

《遵生八箋》舊版造字轉碼說明

中研院資訊所文獻處理實驗室

中研院史語所漢籍電子文獻工作小組

2008/3/20 陳建安 製作

一、《遵生八箋》一書(遵生八箋.xml)使用舊版造字 72 個，字頻 128 次，詳如附件一。這 72 個造字中，53 個可轉成 Windows XP 能顯示的字，字頻 100 次；另外 19 個字必須轉成構字式，字頻 28 次。

二、附件一的造字分析表說明如下：

甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。

乙、造字：舊版造字

丙、字頻(txt)：造字在「.txt」文件的出現次數

丁、字頻(xml)：造字在「.xml」文件的出現次數

戊、Big5：造字的 Big5 碼

己、Unicode：造字所對應的 Unicode 碼

庚、WinXP：造字在 Windows XP 的對應字形

辛、構字式：Windows XP 若無對應字形，則改採用構字式

壬、備註凡例：

- 1、校對問題，舊版漢籍錯字，可用程式全部取代：在舊版漢籍電子文獻中即存在的錯字，因校對時的疏漏而未更正，持續留存在新版漢籍電子文獻中；若該造字的所有頻次，皆屬於錯誤使用的錯字情形，可以用程式全部取代為正確字形。如編號 13 的「衤 𠂔 因」字，原字為「裊」。
- 2、重碼問題，採用部件外字，不轉碼：舊漢籍造字與構字式部件外字，皆是建立在 Big5 的使用者造字區中，兩者所使用的碼區是重疊的，在同時使用構字式和舊漢籍造字的新版漢籍中，這些重疊的字碼可能被用以表達兩個不同字形，我們稱之為「重碼問題」。然而漢籍在舊轉新的過程中，若已將原舊漢籍造字的部分全數改為 Unicode 或構字式，則在 xml 出現的造字頻次皆採用部件外字，此時便不需轉碼。如編號 4904，字碼 8864，對應的舊漢籍造字為「尅」，部件外字為「𠂔」，對照原書發現書中皆取部件外字，故不需轉碼，但由於書中字義為八卦符號之「𠂔」，故手動改為「𠂔」。
- 3、校對問題，新舊版缺字並存：漢籍在舊轉新的過程中，以人工修改方式取代舊漢籍造字，雖然修改的結果是正

確的，但是只修改到部分造字，沒有全部修改到，導致舊漢籍造字在新版漢籍檔案中，一部分已被取代成 Unicode 或構字式，另一部分仍為舊漢籍造字的並存情形。處理方式為配合漢籍工作室的修改，將沒被修改到的舊漢籍造字，以對應的 Unicode 或構字式取代。如編號 4515，舊漢籍造字為「呪」，比對 txt 檔及 xml 檔後，發現原先在 txt 檔使用舊漢籍造字「呪」者，在 xml 檔中部分已人工修改為「咒」字，故配合修改，將編號 4515「呪」字，手動取代為「咒」字。

- 4、異體字問題：新版漢籍考量到使用者檢索及使用時的便利性，將用字原則改為除專詞等特殊情形之外，一律改用標準字呈現。如編號 1332 的「扌𠂔」係「攜」字之異體，故以「攜」字取代。又編號 1434 的「笔」字，係「筆」字之簡體字，亦以正體的「筆」字取代。
- 5、Unicode 字型呈現差異：Unicode 字型與舊漢籍造字有些微差異，但只是字體風格差異，實際上仍為同一個字，因此仍取 Unicode 字型。如編號 4259 的「𩇛」字，Unicode 字型呈現為「𩇛」，實際上仍為同一字。
- 6、待造字：Unicode 及漢字構形資料庫皆未收錄的舊漢籍造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 4005 的「鬼𠂔天」。

三、 Unicode 目前收錄的漢字總數為 70194，分屬於三個不同區段，詳如表一。目前 Windows XP 只支援 CJK 認同表意文字區的 20902 個字，內碼為 4E00-9FFF。所以造字編號 823 的「僑」字，Unicode 編碼為 2256A，由於 Windows XP 並不支援，仍須使用構字式「彳𠂔僑」。

表一、Unicode 的字數及編碼區段

Unicode	字集子集合	新增字數	新增編碼區段	總字數	WinXP
1.1 版	CJK 認同表意文字區	20902	4E00-9FFF	20902	支援
3.0 版	CJK 認同表意文字擴充 A 區	6582	3400-4DFF	27484	不支援
3.1 版	CJK 認同表意文字擴充 B 區	42710	20000-2A6D6	70194	不支援

四、 txt 檔與 xml 檔頻次不同且造字分析表「備註」欄有備註問題情況者，詳細取代內容可參考附件二的手動取代表。附件二的手動取代表的欄位說明如下：

甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。

乙、取代：當實際使用字形與造字不同時，在此標出。

丙、檔案原文摘錄：摘錄檔案(遵生八箋.xml)中包含該造字的文句段落，其中紅字底線的部分，為需要手動取代的原文。

丁、手動取代結果：變更過後的檔案內容，其中藍字底線的部分，為修改後的結果。

附件一、《遵生八箋》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
13	裯	1	1	FA4C	4104		衤 𠂔 因	裯，校對問題，舊版漢籍錯字，可用程式全部取代。
304	𠂔	2	2	FBF4			𠂔	
763	𠂔	1	1	FEE8			𠂔	
823	僑	2	2	8E65	2256A		彳 𠂔 喬	
912	𪚩	2	2	8EE0	9DC0	𪚩		
950	蘭	1	1	8F47	4561		艹 𠂔 閭	
1010	珏	1	1	8FA5	73CF	珏		
1332		1	1	91AD			扌 𠂔 嵩	攜，異體字問題。
1354	褱	1	1	91C3	79A9	褱		
1434	笔	1	1	9254	7B14	笔		筆，異體字問題。
1497		1	1	92B5			𠂔 米 ノ 凡 𠂔	𠂔，異體字問題。
1688	魷	5	5	93D7	9B6B	魷		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1893	袞	1	1	9548	886E	袞		
1905	視	1	1	9554	88E9	視		
2046	踪	4	4	9644	8E2A	踪		
2147	蟾	1	1	96CB	87EE	蟾		
2275	鯪	2	2	97AE	4C49		魚厶且	
2377	訢	1	1	9855	279AC		形工几言口	
2395	鐸	2	2	9867	947B	鐸		
2461	葱	1	1	98CB	44E4		艹厶忽	葱，校對問題，舊版漢籍錯字，可用程式全部取代。
2524	饒	1	1	994B	994D	饒		
2601	岢	1	1	99BA	21D7B		山厶亢	
2731	雀	5	5	9A7D			宀厶隹	雀，異體字問題。
2737	嶢	1	1	9AA5	5D66	嶢		
2874	饒	1	1	9B6F	9939	饒		
3005	糍	1	1	9C55	7CCD	糍		
3016	闌	1	1	9C60	28D91		門厶臭	闌，異體字問題。
3161	壘	1	1	9D54	58DC	壘		
3788	慙	2	2	8153	22871		殼厶心	慙，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3794	湊	9	9	8159	51D1	湊		湊，異體字問題。
3801	剗	1	1	8160	5257	剗		
3822	叁	1	1	8175	53C1	叁		參，異體字問題。
3836	喻	1	1	81A5	564F	喻		
3851	姤	1	1	81B4	59F9	姤		
3906		1	1	81EB	6409		𢦏扌 ㄣ隹 𠂔	推，異體字問題。
3927	柏	2	2	8241	6855	柏		
3981	珉	1	1	8277	73C9	珉		
4004	瘻	1	1	82B0	763B	瘻		
4005		1	1	82B1			鬼 天	待造字。
4025	礪	1	1	82C5	7934	礪		
4034	穉	8	8	82CE	7A49	穉		
4036	窰	1	1	82D0	7AB0	窰		
4041	豎	6	6	82D5	7AEA	豎		
4063	蘖	1	1	82EB	7CF5	蘖		
4069	綳	1	1	82F1	7DB3	綳		
4114	蘖	2	2	835F	8616	蘖		
4129	虬	3	3	836E	866C	虬		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4139	袴	1	1	8378	88B4	袴		
4167	賁	1	1	83B6	8CEB	賁		
4219	類	1	1	83EA	982E	類		
4221	頰	1	1	83EC	9833	頰		
4250	鰭	1	1	844A	9C0C	鰭		
4259	麇	1	1	8453	9E95	麇		Unicode 呈現差異。
4269	鼪	1	1	845D	9F39	鼪		
4278	亘	2	2	8466	4E98	亘		
4286	从	4	4	846E	4ECE	从		
4437	廩	1	1	8568	5EEA	廩		廩，異體字問題。
4483	嵯	1	1	85B8	5D70	嵯		
4501	獮	1	1	85CA	7356	獮		
4515	呪	2	1	85D8	546A	呪		咒，校對問題，新舊版缺字並存。
4558	廐	1	1	8644	53A9	廐		廐，異體字問題。
4607	狔	2	2	8675	3E60		狔 𤝵 屯	
4649	蕒	3	3	86C1	8552	蕒		
4785	贗	4	4	87AC	8D0B	贗		
4870	壑	4	4	8842	97F2	壑		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4904	尅	2	1	8864	5C05	尅		與三重碼，重碼問題，採用部件外字，不轉碼，但因校對問題，人工取代為三。
5108	惛	1	1	89B5	60FD	惛		
5327	崱	1	1	8AF3	5D2A	崱		
5543	𩺰	3	3	8C6F	9B9D	𩺰		
5619	繇	1	1	8CDD	470C		繇	
5743	𩺰	1	1	8DBC	293F6		𩺰米韋隹火	
5776	漚	1	1	8DDD	7002	漚		

附件二、《遵生八箋》手動取代表

編號	取代	檔案原文摘錄	手動取代結果
4515	咒	經 素 完畢方禮拜	經 咒 完畢方禮拜
4904	三	三 坎卦正北	三 坎卦正北