

# Current status and future prospects of the Hanzi Normative Glyphs (HNG) Database

ISHIZUKA Harumichi

(Summary)

Following a presentation at the 2004 Autumn meeting of the Society for Japanese Linguistics, the Internet version of the Hanzi Normative Glyphs (HNG) database (headed by ISHIZUKA Harumichi) was launched in March 2005 (<http://www.joao-roiz.jp/HNG/>). Since then, every year new texts and relevant data have been added to the database. The objectives and methodology of this work was first published, with Ishizuka as the first author, in *Nihongo no kenkyū* 日本語の研究 (2005, vol. 1, no. 4), the official journal of the Society for Japanese Linguistics. Following the increasing amount of texts and data (62 texts, 4,554 unique characters, 432,596 character forms), this paper is an introduction to the current status of the project, its findings and future prospects.

In ISHIZUKA (2005), we documented the existence of an early Tang standard of writing and described how this standard shows significant changes in the Kaicheng Stone Classics of the middle Tang, which eventually served as the basis for Song printed works; in Japan the early Tang standard was established as the Japanese standard of writing and, despite the changes that occurred in China, was preserved until the advent of modern printing culture. With the addition of new texts and data to the database, this view basically remains the same. In addition to this, we now have shown the standard of writing in texts of the Nanbeichao, the Sui and early Tang dynasties; the standard in Chinese language material written by the non-Chinese peoples around China; as well as the standard in Chinese language manuscripts written in Japan, starting from earliest times to the beginning of the modern era. For the sake of comparison, the author also includes informal documents in the database and, using the rate of character variants, attempts to show that while a standard is clearly detectable in standard texts, it is much weaker in writings of informal nature. This will also serve to demonstrate that the standard of writing is applicable for ascertaining the formal vs. informal nature of a text.

Furthermore, since the HNG database in itself does not offer any particular conclusions, the author suggests possible research topics, such as the comparison of Dunhuang and Shōsōin manuscripts, as examples for a meaningful application of data.

## 1. Definitions

Since the terminology related to Chinese characters is defined by researchers in a variety of ways, we are using the terms "form" 書体, "glyph" 字体, and "shape" 字形, as set forth in ISHIZUKA (1984):

Form 書體 (*shuti*): The conventional style of the shape of characters. It is usually defined in reference to a corpus. (E.g. *kaishu* 楷書, *caoshu* 草書)

Glyph 字體 (*ziti*): Within the scope of one form, the conventional norm of writing each character.

Shape 字形 (*zixing*): Within the scope one glyph, the physical appearance of how a particular character was written (or printed).

The terms “form,” “glyph,” and “shape” represent three different levels, each one of which is completely independent of the other. On the basis of this understanding, we propose the definition of character “type” 字種 as follows:

Type 字種 (*zizhong*): The sum total of glyphs recognized by society as one character, which are interchangeable and usually have the same pronunciation and meaning.

Character type is what in everyday life people recognize and understand as an individual character, even though it is rarely defined in specific terms. In this paper, “type” refers to the variety of glyphs collectively.

HNG is a database that provides information on character glyphs. The characters are grouped according to their type (*zizhong*), with their shapes (*zixing*) representing the way they actually occur in manuscripts.

## 2. Overview of HNG

The “Database of the Normative Glyphs in Hanzi Script” (abbreviated as HNG) is a useful tool for observing the standard of writing in each time period and geographical regions. It was initially based on the “Ishizuka Register of Chinese Character Glyphs” 石塚漢字字体資料, the result of the author’s work of twenty some years, which was built by volunteers from the Department of Computational Linguistics at the University of Hokkaido. The Institute for Asian and African Linguistics at the Tokyo University for Foreign Studies provided support for developing the online version of the database (<http://www.joao-roiz.jp/HNG/>) that was opened to the public in 2004.

The “Ishizuka Register of Chinese Character Glyphs” (hereafter abbreviated as “Ishizuka Register”) was developed with the aim of documenting the standard glyphs in different time periods and geographical regions, including their changes in different times and regions. In the course of analyzing standard texts belonging to the cultural sphere of Chinese characters, character types (*zizhong*), glyphs (*ziti*) and their number of occurrences were accumulated on paper cards, amounting to a total number of 500 thousand examples derived from 79 texts of Chinese classical writings, Buddhist scriptures, Japanese manuscripts, etc. In addition, in order to study the standards of writing, not only official but a number of informal documents were included as well.

In order to provide information on character types, glyphs and their standardization for users with different backgrounds, the “Ishizuka Register” was digitized, thus creating HNG as an open

access database. Today, in April 2008, searchable data comprises information on 4,554 unique characters (i.e. types) with 432,596 occurrences from a total of 62 texts from different time periods and geographical regions. The database provides access not only to character types and their number of occurrences but also to other bibliographical information, such as geographical region, time period, edition or manuscript version. This is a unique tool with which users can specify their criteria to query data from all cultures and regions where Chinese characters had been used. The Japan Society for the Promotion of Science provided a research grant for opening the database to the public, and the Institute for Asian and African Linguistics at the Tokyo University for Foreign Studies supplied technical support for this task.<sup>1</sup>

### 3. Opening the HNG to the public

HNG annually publishes new data in batches, according to specific topics. In 2004, these consisted of demonstrating the existence of an early Tang standard of writing; the progression from the early Tang standard to that of the Kaicheng stone classics; the adoption of the Kaicheng standard as the Southern Song printed standard; the patterns of implementing the Chinese standard in old Japanese manuscripts. In 2005, the database documented the Nanbeichao standard of writing prior to that of early Tang; and recorded the specific changes in the standard in Japanese manuscripts. In 2006, the main emphasis was laid on documenting the spread of standard glyphs from China to its neighbours, and for this purpose Korean and other non-Chinese material were also included. In 2007, our objective was to grant open access to an even larger body of material.

In the course of making data to available to the public, we are using the “rate of character variants” as a criterion for determining the nature of the texts. The rate of character variants is the rate at which multiple glyphs appear within the same text. It is calculated according to the following formula:

$$\text{Rate of character variants} = \frac{\text{Total No. of variants}}{\text{Total No. in text} - \text{Total No. of sole occurrences}} \times 100$$

“Sole occurrences” are character types (*zizhong*) that appear in a text only once and since this way no information is available regarding their variation, they are excluded from the calculation. “Character variants,” on the other hand, refer to characters that appear in the same text as different glyphs. A low number of variants in a text is considered significant.

For texts open to the public, HNG currently displays the number of character types, the number of glyphs, the total number of occurrences, as well as the number and rate of variants.<sup>2</sup> The data

---

<sup>1</sup> The HNG Editing Committee consists of ISHIZUKA Harumichi (Chairman, Professor Emeritus at Hokkaido University), TOYOSHIMA Masayuki (Associate Professor, Institute for Asian and African Linguistics, Tokyo University for Foreign Studies), IKEDA Shoju (Professor, Graduate School of Letters and Faculty of Letters, Hokkaido University), SHIRAI Jun (Lecturer, Faculty of Arts, Shinshu University), ITO Chiyuki (Assistant Professor, Institute for Asian and African Linguistics, Tokyo University for Foreign Studies). Cooperators include TAKADA Tomokazu (National Institute of Japanese Language), YAMAGUCHI Keita (Researcher, Graduate School of Letters and Faculty of Letters, Hokkaido University), OKAGAKI Hirotaka (same as above), TAKAGI Yui (Postdoctoral Program, Graduate School of Letters and Faculty of Letters, Hokkaido University), SAIKI Masanao (same as above).

<sup>2</sup> Because of the ongoing maintenance and corrections, the current figures may be different from those in ISHIZUKA et al.(2005), TAKATA and OKAGAKI (2006), etc. For detailed information on each text, see OKAGAKI (2008) and

reveals the low rate of character variants in the Kaicheng Stone Classics, as well as the differences in number and rate of variants within two versions of the same volume (*juan*) of the *Nihonshoki*. Since the rate of character variants is an indicator of the degree of attention paid to the standardization of glyphs, and the low rate of variants demonstrates a strong awareness, we set the criterion for determining standard (i.e. formal) vs. non-standard (i.e. informal) texts at 1.00%. The Kaicheng Stone Classics exhibit a low rate of character variants, confirming that these were highly standardized texts. In contrast with this, the inconsistency of glyphs would be the indicator of the inconsistency of the standard, leading to a higher rate of variants. We can see such high rate of variant characters in informal documents such as *Hanshu Yang Xiong* 漢書楊雄.

Moreover, the 432,596 total occurrences of the 4,554 unique characters in 62 texts, once again, not only confirm the existence of a standard in different time periods and geographical regions, but also show that the standard changed according to different times and regions.

Data opened to public in 2004

Category	Name of text (date)	Abbreviation	Number (type, glyph, total number)	Variants (%)
初唐写本	今西本妙法蓮華經卷五(671)	<宮廷今西>	633字種 645字体 4344字	28字 (0.64%)
"	守屋本妙法蓮華經卷三(675)	<宮廷守屋>	585字種 592字体 5685字	46字 (0.81%)
"	S2577妙法蓮華經卷八(7C末)	<S2577>	780字種 823字体 5605字	114字 (2.03%)
"	上野本漢書楊雄伝(初唐)	<漢書楊雄>	1573字種 1701字体 4510字	206字 (4.57%)
開成石經	論語(837)	<開成論語>	1322字種 1328字体 14325字	5字 (0.03%)
"	周易(837)	<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卷百七(1100)	<東禪毘婆>	357字種 368字体 6979字	42字 (0.60%)
"	齊民要術卷五(12C初)	<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經卷四(1126)	<開元神足>	674字種 692字体 5528字	57字 (1.03%)
南宋版	華嚴經內章門等雜孔目卷一(1146)	<華嚴孔目>	779字種 814字体 16967字	107字 (0.63%)
日本書紀(写本)	岩崎本卷二十四(10C)	<岩崎紀24>	1099字種 1173字体 5401字	116字 (2.15%)
"	兼方本卷二(1286)	<兼方紀2>	1143字種 1166字体 10006字	55字 (0.55%)
日本書紀(版本)	慶長勅版卷二(1599)	<勅版紀2>	1141字種 1163字体 9920字	65字 (0.66%)
日本写本	和銅經大般若經卷二百五十(712)	<和銅250>	161字種 166字体 7476字	10字 (0.13%)
"	高山寺本大教王經卷一(815)	<金剛大教>	495字種 508字体 6645字	52字 (0.78%)
"	東禪寺版写大教王經卷一(12C)	<佛說大教>	794字種 845字体 4291字	118字 (2.75%)

Data opened to public in 2005

敦煌南北朝写本	P2179誠實論卷八(514)	<P2179>	556字種 565字体 6138字	40字 (0.65%)
"	S2067華嚴經卷十六(514)	<S2067>	629字種 643字体 7528字	37字 (0.49%)
"	S81大般涅槃經卷十一(506)	<S81>	928字種 959字体 6661字	58字 (0.87%)
"	P2160摩訶摩耶經卷上(586)	<P2160>	1046字種 1088字体 6008字	54字 (0.90%)
隋写本	P2413大樓炭經卷三(589)	<P2413>	547字種 574字体 4626字	49字 (1.06%)
"	隋經賢劫經卷二(610)	<賢劫經二>	884字種 927字体 7762字	86字 (1.11%)
"	P2334妙法蓮華經卷五(617)	<P2334>	632字種 647字体 5672字	23字 (0.41%)
高昌写本	大品經卷二十八(高昌期)	<京博大品>	271字種 273字体 1547字	2字 (0.13%)

則天写本	守屋本花嚴經卷八(則天期)	<花嚴守屋>	443字種	467字体	5166字	64字 (1.24%)
盛唐写本	S2423 瑜伽法鏡經(712)	<S2423>	939字種	965字体	7733字	69字 (0.89%)
日本書紀(写本)	凶書寮本卷二十四(1142頃)	<圖書紀24>	1079字種	1147字体	5260字	93字 (1.77%)
〃	兼右本卷二十四(1540)	<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(686)	<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(738)	<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(740)	<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(1088)	<成唯識10>	467字種	490字体	7290字	103字 (1.41%)

#### Data opened to public in 2006

開成石經	孝經(837)	<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(857)	<S5309>	709字種	800字体	7499字	223字 (2.97%)
北宋版	通典卷一(11C)	<通典卷一>	1126字種	1147字体	6483字	57字 (0.88%)
南宋版	法藏和尚伝(1149)	<法藏和尚>	1577字種	1613字体	6967字	53字 (0.76%)
〃	後漢書光武帝紀(1198)	<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(754-755)	<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(10C)	<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(11C)	<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(13C)	<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(9-10C)	<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(1149)	<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(1669)	<寛九紀24>	1091字種	1178字体	5429字	149字 (2.74%)
〃	慶長十五年版卷二(1610)	<慶長紀2>	1140字種	1228字体	9998字	282字 (2.82%)
〃	寛文九年版卷二(1669)	<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(1221)	<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(1224)	<教行信証>	612字種	633字体	6149字	55字 (0.89%)

#### Data opened to public in 2007

初唐写本	P2195 妙法蓮華經卷六(675)	<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(8C初)	<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(8C初)	<正毘178>	646字種	685字体	6133字	111字 (1.88%)
〃	唐經四分律卷第二十(740頃)	<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(970年代)	<宝篋天理>	615字種	679字体	2621字	107字 (4.41%)
〃	金剛般若經(北宋期?)	<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(9-10C)	<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(9-10C)	<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(1236)	<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(740頃)	<正四分16>	436字種	469字体	9824字	94字 (0.97%)
〃	東禪寺版写最上秘密那拏天經(12C)	<最上秘密>	435字種	466字体	2853字	64字 (3.63%)
〃	守屋本藥師功德經(1412)	<藥師功德>	832字種	884字体	4927字	109字 (2.37%)
〃	金剛大教王經卷第二(12C初)	<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(13C)	<春日般若>	374字種	380字体	7677字	34字 (0.45%)

In order to be able to use HNG for research, it is important to understand the nature and objective of the database. Firstly, in terms of its basic character set and arrangement, HNG is primarily

based on UEDA Kazutoshi's 上田万年 *Daijiten* 大字典. Characters listed in the *Daijiten* as identical characters 同字, popular characters 俗字, etc. are treated in the database under the same character type (*zizhong*) and are allocated the same data position. There are many such cases; the characters 虫→蟲 and 豎→豎 are typical examples. At the same time, there are cases when experience tells us that it is appropriate to deviate from our initial criteria. For example, although the *Daijiten* lists the character 咲 as the ancient form of 笑, in Japan they are often considered different characters, and thus they are also differentiated in HNG. Similarly, the characters 埤 and 垂, 州 and 洲 are also treated as separate character types (*zizhong*). At the same time, in order to examine glyph standardization more efficiently, the characters 无 and 無 are merged into 無, the characters 修 and 脩 into 脩. (This rule is observed with only a few exception in specific texts.) In these points HNG differs from the *Daijiten* and applies its own criteria, and thus it is worth keeping in mind that in certain cases character variants may be interpreted in a different way from how they are ordinarily understood.

Next, let us look at the definition of variation at the level of glyphs. HNG encompasses material from the entire cultural sphere of Chinese characters, with a spread of over a thousand years, beginning with 6<sup>th</sup> century Nanbeichao manuscripts found in Dunhuang. This material is arranged and presented in a variety of different ways. Distinguishing variants is vital for showing the standardization of glyphs, but there are also exceptional cases when concrete occurrences of characters differ from each other visually and yet they are not accepted as variants.

In contrast with printed editions, in manuscripts the concrete character shapes (*zixing*) of the same glyph may show a considerable degree of discrepancy according to different handwritings or calligraphic forms (*shuti*). To illustrate this point, consider the case of 為 and 為 which are two concrete examples and which differ from each other in that their last portion is written as 𠂇 or 一, respectively. However, since 一 is the abbreviation of 𠂇, we can ascertain that this is a difference of calligraphic forms (*shuti*), and thus consider the two as a single glyph. Beside this, differences in length and intersection of strokes, location of components, and other cases where objective distinction is problematic are not considered as multiple glyphs.

In addition, as shown on Figure 1, when characters can be confirmed as mistakes by comparing the text with parallel sections on other manuscripts, or on the basis of context, they are treated as “erroneous characters” 誤字 without counting them as variants.

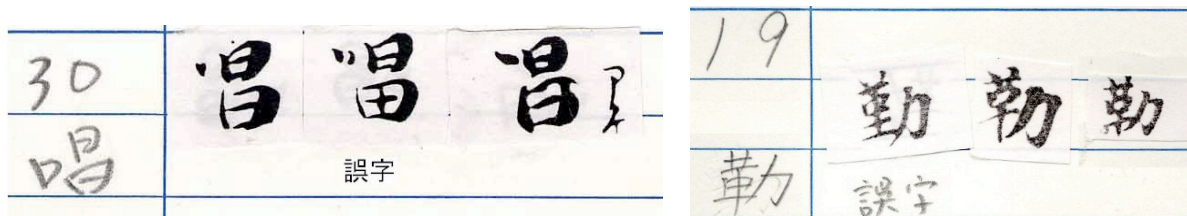


Figure 1: Erroneous characters (left: non-existing glyph in the Zushoryō version of the *Nihonshoki* 圖書紀24; right: typo in the Japanese manuscript of the *Gaosengzhuan* from the Moriya Collection 五一續高)

The above sums up the way information is processed in HNG. Within the search results, each glyph is then displayed using a single representative example -- the rest of the examples recorded



on the original paper cards at this point are not included in the database.

#### 4. Possible research topics

Although HNG in itself does not offer any particular conclusions, it provides the possibility for exploring a variety of research topics. Below are a couple of examples.

##### 4.1 HNG and traditional character dictionaries

In the *Kangxi zidian* and other character dictionaries, the glyph 高 is listed as the standard character (*zhengzi*) and 𠄎 as the popular (*suzi*). At the time HNG was first opened to the public, we have already pointed out (Ishizuka 2005, etc) that this distinction could not be applied uniformly to all time periods and geographical regions. Now, looking over the 62 texts included in HNG, this observation appears just as valid as before (Figure 2).

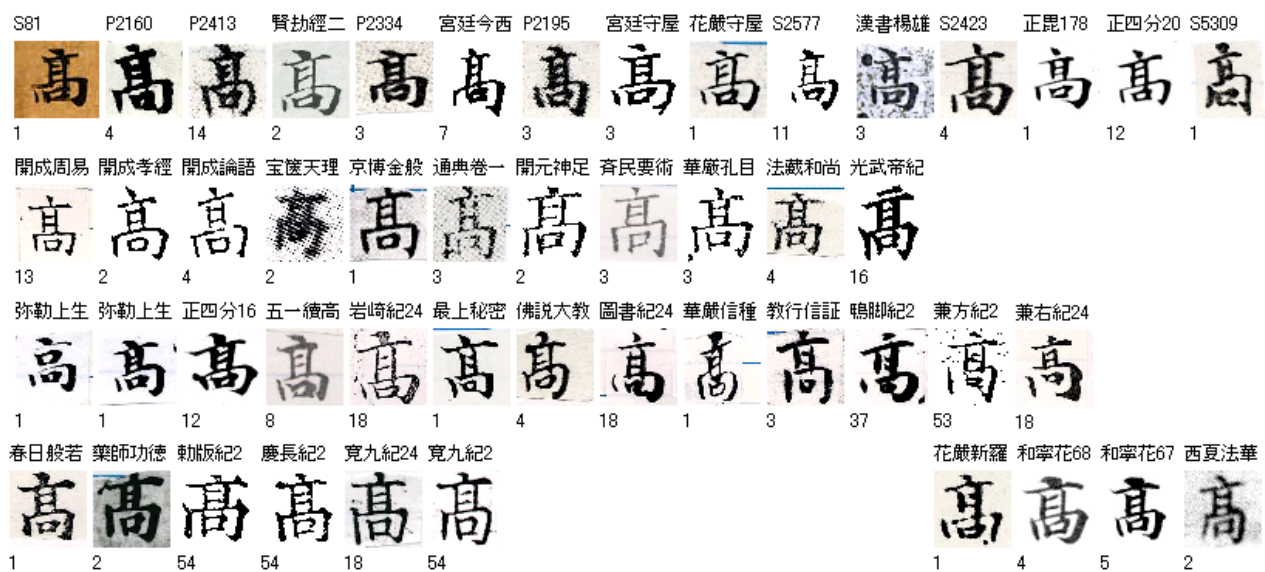


Figure 2: The character 高 in HNG

The dictionary descriptions are generally based on the *Shuowen jiezi* □□□□. For example, the *Xin jia jiu jing ziyang* 新加九經字樣 says:

高高 崇也象臺觀之形上說文／下隸省亭毫等字並從高省  
*Gao* means to esteem highly; the character visually resembles the shape of a raised platform used for observation; the second one is an abbreviated glyph from the clerical script; the characters 亭, 毫, etc all derive from 高 by means of abbreviation.

Accepting the same point of view, the Japanese manuscript of the *Ruiju myōgishō* 類聚名義抄 from the Kanchiin temple's collection sees 高 as the standard character. HNG, on the other hand, lists 高 as the standard in the Kaicheng Stone Classics, the Song prints, and the *Nihonshoki* printed on imperial commission during the Keichō reign. However, the database also shows that in some cases 𠄎 was the standard glyph, as in the Japanese manuscript of the *Mile shangsheng jing* 弥勒上生經 (738). In the course of the transition from the early Tang standard to that of the Kaicheng Stone Classics, there were also smaller divergences that disappeared in later periods, as

it is demonstrated by the Japanese manuscript from the Tempyō era.

#### 4.2 Comparison of the Dunhuang and Shōsōin manuscripts

Before the standard of the early Tang, a series of standards existed during the Nanbeichao period, as it can be learned from HNG by comparing manuscripts S81 and P2160, S2067 and P2179. At the same time, one cannot document this phenomenon in full confidence on the basis of Dunhuang manuscripts alone. Since manuscripts from the Shōsōin Shōgozō collection have been recently published on CD and DVD, this is a newly opened direction for research. Figure 4 shows the glyphs of character 最 in manuscripts S81 (Southern dynasty), S2067 (Northern dynasty), S2423 (Dunhuang), *Mile shangsheng* (Tempyō), *Zhengzifen* 20 (Tang), *Zhengzifen* 16 (Tempyō), and *Huayan Xinluo* (Simla). Manuscript *Zhengzifen* 20, brought back by the monk Ganjin, not only reveals the transitional changes of the high Tang period, which are likewise seen in the *Mile shangsheng* and *Huayan Xinluo* manuscripts, but also raises the possibility that its text originally derived from a manuscript which had been -- similar to S81 but unlike *Zhengzifen* 16 that was based on Genbō's manuscript (i.e. *Chang'an jing* 長安經) -- written in the Nanbeichao standard. (Ganjin was a native of Jiangyang county in Yangzhou and stayed at the Dayun monastery in Yangzhou.) The character 突 illustrates this point. On the other hand, the character 惡 is not a clear-cut case and shows that a number of factors are at play here. The characters 因, 耶, etc show that the glyphs of *Zhengzifen* 20 possess new elements, yet the character 正 is an indication that this is not true for all characters.

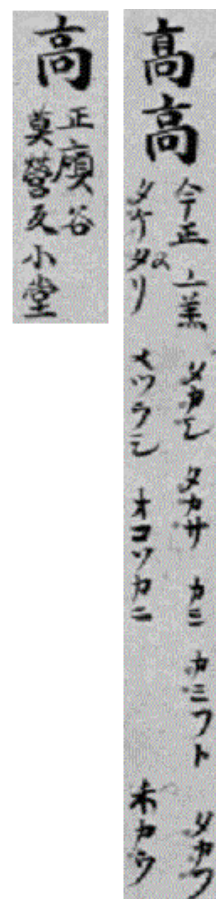









Figure 3: The character 高 in the Kanchiin version of the *Ruiju myōgishō* 類聚名義抄

This kind of comparison once again confirms the uniqueness of the *Mile shangsheng* manuscript. This manuscript is representative of the transition from the early Tang standard to that of the Kaicheng Stone Classics, whereas such small distinctions cannot be detected in the *Huayan Xinluo*. The Dunhuang manuscript S2423 is also very interesting. Generally speaking, the Dunhuang manuscripts are most representative of the culture of Central China for the 70 year period between late 7<sup>th</sup> and early-mid 8<sup>th</sup> centuries. Manuscripts dating earlier or later this period do not always reflect the culture of China proper. But even during this period, when compared with the numerous copies of Nara-period *Prajnaparamita sutras* that survived in Japan, the Dunhuang copies of the same sutra exhibit a certain degree of regional peculiarities.

In either case, looking at the Dunhuang and Shōsōin manuscripts together is also meaningful from the point of textual theory.

Char.	S81 (total number)	S2067	S2423	弥勒上生	正四分20	正四分16	花嚴新羅
因 01611	 (15)	 (9)	 (2)	 (1)	 (47)	 (44)	 (1)



						目 <sub>(3)</sub>	目 <sub>(1)</sub>
惡 03367	惡 <sub>(11)</sub>	惡 <sub>(4)</sub>	惡 <sub>(48)</sub>	惡 <sub>(3)</sub>	惡 <sub>(2)</sub> 惡 <sub>(2)</sub>	惡 <sub>(4)</sub>	
最 04664	最 <sub>(2)</sub>	最 <sub>(9)</sub>	最 <sub>(9)</sub>	最 <sub>(2)</sub>	最 <sub>(31)</sub> 最 <sub>(2)</sub>	最 <sub>(31)</sub>	最 <sub>(5)</sub> 最 <sub>(3)</sub>
正 05643	正 <sub>(14)</sub>	正 <sub>(12)</sub>	正 <sub>(12)</sub>	正 <sub>(6)</sub>	正 <sub>(60)</sub>	正 <sub>(59)</sub> 正 <sub>(1)</sub>	正 <sub>(1)</sub>
突 08303	突 <sub>(3)</sub>				突 <sub>(112)</sub> 突 <sub>(1)</sub>	突 <sub>(113)</sub>	
耶 09292	耶 <sub>(5)</sub>		耶 <sub>(11)</sub>	耶 <sub>(3)</sub>	耶 <sub>(12)</sub> 耶 <sub>(1)</sub>	耶 <sub>(7)</sub> 耶 <sub>(2)</sub>	

Figure 4: Comparison of manuscripts S81, S2067, S2423, *Mile shangsheng*, *Zhengshifen 20*, *Zhengshifen 16*, and *Huayan Xinluo*

## 5. Summary

Above the author described the current status (as of April 2008) of the HNG data that has been opened to the public, and demonstrated the meaning of the rate of character variants within this material. Using this rate as a criterion, it is possible to determine the nature of a text. The standard texts of different time periods and geographical regions reveal the presence of a standard of writing, and the database can demonstrate the changes of this standard in time and space.

Finally, although HNG in itself does not offer any particular conclusions, it provides the possibility for exploring a variety of research topics, such as the relationship of the data with traditional character dictionaries, or the implications of the comparison of Dunhuang and Shōsōin manuscripts for textual theory.

## Bibliography

ISHIZUKA Harumichi 石塚晴通 (1984). *Zushoryōbon Nihonshoki: Kenkyūhen* 圖書寮本日本

- 書紀, 研究篇. Sapporo: Kyūko Shoin.
- ISHIZUKA Harumichi 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
- ISHIZUKA Harumichi 石塚晴通 (2002), ed. “Nihon ni okeru kanji jitai kihan seiritsu no jissōteki kenkyū” 日本に於ける漢字字体規範成立の実証的研究. *Heisei 12-13 nen kagaku kenkyūhi hojokin kiban kenkyū (B)(2) Kenkyū seika hōkokusho* 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. (Project No: 12410109).
- IKEDA Shoju 池田証寿, SHIRAI Jun 白井純, TAKADA Tomokazu 高田智和 (2002). “Sōhan kanji jitai no shori 宋版漢字字体の処理. Dai 69 kai kenkyū seminā hōkoku 第69回研究セミナー報告. Kyōto Daigaku Daikei Kisenki Sentā 京都大学大型計算機センター.
- ISHIZUKA Harumichi 石塚晴通, IKEDA Shoju 池田証寿, SHIRAI Jun 白井純 (2003). “Hanzi ziti ziliao huaxiang shujuku” 漢字字體資料畫像數據庫. *Dunhuang xieben yanjiu, yishu xiufu ji shuzihua guoji yantaohui huiyi shouce* 敦煌写本研究, 遺書修復及数字化国際研討会會議手冊. Beijing: National Library of China.
- ISHIZUKA Harumichi 石塚晴通, TOYOSHIMA Masayuki 豊島正之, IKEDA Shoju 池田証寿, SHIRAI Jun 白井純, TAKADA Tomokazu 高田智和, YAMAGUCHI Keita 山口慶太 (2005). “Shiryō jōhō kanji jitai kihan deitabeisu” 《資料・情報》漢字字体規範データベース. *Hihongo no kenkyu* 日本語の研究. Vol. 1, No. 4.
- TAKADA Tomokazu 高田智和 (2005). “Mojo bangō oyobi bushu bangō no kigen to ōyō: Daijiten to Kaei jiten to Rose-Innes” 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. *Nihongaku, Tonkōgaku, Kanbun kundoku no shin tenkai* 日本学・敦煌学・漢文訓読の新展開. Tōkyō: Kyūko shōin.
- ISHIZUKA Harumichi 石塚晴通, IKEDA Shoju 池田証寿, OKAGAKI Hirotaka 岡墻裕剛 (2006). “Kanji jitai kihan deitabeisu to sono ōyō” 漢字字体規範データベースとその応用. *Tōyōgaku he no konpyūta riyō, Dai 17 kai kenkyū seminā* 東洋学へのコンピュータ利用第17回研究セミナー. Kyōto: Kyōto Daigaku Jimbun kagaku kenkū fuzoku kanji jōhō kenkyū sentā 京都大学人文科学研究所附属漢字情報研究センター.
- ISHIZUKA Harumichi 石塚晴通, IKEDA Shoju 池田証寿, OKAGAKI Hirotaka 岡墻裕剛 (2006). “Kanji jitai kihan deitabeisu no kōchiku to kōkai: HNG purojekuto” 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. *Kokusai gakujutsu shinpojiumu 'Bukku rōdo to bunka kōryū* 国際学術シンポジウム「ブックロードと文化交流. Hangzhou, China.
- TAKADA Tomokazu 高田智和, OKAGAKI Hirotaka 岡墻裕剛 (2006). “Kanji jitai kihan deitabeisu no genjō” 漢字字体規範データベースの現状. *Kokusai wākushoppu 'Tenseki kōryū (kundoku) to kanji jōhō* 国際ワークショップ「典籍交流(訓読)と漢字情報. Hokkaidō University.
- OKAGAKI Hirotaka 岡墻裕剛, ISHIZUKA Harumichi 石塚晴通, IKEDA Shoju 池田証寿, TAKAGI Yui 高木維, SAIKI Masanao 齋木正直 (2008). “HNG-DB (deitabeisu) no igi to kongo no tenbō” HNG-DB (データベース)の意義と今後の展望. *Ajia Afurika gengo bunka kenkyū tsūshin* アジア・アフリカ言語文化研究所通信 122号. Tokyo University of Foreign Studies.
- OKAGAKI Hirotaka 岡墻裕剛, ISHIZUKA Harumichi 石塚晴通, SAIKI Masanao 齋木正直 (2008). “HNG de miru jitaisū no hensen” HNGで見る字体数の変遷. *Dai 78 kai jimbun kagaku to konpyūta kenkyū happyōkai* 第78回人文科学とコンピュータ研究会発表会.

- TOYAMA Hideo 當山日出夫 (2008). “Kanezawa bunko bon Shirai bunshū ‘Chōgonka’ no kanji jitai no jittai: Kanji jitai kihan deitabeisu wo riyō shite” 金沢文庫本白氏文集『長恨歌』の漢字字体の実態—漢字字体規範データベースを利用して. *Ritsumeikan Shirakawa Shizuka kinen Tōyō moji bunka kenkyū kiyō* 立命館白川静記念東洋文字文化研究所紀要 2.
- ISHIZUKA Harumichi (2004). “Japanese standards of writing Chinese characters—from ancient to modern printings.” *37<sup>th</sup> International Congress of Asian and North African Studies*, “JAPANIZATION—from ancient to modern times—(3).”
- ISHIZUKA Harumichi, IKEDA Shoji, 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.