することより、すでに作成されたデータを利用するのが効率的である。

公開されたデータベースの中には、検索インターフェースのみ、プレーンテキスト<sup>2</sup>や API<sup>3</sup>が提供されていないものもある。これらのデータを分析するには、抽出方法の工夫や作成者との相談が必要である。

### インターネット公開され、権利関係上に利用しやすいもの

すでに公開されたデータでも、権利関係が適切に書いてないもの、あるいは再配布しにくいものもある。

以上のことを条件として、次の三つの辞典のテキストデータを利用資料とした。

#### 1.3.2 観智院本『類聚名義抄』

観智院本『類聚名義抄』は院政期に成立した漢和辞書『類聚名義抄』の改編本系の唯一の完本として国語学においてよく研究されている。仏・法・僧の三部に分かれ、合計 10 帖からなる。約 32,000 項目に約 42,000 掲出字があり、20 万字以上の注文が収録されている。

一つの項目にある掲出字は、現代の典型的な漢和辞書と同様に 1 字であるものと、熟語または異体字関係である 2 字以上のものが掲出されている。その下にはより小さい文字で掲出字に対する

形・音・義を注記する注文がある。

観智院本『類聚名義抄』のテキストデータは HDIC の観智院本類聚名義抄全文テキストデータベースを利用する。

#### 1.3.3 夢梅本『和玉篇』

『和玉篇』は室町時代に成立した部首分類体の漢和辞書である。現存最古は「延徳三年写本」(1491)である。異本が多いが、本論文は漢文注記と和訓が多く掲載された夢梅本を取り上げる。

夢梅本『和玉篇』は約 24,000 字が掲出され、掲出字の左右に仮名音注、下に和訓・漢文義注が注記されている。

夢梅本『和玉篇』のテキストデータは筆者が所属する講座の大学院生であった高橋大稀氏によって作成されているが、現時点ではまだ一般に公開されていない [9]。

#### 1.3.4 上田万年『大字典』

『大字典』は『康熙字典』と同じ部首画数順で配列する漢和辞典である。明治以降刊行された数多く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を持っている。

『大字典』の親字に対して、高田智和氏 [10][11] は一連の研究を行っており、全親字を収録した『『大字典』データベース』を作成した。筆者はそれを利用して、『大字典』の重要和訓をすべて収録した大字典和訓データベースを作成した [12][13]

また、本研究で利用した辞典以外のデータ資料をその他の資料として下記に示す。

#### 1.3.5 その他の資料

漢字の周辺情報や字形構造などは以下のデータベースを利用した。

- UniHan<sup>4</sup>
- CHISE<sup>5</sup>

<sup>2</sup> 文字情報以外の情報は含まず、テキストデータのみで構成されているファイル形式である。

<sup>3</sup> Application Programming Interface の略称、ソフトウェア同士が互いに情報を共有するインターフェースである。

<sup>4</sup> <https://unicode.org/charts/unihan.html>

<sup>5</sup> <http://www.chise.org/>

- [cjkvi/cjkvi-ids](https://github.com/cjkvi/cjkvi-ids)<sup>6</sup>

## 1.4 本論文の構成

本論文は、全7章からなる。

第1章では、序論として研究の背景、目的、方法、利用資料を述べる。

第2章では、テキストデータ研究資料の基礎統計を行い、各資料の漢字と和訓との対応関係、和訓の分布を算出した結果について述べる。

第3章では、各利用資料の基本情報を述べ、テキストデータとしての研究資料をどのように利用するか、日本漢字辞書の注文を構造化して掲載している字音注記、字形注記、和訓、漢文義注を抽出していく。そのために開発されたマークアップ・ツール「tagzuke」の機能を述べる。

第4章では、構造化した注文を用いて、各資料に対してさらに高度な分析を行ったことについて述べる。

第5章では、各資料から抽出した和訓、義注、字体注を用いて、漢字と和訓の総合データベースを作成することについて述べる。

第6章では、研究成果を公開するプラットフォームの開発について述べる。

第7章では、本論文をまとめ、成果と今後の課題を述べる。

---

<sup>6</sup> <https://github.com/cjkvi/cjkvi-ids>

## 第2章

# テキストデータに基づく日本漢字辞書の 和訓に関する基礎統計

### 2.1 観智院本『類聚名義抄』の和訓の基礎統計

#### 2.1.1 HDIC 観智院本『類聚名義抄』全文テキストデータ

観智院本『類聚名義抄』は類聚名義抄の唯一の完本として、国語学において研究が盛んである。観智院本の構成は仏・法・僧の3部に分けられ、仏上・仏中・仏下本・仏下末・法上・法中・法下・僧上・僧中・僧下の合計10帖からなる。120部首の配列は表2.1を参照。

表 2.1 観智院本『類聚名義抄』の部首配列

巻	部首
仏上	人彳辵匚走麥一丨十身
仏中	耳女舌口目鼻見日田肉
仏下本	舟骨角貝頁彡髟手木犬
仏下末	牛片豸乙儿収八大火黒
法上	水彳言足立豆卜面齒山
法中	石玉色邑阜土心巾糸衣
法下	示禾米、宀勹穴雨門口尸 虎广鹿疒歹子斗軌寸
僧上	艸竹力刀羽毛食金亼瓜
僧中	网皿瓦缶弓舛矢斤矛戈欠 又支攴皮革韋車羊馬
僧下	鳥隹魚虫鼠龜鬼風酉雜

観智院本『類聚名義抄』全文テキストデータベース (KRM) は、北海道大学大学院文学研究院の池田証壽教授の主導のもと、筆者が所属する同大学院の言語科学講座 (以下、研究室) が推進しているプロジェクトである [14][15][8][16][17]。現在では、掲出字・注文の全文入力を完成して、校正段階に入っている。掲出字と一部分の注文テキストは筆者が開発した HDIC Viewer で公開している。最終校正後の本公開が待たれる。

この全文テキストデータベースは Unicode によって符号化されている (テーブルの構成は表 2.2 を参照)。合計約 3 万 2 千項目、注文と声点などを含めて約 43 万字となる膨大なデータである。

表 2.2 観智院本『類聚名義抄』全文テキストデータベースの構造

フィールド	内容
KRID	掲出字 ID
KR2ID	風間書房版の項目 ID
KR_Tenri_p	八木書店版の掲出字 ID
KR_vol_radical	巻・部首番号
KR_vol_name	巻名
KR_radical	部首名
Entry	掲出字
Entry_original	原字形に近い掲出字
KR_def	注文

KRM の項目は掲出字と注文からなる。掲出字が一字である項目を単字項目、掲出字が複数字である項目を複字項目と呼ぶ（詳細は第 4 章を参照）。注文は字音注記・字体注記・漢文義注・和訓の 4 種類に大別できる（詳細は第 3 章を参照）。本節は注文に記載された和訓のみを抽出して、分析を行う。

なお、本論文で用いる KRM は 2019 年末までに校正されたバージョンであり、今後校正が進んで本公開されるデータとの間で多少相違が生じることが予想される。

### 2.1.2 掲出字と和訓の分布

表 2.3 に、KRM の部首ごと項目・掲出字・和訓の分布を示す。

表 2.3 KRM 部首ごとの掲出字・和訓の分布

順	部首	掲出字数	和訓数	1 掲出字あたりの和訓数
1	人	1,149	1,626	1.415
2	彳	229	503	2.197
3	辵	575	1,002	1.743
4	匸	64	65	1.016
5	走	201	146	0.726
6	麥	116	18	0.155
7	一	216	259	1.199
8	丨	96	100	1.042
9	十	106	170	1.604
10	身	88	54	0.614
11	耳	153	146	0.954
つづく				

順	部首	掲出字数	和訓数	1 掲出字あたりの和訓数
12	女	650	598	0.920
13	舌	51	42	0.824
14	口	1,274	1,047	0.822
15	目	551	418	0.759
16	鼻	45	32	0.711
17	見	137	140	1.022
18	日	676	791	1.170
19	田	193	205	1.062
20	肉	888	731	0.823
21	舟	125	83	0.664
22	骨	141	59	0.418
23	角	150	72	0.480
24	貝	251	317	1.263
25	頁	352	255	0.724
26	彡	32	44	1.375
27	髟	221	124	0.561
28	手	1,465	2,106	1.438
29	木	1,676	1,057	0.631
30	犬	375	183	0.488
31	牛	202	73	0.361
32	片	74	47	0.635
33	豸	144	18	0.125
34	乙	42	70	1.667
35	儿	302	203	0.672
36	収	125	166	1.328
37	八	152	217	1.428
38	大	166	143	0.861
39	火	648	524	0.809
つづく				

順	部首	掲出字数	和訓数	1 掲出 字あたりの和 訓数
40	黒	108	86	0.796
41	水	1,568	1,646	1.050
42	彳	94	168	1.787
43	言	796	1,095	1.376
44	足	589	527	0.895
45	立	113	160	1.416
46	豆	111	33	0.297
47	卜	152	102	0.671
48	面	62	23	0.371
49	齒	157	66	0.420
50	山	584	295	0.505
51	石	472	284	0.602
52	玉	496	200	0.403
53	色	37	17	0.459
54	邑	359	143	0.398
55	阜	330	474	1.436
56	土	700	495	0.707
57	心	1,111	1,387	1.248
58	巾	317	115	0.363
59	糸	856	908	1.061
60	衣	519	350	0.674
61	示	274	285	1.040
62	禾	593	380	0.641
63	米	321	188	0.586
64	丶	169	230	1.361
65	㇇	415	513	1.236
66	勺	37	29	0.784
67	穴	222	232	1.045
				つづく

順	部首	掲出字数	和訓数	1 掲出 字あたりの和 訓数
68	雨	290	237	0.817
69	門	306	325	1.062
70	口	106	101	0.953
71	尸	260	228	0.877
72	虍	127	93	0.732
73	广	392	402	1.026
74	鹿	91	22	0.242
75	犴	603	225	0.373
76	歹	266	149	0.560
77	子	168	101	0.601
78	斗	40	17	0.425
79	軌	29	19	0.655
80	寸	60	89	1.483
81	艸	2,362	1,672	0.708
82	竹	723	365	0.505
83	力	139	193	1.388
84	刀	340	317	0.932
85	羽	147	140	0.952
86	毛	173	42	0.243
87	食	333	204	0.613
88	金	922	861	0.934
89	亼	115	142	1.235
90	瓜	78	48	0.615
91	网	188	116	0.617
92	皿	163	120	0.736
93	瓦	176	77	0.438
94	缶	47	15	0.319
95	弓	150	110	0.733
				つづく

順	部首	掲出字数	和訓数	1 掲出字あたりの和訓数
96	𠂔	93	115	1.237
97	矢	45	55	1.222
98	斤	72	86	1.194
99	矛	53	51	0.962
100	戈	175	265	1.514
101	欠	164	153	0.933
102	又	125	138	1.104
103	支	305	307	1.007
104	殳	104	72	0.692
105	皮	130	66	0.508
106	革	328	158	0.482
107	韋	75	36	0.480
108	車	335	255	0.761
109	羊	119	34	0.286
110	馬	419	245	0.585
111	鳥	809	374	0.462
112	隹	129	128	0.992
113	魚	571	277	0.485
114	虫	1,041	319	0.306
115	鼠	96	21	0.219
116	龜	61	10	0.164
117	鬼	121	41	0.339
118	風	160	112	0.700
119	酉	289	174	0.602
120	雜	2,056	1,151	0.560
合計		42,332	35,358	0.835

玉篇』は、和訓と漢文注記が豊富に記載された部首分類体漢字辞書である。5巻に196部首を収録する。それらの配列を表2.4に示す。

夢梅本『和玉篇』全文データベース(MBWG)は筆者が所属している研究室の高橋大希氏が作成したデータベースである。高橋[9]によると、夢梅本『和玉篇』のすべての掲出字・和訓・仮名音注・漢文注記を収録している。データベースは掲出字テーブルと仮名注テーブルに分かれる。掲出字テーブルはID、部首番号、部首、掲出字、漢字注、仮名音注、仮名注、備考の八つのフィールドから成る。仮名注テーブルは掲出字ID、仮名注ID、仮名注、仮名修正、掲出字、漢字注、部首、部首番号、備考の九つのフィールドから成る。

MBWGの全文掲出字数について、高橋・劉・池田[18]は表2.5のとおり指摘している。

### 2.2.2 漢字と和訓の分布

表2.6に、MBWGの部首ごと項目・掲出字・和訓の分布を示す。

## 2.2 夢梅本『和玉篇』の和訓の基礎統計

### 2.2.1 夢梅本『和玉篇』全文データベース

和玉篇<sup>1</sup>は室町時代に成立して以来、数多くの諸本が流布されている。その中でも、夢梅本『和

<sup>1</sup> 倭玉篇ともいう。

表 2.4 夢梅本『和玉篇』部首リスト

巻	部首
一	日月人言木火土金水白風雨目耳鼻舌身王
二	骨手足鳥广庀心豸牛馬羊龜魚
三	虫貝鬼夕發欠司墓用辵邑阜口宀穴彳行頁車舟尸舛永子女衣
四	示酉玉疒口見斤此収齒弓弟兄隶缶癸色一 冂矛戈寸先軌犬夷香斗竹糸艸革韋角皮厂 長髟支刀夕死戸
五	走門石食禾麥黍立山受皿网華毛彡 冂大共受佳田矢巾爪鬥又瓜力羽多 𠂔書束儿赤勺 旬上生臥臣十 冂黑乃辛九瓦鼠老喜豆鼎片弋白齋韋柔京面自彡止血癰日音卍龠甘鹿 叔東蓐市弓冂鹵鬲冂林谷夕矢高豕毳它齋褰八甲了黃米

表 2.5 夢梅本『和玉篇』全文データベース収録字数

	字数
掲出字	22,423
注文（漢字のみ）	79,564
合計	101,987

表 2.6 MBWG 部首ごとの掲出字・和訓の分布

順	部首	掲出字数	和訓数	1 掲出字あたりの和訓数
1	日	169	249	1.473
2	月	409	394	0.963
3	人	531	587	1.105
4	言	513	553	1.078
5	木	798	515	0.645
6	火	202	213	1.054
7	土	350	367	1.049
つづく				

順	部首	掲出字数	和訓数	1 掲出字あたりの和訓数
8	金	459	410	0.893
9	水	938	840	0.896
10	白	42	55	1.310
11	風	98	91	0.929
12	雨	160	161	1.006
13	目	335	353	1.054
14	耳	97	93	0.959
15	鼻	26	25	0.962
16	舌	13	12	0.923
17	身	37	37	1.000
18	王	5	7	1.400
19	骨	118	108	0.915
20	手	644	822	1.276
21	足	299	305	1.020
22	鳥	416	192	0.462
23	广	165	187	1.133
つづく				

順	部首	掲出字数	和訓数	1 掲出 字あたりの和 訓数
24	虍	44	44	1.000
25	心	625	837	1.339
26	豸	65	57	0.877
27	牛	145	150	1.034
28	馬	275	260	0.945
29	羊	74	66	0.892
30	龜	21	15	0.714
31	魚	320	203	0.634
32	虫	538	332	0.617
33	貝	137	165	1.204
34	鬼	69	50	0.725
35	欠	33	37	1.121
36	殺	9	9	1.000
37	欠	148	170	1.149
38	司	8	12	1.500
39	墓	12	12	1.000
40	用	14	21	1.500
41	辵	250	425	1.700
42	邑	276	17	0.062
43	阜	199	253	1.271
44	口	517	585	1.132
45	宀	136	212	1.559
46	穴	116	151	1.302
47	彳	110	160	1.455
48	行	21	47	2.238
49	頁	203	262	1.291
50	車	244	270	1.107
51	舟	111	109	0.982
つづく				

順	部首	掲出字数	和訓数	1 掲出 字あたりの和 訓数
52	尸	76	110	1.447
53	凵	41	43	1.049
54	永	8	11	1.375
55	子	40	54	1.350
56	女	415	466	1.123
57	衣	297	384	1.293
58	示	142	168	1.183
59	酉	175	206	1.177
60	玉	278	262	0.942
61	疒	268	303	1.131
62	口	63	86	1.365
63	見	104	118	1.135
64	斤	31	45	1.452
65	此	16	26	1.625
66	収	41	45	1.098
67	齒	94	112	1.191
68	弓	79	109	1.380
69	弟	5	7	1.400
70	兄	3	3	1.000
71	隶	4	3	0.750
72	缶	43	48	1.116
73	癸	3	1	0.333
74	色	13	11	0.846
75	一	8	13	1.625
76	卩	17	23	1.353
77	矛	28	27	0.964
78	戈	51	81	1.588
79	寸	15	29	1.933
つづく				



順	部首	掲出字数	和訓数	1 掲出 字あた りの和 訓数
80	先	14	26	1.857
81	軌	6	16	2.667
82	犬	260	309	1.188
83	夷	6	14	2.333
84	香	26	29	1.115
85	斗	22	27	1.227
86	竹	518	564	1.089
87	糸	490	696	1.420
88	艸	1,058	956	0.904
89	革	201	196	0.975
90	韋	59	60	1.017
91	角	106	87	0.821
92	皮	52	51	0.981
93	厂	66	92	1.394
94	長	17	36	2.118
95	彡	108	107	0.991
96	支	178	286	1.607
97	刀	203	327	1.611
98	歹	108	136	1.259
99	死	14	12	0.857
100	戸	23	31	1.348
101	走	155	194	1.252
102	門	132	165	1.250
103	石	297	308	1.037
104	食	214	274	1.280
105	禾	241	304	1.261
106	麥	54	51	0.944
107	黍	24	31	1.292
				つづく

順	部首	掲出字数	和訓数	1 掲出 字あた りの和 訓数
108	立	53	74	1.396
109	山	306	275	0.899
110	爻	38	46	1.211
111	皿	75	73	0.973
112	网	93	92	0.989
113	華	24	36	1.500
114	毛	83	99	1.193
115	彡	54	75	1.389
116	冂	17	13	0.765
117	大	58	85	1.466
118	共	17	26	1.529
119	爻	21	38	1.810
120	隹	93	121	1.301
121	田	88	111	1.261
122	矢	32	29	0.906
123	巾	178	204	1.146
124	爪	16	25	1.563
125	鬥	14	17	1.214
126	又	36	42	1.167
127	瓜	35	36	1.029
128	力	90	130	1.444
129	羽	138	175	1.268
130	多	52	71	1.365
131	𠂔	24	18	0.750
132	書	7	12	1.714
133	束	31	23	0.742
134	儿	21	28	1.333
135	赤	19	25	1.316
				つづく

順	部首	掲出字数	和訓数	1 掲出字あたりの和訓数
136	勺	38	56	1.474
137	句	22	48	2.182
138	上	9	16	1.778
139	生	8	15	1.875
140	臥	9	10	1.111
141	臣	26	16	0.615
142	十	16	14	0.875
143	匸	24	24	1.000
144	黑	88	86	0.977
145	乃	34	48	1.412
146	辛	32	53	1.656
147	九	7	5	0.714
148	瓦	105	103	0.981
149	鼠	61	60	0.984
150	老	19	22	1.158
151	喜	41	48	1.171
152	豆	41	48	1.171
153	鼎	5	6	1.200
154	片	54	43	0.796
155	弋	14	18	1.286
156	白	29	38	1.310
157	齋	46	55	1.196
158	亶	8	12	1.500
159	耒	54	61	1.130
160	京	21	26	1.238
161	面	42	40	0.952
162	自	31	32	1.032
163	彡	48	46	0.958
つづく				

順	部首	掲出字数	和訓数	1 掲出字あたりの和訓数
164	止	35	51	1.457
165	血	24	29	1.208
166	癩	18	21	1.167
167	日	12	17	1.417
168	音	18	21	1.167
169	卍	43	54	1.256
170	龠	13	19	1.462
171	甘	46	75	1.630
172	鹿	63	42	0.667
173	叡	15	19	1.267
174	東	26	18	0.692
175	蓐	14	13	0.929
176	市	34	33	0.971
177	弓	57	43	0.754
178	冂	34	47	1.382
179	鹵	45	48	1.067
180	鬲	52	66	1.269
181	匸	69	59	0.855
182	冂	41	49	1.195
183	林	31	34	1.097
184	谷	66	93	1.409
185	夕	29	49	1.690
186	夂	73	79	1.082
187	高	13	16	1.231
188	豕	100	85	0.850
189	毳	58	70	1.207
190	它	36	38	1.056
191	鬻	62	115	1.855
つづく				

順	部首	掲出字数	和訓数	1 掲出字あたりの和訓数
192	橐	24	19	0.792
193	八	23	63	2.739
194	甲	24	34	1.417
195	了	44	90	2.045
196	黄	41	45	1.098
197	米	164	158	0.963
合計		22,646	24,446	1.079

## 2.3 『大字典』の和訓の基礎統計

### 2.3.1 『大字典』和訓データベース

『大字典』は『康熙字典』と同じ部首画数順で配列する漢和辞典である。明治以降刊行された数多く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（以下、重要視和訓）にその品詞を示すことなどの特徴を持っている。『大字典』和訓データベースは『大字典』に掲載する 10,517 個の重要視和訓を収録したデータベースである [12]。

『大字典』和訓データベース (DJT) を開発するにあたって当初に念頭に置いたのは、

- (a) 近世の国語の基礎資料として蓄積する
  - (b) HDIC の収録の和訓同定に役立つ
  - (c) 自己責任でゼロからデータベースを作成して、自由に実践できる場所がほしい
  - (d) 作成したデータベースをインターネットに公開して、さらに多くの人の利用に提供したい
- などであった。

なお、『大字典』の親字に対して、高田 [10][11] に一連の研究があり、全親字を収録した「『大字典』データベース」を作成した。和訓データベース構築段階で、筆者はそれを利用して、入力時間を大幅に減少させることができた。

DJT の構造は表 2.7 のとおりである。

表 2.7 『大字典』和訓データベースの構造

フィールド	内容
DJT_ID	『大字典』番号
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
Def	和訓とその品詞

### 2.3.2 漢字と和訓の分布

表 2.8 に、DJT の部首ごと掲出字・和訓の分布を示す。

表 2.8 DJT 部首ごとの掲出字・和訓の分布

順	部首	掲出字数	和訓数	1 掲出字あたりの和訓数
1	一	22	47	2.136
2	丨	9	17	1.889
3	丶	4	12	3.000
4	丿	17	34	2.000
5	乙	11	18	1.636
6	丿	4	8	2.000
7	二	13	23	1.769
8	十	12	23	1.917
9	人	272	481	1.768
10	儿	20	47	2.350
11	入	6	9	1.500
12	八	11	20	1.818
				つづく

順	部首	掲出字数	和訓数	1 掲出 字あたりの和 訓数
13	冂	13	18	1.385
14	冂	9	9	1.000
15	彳	22	33	1.500
16	几	9	17	1.889
17	凵	8	14	1.750
18	凵	88	139	1.580
19	力	39	61	1.564
20	勹	15	22	1.467
21	匕	5	9	1.800
22	匚	15	17	1.133
23	匚	4	9	2.250
24	十	11	16	1.455
25	卜	6	8	1.333
26	冂	15	23	1.533
27	厂	19	27	1.421
28	厶	4	7	1.750
29	又	18	31	1.722
30	口	308	464	1.506
31	口	31	64	2.065
32	土	163	229	1.405
33	土	7	9	1.286
34	攴	2	2	1.000
35	攴	4	4	1.000
36	夕	9	20	2.222
37	大	29	60	2.069
38	女	137	193	1.409
39	子	30	56	1.867
40	宀	67	183	2.731
				つづく

順	部首	掲出字数	和訓数	1 掲出 字あたりの和 訓数
41	寸	14	42	3.000
42	小	10	31	3.100
43	尢	7	14	2.000
44	尸	30	65	2.167
45	巾	5	5	1.000
46	山	73	93	1.274
47	巛	6	8	1.333
48	工	7	25	3.571
49	己	7	13	1.857
50	巾	67	100	1.493
51	干	9	29	3.222
52	幺	4	10	2.500
53	广	47	79	1.681
54	廴	4	8	2.000
55	井	4	5	1.250
56	弋	6	10	1.667
57	弓	22	36	1.636
58	冫	5	5	1.000
59	彡	12	17	1.417
60	彳	38	70	1.842
61	心	219	325	1.484
62	戈	23	38	1.652
63	戶	12	22	1.833
64	手	264	375	1.420
65	支	3	6	2.000
66	攴	40	81	2.025
67	文	5	7	1.400
68	斗	4	4	1.000
				つづく

順	部首	掲出字数	和訓数	1 掲出 字あたりの和 訓数
69	斤	9	19	2.111
70	方	17	33	1.941
71	无	3	4	1.333
72	日	101	169	1.673
73	月	15	27	1.800
74	木	393	545	1.387
75	欠	26	45	1.731
76	止	13	35	2.692
77	歹	25	38	1.520
78	殳	11	23	2.091
79	母	5	7	1.400
80	比	2	4	2.000
81	毛	8	9	1.125
82	氏	3	3	1.000
83	气	1	2	2.000
84	水	255	442	1.733
85	火	94	144	1.532
86	爪	7	12	1.714
87	父	2	3	1.500
88	爻	3	7	2.333
89	月	3	4	1.333
90	片	9	13	1.444
91	牙	1	1	1.000
92	牛	20	30	1.500
93	犬	50	66	1.320
94	玄	3	10	3.333
95	玉	41	59	1.439
96	瓜	3	4	1.333
つづく				

順	部首	掲出字数	和訓数	1 掲出 字あたりの和 訓数
97	瓦	17	24	1.412
98	甘	3	4	1.333
99	生	4	14	3.500
100	用	3	6	2.000
101	田	33	81	2.455
102	疋	4	9	2.250
103	疒	61	87	1.426
104	夂	4	7	1.750
105	白	13	22	1.692
106	皮	5	11	2.200
107	皿	23	41	1.783
108	目	59	105	1.780
109	矛	2	4	2.000
110	矢	9	16	1.778
111	石	42	61	1.452
112	示	36	73	2.028
113	肉	2	2	1.000
114	禾	47	84	1.787
115	穴	30	46	1.533
116	立	12	29	2.417
117	竹	94	149	1.585
118	米	31	45	1.452
119	糸	159	342	2.151
120	缶	10	19	1.900
121	网	20	39	1.950
122	羊	9	16	1.778
123	羽	22	47	2.136
124	老	6	10	1.667
つづく				

順	部首	掲出字数	和訓数	1 掲出 字あた りの和 訓数
125	而	3	8	2.667
126	耒	7	10	1.429
127	耳	18	31	1.722
128	聿	4	8	2.000
129	肉	89	160	1.798
130	臣	6	11	1.833
131	自	3	9	3.000
132	至	5	5	1.000
133	白	6	7	1.167
134	與	4	26	6.500
135	舌	4	13	3.250
136	舛	2	4	2.000
137	舟	15	20	1.333
138	艮	3	7	2.333
139	色	2	2	1.000
140	艸	181	297	1.641
141	虍	9	20	2.222
142	虫	96	118	1.229
143	血	5	6	1.200
144	行	11	30	2.727
145	衣	71	126	1.775
146	酉	5	10	2.000
147	見	13	29	2.231
148	角	9	18	2.000
149	言	148	306	2.068
150	谷	3	5	1.667
151	豆	5	8	1.600
152	豕	9	17	1.889
				つづく

順	部首	掲出字数	和訓数	1 掲出 字あた りの和 訓数
153	豸	7	8	1.143
154	貝	62	148	2.387
155	赤	5	8	1.600
156	走	12	24	2.000
157	足	69	96	1.391
158	身	8	9	1.125
159	車	47	73	1.553
160	辛	10	20	2.000
161	辰	2	6	3.000
162	辵	100	246	2.460
163	邑	18	34	1.889
164	酉	36	66	1.833
165	采	3	9	3.000
166	里	5	23	4.600
167	金	111	193	1.739
168	長	1	8	8.000
169	門	35	87	2.486
170	阝	52	131	2.519
171	隶	2	8	4.000
172	隹	22	65	2.955
173	雨	28	56	2.000
174	青	4	11	2.750
175	非	3	10	3.333
176	面	4	10	2.500
177	革	28	43	1.536
178	韋	9	13	1.444
179	韭	1	1	1.000
180	音	4	8	2.000
				つづく

順	部首	掲出字数	和訓数	1 掲出字あたりの和訓数
181	頁	42	84	2.000
182	風	9	10	1.111
183	飛	2	3	1.500
184	食	39	91	2.333
185	首	2	9	4.500
186	香	3	16	5.333
187	馬	46	78	1.696
188	骨	7	8	1.143
189	高	1	5	5.000
190	髟	13	16	1.231
191	鬥	4	7	1.750
192	鬯	1	2	2.000
193	鬲	3	6	2.000
194	鬼	10	17	1.700
195	魚	38	59	1.553
196	鳥	57	82	1.439
197	鹵	4	7	1.750
198	鹿	11	20	1.818
199	麥	6	8	1.333
200	麻	3	5	1.667
201	黄	1	2	2.000
202	黍	4	10	2.500
203	黑	14	27	1.929
204	鬻	3	6	2.000
205	黽	5	11	2.200
206	鼎	1	1	1.000
207	鼓	3	4	1.333
208	鼠	4	5	1.250
				つづく

順	部首	掲出字数	和訓数	1 掲出字あたりの和訓数
209	鼻	3	4	1.333
210	齊	4	10	2.500
211	齒	10	15	1.500
212	龍	2	3	1.500
213	龜	2	2	1.000
214	龠	2	2	1.000
合計		6101	10,517	1.156

## 2.4 三つのデータベースの掲出字と和訓の基礎統計

### 2.4.1 共通する掲出字

本章に挙げた三辞書に共通する掲出字数は5,194字である。表2.9は共通する掲出字の各データベースの延べ和訓数を示す。

表2.9 三辞書に共通する掲出字の延べ和訓数

	KRM	MBWG	DJT
和訓数	13,701	5,228	5,489

### 2.4.2 共通する漢字と和訓の組合せ

本章に挙げた三辞書に掲出する和訓は延べ61,864語、異なり16,152語である。KRMでは和訓フィールドがないが、筆者が開発したマークアップツールを用いて、和訓の抽出が可能である(その詳細は第3章を参照)。

共通する和訓付き掲出字が2,802字、共通する和訓数が977語、共通する漢字一和訓の組合せが965組である。なお、観智院本『類聚名義抄』における掲出字が複数字である項目を除外している。最後に、KRMの出現順に共通する漢字と和訓の組合せを表2.10に示す。なお、品詞は『大字典』によるものである(名:名詞, 動:動詞, 形:形容詞, 副:副詞, 代:代名詞, 接:接続詞, 数:数

詞, 助: 助詞, 感: 感嘆詞).

以下, 多くの紙幅をとるが, 今後の漢字辞書研究の基礎資料として重要であるので, 煩をいとわず前例を示すものである.

表 2.10 三辞書に共通する漢字-和訓の組合せ

順	掲出字	和訓	品詞
1	人	ヒト	名
2	仕	ツカフ	動
3	任	マカス	動
4	佳	ヨシ	形
5	位	クラキ	名
6	倚	ヨル	動
7	傍	カタハラ	名
8	傍	ソフ	動
9	攸	トコロ	接
10	付	ツク	動
11	何	ナニ	代
12	俛	フス	動
13	儼	トシ	形
14	仲	ナカ	名
15	代	ヨ	名
16	代	カハル	動
17	伐	ウツ	動
18	伐	キル	動
19	伐	ホコル	動
20	伏	フス	動
21	仗	ヨル	動
22	倣	ナラフ	動
23	倣	イマシム	動
24	像	カタチ	名
25	俟	マツ	動
26	俛	ヤスシ	形
27	侑	ススム	動
つづく			

順	掲出字	和訓	品詞
28	側	カタハラ	名
29	偕	トモニ	副
30	傷	イタム	動
31	傷	ソコナフ	動
32	値	アフ	動
33	供	ソナフ	動
34	佐	タスク	動
35	佑	タスク	動
36	依	ヨル	動
37	依	タノム	動
38	促	セマル	動
39	使	ツカヒ	名
40	使	ツカフ	動
41	信	マコト	名
42	保	タモツ	動
43	伊	コレ	代
44	儀	ノリ	名
45	俚	イヤシ	形
46	伴	トモ	名
47	偈	セマル	動
48	傚	ムカフ	動
49	往	ユク	動
50	後	ノチ	名
51	後	ウシロ	名
52	徵	メス	動
53	徵	モトム	動
54	彼	カレ	代
55	御	ヲサム	動
56	傒	マツ	動
57	很	モトル	動
58	徂	ユク	動
59	徠	キタル	動
60	待	マツ	動
61	征	ユク	動
つづく			



順	掲出字	和訓	品詞
62	行	ユク	動
63	衢	チマタ	名
64	街	テラフ	動
65	衝	ツク	動
66	道	ミチ	名
67	迅	トシ	形
68	迄	イタル	動
69	這	ムカフ	動
70	速	マネク	動
71	速	トシ	形
72	逖	トホシ	形
73	適	タマタマ	副
74	途	ミチ	名
75	邁	ユク	動
76	通	トホル	動
77	遺	ワスル	動
78	遺	スツ	動
79	遣	ヤル	動
80	遣	ツカハス	動
81	遇	アフ	動
82	逝	ユク	動
83	迹	アト	名
84	造	ツクル	動
85	造	イタル	動
86	遠	トホシ	形
87	進	ススム	動
88	遐	トホシ	形
89	遭	アフ	動
90	迎	ムカフ	動
91	近	チカシ	形
92	之	コレ	代
93	之	ユク	動
94	匠	タクミ	名
95	區	マチマチ	副
			つづく

順	掲出字	和訓	品詞
96	赴	タケシ	形
97	赴	オモムク	動
98	起	タツ	動
99	一	ヒトツ	数
100	下	シモ	名
101	下	シタ	名
102	亘	サカヒ	名
103	至	イタル	動
104	工	タクミ	名
105	丘	ヲカ	名
106	晝	ヒル	名
107	旦	アシタ	名
108	世	ヨ	名
109	哥	ウタフ	動
110	西	ニシ	名
111	中	ナカ	名
112	中	アタル	動
113	申	サル	名
114	引	ヒク	動
115	辛	カラシ	形
116	博	ヒロシ	形
117	赤	アカシ	形
118	幸	サイハヒ	名
119	去	サル	動
120	在	アリ	動
121	左	タスク	動
122	右	タスク	動
123	有	アリ	動
124	喪	ウシナフ	動
125	逋	トシ	形
126	躬	ミ	名
127	軀	ミ	名
128	射	イル	動
129	耳	ミミ	名
			つづく

順	掲出字	和訓	品詞
130	聲	コエ	名
131	取	トル	動
132	聆	キク	動
133	如	ユク	動
134	孀	ヤモメ	名
135	妻	メ	名
136	妻	ツマ	名
137	娘	ムスメ	名
138	娠	ハラム	動
139	嫌	キラフ	動
140	嫉	ソネム	動
141	奴	ヤツコ	名
142	好	ヨシ	形
143	好	コノム	動
144	娶	メトル	動
145	妊	ハラム	動
146	委	クハシ	形
147	舌	シタ	名
148	憩	イコフ	動
149	憩	ヤスム	動
150	口	クチ	名
151	晒	ワラフ	動
152	叶	カナフ	動
153	咬	カム	動
154	兄	アニ	名
155	叩	タタク	動
156	唱	トナフ	動
157	吉	ヨシ	形
158	吹	フク	動
159	咏	ウタフ	動
160	召	メス	動
161	陷	クラフ	動
162	喫	クラフ	動
163	占	ウラナフ	動
			つづく

順	掲出字	和訓	品詞
164	善	ヨシ	形
165	哺	フクム	動
166	名	ナ	名
167	吐	ハク	動
168	吾	ワレ	代
169	鳴	ナル	動
170	鳴	ナク	動
171	目	メ	名
172	睨	ニラム	動
173	瞻	ミル	動
174	睽	ソムク	動
175	瞽	メシヒ	名
176	睛	ヒトミ	名
177	睦	ムツマシ	形
178	瞰	ミル	動
179	相	タスク	動
180	相	ミル	動
181	眉	マユ	名
182	自	ヨリ	助
183	鼻	ハナ	名
184	見	ミル	動
185	見	アラハル	動
186	覺	サトル	動
187	覲	マミユ	動
188	規	ノリ	名
189	曰	イハク	名
190	冒	オホフ	動
191	晶	アキラカ	名
192	星	ホシ	名
193	晨	アシタ	名
194	春	ハル	名
195	時	トキ	名
196	曠	ムナシ	形
197	旭	アサヒ	名
			つづく

順	掲出字	和訓	品詞
198	暑	アツシ	形
199	普	アマネシ	形
200	昔	ムカシ	名
201	易	カハル	動
202	易	ヤスシ	形
203	暗	クラシ	形
204	暗	ヤミ	名
205	是	コレ	代
206	替	スタル	動
207	杲	アキラカ	名
208	書	カク	動
209	書	シルス	動
210	書	フミ	名
211	曉	アカツキ	名
212	昧	クラシ	形
213	昨	キノフ	名
214	齊	ヒトシ	形
215	習	ナラフ	動
216	白	シロシ	形
217	的	マト	名
218	魄	タマシヒ	名
219	町	マチ	名
220	暇	ナハテ	名
221	里	サト	名
222	野	ノ	名
223	畔	ソムク	動
224	畔	クロ	名
225	卑	イヤシ	形
226	畜	ヤシナフ	動
227	肉	シシ	名
228	膺	ムネ	名
229	膾	ナマス	名
230	股	モモ	名
231	腹	ハラ	名
			つづく

順	掲出字	和訓	品詞
232	脅	ワキ	名
233	肚	ハラ	名
234	肴	サカナ	名
235	肝	キモ	名
236	勝	カツ	動
237	勝	マサル	動
238	背	ソムク	動
239	肩	カタ	名
240	腋	ワキ	名
241	胎	ハラム	動
242	腰	コシ	名
243	臀	シリ	名
244	用	モチフ	動
245	期	アフ	動
246	舟	フネ	名
247	船	フネ	名
248	舳	ヘ	名
249	鱸	トモ	名
250	骨	ホネ	名
251	髀	モモ	名
252	角	ツノ	名
253	賢	マサル	動
254	財	タカラ	名
255	貨	タカラ	名
256	賄	タカラ	名
257	賄	タマフ	動
258	賈	アキナフ	動
259	賈	タカラ	名
260	賈	タマフ	動
261	買	カフ	動
262	負	ソムク	動
263	質	スナホ	名
264	賤	イヤシ	形
265	贓	カクス	動
			つづく

順	掲出字	和訓	品詞
266	則	スナハチ	接
267	賜	タマフ	動
268	頭	カシラ	名
269	額	ヌカ	名
270	形	カタチ	名
271	形	アラハル	動
272	久	ヒサシ	形
273	長	マス	動
274	手	テ	名
275	指	サス	動
276	椿	ツク	動
277	把	トル	動
278	搦	クム	動
279	撃	ヒク	動
280	撃	ウツ	動
281	掃	ハラフ	動
282	據	ヨル	動
283	拍	ウツ	動
284	標	ウツ	動
285	援	タスク	動
286	援	ヒク	動
287	披	ヒラク	動
288	扶	タスク	動
289	拽	ヒク	動
290	掩	オホフ	動
291	攫	ツカム	動
292	捻	ヒネル	動
293	撲	ウツ	動
294	操	ミサヲ	名
295	操	トル	動
296	操	アヤツル	動
297	招	マネク	動
298	批	ウツ	動
299	摘	ツム	動
			つづく

順	掲出字	和訓	品詞
300	捷	カツ	動
301	搥	ウツ	動
302	挺	ヌク	動
303	承	ウケタマハル	動
304	揆	ハカル	動
305	擯	トル	動
306	杼	クム	動
307	撫	ヒロフ	動
308	撞	ツク	動
309	抵	アタル	動
310	格	ウツ	動
311	擗	トル	動
312	掘	ホル	動
313	拊	ウツ	動
314	揮	フルフ	動
315	振	フルフ	動
316	擠	オス	動
317	攘	ハラフ	動
318	攘	ヌスム	動
319	擘	サク	動
320	推	オス	動
321	擔	ニナフ	動
322	擥	ツラヌク	動
323	持	モツ	動
324	撥	ヲサム	動
325	掇	トル	動
326	擣	ツク	動
327	撚	ヒネル	動
328	採	トル	動
329	括	ククル	動
330	栖	スム	動
331	攝	ヲサム	動
332	攝	トル	動
333	挾	ハサム	動
			つづく

順	掲出字	和訓	品詞
334	抑	オサフ	動
335	抑	ソモソモ	接
336	擲	カラム	動
337	拂	ハラフ	動
338	拙	ツタナシ	形
339	摺	サシハサム	動
340	摺	ヒク	動
341	搔	カク	動
342	木	キ	名
343	樹	キ	名
344	櫻	サクラ	名
345	格	イタル	動
346	棟	アフチ	名
347	松	マツ	名
348	桃	モモ	名
349	李	スモモ	名
350	檜	ナラ	名
351	桁	ケタ	名
352	梭	ヒ	名
353	檜	ヒノキ	名
354	楹	ハシラ	名
355	板	イタ	名
356	椋	タルキ	名
357	村	ムラ	名
358	棧	カケハシ	名
359	梁	ハシ	名
360	桂	カツラ	名
361	棚	タナ	名
362	椽	タルキ	名
363	椎	ツチ	名
364	樵	キコリ	名
365	橋	ハシ	名
366	核	サネ	名
367	構	カマフ	動
			つづく

順	掲出字	和訓	品詞
368	果	ハタス	動
369	柱	ハシラ	名
370	桶	ヲケ	名
371	柄	エ	名
372	柰	カラナシ	名
373	櫛	カシハ	名
374	本	モト	名
375	櫛	クシ	名
376	牀	トコ	名
377	犬	イヌ	名
378	猛	タケシ	形
379	獨	ヒトリ	名
380	獄	ヒトヤ	名
381	猜	ソネム	動
382	猫	ネコ	名
383	牛	ウシ	名
384	牧	ヤシナフ	動
385	物	モノ	名
386	將	マサニ	副
387	牀	ユカ	名
388	乙	キノト	名
389	孔	アナ	名
390	己	ツチノト	名
391	也	ナリ	助
392	元	モト	名
393	免	マヌカル	動
394	无	ナシ	形
395	克	カツ	動
396	先	サキ	名
397	九	ココノツ	数
398	孰	タレ	代
399	就	ツク	動
400	奉	タテマツル	動
401	其	ソレ	代
			つづく

順	掲出字	和訓	品詞
402	具	ソナフ	動
403	與	アタフ	動
404	與	トモニ	副
405	與	クミス	動
406	輿	コシ	名
407	異	コトナリ	動
408	失	ウシナフ	動
409	共	トモニ	副
410	八	ヤツ	数
411	危	アヤフシ	形
412	羞	ススム	動
413	若	モシ	副
414	若	ワカシ	形
415	美	ヨシ	形
416	募	ツノル	動
417	勇	イサム	動
418	谷	タニ	名
419	天	アメ	名
420	奮	フルフ	動
421	奄	オホフ	動
422	夫	ソレ	代
423	火	ヒ	名
424	烝	ムス	動
425	燔	ヤク	動
426	燒	ヤク	動
427	然	シカリ	動
428	煎	イル	動
429	營	イトナム	動
430	熬	イル	動
431	無	ナシ	形
432	炭	スミ	名
433	墨	スミ	名
434	黧	クロシ	形
435	洞	ホラ	名
			つづく

順	掲出字	和訓	品詞
436	法	ノリ	名
437	海	ウミ	名
438	源	ミナモト	名
439	河	カハ	名
440	汁	シル	名
441	池	イケ	名
442	滌	アラフ	動
443	洒	アラフ	動
444	洗	アラフ	動
445	溢	アフル	動
446	潭	フチ	名
447	瀨	セ	名
448	濟	スクフ	動
449	淺	アサシ	形
450	淵	フチ	名
451	泣	ナク	動
452	淡	アハシ	形
453	洽	アマネシ	形
454	深	フカシ	形
455	渝	カハル	動
456	渥	アツシ	形
457	漏	モル	動
458	渡	ワタル	動
459	湄	ホトリ	名
460	汲	クム	動
461	溪	タニ	名
462	液	シル	名
463	浪	ナミ	名
464	波	ナミ	名
465	湊	ミナト	名
466	湊	アツマル	動
467	漬	ヒタス	動
468	濯	アラフ	動
469	萍	ウキクサ	名
			つづく

順	掲出字	和訓	品詞
470	潜	カクル	動
471	漆	ウルシ	名
472	治	ヲサム	動
473	湯	ユ	名
474	瀉	カタ	名
475	涌	ワク	動
476	澤	サハ	名
477	測	ハカル	動
478	洲	シマ	名
479	満	ミツ	動
480	薄	ウスシ	形
481	酒	サケ	名
482	冬	フユ	名
483	凄	サムシ	形
484	言	イフ	動
485	訊	トフ	動
486	識	シル	動
487	諧	トトノフ	動
488	託	ヨル	動
489	誰	タレ	名
490	誌	シルス	動
491	譏	ソシル	動
492	諱	イミナ	名
493	諱	イム	動
494	諍	アラソフ	動
495	語	カタラフ	動
496	語	カタル	動
497	諒	マコト	名
498	請	コフ	動
499	議	ハカル	動
500	訂	ハカル	動
501	計	ハカル	動
502	許	ユルス	動
503	討	ウツ	動
			つづく

順	掲出字	和訓	品詞
504	謗	ソシル	動
505	諸	モロモロ	名
506	謳	ウタフ	動
507	訾	ソシル	動
508	記	シルス	動
509	譽	ホマレ	名
510	諮	トフ	動
511	訕	ソシル	動
512	誹	ソシル	動
513	謹	ツツシム	動
514	詠	ウタフ	動
515	蹂	フム	動
516	蹈	フム	動
517	踏	フム	動
518	跣	スアシ	名
519	踏	ヌキアシ	名
520	躡	フム	動
521	蹤	アト	名
522	路	ミチ	名
523	踐	フム	動
524	立	タツ	動
525	端	ハシ	名
526	壹	ヒトツ	数
527	ト	ウラナフ	動
528	与	アタフ	動
529	臥	フス	動
530	眞	マ <sub>1</sub>	#N/A
531	北	キタ	名
532	此	コレ	代
533	齒	ハ	名
534	齒	ヨハヒ	名
535	齡	ヨハヒ	名
536	齧	カム	動
537	峯	ミネ	名
			つづく

順	掲出字	和訓	品詞
538	嶼	シマ	形
539	崖	キシ	名
540	岸	キシ	名
541	岫	クキ	名
542	嶺	ミネ	名
543	豈	アニ	副
544	石	イシ	名
545	斫	キル	動
546	碓	カラウス	名
547	磯	イソ	名
548	玉	タマ	名
549	皇	キミ	名
550	璞	アラタマ	名
551	理	ヲサム	動
552	珠	タマ	名
553	環	タマキ	名
554	琴	コト	名
555	色	イロ	名
556	鄙	イヤシ	形
557	鄰	トナリ	名
558	部	ワカツ	動
559	阜	ヲカ	名
560	阿	クマ	名
561	阻	ケハシ	形
562	陲	ホトリ	名
563	陂	ツツミ	名
564	階	ハシ	名
565	階	シナ	名
566	陵	ツカ	名
567	陵	ヲカ	名
568	隅	スミ	名
569	陬	スミ	名
570	險	ケハシ	形
571	附	ツク	動
			つづく

順	掲出字	和訓	品詞
572	坑	アナ	名
573	壁	カキ	名
574	均	ヒトシ	形
575	垢	アカ	名
576	壻	ムコ	名
577	渥	ヌル	動
578	堞	カキ	名
579	塊	ツチクレ	名
580	壓	オス	動
581	埜	チリ	名
582	坡	ツツミ	名
583	垣	カキ	名
584	堵	カキ	名
585	埒	アナ	名
586	堤	ツツミ	名
587	垠	キシ	名
588	埃	チリ	名
589	域	サカヒ	名
590	塵	チリ	名
591	堯	タカシ	形
592	堊	シラツチ	名
593	坂	サカ	名
594	墳	ツカ	名
595	塗	ヌル	動
596	塗	ミチ	名
597	境	サカヒ	名
598	心	ココロ	名
599	忡	ウレフ	動
600	傷	イタム	動
601	忘	ワスル	動
602	忌	イム	動
603	息	イキ	名
604	思	オモフ	動
605	患	ウレフ	動
			つづく



順	掲出字	和訓	品詞
606	慣	ナラフ	動
607	忤	サカフ	動
608	悍	イサム	動
609	恪	ツツシム	動
610	憶	オモフ	動
611	意	ココロ	名
612	慘	イタム	動
613	忱	マコト	名
614	慙	ウラム	動
615	協	カナフ	動
616	懟	ウラム	動
617	忽	タチマチ	副
618	恂	マコト	名
619	慎	ツツシム	動
620	慵	モノウシ	形
621	怒	イカル	動
622	忿	イカル	動
623	惟	コレ	代
624	懣	ツトム	動
625	愴	イタム	動
626	悛	アラタム	動
627	恨	ウラム	動
628	慊	ウラム	動
629	情	ココロ	名
630	情	ナサケ	名
631	想	オモフ	動
632	憲	ノリ	名
633	應	コタフ	動
634	憑	ヨル	動
635	悟	サトル	動
636	悽	イタム	動
637	布	ヌノ	名
638	布	シク	動
639	絲	イト	名
			つづく

順	掲出字	和訓	品詞
640	經	ノリ	名
641	緯	ヌキ	名
642	維	コレ	感
643	纂	アツム	動
644	索	モトム	動
645	袴	ハカマ	名
646	績	ウム	動
647	績	ワタ	名
648	紋	アヤ	名
649	級	シナ	名
650	紀	シルス	動
651	綿	ワタ	名
652	細	ホソシ	形
653	絢	アヤ	名
654	編	アム	動
655	綏	ヤスシ	形
656	綾	アヤ	名
657	純	イト	名
658	組	クミ	名
659	絮	ワタ	名
660	繩	ナハ	名
661	給	タマフ	動
662	納	ヲサム	動
663	紫	ムラサキ	名
664	紙	カミ	名
665	表	アラハス	動
666	裏	ツツム	動
667	袂	タモト	名
668	襜	ムツキ	名
669	褥	シトネ	名
670	裨	タスク	動
671	褰	ハカマ	名
672	裂	サク	動
673	袴	ハカマ	名
			つづく

順	掲出字	和訓	品詞
674	袋	フクロ	名
675	示	シメス	動
676	神	カミ	名
677	神	タマシヒ	名
678	祀	トシ	名
679	祝	イハフ	動
680	祭	マツリ	名
681	稻	イネ	名
682	私	ワタクシ	名
683	秋	アキ	名
684	稗	ヒエ	名
685	香	カ	名
686	稔	トシ	名
687	科	シナ	名
688	種	タネ	名
689	利	トシ	
690	粃	チマキ	名
691	糧	カテ	名
692	糖	アメ	名
693	粕	カス	名
694	料	ハカル	動
695	齋	モノイミ	名
696	市	イチ	名
697	京	ミヤコ	名
698	良	ヨシ	形
699	良	ヤヤ	副
700	秉	トル	動
701	亦	マタ	副
702	向	ムカフ	動
703	夜	ヨル	名
704	盾	タテ	名
705	亥	ヰ	名
706	重	オモシ	名
707	云	イフ	動
			つづく

順	掲出字	和訓	品詞
708	云	イハク	名
709	乖	ソムク	動
710	宜	ヨロシク	副
711	寡	スクナシ	形
712	寡	ヤモメ	名
713	寮	ツカサ	名
714	寒	サムシ	形
715	牢	カタシ	形
716	安	ヤスシ	形
717	家	イヘ	名
718	室	ムロ	名
719	宛	アタカモ	副
720	守	マモル	動
721	宗	ムネ	名
722	實	マコト	名
723	復	フス	動
724	富	トム	動
725	軍	イクサ	名
726	冥	クラシ	形
727	勻	ヒトシ	形
728	匈	ムネ	名
729	空	ムナシ	形
730	空	ソラ	名
731	窟	イハヤ	名
732	窟	アナ	名
733	竇	アナ	名
734	雨	アメ	名
735	霞	カスミ	名
736	霜	シモ	名
737	霧	キリ	名
738	雹	アラレ	名
739	霄	ソラ	名
740	震	フルフ	動
741	露	ツユ	名
			つづく

順	掲出字	和訓	品詞
742	雲	クモ	名
743	開	ヒラク	動
744	關	セキ	名
745	闡	ヒラク	動
746	聞	キク	動
747	閑	ナラフ	動
748	間	ヒマ	名
749	闢	ヒラク	動
750	闊	ヒロシ	形
751	閏	ウルフ	名
752	闇	クラシ	形
753	闇	ヤミ	名
754	闕	ヲハル	動
755	問	トフ	動
756	囑	カハヤ	名
757	圈	ヲリ	名
758	囿	ソノ	名
759	圖	ハカル	動
760	因	チナミ	名
761	團	カハヤ	名
762	固	マコト	名
763	固	カタシ	形
764	屠	ホフル	動
765	屠	コロス	動
766	屁	ヘ	名
767	尻	シリ	名
768	履	フム	動
769	雇	ヤトフ	動
770	启	ヒラク	名
771	所	トコロ	接
772	虎	トラ	名
773	庇	オホフ	動
774	麾	サシマネク	動
775	廢	スタル	動
つづく			

順	掲出字	和訓	品詞
776	廟	ヤシロ	名
777	摩	スル	動
778	靡	ナシ	名
779	庫	クラ	名
780	廂	ヒサシ	名
781	廁	カハヤ	名
782	慶	ヨシ	形
783	床	ユカ	名
784	辰	トキ	名
785	曆	コヨミ	名
786	疾	トシ	形
787	痕	アト	名
788	瘡	カサ	名
789	殘	ソコナフ	動
790	外	ホカ	名
791	夢	ユメ	名
792	子	コ	名
793	孚	マコト	名
794	了	ヲハル	動
795	斟	クム	動
796	寺	テラ	名
797	尊	タフトシ	形
798	慕	シタフ	動
799	花	ハナ	名
800	苛	カラシ	形
801	藥	クスリ	名
802	剪	キル	動
803	葦	アシ	名
804	蘆	アシ	名
805	葭	アシ	名
806	苔	トマ	名
807	芹	セリ	名
808	藍	アキ	名
809	蓑	ミノ	名
つづく			

順	掲出字	和訓	品詞
810	蘋	ウキクサ	名
811	藏	クラ	名
812	菲	ウスシ	形
813	薤	ニラ	名
814	菰	コモ	名
815	蒼	アヲシ	形
816	蓐	シトミ	名
817	苟	イヤシクモ	副
818	蔽	オホフ	動
819	蒔	マク	動
820	藿	イラカ	名
821	篁	タカムラ	名
822	箴	ヒル	動
823	篠	シノ	名
824	筒	ツツ	名
825	笠	カサ	名
826	範	ノリ	名
827	節	フシ	名
828	箱	ハコ	名
829	等	ヒトシ	形
830	筵	ムシロ	名
831	笛	フエ	名
832	力	チカラ	名
833	力	ツトム	動
834	助	タスク	動
835	幼	イトケナシ	形
836	務	ツトム	動
837	勸	ススム	動
838	勉	ツトム	動
839	加	クハフ	動
840	刀	カタナ	名
841	剖	サク	動
842	刖	アシキル	動
843	剋	カツ	動
			つづく

順	掲出字	和訓	品詞
844	剗	ソル	動
845	到	イタル	動
846	製	ツクル	動
847	翼	タスク	動
848	翦	キル	動
849	翕	アフ	動
850	翱	カケル	動
851	翔	カケル	動
852	毛	ケ	名
853	飴	アメ	名
854	釣	ツリ	名
855	鑠	ケス	動
856	錐	キリ	名
857	鍛	キタフ	動
858	鑄	イル	動
859	鉞	マサカリ	名
860	銘	シルス	動
861	鎧	ヨロヒ	名
862	錯	オク	動
863	鋪	シク	動
864	錦	ニシキ	名
865	欽	ツツシム	動
866	今	イマ	名
867	會	アフ	動
868	倉	クラ	名
869	禽	トリ	名
870	命	イノチ	名
871	念	オモフ	動
872	爪	ツメ	名
873	罪	ツミ	名
874	盤	サラ	名
875	監	ミル	動
876	畫	ハカル	動
877	四	ヨツ	数
			つづく

順	掲出字	和訓	品詞
878	血	チ	名
879	璧	カハラ	名
880	甌	コシキ	名
881	弓	ユミ	名
882	張	ハル	動
883	旗	ハタ	名
884	旛	ハタ	名
885	族	ヤカラ	名
886	放	ハナツ	動
887	知	シル	動
888	矩	ノリ	名
889	新	アタラシ	形
890	斬	キル	動
891	載	トシ	名
892	戦	タタカフ	動
893	式	ノリ	名
894	咸	ミナ	副
895	戕	ソコナフ	動
896	武	タケシ	形
897	戢	ヲサム	動
898	我	ワレ	代
899	裁	タツ	動
900	歡	ススル	動
901	歌	ウタ	名
902	歌	ウタフ	動
903	歐	ハク	動
904	又	マタ	接
905	父	チチ	名
906	爰	ココニ	副
907	敵	カタキ	名
908	敦	アツシ	形
909	救	スクフ	動
910	救	タスク	動
911	文	フミ	名
			つづく

順	掲出字	和訓	品詞
912	文	アヤ	名
913	毀	ソシル	動
914	毆	ウツ	動
915	韋	ソムク	動
916	車	クルマ	名
917	輔	タスク	動
918	輓	ヒク	動
919	輳	アツマル	動
920	駒	コマ	名
921	馳	ハス	動
922	驗	シルシ	名
923	鷗	カモメ	名
924	鷺	ウ	名
925	鳩	ハト	名
926	鶴	ヌエ	名
927	鷹	タカ	名
928	離	ツク	動
929	難	カタシ	形
930	雞	ニハトリ	名
931	雁	カリ	
932	鮫	サメ	名
933	鮭	サケ	名
934	鯉	カツヲ	名
935	鮒	フナ	名
936	鱒	マス	名
937	鮓	タコ	名
938	虫	ムシ	名
939	蟬	セミ	名
940	蚊	カ	名
941	螢	ホタル	名
942	蟹	カニ	名
943	蛭	ヒル	名
944	蝨	シラミ	名
945	鼬	イタチ	名
			つづく

順	掲出字	和訓	品詞
946	龜	カメ	名
947	酌	クム	動
948	醜	ミニクシ	形
949	辟	キミ	名
950	少	スクナシ	形
951	少	ワカシ	形
952	隸	ツク	動
953	民	タミ	名
954	釋	トク	動
955	釋	ユルス	動
956	甘	アマシ	形
957	束	ツカヌ	動
958	奏	ススム	動
959	疆	サカヒ	名
960	哿	ヨシ	形
961	舞	マフ	動
962	嘉	ヨシ	形
963	求	モトム	動
964	乃	スナハチ	接
965	氣	イキ	名

表 2.11 共通する漢字と和訓の組合せの品詞統計

品詞	数	パーセンテージ
名詞	422	43.7%
動詞	415	43.0%
形容詞	85	8.8%
副詞	16	1.7%
代名詞	12	1.2%
接続詞	6	0.6%
数詞	5	0.5%
助詞	2	0.2%
感嘆詞	1	0.1%
空白 <sup>2</sup>	1	0.1%
総計	965	100.0%

#### 2.4.3 品詞から見た共通する漢字とその和訓

前節で集計した三つの辞書に共通する漢字と和訓の組合せの品詞の統計を表 2.11 に示す。名詞と動詞は約各半分、全体のほとんどを占めている。名詞と動詞の次に、形容詞は約 9% を占めている。

#### 2.4.4 和訓数から見た共通する漢字とその和訓の対応

三つの辞書に共通する漢字と和訓の組合せにおいて、一つの漢字が複数の和訓を持つ場合もあるが、ほとんどが 1 漢字と 1 和訓の対応関係である(表 2.12)。原因としては、明治以降、漢字に対する読み方が定着されることがあげられる [19]。

表 2.12 共通する漢字—和訓組合せの漢字和訓対応

和訓数	数	パーセンテージ
1	847	93.7%
2	53	5.9%
3	4	0.4%
総計	904	100.0%

<sup>2</sup> 掲出字“雁”と和訓“カリ”の組合せである。『大字典』の項目に品詞が掲載されていない。

## 第3章

# 日本漢字辞書の項目構造の多様性に対応するマークアップ

### 3.1 はじめに

部首分類体の日本古辞書の項目構造は、掲出字と注文からなるが、注文は大別して音注、意義注、字体注、和訓の4要素からなる。音注は字音注、意義注は漢文意義注、義注、漢文義注と呼ぶことがある。さらにこの4要素には多様な形式で注記が施され、それぞれの要素に数種のタイプが認められる。さらに訓点、傍訓などデータ化しにくい情報が存する。したがって、日本古辞書のデータベース化には、項目構造の多様性に対応したタグセットとマークアップ・ツールの開発が求められる。本章では、観智院本『類聚名義抄』を対象にして、タグセットとそれに対応させた日本古辞書マークアップ・ツール「tagzuke」について述べる。

「平安時代漢字字書総合データベース」(HDIC)は日本平安時代に編纂された漢字字書の総合データベースである[20][15][8]。データベースの対象としては、高山寺本『篆隸万象名義』、天治本『新撰字鏡』、図書寮本『類聚名義抄』、観智院本『類聚名義抄』の4点となる。これらの内、図書寮本と観智院本の『類聚名義抄』は項目構造が複雑であり、多様性が顕著である。音注には、反切・類音注、仮名音注、声点などがあり、意義注は漢字1文字で注記されることが多いが、2文字以上の

こともある。字体注は、「正」「通」「俗」などが注記される。和訓は、万葉仮名や片仮名で注記され、声点が施されることも多い。詳細は後述する。

国語学の研究では、このような項目構造の多様性に対応した検索と表示が求められており、それを可能にするには、適切なマークアップが必要である。しかし、既存のマークアップ・ツールは効率性の点で満足できないところが少なくない。そこで、JavaScriptとHTMLでマルチデバイス対応のマークアップ・ツール「tagzuke」を独自に開発した。

なお、本章は文献の全体をマークアップすることよりも、日本古辞書の項目の内、特に注文の構造をマークアップする方法に注目するものである。

### 3.2 観智院本『類聚名義抄』全文データベースと注文項目の構造

図3.1を例にすると、注文の「音工 コウ クウ」は音注の類音注と仮名音注、「續也 事也成也」は義注、「タシカニ」は和訓、「功歟」は字体注である。字の周りの点は声調を表す声点であり、「ㄱ」と「、」はそれぞれの「音」と「也」の略符である。これらの要素は4大要素(レベル1)の下位要素(レベル2)となるので、音級、義級、字級という下位要素を設定する。対応関係をまと

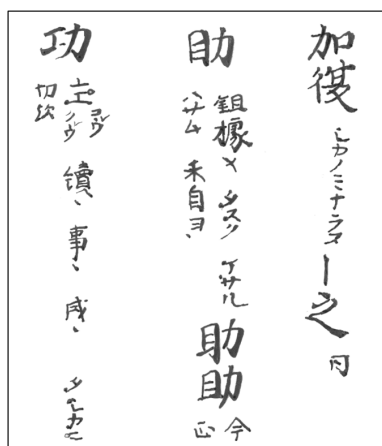


図 3.1 観智院本『類聚名義抄』記載例（原本模写）

めると、次の図 3.2 となる。

### 3.3 段階的なタグセット拡張

#### 3.3.1 辞書類を対象にしたマークアップの前例

マークアップ言語 XML は幅広く用いられ、メタ言語として非常に自由度が高いが、一方では互換性の低さをもたらす。人文系資料をデジタル化するための Text Encoding Initiative (TEI) が提唱され、コーパス化に多く利用されており、辞書のデータベースにも応用例がある [21][22]。資料の可視化によって、研究作業の負担低減もすでに実践されている [23]。詳細な階層構造を付けて、タグとそれの中に属性を付与することによって、資料に対する忠実度を高めることができる。

中国辞書のデータベースにおいて、漢字データベースプロジェクトの広韻データベースが XML の形式で公開されており、これは周祖謨『校正宋本廣韻』を用いて UCS 符号化したデータベースである [24]。目次、掲出字とそれらの注文は構造化されているが、マークアップは項目レベルに止まっており、注文中の要素については及んでいない。

日本の古辞書に対しては、申が図書寮本『類聚名義抄』をマークアップしており [25]、目次(篇)、掲出字、注文からなるすべての内容を構造化でき

る仕様を提案した。これは体系的なタグセットを目指しているが、図書寮本『類聚名義抄』には未読の箇所も少なくはなく、実際のマークアップ作業の実現性はまだ分からない。

観智院本『類聚名義抄』のマークアップの難易度について、『校正宋本廣韻』と図書寮本『類聚名義抄』を比較してまとめると表 3.1 となる。

観智院本『類聚名義抄』の分量は他の二書より著しく多いが、楷書で書写され、注文の間は空白で区切られていることから、注文要素の弁別が容易である。

#### 3.3.2 マークアップの準備

観智院本『類聚名義抄』の注文をマークアップする前に、まず、影印本のコピーに消せる蛍光ペンで仮マークアップを付ける作業を行い、マークアップの際に生じる可能性がある問題点を明らかにした。

仮マークアップの作業では、消せる蛍光ペンが見えやすいよう白地に黒字の風間書房版のコピーを用い、注文の 4 大要素ごとに色分けをした。

筆者と同じ研究室に所属し、観智院本『類聚名義抄』を研究資料として利用することの多い 3 人(李媛・鄭門鎬・張馨方)に依頼して、それぞれが分担して 1 人 10 ページずつ実験的に作業をした。その結果、次の 2 点の問題点が明らかになった。

1. 字音注の仮名音注と和訓との区別や、音・訓の出典と義注との区別は判断が難しい箇所があり、作業者の経験と知識によるところが大きい。
2. 蛍光ペンでの色分け作業はペンの切り替えに一定の時間がとられる。

この 2 点を解決するため、残りの部分の作業では次の 2 つの対策を講じた。

1. 風間書房版に付された和訓索引と広韻などの辞書類を参照しながら作業を行った。
2. 4 大要素の色付けは一人の作業者に 1 つの要



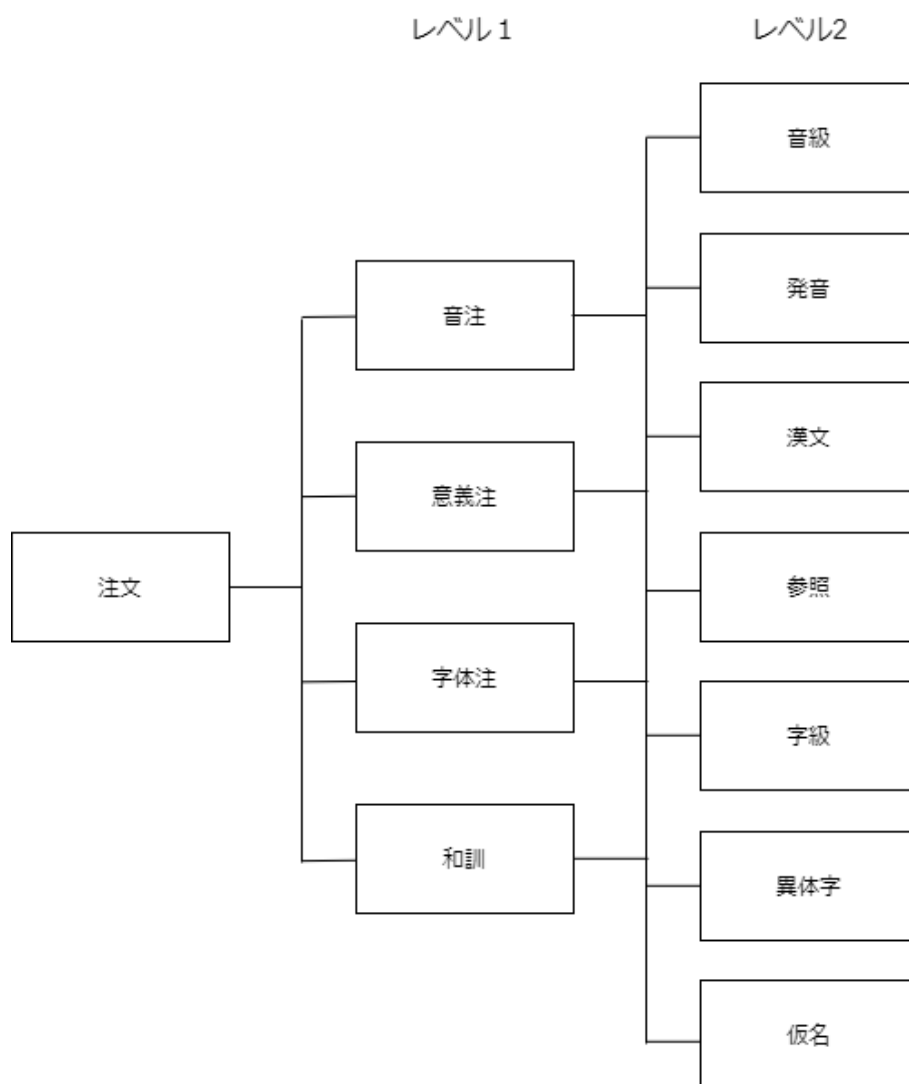


図 3.2 観智院本『類聚名義抄』の項目構造

表 3.1 マークアップの難易度の比較

	宋本広韻	図書寮本	観智院本
書体	楷書（印刷）	行書を交えた楷書（筆写）	楷書（筆写）
文字	漢字のみ	漢字が多く、仮名注が少ない	漢字が少なく、仮名注が多い
分量	554 頁	346 頁	1,331 頁
符号類	僅少	種類・量ともに多い	種類少なく、量多い
句読・種類・可読性	無	有・朱点・やや難	有・空白・容易

素を分担し、作業が終わると次の担当者と交代するようにした。なお、北海道大学教授の池田証壽は疑問箇所の検討を担当した。

結果として、合計の作業時間は40時間となり、平均して1項目に4.5秒かかった。

### 3.3.3 段階的なタグセット拡張

大規模なデータベースを構築する際には、共同で作業を分担することがよく行われている。しかし、XML言語の利用とTEIの応用にはいずれも一定の前提知識が必要となり、多人数の共同作業を行うのはかえって負担になると考えられる。

TEI P5の9. Dictionary (辞書類) 部分に記載されているタグセットと構造は日本古辞書に適用しにくい点が多い。例えば、注文の和訓を、定義を表す<def>と発音を表す<pron>のどちらに入れるべきかという問題がある。注釈<note>に入れて、属性@typeに注文の種類を表すこともできるが、<note>の下位タグに制限され、埋め項目の処理は難しい。そのため、本章はTEI標準に依拠せず、独自のタグセットを設定することにした。

注文は多様な要素からなるだけでなく、複雑な構造を有する。本章で設定したタグセットは浅い構造から深い構造への進行においても実現性が高い。特に字音注・和訓においては未解明な箇所もあり、研究の進行に合わせてタグが変更できる柔軟性が求められる。

また、既存のシステムとの互換性も考慮すべきである。HDICデータベースは関係データベースで構築されたため、既存のスキーマのまま注文の構造化を行うことにした。

したがって、本章は観智院本『類聚名義抄』のすべてをXML化するのではなく、既存のデータベースにある注文フィールドのみのマークアップから始める。マークアップのタグセットの命名には次の2つの条件を満たす必要がある。

1. 簡単な構造：タグ名が短く、入力の手間を低

減する。

2. 可読性：機械が読みやすく、人間が読みやすい。タグは一定の情報性が見えるものとし、<s></s>、<q></q>などは短すぎるので採用しない。

1と2は矛盾するが、バランスをとって、次の表3.2のとおり4大要素の基本タグセットを作成した。

表3.2 タグと要素の対応

タグ	要素
<jion>	音注
<jitai>	字体注
<wakun>	和訓
<kanbun>	(漢文) 義注

長さに対応性を考慮してタグ名を付与することにした。例えば、「和訓」の英訳に当たる<japanese-reading>などをやめ、日本語ローマ字のまま<wakun>をタグ名にした。「義注」は、「和訓」との区別を強調するため、<kanbun>にした。

研究対象の解読と研究を進めながら、段階的にマークアップを行う。3.2節の図3.2で述べたレベル1、レベル2、レベル3に対応するよう3つのステップに分けて表3.2に基づくマークアップの作業を行う。

ステップ1：分類

音注：

---

```
<jion>音工(L-R)</jion>
<jion>コウ(@N)</jion>
<jion>クウ(@N)</jion>
```

---

漢文注：

---

```
<kanbun>續也</kanbun>
<kanbun>事也</kanbun>
<kanbun>成也</kanbun>
```

---

和訓：

---

```
<wakun>タシカニ(LHLH)</wakun>
```

---

字体注：

---

```
<jitai>功歟</jitai>
```

---

ステップ 2：属性

音注：

---

```
<jion type="ruion">音工(L-R) </jion>
<jion type="kana">コウ(@N) </jion>
<jion type="kana">クウ(@N)</jion>
```

---

漢文注：

---

```
<kanbun type="single_char">續也</kanbun>
<kanbun type="single_char">事也</kanbun>
<kanbun type="single_char">成也</kanbun>
```

---

和訓：

---

```
<wakun>タシカニ(LHLH)</wakun>
```

---

字体注：

---

```
<jitai>功歟</jitai>
```

---

ステップ 3：階層

音注：

---

```
<jion type="ruion">
  <hojo>音</hojo>工 (L-R)
</jion>
<jion type="kana">コウ(@N) </jion>
<jion type="kana">クウ(@N)</jion>
```

---

漢文注：

---

```
<kanbun type="single_char">續<hojo>也</
  hojo></kanbun>
<kanbun type="single_char">事<hojo>也</
  hojo></kanbun>
<kanbun type="single_char">成<hojo>也</
  hojo></kanbun>
```

---

和訓：

---

```
<wakun>タシカニ(LHLH)</wakun>
```

---

字体注：

---

```
<jitai>功<hojo>歟</hojo></jitai>
```

---

ステップ 2 の属性名とステップ 3 のタグ名は現時点で仮とし、今後再検討する必要がある。

## 3.4 マークアップツールの必要性と現状

### 3.4.1 文系研究資源の作成について

これまでの文系の研究では一次資料（原文データ）から二次資料（抽出・加工済みの紙カード）、最後は三次資料としての研究成果という流れであった。最近では文系研究でもコンピュータを利用して、データベースやコーパスなどを作成して研究を進めている。つまり、二次資料はデジタルである場合が多くなっている。研究資料のデジタル化は選択的であり、デジタル化する要素があらかじめ標準化されている場合と標準化されていない場合の両方がある。

古辞書は日本語史の資料として有益であり、例えば和訓アクセント、字音の声点、字体注記がある。しかし各研究者の間で統一的な記載方法は存在しない。前述のように、部首分類体の日本古辞書の項目構造は複雑である。それを研究するため、前述のとおり蛍光ペンなどで標識して統計すること試みた。図 3.4.1 は観智院本『類聚名義抄』の注文に出現する要素を蛍光ペンで分類した例である。また、鉛筆などの筆記具で文献影写の隅に備考を追加したり、関係ある資料をカットしてまとめたりすることもある。これらの二次資料は原本側から見ると、原本に対して研究者の理解を示した一種の情報追加ともいえる。各研究者がそれぞれに工夫にしてエクセルなどでデータ処理を行っており、現段階では共通の入力・加工方法を行うことはなされていない。

二次資料をデジタル化して作成する際には、どのような方法で情報を追加して、どのような形式で追加する情報を記述したか、その過程を逐一記録しておくことが重要な課題となる。一方、マークアップツールを開発するには全体像がなければ進むことができない。大きな枠組の設定ができれば細部に入れる。そこでまず大きな枠組みを提案し、それを踏まえて、さらに細部の検討を進める

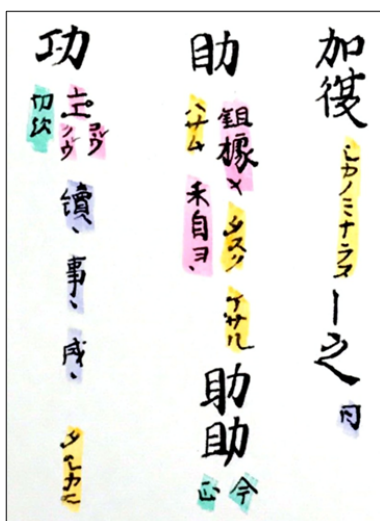


図 3.3 観智院本『類聚名義抄』の模写による注文マークアップ一例

ことにする。

### 3.4.2 データベースへのながれ

古辞書の研究の場合、文献に記載されている内容の符号化、およびその後の ID 付与やスキーマ設計などの構造化によって、データベースを構築すること自体はすでに原本画像以上の情報追加を行ったことになる。画像データベースでは、文献を画像化して、画像とテキストデータとの連携によって構築がされる。図 3.4.2 はそれらのイメージである。

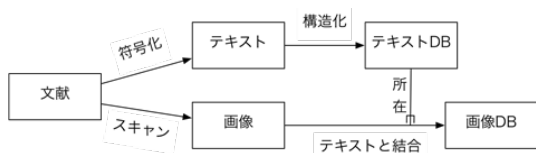


図 3.4 古辞書系データベース作成のながれ

以下では主としてテキストデータベースの構造化に際して必要となる二次資料の情報の追加方法を検討する。具体的には、古辞書に記載されている注文要素の分類について述べる。

テキストデータベースの実装では主にスキーマ

を事前に定義する RDB<sup>1</sup>とそうではない NoSQL DB に分けられる。古辞書は資料としてすでに一定の構造を持っており、RDB との相性が良いと言えるが、注文を扱う段階に入ると、スキーマの変動が多く生じるため、XML や JSON-LD などで実装するのが理想的であろう。現時点では共有形式の統一がまだなされていないため、本研究で XML 風の簡易タグを用いて古辞書の注文要素にマークアップを行う。

### 3.4.3 tagzuke の必要性

観智院本『類聚名義抄』の掲出字は約 42,000 字、注文は約 200,000 字と膨大である。その注文要素のマークアップには効率的な手法が強く求められる。

市販の XML エディターを利用してテキストデータをマークアップしたり、タグの編集をしたりすることができるが、ソフトのライセンス、学習コストを考えると、古辞書に特化したマークアップ・ツールの開発が必要である。

### 3.4.4 「tagzuke」の開発と利用

筆者は HTML と JavaScript でローカルの CSV ファイルをマークアップするツール「tagzuke」を開発した。マークアップ・ツールの効率化、簡易化を目指して、次の機能を実装した。

1. CSV (TSV) ファイルの読み込み
2. CSV (TSV) ファイルの書き出し
3. クリック操作による注文のタグの付与
4. 注文の修正

実際の操作画面は図 3.5 のとおりである。

作業の手順としては、ファイルを読み込み、画面の上にあるメニューのなかからタグを選択して、対応する注文要素をクリックすることによってマークアップする。マークアップする要素はオレンジ (和訓)、赤 (字音注)、緑 (字体注)、青

<sup>1</sup> 関係データベース

**Tag Zuke!**  
Version:0.2.0 ©LIU Guanwei

**Step 1**  
下にある「ファイルを選択」ボタンをクリック、CSVデータファイルをアップロードしてください。[サンプルデータ](#)  
ファイルを開く  
参照... sample.csv  
CSVファイルをアップロードしてください。  
- FileType: application/x-mdownload  
- FileSize: 272720 bytes  
- LastModified: 2017/11/14  
- FileType: application/vnd.ms-excel  
- FileSize: 1853 bytes  
- LastModified: 2017/11/14

**Step 2**  
ファイルの読み込みが終了後、小さいタグボタンをクリックしてタグの種類を選択できます。選択されたタグは大きいボタンで示します。タグボタンが選択された上で、注文をクリックすればタグを付けることができます。  
字音  
和訓 字音 字体 漢文 クリア  
修正

**Step 3**  
タグを付けた後、「保存」ボタンをクリックするとタグ付けたCSVファイルをダウンロードできます。ダウンロードしたファイルをもう一度アップロードすると続けて作業できます。  
保存ファイル

Status	KRID	KR2ID	KR_Tenri_p	KR_vol_radical	KR_vol_name	KR_radical	Entry	Entry_original	KR_def
✓完成	S00001	K0100131	Ta023310	v1#1	仏上	人	人		音仁(L)ニ、ヒト(H)、フレ(L)、サニ、マホル、ユク
✓完成	S00002	K0100133	Ta023331	v1#1	仏上	人	一(人)		ヒトリ(LH@)
✓完成	S00003	K0100133	Ta023332	v1#1	仏上	人	人		ヒトリ(LH@)
✓完成	S00004	K0100134	Ta023341	v1#1	仏上	人	二		フタリ(HHL)
✓完成	S00005	K0100134	Ta023342	v1#1	仏上	人	人		フタリ(HHL)
✓完成	S00006	K0100141	Ta023411	v1#1	仏上	人	五		イトリ(LHL)
✓完成	S00007	K0100141	Ta023412	v1#1	仏上	人	一(人)		イトリ(LHL)
✓完成	S00008	K0100142	Ta023421	v1#1	仏上	人	一(人)		ヒト、モ(HHHFH)
✓完成	S00009	K0100142	Ta023422	v1#1	仏上	人	等		ヒト、モ(HHHFH)
✓完成	S00010	K0100143	Ta023431	v1#1	仏上	人	漏		ユエ(L@)

Showing 1 to 10 of 27 rows 10 rows per page

図 3.5 tagzuke 操作画面

(義注) で示す。「保存」ボタンから作業ファイルをダウンロードして保存する。

### 3.5 ツールの有用性

前述のように、本マークアップ・ツールは HTML と JavaScript によって作成したもので、インターフェースはレスポンス・デザインを応用することによって、スマートフォンとタブレット端末でも作業できる。

マークアップ作業は、主に筆者が所属する研究室の 3 名の院生が担当し (3.3.2 を参照)、改善すべき点が発見され次第、修正して補った。本ツールの有用性について、各作業員 1 人あたり約 200 項目のマークアップで実験した。実験の結果として、1 項目の注文をマークアップするには平均約

6 秒かかった。前述の蛍光ペンによるマークアップ作業よりは遅いが、実用できる程度に達したと考えられる。ツールの改善と作業員の熟練度によって、作業のスピードアップが期待できる。

このツールには試作した旧バージョンとそれを改善した新バージョンがある。以下では新バージョンにおける改善点を中心に述べる。

### 3.6 ツールの改善

#### 3.6.1 改善点

本ツールを使えば、翻刻テキストの CSV ファイルを読み込み、マークアップした要素に色を付けて表示し、マークアップしていない要素にマウスのクリック操作によってタグ付けし、CSV ファイルとして書き出す、という一連の操作を効

率的に行うことができる。これまで観智院本『類聚名義抄』の注文をサンプルとして、ツールの有用性を検証してきたが、改善すべき点をいくつか見出したので、以下では、操作性、汎用性、維持性の三つの側面から述べていきたい。

- 操作性では、キーボードのカーソル操作を加え、予備に仮タグを付与することで、ユーザの操作量を減らした。
- 汎用性ではエクセルファイルの読み込み・書き出しに対応させ、フィールド・句切り符号の指定を可能にすることなど、従来、マークアップ前に必須な作業を軽減した。
- 維持性ではオープンソースのフレームワーク VueJS<sup>2</sup>を導入することによって、インターフェースをコンポーネントに分散して、開発を効率的に行うことで、オープンソースソフトウェアとして共有性を高めることができた。

順序として、3.6.2 では tagzuke の必要性と現状、3.6.3、3.6.4、3.6.5 では tagzuke の改善点について述べる。

### 3.6.2 新しい tagzuke の全体像

旧バージョン (3.4.4 を参照) では CSV ファイルを読み込み、マークアップした要素に色を付けて表示し、マークアップしていない要素にマウスのクリック操作によってタグ付けし、CSV ファイルとして書き出す、という操作の流れであった。

この旧バージョンを用いて、3名の作業者が約200項目をマークアップする実験を行った。実験の結果として、1項目の注文をマークアップするには平均約6秒かかった。また、原本のコピーに蛍光ペンで分類する作業では1項目に4.5秒かかった。旧バージョンによる作業は蛍光ペンの作業より遅いが、デジタルデータを作成することによって、数量を統計する作業効率の向上に一定の

成果があった。

旧バージョンに対する改善としては、可能な限り、タグ付与の自動化を目指した。新バージョンの操作の流れはつぎのどおりである。

1. ファイルを開く (図 3.6)
2. フィールド・区切り文字などの設定 (図 3.7)
3. タグ付け作業 (図 3.8)
4. 確認と保存 (図 3.9)

部首分類体の日本古辞書における注文要素は一定のパターンがある。そのパターンを用いて注文要素の自動マークアップを実現した。詳細は後述する (3.6.3 を参照)。

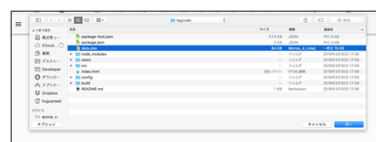


図 3.6 ファイルを開く画面



図 3.7 読取りの設定画面



図 3.8 タグ付けの操作画面

tagzuke は左のメニューパネルと右の操作パネルからなる。操作パネルの上部分はタイトル・ツールバー、中央部分はテキストデータ、下部分

<sup>2</sup> <https://vuejs.org/>



図 3.9 確認画面

はタグ操作区域である。マルチデバイスに対応するため、画面のサイズによってメニューパネルを自動的に隠すなどのレスポンスデザインを応用している。旧バージョンと比べた相違点を次の表 3.3 に示す。

表 3.3 新旧 tagzuke の機能比較

機能	旧	新
仮タグ	なし	あり
操作	マウス	キーボード・マウス
対応ファイル	CSV	CSV・エクセル

### 3.6.3 操作性に関する改善

操作性では、事前に仮タグ付与とキーボードのカーソル操作機能の追加を行なった。観智院本『類聚名義抄』の注文要素は漢字のみ、仮名のみ、漢字と仮名の混在の三つのパターンしかない。さらに、各要素は独自の特徴がある。例えば、字音注記の要素は「音」字で始まるか「反」字で終わるかのものがほとんどである。これらの特徴を利用してファイルを読み取るとともに注文要素の区別を文字列で判定し、特徴に合致したタグを事前につけるといったアイデアである。表 3.4 は注文要素の特徴を整理したものである。

また、キーボードのカーソル操作を加えて、仮タグを確認しながらタグを選択する。Tab キーと Shift キー + Tab キーで前後の項目を切り替える。カーソル移動キーの左方向キーと右方向キーで注文の要素を切り替える。上方向キーと下方向キーで要素のタグを選択する (図 3.10)。



図 3.10 tagzuke のタグ操作区域

### 3.6.4 汎用性に関する改善

汎用性ではエクセルファイルの読み込み・書き出しの対応を行なった。これによって、エクセルから CSV テキストファイルまでの変換は不要となる。文字化けや変換のミスで符号ずれなどを避けることができる。

さらにフィールド・句切り符号の指定が可能とした。以前の tagzuke では ID, 見出し, 注文は必ず既定のフィールドに格納しないとデータを認識しなかった。要素の間の区切り符号もユーザが自分で入力できるようにした。



図 3.11 高山寺本『篆隸万象名義』をマークアップする画面

また、観智院本『類聚名義抄』だけではなく、ほかの古辞書データセットにも対応できるように改良を加えた。例えば、HDIC ではすでに公開されている高山寺本『篆隸万象名義』データベースのテキストファイルを読み込むと、図 3.11 で示したように、マークアップすることができる。高山寺本『篆隸万象名義』には和訓がないため、字音注・字体注・意義注の 3 要素をマークアップする。

表 3.4 観智院本『類聚名義抄』の注文要素の特徴

要素	特徴
字音注	①「音」で始まる漢字のみの要素 ②「反」で終わる漢字のみの要素
和訓注	仮名のみの要素
字体注	「正」「俗」「籀」などの字体注記がある漢字のみの要素
漢文注	上記した字音注・字体注以外の漢字のみ要素

### 3.6.5 維持性に関する改善

維持性ではオープンソースのフレームワーク VueJS<sup>3</sup>を導入することによって、インターフェースをコンポーネントに分散して、開発を効率的に行い、コードの難読性を減らした。オープンソースソフトウェアとして共有性が高めることができる。ソースコードはオープンソースコード共有プラットフォーム GitHub の <https://github.com/toyjack/tagzuke> に格納して公開している。

### 3.6.6 改善後の効率

新しい tagzuke を利用して、2名の作業者がそれぞれ観智院本『類聚名義抄』の約 200 項目をマークアップするのに要した時間を記録した。仮タグによって作業の速度はだいぶ上がって、項目数と時間を加算して、合計 408 項目のマークアップ作業は 19 分で完成した。原本コピーに蛍光ペンでマークアップした場合と旧バージョンの tagzuke でマークアップした場合、そして新バージョンの tagzuke でマークアップした場合の 1 項目にかかるマークアップ作業の平均時間を次の表 3.5 に比較して示す。

表 3.5 1 項目あたりのマークアップ平均時間の比較

方法	時間
蛍光ペン	4.5 秒
旧 tagzuke	6 秒
新 tagzuke	2.8 秒

<sup>3</sup> <https://vuejs.org>



## 第4章

# 構造化した注文に基づく分析

### 4.1 観智院本『類聚名義抄』の熟語項目と異体項目

#### 4.1.1 はじめに

前章に述べたように、観智院本『類聚名義抄』は複雑な項目構造を持っている。掲出字は字数により、単字の形式（単字形式）と2字以上の掲出字の形式（複字形式）に分けられる。複字形式はさらに熟語に関する項目（熟語項目）、異体字に関する項目（異体項目）、その他の項目の三種類に分けることができる。図4.1は左から順に単字、異体字、熟語の例を挙げている。本章では、主に前章で構造化した注文データを用いて、観智院本『類聚名義抄』の項目を分析していきたい。

なお、注文が空白、「同」、「未詳」などで内容を判別しにくい項目をその他に分類する。

#### 4.1.2 先行研究

項目数については、表4.1に示した通り、酒井[26]が、観智院本の項目総数が約3万2千、熟字項目が約3千6百であるとした。一方、池田[20]は、項目総数がほぼ同様であることを確認したうえで、複字形式が約8千項目であるとした。この結果から、複字形式の約8千項目は、約3千6百の熟語項目とそれ以外の約4千4百項目とに分けられる[17]。

#### 4.1.3 観智院本『類聚名義抄』の項目数

次の表4.2は項目の字数によって統計した結果である。

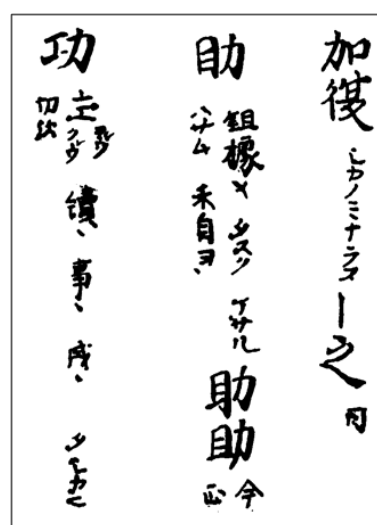


図4.1 観智院本『類聚名義抄』記載例（原本模写）

項目の文字数は1字と2字がほとんどを占めている。2字以上の項目には異体字と熟語の両方が含まれている。「力」部をサンプルとして、合計105項目に対して、次の表4.3のとおり項目種類と項目数を統計した。

#### 4.1.4 観智院本『類聚名義抄』の項目構造と要素

項目の性質は注文の種類に依存するため、注文の種類と内容の検討を優先的に行う。注文は音注、義注、字体注、和訓の4大要素からなる。1項目に4大要素がすべて用いられることは少ない。どの注記の種類が用いられるかは、それぞれ異なっており、4大要素の有無のパターンとしては16通りの可能性が想定される。「力」部を対象にし、注文のパターンをまとめると表4.4のお

表 4.1 観智院本項目数先行研究

	酒井憲二 (1967)	池田証壽 (2017 時点データ)
単字形式	不明	24,703
複字形式	不明	7911
熟語項目	3,589	不明
異体項目	不明	不明
その他	不明	不明
総項目数	32,613	32,614

表 4.2 観智院本『類聚名義抄』項目文字数

	仏	法	僧	項目数
1 字	8,671	8,776	6,068	23,515
2 字	2,138	2,025	2,899	7,062
3 字	270	324	516	1,110
4 字	65	42	81	188
5 字	21	24	16	61
6 字	3	9	10	22
7 字	2	4	2	8
8 字	2	2	2	6
9 字	2	1	3	6
12 字	1	0	0	1
計	11,175	11,207	9,597	31,979

表 4.3 「力」部項目数

字数	種類	項目数
1 字	単字	73
2 字以上	異体字	16
	熟語	16
総計		105

りになる（音＝音注，意＝意義注，字＝字体注，和＝和訓注，○は有，×は無）。

A はすべての 4 大要素を備えるもので，9 例しかない。B は 3 種類の要素を持っているものであるが，すべての場合で字音注を有する。C は 2 種類の要素を持っているものであり，6 の音注と義

注，8 の音注と和訓がほとんどを占める。D は 1 種類の要素しか持っていないものであり，15 の和訓のみが約 3 分の 2 を占める。注文の 4 大要素の組み合わせは注文の多様性を示している。

また，各々の注文の形式も多岐に亘ることがわかっており，更に注文を被注字にした場合があるため，非常に複雑な体裁で構成されている。

図 4.1 を例にすると，注文の「音工 コウ クウ」は音注の類音注と仮名音注，「續也 事也 成也」は義注，「タシカニ」は和訓，「功歟」は字体注である。字の周りの点は声調を表す声点であり，「ㄱ」と「、」はそれぞれの「音」と「也」の略符である。これらの要素は 4 大要素（レベル 1）の下位要素（レベル 2）となるので，音級，義級，字級という下位要素を設定する。それらの対応関係をまとめると，次の図 4.2（前節の図 3.2 を参照）となる。

#### 4.1.5 掲出字の形式と注文内容との対応

本節では，掲出字の「一」と注文の文字種・その順序などの記載方法に注目して，熟語項目と異体項目とをどこまで判別できるかを検証する。検証は複字形式の項目を中心に行う。

熟語項目と異体項目との判別の方法としては，次の三段階で進める。

1. 掲出字における前出字を省略して示す「一」の有無
2. 注文における文字種（漢字・仮名）の有無

表 4.4 「力」部注文の分類

分類	No	音	意	字	和	計	分類計
A	1	○	○	○	○	9	9(8.6%)
B	2	○	○	○	×	9	18(17.1%)
	3	○	○	×	○	6	
	4	○	×	○	○	3	
	5	×	○	○	○	0	
C	6	○	○	×	×	15	41(39.0%)
	7	○	×	○	×	4	
	8	○	×	×	○	18	
	9	×	○	○	×	0	
	10	×	×	○	○	0	
	11	×	×	○	○	4	
D	12	○	×	×	×	6	31(30.1%)
	13	×	○	×	×	0	
	14	×	×	○	×	7	
	15	×	×	×	○	18	
E	16	×	×	×	×	6	6(5.8%)
計						105	105(100%)

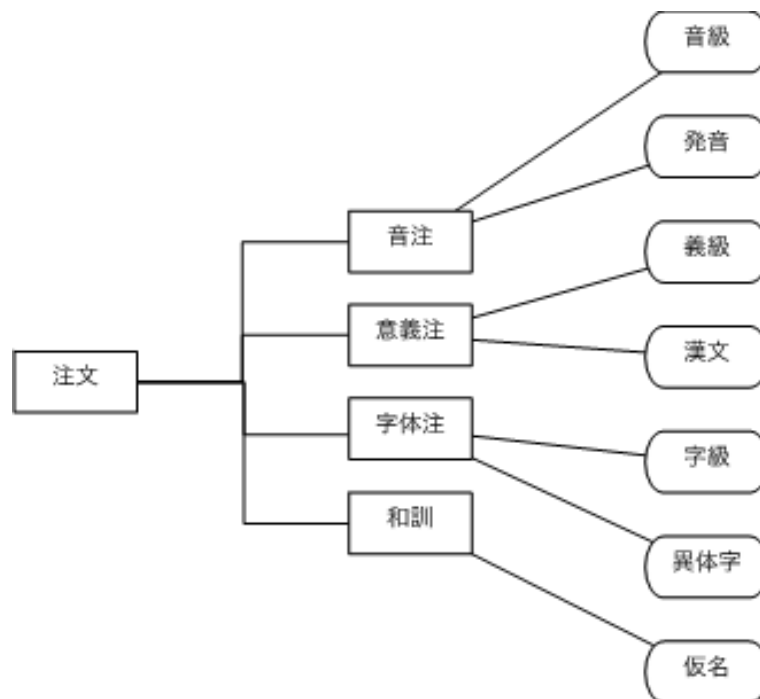


図 4.2 観智院本『類聚名義抄』の項目構造

## 3. 注文における文字種の配列順

まず、1から複字形式の約8千項目は「一」有の約2千項目と「一」無の約6千項目に二分できる(表4.5).

表4.5 掲出字における「一」の有無

縦棒	総計
あり	2,081
なし	5,834
総計	7,915

次に、2からa漢字のみの約3千2百項目、b仮名のみの約2千2百項目、c漢字と仮名併用の約2千4百項目、d空欄の約160項目に分けられる(表4.6).

表4.6 注文の文字種

	項目数
a漢字のみ	3,180
b仮名のみ	2,154
c漢字と仮名併用	2,417
d空欄	164
合計	7,915

最後に3からc漢字と仮名併用は①漢字で始まる約2千項目と②仮名で始まる約4百項目に分けられる(表4.7).

表4.7 文字種の配列順

	項目数
①漢字で始まる	1,985
②仮名で始まる	432
合計	2,417

以上の基準によって、複字形式の全体を通覧すると、熟語項目は、「一」有の約2千項目の8割以上を占めるのに対して、「一」無の2割弱を占めるに過ぎないことが判明した(表4.8).

表4.8 「一」の有無と項目分類

縦棒	熟語	異体	その他	合計
あり	1,788	22	271	2,081
なし	1,443	4,243	148	5,834
合計	3,231	4,265	419	7,915

さらに、注文の文字種と順序から、複字形式の掲出項目を熟語項目、異体項目、その他の項目に分類できる(表4.9).

表4.9 文字種出現順と項目分類

	熟語	異体	その他	合計
漢字先出	448	1,536	1	1,985
仮名先出	398	34	0	432
合計	846	1,570	1	2,417

分類した項目を対照すると、図4.3の対応関係が得られる。

## 4.1.6 おわりに

ここまでの結論は以下となる。

1. 仮名のみである項目は、ほぼすべてが熟語項目である。
2. 漢字のみである項目は、8割が異体項目である。
3. 漢字仮名併用の項目は、仮名が先に出現する例のほとんどが熟語項目である。
4. 掲出字に「一」がある項目は、ほとんどが熟語項目である。

## 4.2 観智院本『類聚名義抄』の基本項目と拡張項目

## 4.2.1 はじめに

4.1節では、掲出字の形式に着目して分析したが、本節では注文形式に着目して、基本項目と拡張項目とをどこまで判別できるかを検証する。基本項目の後ろに出現する拡張項目はその基本項目

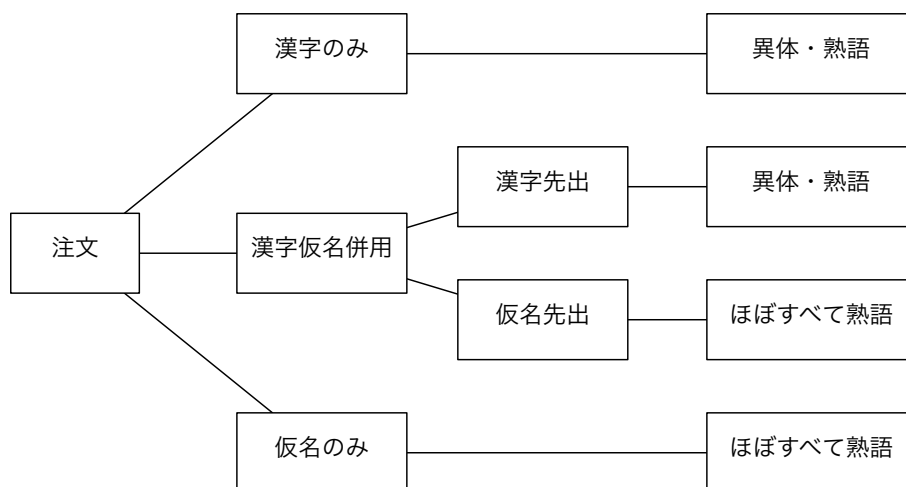


図 4.3 注文表記と内容の関係

の拡張項目とする。基本項目と拡張項目の定義を以下に示す。

- 基本項目：注文は掲出字の形・音・義のすべてが記入されている項目
- 拡張項目：注文は掲出字の形・音・義の一部が記入されている項目

観智院本『類聚名義抄』掲出項目の注文には字音注、字体注、漢文意味注、和訓の四つの要素が見える。掲出字の形は掲出字自身と字体注、音は字音注、義は漢文意味注と和訓によって表す。基本項目と比べると、拡張項目は掲出字に対して形・音・義の一部が記入される。

#### 4.2.2 判別方法

基本項目と拡張項目との判別の方法としては、次の三段階を進める。

1. 注文の形式上の特徴によって、要素を分類する。
2. 各要素の有無によって、基本項目を抽出する。
3. 基本項目の間に出現している拡張項目を抽出する。

段階 1 は 3.4.4 で述べた古辞書マークアップ・ツールを用いて分類した。

#### 4.2.3 要素形式を用いて自動判別

表 4.10 は各要素の形式上特徴を示している。これらの特徴を用いて、一括ですべての要素を分類することができる。

表 4.10 注文要素の形式上特徴

要素	特徴
字音注	①「音」から始まる漢字のみの要素 ②「反」で終わる漢字のみの要素
和訓注	仮名のみ要素
字体注	「正」「俗」「籀」などの字体表記がある漢字のみの要素
義注	上記以外の漢字のみの要素

#### 4.2.4 基本項目と拡張項目の統計

以上の方法で基本項目と拡張項目の抽出・分類ができる。分類した項目の掲出字形式を統計すると、表 4.11 となる。観智院本『類聚名義抄』に掲載している約 3 万 2 千項目の内、基本項目は約 1 万 5 千、拡張項目は残りの約 1 万 7 千 5 百である。基本項目の後ろに出現する拡張項目はその基本項目から拡張したものとみなされる。

単字項目は基本項目と拡張項目に半々の比率で分布している。複字項目は拡張項目であることが多いことが分かる。

次に、各巻の分布を表 4.12, 4.13 に示す。

表 4.11 掲出字と項目の分類

	単字	複字	合計
基本項目	12,324	2,771	15,095
拡張項目	12,373	5,144	17,517
合計	24,697	7,915	32,612

表 4.12 基本項目の分布

巻	基本項目数
仏上	948(6.3%)
仏中	1,560(10.3%)
仏下本	1,797(11.9%)
仏下末	721(4.8%)
法中	2,012(13.3%)
法上	1,581(10.5%)
法下	1,623(10.8%)
僧上	1,842(12.2%)
僧中	1,664(11.0%)
僧下	1,347(8.9%)
総計	15,095(100%)

表 4.13 拡張項目の分布

巻	拡張項目数
仏上	1,248(7.1%)
仏中	2,226(12.7%)
仏下本	871(5.0%)
仏下末	2,029(11.6%)
法上	1,995(11.4%)
法中	2,323(13.3%)
法下	2,014(11.5%)
僧上	1,706(9.7%)
僧中	1,416(8.1%)
僧下	1,689(9.6%)
総計	17,517(100%)

分布は基本に各巻の項目数と相当にするが、基本項目と拡張項目の比率分布を見てみると、総計した平均比率と比べ、仏下本と仏下末との両巻が異常な数値を示している(表 4.14)。

表 4.14 基本項目あたりの平均拡張項目数分布

巻	基本項目数	拡張項目数	比率
仏上	948	1,248	1.32
仏中	1,560	2,226	1.43
仏下本	1,797	871	0.48
仏下末	721	2,029	2.81
法中	2,012	1,995	0.99
法上	1,581	2,323	1.47
法下	1,623	2,014	1.24
僧上	1,842	1706	0.93
僧中	1,664	1,416	0.85
僧下	1,347	1,689	1.25
総計	15,095	17,517	1.16

表 4.15 は基本項目とそれに所属する拡張項目数の対照である。

基本項目の約 1 万 5 千は、拡張項目の有無により、拡張項目を持たない約 7 千 4 百、一つの拡張項目を持つ約 3 千 7 百、二つの拡張項目を持つ約 1 千 8 百、三つの拡張項目を持つ約 9 百、四つの拡張項目を持つ約 5 百、五つ以上の拡張項目を持つ約 7 百に分類できる。

#### 4.2.5 おわりに

基本項目は注文に 4 大要素が揃っている項目であり、観智院本『類聚名義抄』の撰者が重要視している項目とも言えるだろう。同時代の日本漢字辞書である新撰字鏡および篆隸万象名義と比較すると、基本項目は共通する項目の約 7 割を占めている(表 4.16)。課題として、これらの共通項目をさらに検討する必要がある。

また、拡張項目と見なされた項目には、熟語項目と異体字項目があり、注文がない項目や注文に「不詳」が記入された項目など、基本項目との関

表 4.15 基本項目とそれに所属する拡張項目数

拡張項目数	基本項目数	拡張項目数	基本項目数
0	7,399	16	4
1	3,781	17	6
2	1,795	18	1
3	897	19	2
4	480	20	1
5	287	22	1
6	141	25	1
7	96	26	1
8	61	27	2
9	41	30	1
10	30	31	1
11	20	32	1
12	11	38	1
13	12	44	1
14	10	48	1
15	8	65	1

表 4.16 他の字書と共通の項目と基本項目の比率

	共通項目	基本項目	パーセント
新撰字鏡	12,044	8,508	70.6%
篆隸万象名義	13,514	9,729	72.0%

表 4.17 複字形式の基本項目

分類	項目数
異体	2,291
熟語	456
不明	21
その他	3
総計	2,771

係を判断しにくいものもある（表 4.17）。これらの項目の性質を把握するため、さらに精査することが必要となる。

## 第5章

# 字形・字体に関する研究

### 5.1 GlyphWiki による観智院本『類聚名義抄』字形の再現

#### 5.1.1 はじめに

国語学研究資料のデジタル化が進んでおり、学術データベースの作成・公開が多く見られる。筆者らの所属する研究室で推進している平安時代漢字字書総合データベース (HDIC) プロジェクトは、近時刊行の高精細カラー版の観智院本『類聚名義抄』によって本文解読の精度を高めた上で、全文テキストデータベースを近日中に公開する予定である。古辞書が収録する掲出字には、次の三つの特徴がある。

(1) 古辞書は膨大な数の漢字を収録し、異体字も多数ある。また、古辞書にしか記録がなく、実用例が存しない字も多々ある。

(2) 古辞書のほとんどが古写本として伝存されているため、筆や書きぶりの影響で筆画の細かい部分を認識、抽象化することが困難である。

(3) 先行文献の引用や転写の過程で誤写の箇所が多い [27]。

研究や教育の目的による古辞書における漢字字形の利用は、前述の三つの特徴に直面するため、古辞書の原本画像を参照しなければならない状況がほとんどである。原本に書かれた字形を画像データとして扱うことは格段に容易になった。しかし、古辞書の原本画像は、所蔵機関の権利関係等により、二次利用の面において制約が課せられ

ている。加えて、古辞書の古写本における漢字字形は、難読の内容が多い。これは、同形異字・類形異字の存在、時間・空間の隔たりによる標準字体の変遷等の影響が関係する [8]。

研究資料としてのオープン性を高めると同時に解読した利便性の高い古写本字形表現をする方法が求められる。可能な限り筆写体である字形の字体上の特徴を残すと同時に、筆写の影響を可能な限り捨象して活字化を行う。そのデータをインターネットで共有する。

そこで、漢字字形共有データベース GlyphWiki による観智院本類聚名義抄字形の再現を考えついた。具体的には、次の三つのステップで進行する。

1. GlyphWiki を用いて、HDIC の公開用インターフェース HDIC Viewer に字形作成プラットフォームを追加する。(5.1.3.2)
2. 観智院本『類聚名義抄』において認められる外字を作成する。(5.1.4.1)
3. 同システム GlyphWiki で Unicode に既存する掲出字に対して、所在情報を付与することで、すべての掲出字字形の表現を試みる。(5.1.4.2)

なお、観智院本『類聚名義抄』に関する作字作業の概要を李・劉・池田 (2020) で指摘したが [27]、本節はその後の進行状況及び改善等を報告するものである。



## 5.1.2 字形情報利用の制限

### 5.1.2.1 符号化による制限

HDIC データベースは Unicode によって符号化されている。Unicode の収録字数は年々増加しているが、その収録方針によって、古辞書に収録された異体字に少なからず未登録のものがある。独自包摂基準を設定し、翻刻文字を原本と全く同様にするには限界がある。さらに、データベースにアクセスする端末には最新の Unicode 漢字を表示させるフォントの実装も期待できず、古辞書データベースのインターフェースとして、画像で漢字を表示することが避けられない。

### 5.1.2.2 著作権による制限

観智院本『類聚名義抄』は現在天理図書館に所蔵され、KRM での翻刻テキスト公開は同館から特別な許可を得て推進している。掲出字の原本画像を掲載・転載する際には、同館の要項に準じなければならない [28]。HDIC プロジェクトとしての利用は、許可を得たが、二次利用に関して、利用者による再度の申請が必要となる。

## 5.1.3 GlyphWiki による字形再現

### 5.1.3.1 GlyphWiki について

GlyphWiki は 2007 年 10 月から公開されている漢字字形自由共有サイトである。GlyphWiki では、ユーザー<sup>1</sup>が漢字字形を登録・編集することによって、インターネット上で漢字字形が共有可能となり、字形画像や外字フォントも同時に自動生成される。字形の編集には、専用エディタを採用し、容易に字形の骨格情報を操作できることから、これまでに大量の漢字・異体字の字形データが登録されている。GlyphWiki は、用途が限定されない。よって、開発者はあるテキストデータに含まれる外字集合をまとめて登録し、GlyphWiki を通じて外字フォントを公開することが可能となる。

### 5.1.3.2 字形データと Kage システム

GlyphWiki では、文字の情報しか登録できない。当該サイトの字形専用エディタを利用して、字形を筆画ごとに直線、曲線などの線種情報と、線の座標情報、および「はね」「ねらい」「とめ」などの端形状情報の 3 種類の情報を組み合わせて記録する [29]。GlyphWiki の開発者による Kage システムでは、これらの情報を利用して、字形の画像とフォントを生成できる。

### 5.1.3.3 作字プラットフォームの実現

作字プラットフォームは、HDIC の公開用インターフェース HDIC Viewer<sup>2</sup>から構築する。HDIC Viewer に格納している観智院本『類聚名義抄』テキストデータベースを用いて、ID と GlyphWiki に作成する字形<sup>3</sup>のリンクを生成して、所在・原本画像・作成した字形・注文などを一つのページにまとめる。現在、HDIC Viewer の “/krm-glyphwiki” から部首の索引にアクセスできる (図 5.1)。

ID	漢字	部	漢字	部	漢字	部	漢字	部
1	一	一	41	一	42	一	43	一
2	二	二	44	二	45	二	46	二
3	三	三	47	三	48	三	49	三
4	四	四	50	四	51	四	52	四
5	五	五	53	五	54	五	55	五
6	六	六	56	六	57	六	58	六
7	七	七	59	七	60	七	61	七
8	八	八	62	八	63	八	64	八
9	九	九	65	九	66	九	67	九

図 5.1 作字プラットフォーム部首索引 (部分)

部首のリンクをクリックすると、該当部首の作字リストページを表示する (図 5.2)。

作字済みの字形は原本画像の右に表示され、未完成の字形は赤色の “×” 符号となる。その符号をクリックすると、作字担当者が事前に登録した GlyphWiki ユーザー名を用いて、字形を登録する予定の GlyphWiki ページへ移動できる。

<sup>1</sup> GlyphWiki の用語に従い、ユーザーと表記する

<sup>2</sup> <https://hdic2.let.hokuda.ac.jp/krm-glyphwiki>

<sup>3</sup> GlyphWiki ではグリフと呼ぶ。

No	Charge	原形ID	原形	Glyph字形 (国)	Glyph字形 (数)	部首	注文
1	Shichuji	Nom-03038610	手	手	手	手	原形 (ア) アツカウ(ア) (手)
2	Shichuji	Nom-03038620	才	才	才	才	今製作此原字之形 ヒサク
3	Shichuji	Nom-03038630	手	手	手	手	タナス(ア) (手)
4	Shichuji	Nom-03038641	手	手	手	手	タナス(ア) (手)
5	Shichuji	Nom-03038642	手	手	手	手	タナス(ア) (手)

図 5.2 作字プラットフォームリストページ

GlyphWiki の字形 URL に関して次の 5.1.4.1 で詳述する。

#### 5.1.4 再現した字形とその利用

##### 5.1.4.1 再現した字形のアクセス

字形を GlyphWiki に登録する際、その名前を最初に決めなければならない。無秩序な登録を防ぐため、名前の登録には様々なルールがある [30]。また、GlyphWiki に登録した字形は誰でも修正・更新できることを基本とするが、確認・管理上の問題を避けるため、现阶段では GlyphWiki に登録する字形は登録者しか修正・更新できない“ユーザー占有グリフ”としている。作字担当者が事前に各自 GlyphWiki のユーザーを登録して、自分の“ユーザー占有グリフ”に字形を作成する。

各担当部分のグリフの登録・作字が完成した後、「hdic」というユーザーのユーザー占有グリフに集成し、字形の確認を行う。

GlyphWiki に登録する名前は“hdic”と、“hkrm-”と所在情報を組み合わせて設定する。次は登録済みの字形 URL の例である。

[https://glyphwiki.org/wiki/hdic\\_hkrm-03038630](https://glyphwiki.org/wiki/hdic_hkrm-03038630)

##### 5.1.4.2 観智院本類聚名義抄データベースの画像代用

HDIC Viewer は著作権とデータ精度を考慮して、現在 KRM のインターフェースは公開版と内部版とに分けている。内部版では、ユーザー登録制を採用して、観智院本の掲出字の原本画像を表

示することができるユーザーを研究グループのメンバーに限定している。一方、公開版では、今まで部分的なテキストしか公開されていなかった。しかし、GlyphWiki に登録した字形を導入することで、掲出字を原本画像の代わりとして表示することができるようになった (図 5.3)。

順	所在(風簡)	所在(天理)	部首	GlyphWiki	掲出字
1	仏下本帖39頁	仏巻301頁	手	駢	駢ノ一 (掌)

図 5.3 HDIC Viewer の KRM 公開版検索結果

##### 5.1.5 IDS を用いて作成した字形の検索

現時点まで、作成済みの字形は KRM のウェブインターフェースを用いて、翻刻された Unicode 漢字または所在情報で検索できる。文字同定や字形集成などの研究活動を行う際、同一漢字部品を含む漢字を特定する機能が求められる。つまり、字形で字形を検索することが必要となる。

しかし、現時点ではまだ KRM の字形データの完成度が低く、カバーできる掲出字数が少ないため、符号レベルで漢字の部品検索する機能を実装することが現実である。

部品検索の実現は以下のステップで行う。

1. 部品データの準備: CHISE 漢字構造情報データベースを用いて、漢字一部品から部品一漢字索引を作成する (図 5.4)。このようなデータ用意すれば、インターフェース側はより少ない計算で漢字を構成する最小部品を検索することができる。
2. インターフェース側の検索機能: オープン

ソース部品筆画漢字検索システム「零時字引」を参考して、部品と残りの筆画数で漢字を検索する機能を実装した。検索結果を KRM の Entry フィールドと照合して、掲出字のみのデータを表示する。

3. 字形の表示：すでに GlyphWiki に字形が登録された掲出字は、画像の字形を表示する。

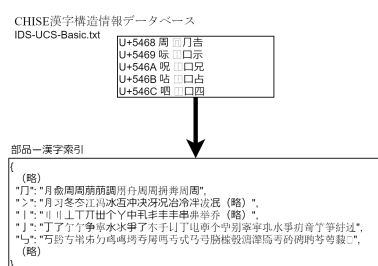


図 5.4 部品一漢字索引の生成

実装したシステムを図 5.5 に示す。



図 5.5 部品字形検索画面

KRM は Unicode を用いて符号化している。漢字構成の IDS データは入手しやすいが、符号と表示するフォントの対応関係は一对多のため、求める字形と異なる検索結果が出てくる可能性がある。このような字体と字形との対応に関して、5.1.7 節で詳しく述べる。

### 5.1.6 字形データ

先述のように、GlyphWiki に保存している字形の情報は Kage システムを用いて記録される (図 5.6)。Kage データは筆画の線の種類、位置、端形状からなっており、さらに別の掲出字に再利用することができる。単純な輪郭データと比べると、

より筆画の本質的な情報を使うため、データサイズは小さい [31]。

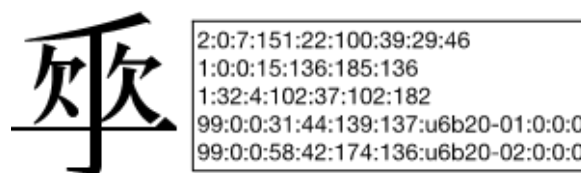


図 5.6 GlyphWiki の「手」の異体字とその Kage データ

写本字形を Kage データで記録することによって、字体が相違する部位を表示したり筆画単位で諸本間を比較したりするなど、写本文字の情報学的分析が可能になる。

### 5.1.7 字形情報と漢字の包摂粒度

IDS 情報は符号化された漢字の構造と利用された部品からなったものである。IDS の差分により、漢文字符の構成相違を比較することができるが、テキストデータ化された漢字文献を利用する際に、漢字の包摂基準を考えないといけない。特に、データ中の符号を字形研究の根拠として扱う場合、どこまで包摂された情報であるか、つまり、漢字の包摂粒度が重要な指標となる。包摂粒度について、守岡知彦氏は以下のように指摘している。

漢字の包摂粒度を考える場合、常用漢字表で用いられている『字種』-『字体』-『字形』という概念と、UCS 等の符号化文字集合で用いられている『抽象文字』-『グリフ』-『グリフイメージ』という概念等があるが、これらをざっくりとまとめると、大まかに言って、『字種』-『抽象文字』-『字体』≒『グリフ』-『字形』≒『グリフイメージ』という 4 階層の包摂粒度に整理することができる。 [32]

KRM の利用者端末における漢字環境は複雑であり、作成者が作成時に意図した字形のままに表示されるとは限らない。そのため、Unicode に登

表 5.1 KRM の包摂粒度

字種	抽象文字	字体 (グリフ)	字形 (グリフイメージ)
Entry	Entry	Entry または Entry_original	利用者端末のフォント

録された字形であっても、GlyphWiki を用いてグリフイメージを用意する。

### 5.1.8 作成字形と Unicode の関係

前述のように、Kage データは漢字の筆画を記述することによって字形を保存するものである。また、複数の筆画をグループとして、別の字や部品を引用することができる。図 5.6 の右にある Kage データの 2 行目、99 から始まるデータが例として挙げられる。

なお、GlyphWiki で字形を作成する際、既存する字形を引用することは強制的ではない。引用せず、一つの筆画を一行のデータにしても、画像的に字形の表現ができる。引用した字形を筆画に分割することができるが、GlyphWiki 上の関連字と名前を失うことになる。したがって、本研究における字形作成では、既存字形を積極的に引用することを方針にした。既存字形がない場合のみ、筆画情報で登録する。

現時点において、GlyphWiki の「hdic」ユーザーが占有する 2,100 個の字形を分析する。KRM に翻刻された Entry 欄に格納する掲出字との相違部分、およびよく出現するが引用できる字形がない部品を算出してみる (表 5.2)。

表 5.2 作成字形と Unicode の関係

分類	字数	
Entry のまま引用	870	
新作	部分 Entry の部品引用	765
	Entry の部品引用せず	75
対象外	IDS	359
	代用符号	31
合計	2,100	

翻刻した掲出字の Entry と GlyphWiki での作字と比較して、部品レベルで字形差があるものはほぼ半分 (対象外以外) を占めていることが分かる。

Entry のまま引用したのは 5.1.7 節で述べたグリフイメージとしての字形である。例を挙げると、観智院本『類聚名義抄』の仏下本 38 頁にある“手：音守，テ，テツカラ”の掲出字「手」は、GlyphWiki に登録されたグリフデータが“99:0:0:0:0:200:200:u624b-j”である (図 5.7)。つまり、Unicode に登録した日本ソースの“手” — “u624b-j” という名前の字形のみを引用したことを示している。

部分 Entry の部品引用は、部品は部分に Entry の掲出字と同様である新作字形である。仏下本 126 頁にある“齒：音葵 (蔡)，芥”の掲出字“齒”は例として挙げられる。作成された字形は，“𪗇”となる。そのグリフデータを次に示す。

99:0:0:8:75:187:194:u708e-j:0:0:0

99:0:0:-1:52:209:200:u5196-03:0:0:0

99:0:0:7:1:195:150:u6797-03:0:0:0

“𪗇”は u708e の“炎”，u5196 の“一”，u6797 の“林”の三つの字形からなる。翻刻された“齒”と比べると，“一”と“口”の相違がみられる。このような字形は、合計 1,760 回に及んで Entry の部品字形を引用している。一字形は平均約 2.3 回引用する。

字形のグリフデータの蓄積が進めば、翻刻された漢字と原本字形との差分を明らかにすることができる。さらに諸本のグリフデータを揃えることができれば、諸本間の字形相違を情報学手法で観察することが期待できる。

147433	hayama_yaenadeshiko	u3013	99:0:0:2:6:199:114:u0279-07@3
147434	hdic_hkrm-03038610	u3013	99:0:0:0:0:200:200:u624b-j
147435	hdic_hkrm-03038620	u3013	1:0:0:31:55:169:5531:0:4:105:
147436	hdic_hkrm-03038630	u200bf	1:0:0:15:136:185:136\$99:0:0:3'
147437	hdic_hkrm-03038641	u3013	99:0:0:0:0:200:200:u624b-j
147438	hdic_hkrm-03038642	u3013	99:0:0:0:0:200:200:u6b50-j
147439	hdic_hkrm-03038711	u3013	99:0:0:0:0:200:200:hdic_hkrm-

図 5.7 GlyphWiki のグリフデータ形式

### 5.1.9 おわりに

以上、GlyphWiki を利用して、観智院本『類聚名義抄』の掲出字を原本字形に近い形式に再現することを試みた。現在までの作業結果は 4 部首（人部・水部・手部・木部）である。将来的には、登録した字形の IDS を追加して、字形の構造と利用した漢字部品の情報を連関させることを考えている。これらの字形データを蓄積することで、漢字規格への符号登録や情報学的字体分析の一助となることを期待している。

## 5.2 漢字辞書から異体字情報の抽出：宋本『玉篇』を例に

### 5.2.1 はじめに

第 3 章では、日本漢字辞書の項目構造の多様性に対応できるマークアップツール tagzuke について述べた。本節では、このツールを用いて、HDIC に公開された宋本『玉篇』（SYP）から異体字情報を抽出してみる。

前述のとおり、日本漢字辞書の注文は字音注記、字体注記、漢文義注、和訓の 4 種類の要素に分けられる。各要素は異なる表現形式を持ち、各分野の研究者が関心寄せる要素も異なる。データベースから迅速かつ正確にこれらの要素を抽出することが強く求められている。さらに、文献データベースを構築・校正する際、各要素はそれぞれの方針で行われている。例を挙げると、字体注記はなるべく原本に近い字形を採用するべきとする一方、検索の利便性のため、字音注記は後世の音韻資料と同一の字体を選べべきだと考えられている。

中国の古典籍分野では、Markus というテキス

トマークアップツールがすでに周知され、Markus を用いれば文献に出現した人名、地名、年代などをマークアップすることができる。しかし、辞書系資料に上記の要素を応用するのは難しい。

筆者が開発したマークアップツール tagzuke は漢字辞書データベースに記載された注文を分析して自動マークアップすることができる。宋本『玉篇』は日本の漢字辞書より要素が 1 種類少ないため、tagzuke で容易に対応できるだろう。

### 5.2.2 宋本玉篇テキストデータベース

『大広益会玉篇』（宋本玉篇）は宋・陳彭年等撰、1013 年に成立した部首分類体漢字辞書である。542 部首に 22,809 字を収録する<sup>4</sup>。宋本玉篇は原本玉篇にあった詳細な訓詁を簡略化して、掲出字を増補したものである [15]。日本漢字辞書の篆隸万象名義は原本玉篇を抄録した辞書であるため、原本玉篇が散逸した現在、宋本玉篇は日本漢字辞書の研究における不可欠な資料である。

本研究は、HDIC で公開された宋本玉篇テキストデータベース（SYP）を用いている。池田 [15] はその構造を述べている。

### 5.2.3 宋本『玉篇』各要素の特徴

宋本玉篇の注文は字音注記（音注）、字体注記（字体注）、意義注記（義注）の 3 要素に分けられる。この 3 要素は表記上に、以下の特徴が見られる（○は一漢字を表す）。

字音注：○○切，○○○○二切，音○，又音○，亦音○，○○二音等。

字体注：俗作○，俗○字，亦作，或作，本作，籀文作，篆文作，本亦作，原作，正作，今作，俗以

<sup>4</sup> HDIC 宋本玉篇テキストデータベース Version 1.13 による。

表 5.3 宋本玉篇テキストデータベースの例

ID	掲出字	注文
a005b011	弌	古文.
a005b012	天	他前切.《説文》曰：天顛也，至高無上，从一大.《爾雅》曰：春爲蒼天，夏爲昊天，秋爲旻天，冬爲上天.《詩傳》云：尊而君之則稱皇天，元氣廣大則稱昊天，仁覆閔下則稱旻天，自上降監則稱上天，據遠視之蒼蒼然則稱蒼天.《呂氏春秋》云：天有九野，東方蒼，東南方陽，南方炎，西南方朱，西方顛，西北方幽，北方玄，東北方變，中央鈞.《太玄經》曰：九天，一爲中，二爲羨，三爲從，四爲更，五爲睟，六爲廓，七爲減，八爲沈，九爲成.《釋名》曰：天豫司堯冀以舌腹言之.天顯也，在上高顯也.青徐以舌頭言之.天坦也，坦然高而遠也.
a005b051	𠄎	並古文.
a005b052	𠄎	並古文.
a005b053	元	五袁切.《説文》曰：元始也.《易》曰：元者善之長也.《春秋傳》曰：狄人歸其元元首也.《韓詩》曰：元長也.《公羊傳》曰：元年春者何君之始年也.何休云：變一言元元者氣也.《左氏傳注》曰：凡人君即位欲其體元以居正故不言一年一月也.《漢書》曰：勸元元.師古曰：元元善意.
a005b071	丕	普邳切.《虞書》曰：嘉乃丕績.孔安國曰：丕大也.或作平.
a005b072	吏	力致切.《説文》曰：治人者也.《夏書》曰：天吏逸德烈于猛火.《周禮》：八則，三曰廢置以馭其吏.《左氏傳》曰：王使委於三吏.杜預曰：三吏三公也.《禮記》曰：五官之長，曰伯.其擯於天子也，曰天子之吏.
a005b091	𠄎	竹瓦切.

璫瑁作玳，古文等。

義注：○○也，○○兒（貌）等。

なお，“同上”“同○也”など，他の項目を参照する注記は本研究の対象外としている。

#### 5.2.4 マークアップした結果

表 5.3 は利用した宋本玉篇テキストデータベースの一部である。それを tagzuke でマークアップした結果を表 5.4 で示す。

#### 5.2.5 異体字情報の抽出

表 5.4 の字体注記のみを抽出した結果を，表 5.5 で示す。

抽出した字体注にある掲出字をまとめると，表 5.6 のとおり，宋本玉篇における異体字関係表を

作成することができる。

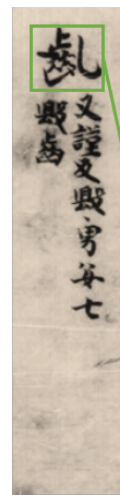
このような異体字関係データを各辞書から抽出して，対照すれば，さらに大規模な異体字データベースを作成することができる。図 5.8 は宋本玉篇と篆隸万象名義との比較を一例として挙げる。

表 5.4 マークアップした注文

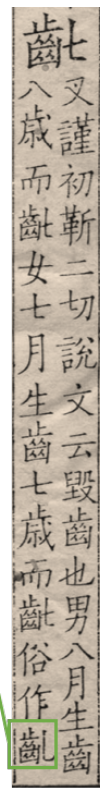
ID	掲出字	注文
a005a101	一	<ziyin>於逸切</ziyin> <shiyi>《說文》曰：惟初太始，道立於一，造分天地，化成萬物</shiyi> <shiyi>《道德經》云：昔之得一者，天得一以清，地得一以寧，神得一以靈，谷得一以盈，萬物得一以生，侯王得一以爲天下正</shiyi> <shiyi>王弼曰：一者數之始也，物之極也</shiyi> <shiyi>又同也，少也，初也</shiyi> <ziti>或作壹</ziti>
a005b011	弌	<ziti>古文</ziti>
a005b012	天	<ziyin>他前切</ziyin> <shiyi>《說文》曰：天顛也，至高無上，从一大</shiyi> <shiyi>《爾雅》曰：春爲蒼天，夏爲昊天，秋爲旻天，冬爲上天</shiyi> <shiyi>《詩傳》云：尊而君之則稱皇天，元氣廣大則稱昊天，仁覆閔下則稱旻天，自上降監則稱上天，據遠視之蒼蒼然則稱蒼天</shiyi> <shiyi>《呂氏春秋》云：天有九野，東方蒼，東南方陽，南方炎，西南方朱，西方顛，西北方幽，北方玄，東北方變，中央鈞</shiyi> <shiyi>《太玄經》曰：九天，一爲中，二爲羨，三爲從，四爲更，五爲睟，六爲廓，七爲減，八爲沈，九爲成</shiyi> <shiyi>《釋名》曰：天豫司堯冀以舌腹言之</shiyi> <shiyi>天顯也，在上高顯也</shiyi> 青徐以舌頭言之 <shiyi>天坦也，坦然高而遠也</shiyi>
a005b051	禿	<ziti>並古文</ziti>
a005b052	禿	<ziti>並古文</ziti>
a005b053	元	<ziyin>五袁切</ziyin> <shiyi>《說文》曰：元始也</shiyi> <shiyi>《易》曰：元者善之長也</shiyi> <shiyi>《春秋傳》曰：狄人歸其元元首也</shiyi> <shiyi>《韓詩》曰：元長也</shiyi> <shiyi>《公羊傳》曰：元年春者何君之始年也</shiyi> <shiyi>何休云：變一言元元者氣也</shiyi> <shiyi>《左氏傳注》曰：凡人君即位欲其體元以居正故不言一年一月也</shiyi> <shiyi>《漢書》曰：勸元元</shiyi> 師古曰：元元善意
a005b071	丕	<ziyin>普邳切</ziyin> <shiyi>《虞書》曰：嘉乃丕績</shiyi> <shiyi>孔安國曰：丕大也</shiyi> <ziti>或作丕</ziti>
a005b072	吏	<ziyin>力致切</ziyin> <shiyi>《說文》曰：治人者也</shiyi> <shiyi>《夏書》曰：天吏逸德烈于猛火</shiyi> <shiyi>《周禮》：八則，三曰廢置以馭其吏</shiyi> <shiyi>《左氏傳》曰：王使委於三吏</shiyi> <shiyi>杜預曰：三吏三公也</shiyi> <shiyi>《禮記》曰：五官之長，曰伯</shiyi> 其擯於天子也，曰天子之吏
a005b091	𠂔	<ziyin>竹瓦切</ziyin>

表 5.5 注文より抽出した字体注記

ID	掲出字	字体注記
a005a101	一	<ziti>或作壹</ziti>
a005b011	弌	<ziti>古文</ziti>
a005b051	𠄎	<ziti>並古文</ziti>
a005b052	𠄎	<ziti>並古文</ziti>
a005b071	丕	<ziti>或作平</ziti>
a006a011	上	<ziti>古文</ziti>
a006a031	帝	<ziti>古文</ziti>
a006a041	𠄎	<ziti>籀文</ziti>
a006a042	𠄎	<ziti>（並古文</ziti>
a006a043	𠄎	<ziti>並古文</ziti>
a006a081	𠄎	<ziti>古文</ziti>
a006a091	祇	<ziti>《周禮》：亦作示</ziti>
a006a102	𠄎	<ziti>籀文</ziti>



篆隸万象名義



宋本玉篇

表 5.6 宋本玉篇異体字関係

掲出字	異体字
一	壹
丕	平
祇	示
柴	柴
𠄎	𠄎
祐	佑
𠄎	𠄎
𠄎	𠄎
𠄎	𠄎
𠄎	𠄎
𠄎	𠄎
𠄎	𠄎
𠄎	𠄎
𠄎	𠄎
𠄎	𠄎

図 5.8 宋本玉篇と篆隸万象名義の異体字対照



## 第 6 章

# 漢字辞書データベース公開用インターフェースの開発

### 6.1 平安時代漢字字書総合データベース検索インターフェース：HDIC Viewer

#### 6.1.1 HDIC Viewer とは

HDIC Viewer は最初 HDIC（平安時代漢字字書総合データベース）を公開・検索するためのインターフェースとして開発された。HDIC 本文の校訂・新技術の検証が増え、HDIC 以外の古辞書データベースを追加した。現在は <https://hdic2.let.hokudai.ac.jp/> で公開している。

#### 6.1.2 Web インターフェースの提供

HDIC Viewer は PC と携帯端末の両方に対応している。PC でアクセスする際の画面を図 6.1 に示す。



図 6.1 HDIC Viewer ホーム画面

上にあるメニューから各データベースをアクセスできる。「HDIC 本編」と「HDIC 拡張篇」メニューに分けている。なお、古辞書データベースを構築・公開するのに有益なツールやユーティリ

ティなどは「ツールなど」にいている。

現時点では、HDIC 本編では HDIC プロジェクトが公開している篆隸万象名義，新撰字鏡，観智院本類聚名義抄<sup>1</sup>，宋本玉篇を収納，HDIC 拡張篇では筆者が作成したかまたは作成に参加したデータベースである大字典和訓データベース，夢梅本和玉篇データベース，本草和名データベース，一切経音義データベースを格納している。

#### 6.1.3 Web API の提供

##### 6.1.3.1 Web API の意義

HDIC は従来の学術データベースと違い、公開と校正を同時に進行している。そのため、常に最新版のデータを入手するには GitHub のレポジトリの動向を監視しなければならない。最新のデータを利用するには不便である。

Web API の提供によって、より大量、快速で最新の情報を引用することが可能となる。

##### 6.1.3.2 HDIC Viewer Web API のスキーマ

HDIC Viewer Web API は HDIC Viewer の Web インターフェースと同様に HTTPS を利用している。レスポンスはすべてが JSON 形式となる。API の構造はバージョン、字書名略称、動作名からなる。現在のバージョンは「v1」である。つまり、エンドポイントは

<sup>1</sup> 一般利用者には部分公開。

<https://hdic2.let.hokudai.ac.jp/api/v1>

となる。利用できるデータベースとそれらの略称は次の表 6.1 を参照。

表 6.1 データベースの略称

データベース	略称
篆隸万象名義	ktb
新撰字鏡	tsj
大広益会玉篇	syp

動作名は各古辞書独自のものと共通のもの両方があり、共通のアクションは検索の“search”と表示の“show”がある。検索のパラメータは掲出字検索の“entry”と注文検索の“def”がある。次は『新撰字鏡』の注文に「俗」が記載している項目を検索する例である。

<https://hdic2.let.hokudai.ac.jp/api/v1/tsj/search?entry=&def=俗>

検索の結果は JSON のアレーとして回答する。一項目を一つのオブジェクトにする。

```
[
  {
    "TSJ2ID": "s0107b505",
    "Entry_word": "昭",
    "SJ_def": "正音：止遙反。平：著也，光明也，明也，覲也。爲俗字。時昭反。去。",
    "SJ_wakun": "",
    "SJ_research": "",
    "remarks": ""
  },
  {
    "TSJ2ID": "s0108a604",
    "Entry_word": "晋",
    "SJ_def": "従口。古吝反。去：進也。□(晋)俗作。",
    "SJ_wakun": "",
    "SJ_research": "",
    "remarks": ""
  },
  (中略)
]
```

表示は ID と Unicode との二つの方法がある。

次は『大広益会玉篇』の掲出字の Unicode が「一」である項目を表示する例である。

<https://hdic2.let.hokudai.ac.jp/api/v1/syp/show/unicode/4e00>

表示の結果も JSON のオブジェクトとして回答する。

```
{
  "SYID": "a005b012",
  "SY_vol_radical": "v1#1",
  "SY_radical": "一",
  "Entry": "天",
  "Entry_original": "",
  "unicode": "&#x5929;",
  "SY_def": "他前切。《説文》曰：天顛也，至高無上，从一大。（中略）天坦也，坦然高而遠也。",
  "KSY_diff": ""
}
```

### 6.1.3.3 画像データの Web API

掲出字の画像も API で取得できる。スキーマは「/img/略称/ID」となる。次の URL で『大広益会玉篇』の「恙」字の画像を表示する。

<https://hdic2.let.hokudai.ac.jp/img/syp/a080a083>

### 6.1.4 サーバーへの実装

開発の効率を向上するため、フレームワークを利用して HDIC Viewer のバックエンドを構築する。開発言語は筆者に馴染みのある PHP を用い、フレームワークは近年、ネット上に流行している Laravel にする。Laravel は、RESTful ルーティングのサポート、豊かなコミュニティなどの特徴があり、他の PHP フレームワークと比較すれば、プログラム初心者でもより容易に利用できる。かつ、MIT ライセンスのもとで配布されているので、利用上の制限が少ない。

ほかには、VueJS, Bootstrap, BulmaCSS などのオープンソースライブラリを積極的に利用している。

## 6.2 多漢字文献の検索効率向上とマルチデバイス対応の試み

### 6.2.1 はじめに

近年、漢籍や仏典に関する多漢字文献データベースが数多く公開され、国語学の研究・学習に非常に有益となっている。また、スマートフォンやタブレット端末などの PC 以外の端末が盛んに利用されており、インターネット上の言語資源もこれらのスマートデバイスへの対応が求められている。そこで、PC にも PC 以外の端末にも使える漢字検索と表示に特化した多漢字文献データベースシステムを開発したいと考えるに至った。その課題は、次の 5 点となる。

(1) 利用に制限のない、多漢字文献のデジタル化データ

(2) 多漢字文献のデータの格納・検索

(3) 難字・異体字の入力・表示

(4) PC 以外のマルチデバイス対応

(5) 開発と実装

まず 6.2.2 節では課題 (1) に関する利用資料の詳細を述べる。6.2.3 節では課題 (2)・課題 (3) に関して漢字検索をめぐる問題とその解決方法を述べる。6.2.4 節では課題 (4) のマルチデバイス対応について、PC との区別を中心に述べる。6.2.5 節では、実際に開発した際に生じた問題点と実装したフレームワークについて述べる。

### 6.2.2 利用資料

HDIC は日本の平安時代に編纂された『篆隸万象名義』『新撰字鏡』『類聚名義抄』の総合データベースであり、池田 [15] によると、この三つの辞書を解読するために、中国の『大広益会玉篇』などの辞書もデータベースに含めている。ほかの多漢字文献と比べると、次の二つの公開特徴がある。

1. GitHub で公開しているため、データの点検・更新がわかりやすい
2. ライセンスが明記されており、利用に制限が

少ない

### 6.2.3 漢字検索の効率化

#### 6.2.3.1 多漢字文献における漢字の検索と入力の問題

前述のように、多漢字文献データベースを利用する際に、かな入力で入れにくい難字・異体字が多い。それに対して、通常の文献の内容を検索するにあたっては、次の三つのステップによって行うことが多い。

(1) 「かな入力」などの入力方式によってかなを入力する [予備入力]

(2) (1) の入力内容によって変換した漢字の候補から一つを選定する [本入力]

(3) システムは選定した漢字を検索する [内容検索]

かな入力以外の手段では、次の三つの漢字検索・入力方法がある。

1. 手書き入力：指や特別なペンを用いて入力する方法。漢字知識を必要としていないが、拡張漢字のサポートが不十分である。速度はキーボードほど期待できない。デバイスやソフトウェアによって認識度が大きく異なるが、すべての Unicode 漢字をサポートするシステムは管見の限りまだ見つからない。
2. 字形入力：「蒼頡」や「五筆」など、本来、中国語専用の入力方式を用いて入力する方法。字形をいくつかの小さなパーツに分け、それぞれに対応させたアルファベットを入力する方式である。特に「蒼頡」はほぼすべての Unicode 漢字をサポートするので、難字・異体字の入力には最も頼れる方式である。しかし、これらの入力方式は運用できるまでの長い学習時間を要し、台湾や香港のソフトウェアであるため、フォントが異なることなどで、国語学研究に応用するのはまだ検討が必要である。
3. 部首画数検索：部首画数を用いて漢字検索す

る方法。利用しやすいが、候補が多数を表示されるため、検索効率が低い。

表 6.2 Unicode 漢字入力方法の比較

	入力可能数	入力効率
かな入力	やや少ない	やや低い
手書き入力	少ない	低い, やや低い(機種による)
字形入力	ほぼすべて	高いが, 学習時間が必要
部首画数検索	すべて	低い

上記の表 6.2 に示したように、入力効率から考えると字形入力が最も優れている。一方、かな入力は学習が容易であるが、次の問題がある。

多漢字文献を内容検索する際は、難字・異体字が多いため、探したい漢字の読み方が不明であったり、変換辞書に未収であったりすると、漢字を内容検索システムに入れられない場合もある。つまり、予備入力ができず、本入力もできないこととなってくる。和訓がわからず、漢字音を用いて入力すると候補が大量となってしまう、本入力の候補を探し出すまでに相当な時間や工数がかかる。

### 6.2.3.2 IDS 漢字検索の必要性

部首画数検索は紙の辞書から継承され、利用しやすい漢字検索方法である。検索効率では次の2点が特に問題となる。

(1) 同部首同画数の字数が多い場合、求める漢字を探すのは難しい。

(2) 所属する部首が分からない場合、利用できない。

部首に依存せず、より小さい漢字構造上の要素によって検索するシステムを作ることによってこの二つの問題は解決できる。そのようなシステムを実現するための漢字記述の方法として、「漢字構成記述

文字列 (Ideographic Description Sequence, IDS と略す)」がある。IDS とは、漢字の構成を文字列で記述したものである。IDS は IDC (Ideographic Description Character) と漢字の部品からなる。符号化されていない漢字を表すことのできる漢字記述言語の一種である。IDS をすでに符号化した漢字に用いて、漢字の検索方法とすることもできる。このような漢字検索システムはいくつか開発されており、もっとも代表的なものは CHISE IDS-FIND である。

表 6.3 入力方式の比較

入力方式	予備入力	本入力
かな入力	かな, または「かな」にあたるローマ字	候補から漢字を探す。エンターキーなどを押して検索欄に入れる
字形入力	字形のパーツにあたるアルファベット	候補から漢字を探す。エンターキーなどを押して検索欄に入れる
IDS-FIND	かな入力などを用いて入力する IDS 情報	候補から漢字を探す。目標漢字をコピーして、検索欄にペーストする

IDS-FIND は、難字・異体字の入力問題を解決したように考えられたが、漢字入力専用のツールではないため、上記の表 6.3 を示したように本入力の前の予備入力での利用にとどまるのでさらに工夫の余地がある。

### 6.2.3.3 IDS 画数漢字検索の必要性とその開発

CHISE の IDS-FIND はすぐれたシステムであるが、予備入力における利便性を考慮すると、次の二つの問題がある。

(1) 漢字のパーツ自体が難字・異体字である場合、IDS 漢字検索が利用できない。

(2) 漢字のパーツを間違えると目標漢字が検索できない。つまり、曖昧検索を実現していない。

例えば、IDS 漢字検索を用いて「錫」(U+935A) 字を入力したい場合には、予備入力が「金易」となる。しかし、そもそも「易」が入力しにくい難字である。「金日」で検索すると候補の数が大量となる。選定するには効率が落ちる。

この二つの問題点を解決するため、予備入力の段階で、漢字の総画数情報を併用する。「IDS」+「残り画数」を予備入力の情報として用いることで、漢字検索の効率の向上をはかる。

上記の例の場合では、「金日」と残りの画数「5」を予備入力すれば、候補漢字を 18 字に絞ることができる。次の図 6.2 に示す。

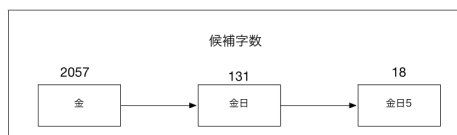


図 6.2 「錫」の候補字数の比較

候補字数の低減によって、目標漢字の選定速度が上がり、漢字検索の効率向上ができるようになる。

また、残り画数の幅をとるために、曖昧検索を実現できるように改善を加える。例えば、上記の残り画数「5」で検索すると、残り画数が 4 と 6 の漢字も候補欄の上下に示される。検索画面は図 6.3 に示している。

利用する漢字総画数データは Unihan<sup>2</sup>の「kTotalStrokes」である。

<sup>2</sup> <http://www.unicode.org/Public/UCD/latest/ucd/UCD.zip>

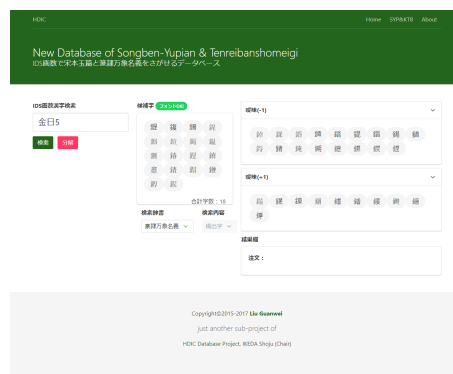


図 6.3 PC 端末の検索画面

## 6.2.4 マルチデバイスへの対応

### 6.2.4.1 マルチデバイスにおける問題点

多漢字文献の利用について、スマートデバイスは PC と大きく異なって、次のような違いがある。

1. 画面のサイズが小さく、同時に表示できる情報が少ない。
2. OS の制限で、一部の漢字が表示できない。
3. アプリケーションとウィンドウの切り替え、テキストのコピー・ペースト操作が難しい。
4. 画面をタッチすることによって操作する。

上記の制限を解消するため、多漢字文献をスマートデバイスで検索する際に、スマートデバイスの特徴に対応するシステムを開発することが必要となる。

しかし、PC と全く異なるシステムとするのは利用者・開発者のいずれにとっても負担となる。

### 6.2.4.2 画面サイズへの対応

レスポンシブデザインを採用して、画面の解像度によって表示の配置が自動的に変わるシステムとする (図 6.4)。

1. ナビメニューを隠す
2. 左右で配置された検索欄と結果欄を上下にする
3. 視認性の高い文字サイズに対応する



CSS フレームワークの Bulma<sup>5</sup>から構成した。全 Unicode 漢字を表示するため、花園フォントの B ファイル「HanaMinB.ttf」をウェブフォントとして利用している。少なくともローディング時間を少しでも軽減するため、TTF<sup>6</sup>ファイルであったフォントファイルを WOFF<sup>7</sup>へ変換した。

### 6.2.5.2 バックエンドについて

このシステムのバックエンドはユーザ管理、HDIC に所属するデータベースの検索・編集などのデータ処理を行う。

PHP フレームワークの Laravel<sup>8</sup>によって構成した。Laravel はドキュメントが詳細であり、非専門家でも比較的の開発しやすいフレームワークである。

4 バイト漢字を MySQL と通信するため、Laravel の /Config/database.php ファイルの MySQL に関する部分にあたって、「charset」と「collation」を修正する必要がある。

フロントエンドと通信のための API は次の表 6.4・表 6.5 のように作成した。検索結果は JSON 形式でレスポンスする。

表 6.4 API 公開アドレス

Scheme	Host	Path
https://	hdic2.let.hokudai.ac.jp	/api/v1/search

次に一例を示す。

リクエスト：

```
/api/v1/search?db=ktb&def=北
```

レスポンス：

```
{
  "TBID": "5_015_A31",
  "TB_vol_radical": "v15#223",
```

表 6.5 API のパラメーター一覧

パラメータ	説明
db	検索するデータベースである。現在では ktb (篆隸万象名義) と syp (大広益会玉篇) のみ利用できる。
entry	検索する親字である。entry か def かのいずれかが必須である。
def	検索する注文である。entry か def かのいずれかが必須である。

```
"TB_radical": "北",
"Entry": "北",
"Entry_type": "Regular",
"Entry_diff": "",
"TB_def": "補墨反。乖也。",
"SYID": "b049b041",
"YYID": "",
"TB_remarks": "\r"
}
```

### 6.2.6 IDS データの利用

Unicode 委員会が公開している Unihan データベースは漢字データベースとして最も広く知られているものであるが、IDS に関する情報は現時点まで公開されていない。

現在公開中の漢字 IDS データの中で、整備状況が一番良好なのは CHISE IDS である。CHISE IDS の README ファイルによると、「<CODEPOINT><CHARACTER><IDS>」三つのフィールドをタブで区切ったデータ構造をとっており、つまり TSV (Tab-separated Values) で示している。CODEPOINT は UCS のコードポイ

<sup>5</sup> <https://bulma.io>

<sup>6</sup> True Type Font

<sup>7</sup> Web Open Font Format, TTF 形式のフォントを圧縮したものである。

<sup>8</sup> <https://laravel.com/>

表 6.6 CHISE IDS データの一例

CODEPOINT	CHARACTER	IDS
U+5B9A	定	𠄎𠄎&CDP-8BCE;
U-0002A76B	儼	𠄎𠄎 𠄎亞取
U-0002B7AA	甚	𠄎甘𠄎&AJ1-04307; ム

表 6.7 CJKVI-IDS データの一例

CODEPOINT	CHARACTER	IDS
U+5B9A	定	𠄎𠄎定
U+2A76B	儼	𠄎𠄎 𠄎一④取
U+2B7AA	甚	𠄎⑤区

表 6.8 簡略 IDS データの一例

CODEPOINT	CHARACTER	IDS
U+5B9A	定	𠄎定
U+2A76B	儼	イ一取
U+2B7AA	甚	区

ントである。拡張漢字 A までは「U+hhhh」のように「U+」と 4 桁の 16 進数で示す。それ以降は「U-hhhhhhhh」のように「U-」と 8 桁の 16 進数で示す。CHARACTER では CODEPOINT が対応する漢字の符号化字形を示す。IDS はその漢字の構成記述情報を示す。次の表 6.6 に例を示す。

また、漢字データベースプロジェクトの「漢字構成データベース」に「字形 IDS データ」があり、現在はオープンソース共有プラットフォームの GitHub<sup>9</sup>を用いて「CJKVI-IDS」という名称で公開している。CJKVI-IDS の構造は主に CHISE IDS と同様で、拡張漢字 B まで CHISE IDS のデータがそのまま利用されているようである。拡張漢字 C・D・E のところでは独自の IDS データを採用している。ただし、CHISE IDS と比べると、CJKVI-IDS は CDP 漢字 など表外漢字をそれらの画数である「①②③…」のような丸数字に変換している。Unicode の表示方法も少し異なる。IDS の問題点は川幡 [33] に詳しい。次の表 6.7 に例を示す。

本節の試作における IDS データは、高速な検索を実現するため、部品から漢字を合成するデータベースと、漢字から部品に分解するデータベース

の二つに分けて作成する。合成データベースは CJKVI-IDS の ids.txt と ids-ext-cde.txt ファイルをベースとして、IDC と符号化されない部品、すなわち入力困難なものや作者の備考などを削除して、さらにシンプルなデータベースを作成する。次の表 6.8 に例を示す。

### 6.2.7 対応するフォント

問題 3 については、ウェブフォント技術などをすれば解決できるが、Unicode 漢字を全すべて含めるとフォントファイルが大きくなる、通信の制限で実現するためにさらに努力が必要と考える。ウェブフォントの圧縮・区分をさらに検討することが必要であるが、これは今後の課題とする。

### 6.2.8 検索速度の向上

検索速度を向上するため、IDS データをローカルに保存する。検索プログラムを JavaScript にして、検索の計算をクライアント側に負担させる。より効率的に IDS データベースを利用するため、オープンソース漢字検索システム刹那字引<sup>10</sup>の部分コードを利用した。「刹那字引」は拡張漢字 E までサポートする漢字検索システムである。特に検索スピードに優れている。ただし、現在では

<sup>9</sup> <https://github.com>

<sup>10</sup> <https://github.com/g0v/z0y>



開発が止まっているようである。「刹那字引」の検索がはやい理由は、検索用 IDS データベースの逆引き索引をクライアント側に再構築されたことである。「刹那字引」では、次に示すデータを逆引きにして利用する。再構築したデータは次のようである。

```
{
  "ㄨ": "...宜弁宓宝室宏宗官宙定宛...",
  (中略)
  "止": "促呢哇噯妮定徒...",
  (中略)
}
```

また, CHISE IDS Find とは異なり, 毎回検索で画面を更新する必要がない。ただし, 画像ファイルの集合のデータ量が大きいので, サーバ側に保存する。アプリの初回利用と画像を請求する時のみサーバと通信する。

作成したウェブアプリは, <http://idstest.toyjack.net/> で公開している。

### 6.2.9 おわりに

本節ではマルチデバイス対応する多漢字文献検索システムの設計と開発に着目した。本来の目的は古辞書を翻字する際の効率的な入力方式の開発であったが, スマートデバイスの普及, オープンデータの活発化によって, さらに利便性が高い多漢字文献の公開手段が求められている。本節はその試みを述べた。

## 6.3 古辞書における横断検索の実現

### 6.3.1 問題点

HDIC Viewer では複数の古辞書データベースを検索できるが, 各辞書の翻刻方針やデータベースの構造が相違するため, 横断検索の実現が難しい。表 6.9 で HDIC で公開している三つのデータベースの構造を示す。

### 6.3.2 古字書の研究が求める横断検索

各古辞書は書写年代, 編纂者によって, 同一漢字を収録しても違う字形で記入することがある。例えば, 『大広益会玉篇』では「齧」字の項目があ

り, 『篆隸万象名義』では「齧」字の項目を収録せず, 異体字である「齧」を項目として立てている(図 6.5)。

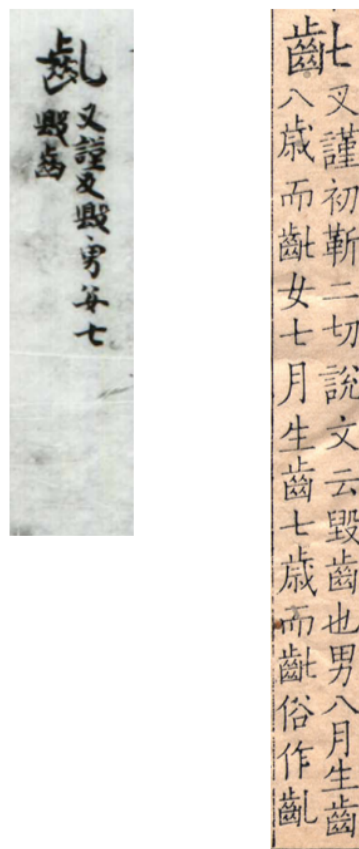


図 6.5 篆隸万象名義の齧(左)と大広益会玉篇の齧(右)

古辞書などの古文書は写本である場合が多く, データベース化する際に, 活字への翻刻は必須となる。その翻刻はデータベース構築の方針によって, 各古辞書に相違がある。従来の国語学研究における翻刻は手書き, または自作フォントで表現できるが, HDIC では Unicode を利用しているため, 利用可能字数や同形異字などの Unicode に既存する問題にも制限されている。

以上の字体における検索問題はおそらくあらゆる多漢字文献データベースで避けられない問題であろう。解決策として, 入力段階で約束したり, 字体の変換表を作ったりすることは今までよく利用されている。

表 6.9 HDIC データベースの構造 (部分)

	篆隸万象名義	大広益会玉篇	新撰字鏡	
ファイル名	KTB.txt	SYP.txt	TSJ_entries.tsv	TSJ_definition.tsv
所在 ID	TBID	SYID	SJID	
			SJ2ID	SJ2ID
巻と部首の通し番号	TB_vol_radical	SY_vol_radical	SJ_vol_radical	
所属部首	TB_radical	SY_radical	SJ_radical	
掲出字	Entry	Entry	Entry	Entry_word
		Entry_original	Entry_original	
注文	TB_def	SY_def		SJ_def
参照 ID	SYID		SYID	
			TBID	
備考	TB_remarks	SY_remarks	SJ_entry_remarks	SJ_def_remarks

HDIC Viewer では、共同での入力、他のデータベースとの連携を考えると、入力段階で字体を約束することは難しい。字体の変換表を用いて、横断検索することは実現性が高い (図 6.6)。

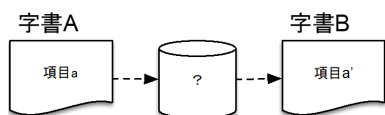


図 6.6 字体変換表による横断検索

HDIC Viewer における横断検索は、次の三つのステップによって横断検索を実現したい。

1. 研究者による判断
2. 構造化された DB の字体注記
3. Unicode 符号位置

その仕組みを次の図 6.7 に示す。

研究者による判断とは、先行研究や索引などを利用して、手動で各辞書の掲出字をグループ化することである。HDIC のテキストデータにある照合 ID はそれらの関係を格納しており、6.3.1 の表 6.9 で示した参照 ID を参照。

照合 ID がない掲出字、また複数の異体字を

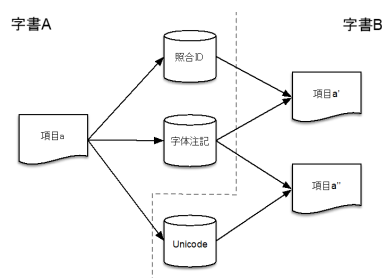


図 6.7 HDIC Viewer 横断検索の仕組み

持っている項目もあるため、これらの項目は注文にある字体を字体変換表として利用すれば、関連する掲出字を検索できる。詳細は 6.3.3 で述べる。

照合 ID と字体注記いずれもない掲出字は Unicode コードが同一である漢字を用いて他字書に検索する。

### 6.3.3 字体注による字体変換表

古辞書の注文には字体注 (字体注記) がある。字体注は、「正」「通」「俗」などの注記によって、漢字と漢字の字体関係を表す。これらの字体注を抽出して、字書ごとの字体変換表を作成できる。作成する手順は、次の三つのステップとした。

1. 古辞書注文テキストデータの構造化
2. 字体注を抽出・加工

## 3. 変換表を作成

古辞書注文テキストデータの構造化とは、古辞書の注文データを一定のタグでマークアップすることである。古辞書のマークアップに関して筆者はすでに専用のツールを開発しており、それを用いて掲出字の字体注を抽出できる（3章を参照）。

抽出した字体注に正規表現を用いて、関連する漢字（字体）のみを残す。それを掲出字とペアにして字体変換表を作成する。図 6.8 は「𪗇」、「𪗈」を例にしたものである。

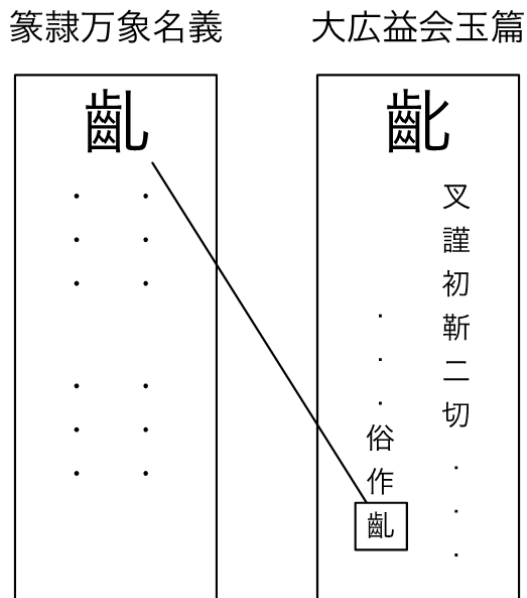


図 6.8 字体注記の利用

## 第7章

# 結論

### 7.1 本研究のまとめ

本研究では、日本漢字辞書の研究における、次の三つの難点に対して解決方法を模索してきた。

1. 注文の構造が複雑で、漢字の形・音・訓・漢文義注と引用出典の分析が十分ではない。
2. 写本資料では、符号化方針が異なり、異体字・形近異字の検索が難しい。
3. 簡単に検索・表示できるインターフェースがなく、応用のシーンが限られている。

これらの問題に対して、第3, 5, 6章では以下の研究について述べた。

1. 注文項目構造の多様性に対応したタグセットとマークアップ・ツールの開発
2. 漢字辞書テキストデータから和訓と異体字関係を抽出して、対応関係を分析
3. ウェブインターフェースを開発して、成果を利用性高い研究資料として公開

また、第2, 4章では、日本漢字辞書のテキストデータを用いて、各辞書の構成を分析してきた。第2章では平安時代から近世時代まで、三つの漢字辞書における漢字と和訓の対応関係、辞書間に共通する漢字和訓を示した。第4章では、上記の新たな方法の検証として、観智院本『類聚名義抄』の項目を分類して、各類の特徴を述べた。

以上のことにより、日本漢字辞書研究における

テキスト資料の利用と分析方法を述べた。

### 7.2 本研究の意義

#### 7.2.1 国語学において

近年、言語研究ではコーパスの利用が多く見られる。国立国語研究所は様々なコーパスを提供しており、日本語の研究でも広く用いられている。特に、国語学においては、同所が開発した日本語歴史コーパス（CHJ）が、日本語史研究の基礎資料として盛んに用いられている。

しかし、これらのコーパスには、国語学における重要な資料である日本漢字辞書が収録されていない。日本漢字辞書がコーパスに収録されていない原因として、新聞記事・文学作品などのテキストに比べ、漢字辞書は、複雑な構造、注文における漢字と仮名の混在、大量の異体字の存在など、正しいコーパス検索に必要なアノテーションの付与が難しいことが考えられる。

本研究は、日本漢字辞書を対象とした高度な検索の実現を目指し、漢字辞書の注文テキストの構造化の方法を述べた。この方法で作成したデータを蓄積することで、コーパスを利用した研究と同じように量的な研究が可能となる。

本研究では、観智院本『類聚名義抄』の注文テキストを構造化し、注文の内容を音注、義注、字体注、和訓の4大要素に分けた。注文の文字種の出現順序と4大要素の有無によって、項目の分類をした。

### 7.2.2 人文情報学において

後藤真は、人文情報学の研究傾向として、発見系・解析系・可視化系の3つの系統があるとした[34]。各系統の特徴は表 7.1 にまとめられる。

表 7.1 人文情報学の研究傾向の系統

発見系	人文系の大規模な情報群・データ群の中から必要なものをいかに効率的に発見することができるか、もしくは研究の目的に沿って見つけることができるかを研究する系統
解析系	テキストや絵画などをデータ化し、それらのデータについて、コンピュータプログラムを用いて分析し、人間の手法と異なるアプローチを行う系統
可視化系	研究成果を中心にいかにわかりやすく研究者以外の人びとに見せるかなどを研究する系統

本研究の研究対象は日本漢字辞書であるが、情報学の技法を多く採用したため、人文情報学の分野に属する。上記の表 7.1 と対照すると、本研究の第 3・5 章は発見系に、本論の第 6 章は可視化系に当てはまる。解析系は従前の研究成果を検証して本研究で遂行している。第 2 章では、三つの辞書の共通する和訓を抽出したことがあげられる。

第 3 章では、日本漢字辞書の多様性に対応したタグセットとマークアップ・ツールの開発した。これを用いて、第 4 章の観智院本『類聚名義抄』の項目分類を完成した。

第 5 章では、漢字字形共有サイト GlyphWiki を利用して、翻刻文字と原本文字の差分を調べた。また、正規表現を利用して、漢字辞書テキストデータから和訓と異体字関係を抽出した。

第 6 章では、PHP と JavaScript を用いて日本漢字辞書検索ウェブインターフェースを開発した。広く漢字の教育と研究において利用できるツールとして公開している。

### 7.2.3 成果の発信において

近年、海外でも日本漢字辞書の研究が盛んに行われており、『篆隸万象名義』の翻刻と校釈は日本国内より先に中国で出版された [35][36]。日本漢字辞書の研究は日本の国内にとどまらず、海外においても格段の進展を見せている。

そこでは、研究成果の国際的な発信が求められる。その発信は精度と速度を両立する必要がある。本研究の第 7 章では、オープンデータ・オープンサイエンスに沿い、研究資源の共有を行っている。その成果の詳細は次の節で述べる。

## 7.3 HDIC Viewer の利用者に関する統計

### 7.3.1 統計方法

HDIC Viewer の開発と運営は 2015 年からであるが、本研究の成果の一環として、筆者が博士課程に入学した 2016 年からの 4 年半の利用者状況を統計していきたい。統計は Google Analytics<sup>1</sup> を利用している。統計期間は 2016 年 4 月 1 日から 2020 年 10 月 31 日までとした。

Google Analytics では、以下のようにユーザーを定義している。

アプリやサイトを操作する人です。そのアクティビティを Google アナリティクスで測定します。<sup>2</sup>

なお、Google Analytics でのユーザーはユニックユーザーを指している。

ユーザーによる操作の時間について、セッションという概念がある。セッションの定義を以下に示す。

サイトやアプリでユーザーが操作を行っている時間のことです。ユーザーが 30 分以上

<sup>1</sup> <https://analytics.google.com/>

<sup>2</sup> <https://support.google.com/analytics/answer/9355670?hl=ja>

操作を行わなかった場合、それ以降の操作は新しいセッションと見なされます。サイトを離れたユーザーが 30 分以内に同じサイトに戻ってきた場合は、同じセッションとして扱われます。<sup>3</sup>

### 7.3.2 統計結果

Google Analytics によって、この 4 年半の期間に、Google Analytics によって合計 1,102 ユーザーが HDIC Viewer を利用したことがわかった。これらのユーザーは 1,923 のセッションを行い、5,327 ページを閲覧した。

地域の統計は表 7.2 に示す。ユーザー数の一位はアメリカである。いわゆる漢字圏（日本、中国、台湾）からのユーザー数は全体の約三分の一を占めている。

表 7.2 ユーザーの地域

国	ユーザー	セッション
United States	381	387
Japan	261	845
United Kingdom	226	226
China	69	109
(not set)	41	44
Austria	34	34
Iraq	17	17
Taiwan	16	30
Brazil	11	11
Italy	8	8
その他	39	212
合計	1103	1923

日本におけるユーザーの地域分布は表 7.3 に示す。ユーザー数の一位は北海道であり、全日本ユーザーの半分に近い。原因としては、筆者が所属する北海道大学では、HDIC Viewer が教育と研究に利用されたためであろう。

ユーザーの利用端末に関する統計は表 7.4 に示

表 7.3 日本におけるユーザー分布

地域	ユーザー	セッション
Hokkaido	109	574
Tokyo	60	117
Kyoto	30	46
Osaka	16	29
Kanagawa	14	32
Hyogo	10	11
(not set)	3	3
Yamagata	3	4
Nara	3	6
Tochigi	2	2
その他	11	21
合計	261	845

す。HDIC Viewer は携帯端末での閲覧や操作を優先的に想定して開発したが、実際にユーザーが利用した端末のほとんどが PC である。

表 7.4 ユーザーが利用した端末

デバイス	ユーザー	セッション
PC	1,010	1,671
携帯端末	81	221
タブレット	12	31
合計	1,103	1,923

研究資源としての学術データベースにおいて、ユーザー利用の実態に関する調査はまだ少ない。モバイル技術の普及によって、PC 以外の携帯端末からの利用が増えると考えられる。タッチパネルと高解像度カメラを生かすことによって、携帯端末における HDIC Viewer の利便性を向上することが可能である。それは今後の課題とする。

<sup>3</sup> <https://support.google.com/analytics/answer/6086069?hl=ja>

## 参考文献

- [1] 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, Vol. 135, pp. 60–41, 2015.
- [2] 李乃琦. 図書寮本『類聚名義抄』における玄心撰『一切経音義』の依拠テキスト：『一切経音義』巻第四を中心に. 訓点語と訓点資料, Vol. 137, pp. 132–115, 2016.
- [3] 申雄哲. 図書寮本『類聚名義抄』の翻字と校注(足部). 訓点語と訓点資料, Vol. 138, pp. 46–71, 2017.
- [4] 中野直樹. 天文本『字鏡鈔』乙部の本文について：合点が付された本文をめぐる. 訓点語と訓点資料, Vol. 140, pp. 22–37, 2018.
- [5] 賈智. 『新撰字鏡』における『干祿字書』の利用について. 訓点語と訓点資料, Vol. 142, pp. 18–40, 2019.
- [6] 武倩. 倭名類聚抄における本草和名の誤引：狩谷椽斎の箋注に注目して. 訓点語と訓点資料, Vol. 142, pp. 83–67, 2019.
- [7] 池田証壽. 日本古辞书研究的现状与课题. 首届跨文化汉字国际研讨会“东亚碑刻汉字及文献研究”论文集, pp. 78–86, 2018.
- [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, Vol. 12, No. 2, pp. 68–75, 2016.
- [9] 高橋大希. 夢梅本『倭玉篇』の研究. Master's thesis, 北海道大学, 2017.
- [10] 高田智和. 『大字典』データベースをつくる. じんもんこん 2001 論文集, Vol. 2001, pp. 221–228, 2001.
- [11] 高田智和. 大字典データベースをつかう. 情報処理学会研究報告. 人文科学とコンピュータ研究会報告, Vol. 62, pp. 45–52, 2004.
- [12] 劉冠偉, 李媛, 池田証壽. 平安時代漢字字書総合データベースの拡張と和訓対応. 情報処理学会研究報告. 人文科学とコンピュータ研究会報告, Vol. 2015, No. 4, pp. 1–8, 2015.
- [13] 劉冠偉. 『大字典』和訓データベース構築の現状と課題. 情報処理学会研究報告. 人文科学とコンピュータ研究会報告, No. 9, pp. 1–4, 2016.
- [14] 池田証壽. 漢字字体研究のための日本古字書データベースの作成：観智院本『類聚名義抄』を例に. 情報処理学会研究報告. 人文科学とコンピュータ研究会報告, Vol. 77, pp. 35–42, 2008.
- [15] 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, No. 142, pp. 79–90, 2014.
- [16] 池田証壽, 劉冠偉, 鄭門鎬, 張馨方. 観智院本『類聚名義抄』全文テキストデータベース構築の方法. 日本語学会 2018 年秋季大会, 2018.
- [17] 池田証壽, 劉冠偉, 鄭門鎬, 張馨方, 李媛. 観智院本『類聚名義抄』全文テキストデータベース—その構築方法と掲出項目数等の計量—(小特集類聚名義抄). 訓点語と訓点資料, Vol. 144, pp. 105–129, 2020.
- [18] 高橋大希, 劉冠偉, 池田証壽. 夢梅本『倭玉篇』全文テキストデータベースの構築. 言語資源活用ワークショップ発表論文集, 第1巻, pp. 49–56. 国立国語研究所, 2017.

- [19] 芮真慧. 漢字とその訓読みとの対応についての研究. PhD thesis, 北海道大学, 2012.
- [20] 池田証寿. 観智院本類聚名義抄の掲出項目数と掲出字数. 北海道大学文学研究科紀要, Vol. 124, pp. 137–151, 2008.
- [21] Charles Muller, Kiyonori Nagasaki, Jean Soulat, Christoph Anderl, Christian Wittern, and Marcus Bingenheimer. The xml-based ddb: The ddb document structure and the p5 dictionary module; new developments of ddb interoperation and access. *Chung-Hwa Buddhist Journal*, Vol. 25, pp. 105–28, 2012.
- [22] 高橋晃一. XML による仏教重要語彙定義集の作成と課題. 情報処理学会研究報告. 人文科学とコンピュータ研究会報告, Vol. 86, pp. 1–5, 2010.
- [23] 青野道彦, 永崎研宣, 下田正弘. 仏教文献における注釈構造の可視化に関する予備的研究—パーリ語仏教文献を事例として—. Technical Report 2, 東京大学文学部, 人文情報学研究所, 東京大学大学院人文社会系研究科, 2017.
- [24] 漢字データベース: 宋本廣韻データ. <http://kanji-database.sourceforge.net/dict/sbgy/index.html>, 2008.
- [25] 申雄哲. 図書寮本類聚名義抄の本文解読とデータベース作成の問題点. 漢デジ 2016. 北海道大学, 2016.
- [26] 酒井憲二. 類聚名義抄の字順と部首排列. 本邦辞書史論叢. 三省堂, 東京, 1967.
- [27] 李媛, 劉冠偉, 池田証壽. 観智院本類聚名義抄に関する作字作業をめぐって. 東洋学へのコンピュータ利用第 32 回研究セミナー, 2020.
- [28] 天理図書館 所蔵資料掲載・転載要項. <https://www.tcl.gr.jp/usage-guidance/copy/copy-essential-point/>, 2020.
- [29] 上地宏一. GlyphWiki: 漢字字形自由共有サイト. 学芸国語国文学, Vol. 48, pp. 181–191, 2018.
- [30] GlyphWiki: グリフを登録する. <https://glyphwiki.org/wiki/GlyphWiki:グリフを登録する>, 2020.
- [31] 上地宏一. 漢字グリフ管理 Wiki システム (GlyphWiki) の構築. じんもんこん 2007 論文集, No. 2017, pp. 237–244, 2007.
- [32] 守岡知彦. CHISE における漢字字体・字形粒度の整理規準について. 東洋学へのコンピュータ利用 第 26 回研究セミナー, 2015.
- [33] 川幡太一. UCS 外漢字の符号化方法の検討. 情報処理学会研究報告. 人文科学とコンピュータ研究会報告, Vol. 81, pp. 41–46, 2009.
- [34] 後藤真, 橋本雄太, 国立歴史民俗博物館. 歴史情報学の教科書: 歴史のデータが世界をひらく. 文学通信, 2019.
- [35] 呂浩. 《篆隸萬象名義》研究. 上海古籍出版社, 2006.
- [36] 呂浩. 篆隸萬象名義校釋. 學林出版社, 2007.



# 謝辞

最初に、研究生として北海道大学に入学してから、8年間にわたってご指導をいただきました池田証壽先生に深く感謝申し上げます。文献の読み方から論文の執筆まで、一からご指導いただきました。また、池田先生から最も影響を受けたのは研究への姿勢です。一日も欠かさずにデータベースの入力・校正をされている池田先生の作業ログを拝見するにつけ、私も、真摯な態度で研究に取り組めるよう、目標としてまいりました。最後までご指導いただき、感謝の念に堪えません。

講座の加藤重広先生、佐藤知己先生、李連珠先生、ご退官された小野芳彦先生に感謝申し上げます。この8年間、講義や研究活動を通じ大変お世話になりました。修士時代から多くのご教示・ご指導を賜りました先生方に厚く御礼を申し上げます。

また、講座の先輩・同輩・後輩にも感謝申し上げます。良い仲間に恵まれ、切磋琢磨しつつ、素晴らしい研究環境で8年間を過ごすことができました。とりわけ、講座の先輩であり酒の友でもある、札幌学院大学の岸本宜久さんに感謝申し上げます。研究内外を含め、いろいろな面で助けていただきました。本論文の日本語のネイティブチェックも引き受けていただきまして、感謝の気持ちでいっぱいです。

妻の姚静君に感謝します。札幌に来て一番幸せなことは、あなたと出会えたことです。

最後に、ここまで育ててくれた両親に感謝いたします。日本での留学生生活を支援していただいたからこそ、本学位論文を提出することができました。

多くのご指導、ご教示、ご支援を賜りました皆様に、あらためまして心から感謝申し上げます。

劉 冠偉

2020年11月

## 付記

本研究は JSPS 科研費 19J11614 の助成を受けたものです。