

## HNG で見る字体数の変遷

岡墻裕剛・石塚晴通・斎木正直

北海道大学大学院文学研究科

「漢字字体規範データベース」(略称 HNG)は、2004 年度よりオンラインで公開中のデータベースである。2008 年 4 月現在、各時代・各地域の標準的な文献 62 点に出現する 4,554 字種 432,596 字について、漢字の字種・字体・字形や字体規範に関する情報が検索可能である。この 62 文献について、各文献における字種数・字体数・総字数・異体字率を集計した。異体字率は文献の性質を示す指針となり、HNG で取り扱っている文献の多くが標準的な文献であることが確認できる。また、HNG では複数の文献間での字体整理は行われていないが、その方法には一定の基準が必要であり、本稿では部首が「心」となる漢字を取り上げ、実例を交え報告を行う。

## A transition of standard of writing in HNG

OKAGAKI Hirotaaka・ISHIZUKA Harumiti・SAIKI Masanao

Hokkaido University, Graduate School of Letters

“Database of the normative glyphs in Hanzi script”(abbreviation HNG) is a big data base, which is opening to the public online from 2005. Now HNG shows information of 4,554 character type 432,596 characters that appear in 62 standard documents. To use HNG, we can get the information of Chinese character, for example Zitai (standard of writing), Zikei (shape itself), and so much. We research the number of characters and character types, Zitai, and the rate of Itaizi (another from). We also report about the way to an adjustment of HNG, by the point of Zitai, giving an example of radical of “心”.

### 1. HNG の概要

漢字字体の各時代・各地域の標準と、その変遷を見る上では、「漢字字体規範データベース」(英語名 Database of the normative glyphs in Hanzi script, 略称 HNG)が有用な資料である。これは発表者の一人である石塚が二十数年来作成してきた「石塚漢字字体資料」を基に、北海道大学言語情報学講座有志が協力して作成し、東京外国語大学アジア・アフリカ言語文化研究所の支援により 2004 年度よりオンラインで公開中のデータベースである(<http://www.joao-roiz.jp/HNG/>)。

「石塚漢字字体資料」は、漢字字体には各時代・各地域によって標準が存在し、その標準が変遷することの証明を目的として作成した漢字資料である。漢字文化圏における標準的な文献に出現する漢字の字種・字体・用例数についての調査結果を紙カードで蓄積したもので、漢籍・仏典・国書等 79 点を対象とし、総用例数は 50 万字に近い。また、漢字字体の標準を調査するために、公的な文献だけでなく私的な文献も取り扱っている。

HNG は、様々な立場の利用者に、漢字の字種・字体や字体規範に関する情報を提供するために、「石塚漢字字体資料」を電子化し公開したデータベースである。2008 年 4 月現在、各時代・各地

域の標準的な文献 62 点に出現する 4,554 字種 432,596 字の情報が検索可能である。文献の字体や用例数だけでなく、当該文献の地域・時代・版本写本の別等の書誌情報を付加し、漢字文化圏全域の幅広い時代の資料を一定の基準で検索できるデータベースは、他に類を見ないであろう。データベース公開にあたっては、日本学術振興会科学研究費補助金の助成と、東京外国語大学アジア・アフリカ言語文化研究所の技術支援を受けている<sup>1</sup>。

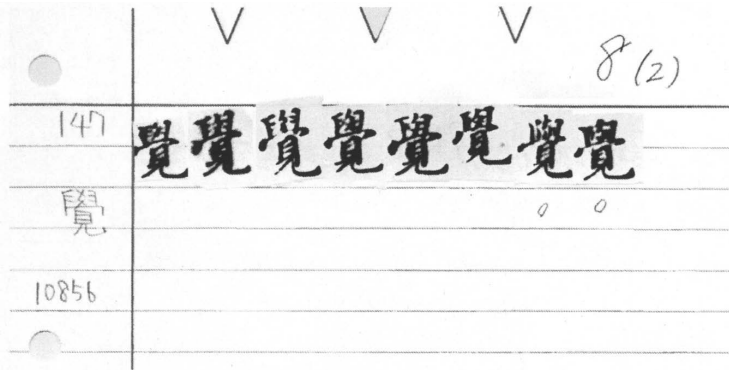


図 1：「石塚漢字字体資料」(S2423 瑜伽法鏡經)

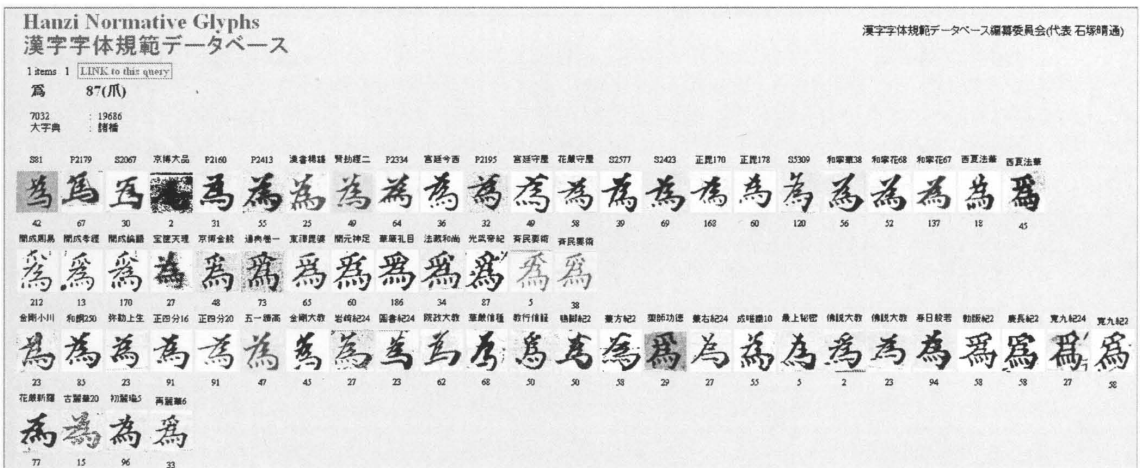


図 2：HNG の検索結果(検索文字「為」、一部画像を加工)

本発表では、この HNG を用いた一研究として、時代・地域・公的私的・版本写本等の性質の異なる複数の文献に出現する漢字字体の整理に関する一方法論を提示し、実際に HNG の「心」部の字体整理を行った結果明らかとなった字体数とその問題点を報告する。

<sup>1</sup> 漢字字体規範データベース編纂委員会の構成委員は、石塚晴通(委員長・北海道大学名誉教授)・豊島正之(東京外国語大学アジア・アフリカ言語文化研究所准教授)・池田証壽(北海道大学大学院文学研究科教授)・白井純(信州大学人文学部講師)・伊藤智ゆき(東京外国語大学アジア・アフリカ言語文化研究所助教)。協力者として、高田智和(国立国語研究所)・山口慶太(北海道大学大学院文学研究科専門研究員)・岡嶋裕剛(同)・高木維(北海道大学大学院文学研究科博士後期課程)・斎木正直(同)となっている。

## 2. 定義

漢字の定義に関する用語は、論者によってさまざまに規定されるが、本発表では、「書体」・「字体」・「字形」を石塚晴通(1984)に従うものとする。

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

楷書・草書等

字体：書体内に於て存在する一々の漢字の社会共通の基準

字形：字体内に於て認識する一々の漢字の書写された形そのもの

また、「書体」・「字体」・「字形」の3体は、「書体」を上位とする階層構造を構成し、それぞれの概念は完全に独立したレベルで認識される。この考えに基づき、次の概念を「字種」と定義する。

字種：社会通念上同一のものとして認識され、一般的に音訓と意味が共通する相互交換可能な漢字字体の総合

「字種」は一般生活で使用され認識・理解される語彙であるが、その定義が明文化されることは少ない。本発表では、「字種」を「字体」レベルの概念として捉え、種々の「字体」が集合化したものを「字種」とする。これらの定義を図示すると図3のようになる。

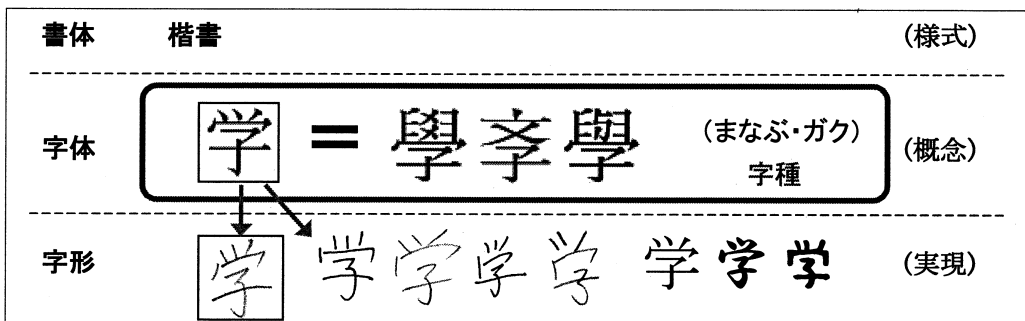


図3：「書体」・「字体」・「字形」・「字種」

HNGは「字体」情報を提供するデータベースではあるが、各漢字は「字種」レベルに基づいて整備しており、文献中に実際に出現する用例は「字形」レベルのものである。また、写本・版本等の文献の性質の差は「書体」の問題として表出する。HNGを適切に使用するためには、これらの概念についての正しい理解が必要である。

## 3. HNGの公開資料

HNGでは、年度によりテーマを設定し資料を公開している。2004年度は、初唐標準字体の存在とその実態、初唐標準字体から開成石経標準字体への移行、開成石経標準字体の南宋版における定着、日本古写本における中国標準字体の導入の典型等を示す資料、2005年度は、初唐標準字体以前の南北朝時代や隋の標準字体の実態、日本資料における字体推移の詳細を示す資料、2006年度は、韓国資料・周辺資料を意図的に含め、中国から漢字文化圏全域への字体の伝播に重点を置いた資料、2007年度は更なる充足を目的とした資料を厳密に選定し公開を行った。

資料の公開に際しては、「異体字率」を資料性質の判断目安としている。異体字率とは、漢字を複数回書く時に字体がゆれる割合を示し、次の式で算定する。

$$\text{異体字率} = \frac{\text{異体字の総字数}}{\text{文献の総字数} - \text{孤例の総字数}} \times 100$$

「孤例」とは、各文献内に1用例のみしか出現しない字種を示し、字体のゆれが把握できないために総字数から除外する必要がある。また、ここでの「異体字」とは、同一文献内に複数の字体が出現する字種のうち、出現数の少ない方の字体を意味することとする。

現在 HNG で公開中の文献について、各文献内の総字種数・総字体数・総字数と異体字数・異体字率を一覧する<sup>2</sup>。開成石経における低い異体字率や、同一巻の日本書紀文献であっても総数や異体字率に相違が見られること等が読み取れる。異体字率は、漢字字体の規範に対する意識を示す指標であり、異体字率の低さは規範に対する強い意識を表し、文献が標準的なものか否かを判断する基準として 1.00% が目安となる。〈開成論語〉等の開成石経資料は異体字率が低く、極めて規範的な文献であることが確認できる。これに対し、字体のゆれは規範のゆるみを意味しており、異体字率の上昇につながる。〈漢書楊雄〉等の私的な資料は、高い異体字率を示すことが分かる。

#### ○2004 年度公開

分類	資料名	略称	全体			異体字(率)
初唐写本	今西本妙法蓮華経巻五	〈宮廷今西〉	633 字種	645 字体	4344 字	28 字 (0.64%)
〃	守屋本妙法蓮華経巻三	〈宮廷守屋〉	585 字種	592 字体	5685 字	46 字 (0.81%)
〃	S2577 妙法蓮華経巻八	〈S2577〉	780 字種	823 字体	5605 字	114 字 (2.03%)
〃	上野本漢書楊雄伝	〈漢書楊雄〉	1573 字種	1701 字体	4510 字	206 字 (4.57%)
開成石経	論語	〈開成論語〉	1322 字種	1328 字体	14325 字	5 字 (0.03%)
〃	周易	〈開成周易〉	1404 字種	1420 字体	23248 字	43 字 (0.18%)
北宋版	東禪寺版阿毘達磨大毘婆沙論巻百七	〈東禪毘婆〉	357 字種	368 字体	6979 字	42 字 (0.60%)
〃	齊民要術巻五	〈齊民要術〉	994 字種	1051 字体	5464 字	97 字 (1.78%)
〃	開元寺版道神足無極變化経巻四	〈開元神足〉	674 字種	692 字体	5528 字	57 字 (1.03%)
南宋版	華嚴経内章門等雜孔目巻一	〈華嚴孔目〉	779 字種	814 字体	16967 字	107 字 (0.63%)
日本書紀(写本)	岩崎本巻二十四	〈岩崎紀 24〉	1099 字種	1173 字体	5401 字	116 字 (2.15%)
〃	兼方本巻二	〈兼方紀 2〉	1143 字種	1166 字体	10006 字	55 字 (0.55%)
日本書紀(版本)	慶長勅版巻二	〈勅版紀 2〉	1141 字種	1163 字体	9920 字	65 字 (0.66%)
日本写本	和銅経大般若経巻二百五十	〈和銅 250〉	161 字種	166 字体	7476 字	10 字 (0.13%)
〃	高山寺本大教王経巻一	〈金剛大教〉	495 字種	508 字体	6645 字	52 字 (0.78%)
〃	東禪寺版写大教王経巻一	〈佛説大教〉	794 字種	845 字体	4291 字	118 字 (2.75%)

#### ○2005 年度公開

敦煌南北朝写本	P2179 誠実論巻八	〈P2179〉	556 字種	565 字体	6138 字	40 字 (0.65%)
〃	S2067 華嚴経巻十六	〈S2067〉	629 字種	643 字体	7528 字	37 字 (0.49%)
〃	S81 大般涅槃経巻十一	〈S81〉	928 字種	959 字体	6661 字	58 字 (0.87%)
〃	P2160 摩訶摩耶経巻上	〈P2160〉	1046 字種	1088 字体	6008 字	54 字 (0.90%)
隋写本	P2413 大樓炭経巻三	〈P2413〉	547 字種	574 字体	4626 字	49 字 (1.06%)
〃	隋経賢劫経巻二	〈賢劫經 2〉	884 字種	927 字体	7762 字	86 字 (1.11%)
〃	P2334 妙法蓮華経巻五	〈P2334〉	632 字種	647 字体	5672 字	23 字 (0.41%)
高昌写本	大品経巻二十八	〈京博大品〉	271 字種	273 字体	1547 字	2 字 (0.13%)
則天写本	守屋本花嚴経巻八	〈花嚴守屋〉	443 字種	467 字体	5166 字	64 字 (1.24%)
盛唐写本	S2423 瑜伽法鏡経	〈S2423〉	939 字種	965 字体	7733 字	69 字 (0.89%)

<sup>2</sup> データの整備と再点検に基づく修正の結果、石塚他(2005)、高田・岡墻(2006)等で発表した数値と一部異同がある。各資料の詳細情報は、岡墻他(2008)やオンラインで公開中の HNG トップページを参照のこと。

日本書紀(写本)	凶書寮本卷二十四	<圖書紀 24>	1079 字種	1147 字体	5260 字	93 字 (1.77%)
"	兼右本卷二十四	<兼右紀 24>	1098 字種	1157 字体	5425 字	102 字 (1.88%)
日本写本	小川本金剛場陀羅尼經	<金剛小川>	501 字種	509 字体	6118 字	18 字 (0.29%)
"	高山寺本弥勒上生經	<弥勒上生>	587 字種	605 字体	3523 字	26 字 (0.74%)
"	守屋本五月一日經統高僧伝	<五一續高>	1400 字種	1463 字体	5928 字	86 字 (1.45%)
日本版本	寛治二年刊本成唯識論卷十	<成唯識 10>	467 字種	490 字体	7290 字	103 字 (1.41%)
○2006 年度公開						
開成石經	孝經	<開成孝經>	478 字種	478 字体	1967 字	0 字 (0.00%)
吐蕃写本	S5309 瑜伽師地論卷三十	<S5309>	709 字種	800 字体	7499 字	223 字 (2.97%)
北宋版	通典卷一	<通典卷一>	1126 字種	1147 字体	6483 字	57 字 (0.88%)
南宋版	法藏和尚伝	<法藏和尚>	1577 字種	1613 字体	6967 字	53 字 (0.76%)
"	後漢書光武帝紀	<光武帝紀>	1192 字種	1225 字体	6622 字	53 字 (0.80%)
韓国写本	新羅本花嚴經卷八	<花嚴新羅>	471 字種	481 字体	6539 字	23 字 (0.35%)
韓国印刻本	晋本華嚴經卷二十	<古麗華 20>	457 字種	476 字体	7682 字	35 字 (0.46%)
"	高麗初彫本瑜伽師地論卷五	<初麗瑜 5>	598 字種	610 字体	6188 字	55 字 (0.89%)
"	高麗再彫本華嚴經卷六	<再麗華 6>	490 字種	494 字体	8063 字	5 字 (0.06%)
大和寧写本	華嚴經卷三十八	<和寧華 38>	590 字種	620 字体	7066 字	91 字 (1.29%)
西夏版	妙法蓮華經卷一	<西夏法華>	834 字種	893 字体	9085 字	141 字 (1.55%)
日本書紀(版本)	寛文九年版卷二十四	<寛九紀 24>	1091 字種	1178 字体	5429 字	149 字 (2.74%)
"	慶長十五年版卷二	<慶長紀 2>	1140 字種	1228 字体	9998 字	282 字 (2.82%)
"	寛文九年版卷二	<寛九紀 2>	1140 字種	1256 字体	10021 字	283 字 (2.82%)
日本写本	明恵自筆華嚴信種義	<華嚴信種>	633 字種	651 字体	6262 字	67 字 (1.07%)
"	親鸞自筆教行信証卷四	<教行信証>	612 字種	633 字体	6149 字	55 字 (0.89%)
○2007 年度公開						
初唐写本	P2195 妙法蓮華經卷六	<P2195>	612 字種	620 字体	4371 字	24 字 (0.58%)
盛唐写本	阿毘達磨大毘婆沙論卷百七十	<正毘 170>	169 字種	196 字体	6366 字	156 字 (2.46%)
"	阿毘達磨大毘婆沙論卷百七十八	<正毘 178>	646 字種	685 字体	6133 字	111 字 (1.88%)
"	唐經四分律卷第二十	<正四分 20>	430 字種	458 字体	9875 字	69 字 (0.71%)
北宋版	宝篋印陀羅尼經	<宝篋天理>	615 字種	679 字体	2621 字	107 字 (4.41%)
"	金剛般若經	<京博金般>	442 字種	449 字体	5414 字	34 字 (0.64%)
大和寧写本	守屋本華嚴經卷六十七	<和寧花 67>	852 字種	899 字体	9975 字	99 字 (1.02%)
"	守屋本華嚴經卷六十八	<和寧花 68>	801 字種	828 字体	7245 字	87 字 (1.25%)
日本書紀(写本)	鴨脚本卷二	<鴨脚紀 2>	1090 字種	1168 字体	8805 字	257 字 (3.04%)
日本写本	五月一日經四分律卷第十六	<正四分 16>	436 字種	469 字体	9824 字	94 字 (0.97%)
"	東禪寺版最上秘密那拏天經	<最上秘密>	435 字種	466 字体	2853 字	64 字 (3.63%)
"	守屋本薬師功德經	<薬師功德>	832 字種	884 字体	4927 字	109 字 (2.37%)
"	金剛大教王經卷第二	<院政大教>	457 字種	493 字体	5711 字	98 字 (1.76%)
日本版本	春日版大般若經卷八十	<春日般若>	374 字種	380 字体	7677 字	34 字 (0.45%)

## 4. 字体の整理

### 4.1. 問題点

HNG を用いた研究のためには、HNG の性質と問題点を理解する必要がある。まず字種について見ると、HNG は主に上田万年編『大字典』に基づき、漢字を区別する。『大字典』で「同字」・「俗字」等とされる字体同士は、同一の字種として同じデータ位置で処理する。この例は、「虫」→「蟲」, 「豎」→「豎」等、数多く存在する。また、多年の経験や蓄積による検討の結果、漢字字体の判定基準からは区別するのが妥当と考える例がある。『大字典』の記述では「咲」は「笑」の「古字」だが、日本では別字として認識されることが多く、HNG 上でも区別している。その他にも、「埵」と





## 5. まとめ

本発表では、HNG で公開されている情報をまとめ、各文献における字種数・字体数・総字数・異体字率を集計した。異体字率からは、HNG で取り扱っている文献の多くが標準的な文献であることが言える。この異体字率を目安として、文献の性質を判断することが可能である。

また、複数の文献間での字体整理について、「心」部を取り上げ、実例を交え報告した。字体の整理を行うためには一定の基準が必要であるが、その一方法論としての提案であると理解して頂きたい。今後は、他の部首でも字体整理を行っていく予定である。

## 参考文献

- 池田証寿・白井純・高田智和(2002)「宋版漢字字体の処理」、『第 69 回研究セミナー報告』京都大学大型計算機センター
- 石塚晴通(1984)『圖書寮本日本書紀 研究篇』汲古書院
- 石塚晴通(1999)「漢字字体の日本的標準」、『国語と国文学』第 76 巻第 5 号
- 石塚晴通編(2002)「日本に於ける漢字字体規範成立の実証的研究」平成 12-13 年科学研究費補助金 基盤研究(B)(2)研究成果報告書 (課題番号: 12410109)
- 石塚晴通・池田証寿・白井純(2003)「漢字字体資料画像データベース」『敦煌写本研究、遺書修復及デジタル国際研究会会議手冊』中国国家図書館 (北京)
- 石塚晴通・豊島正之・池田証寿・白井純・高田智和・山口慶太(2005)「《資料・情報》漢字字体規範データベース」、『日本語の研究』第 1 巻 4 号, 日本語学会
- 石塚晴通・池田証寿・岡崎裕剛(2006)「漢字字体規範データベースとその応用」、『東洋学へのコンピュータ利用 第 17 回研究セミナー』京都大学人文科学研究所附属漢字情報研究センター
- 石塚晴通・池田証寿・岡崎裕剛(2006)「漢字字体規範データベースの構築と公開—HNG プロジェクト—」, 国際学術シンポジウム「ブックロードと文化交流」
- 岡崎裕剛・石塚晴通・池田証寿・高木維・斎木正直(2008)「HNG-DB (データベース) の意義と今後の展望」、『アジア・アフリカ言語文化研究所通信』122 号, 東京外国語大学
- 高田智和(2005)「文字番号および部首番号の起源と応用—『大字典』と華英辞典と Rose-Innes—」『日本学・敦煌学・漢文訓読の新展開』汲古書院
- 高田智和・岡崎裕剛(2006)「漢字字体規範データベースの現状」,(国際ワークショップ「典籍交流(訓読)と漢字情報」)
- ISHIZUKA Harumichi(2004) “Japanese standards of writing Chinese characters—from ancient to modern printings”, 37th International Congress of Asian and North African Studies, “JAPANIZATION—from ancient to modern times—(3)”
- ISHIZUKA Harumichi・IKEDA Shoju・SHIRAI Jun・TAKADA Tomokazu(2003) “The data-base focusing on the standard of writing Chinese characters in Dunhuang manuscripts”, Proceedings of the Nara Symposium for Digital Silk Roads, National Institute of Informatics.