

《四庫全書總目提要》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
2008/1/9 陳建安 製作

一、《四庫全書總目提要》一書(四庫全書總目提要.xml)使用舊版造字 49 個，字頻 602 次，詳如附件一。這 49 個造字中，29 個可轉成 Windows XP 能顯示的字，字頻 383 次；另外 20 個字必須轉成構字式，字頻 219 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、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 5844 原無對應字形，比對原書後發現書上字形皆為「冂」，故以「冂」全部取代。
- 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 4262，字碼為 8456，對應的舊漢籍造字為「麪」，部件外字為「昏」，對原書發現書中皆取部件外字「昏」，故不需轉碼。
- 3、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 5602 的「粘」字，

Unicode 字型呈現為「𣎵」，實際上仍為同一字。

- 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 4366 的「𣎵」字，Unicode 編碼為 48BC 由於 Windows XP 並不支援，仍須使用構字式「共𠂇𠂇」。

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、 《四庫全書總目提要》中未確定字形「」共 166 處，詳細處理內容可參考附件二的未確定字形手動取代表。附件二的手動取代表說明如下：

- 甲、 檔案原文摘錄：摘錄檔案(四庫全書總目提要.xml)中包含未確定字形「」的文句段落，其中紅字底線的部分，為需要手動取代的字形。
- 乙、 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。

- 五、 對於在缺字之外發現的錯字，其詳細內容可參考附件三的校出錯字表。附件三校出錯字表的說明如下：

- 甲、 檔案原文摘錄：摘錄檔案(四庫全書總目提要.xml)中包含該錯字的文句段落，其中紅字底線的部分，為需要手動取代的原文。文句後方標示(全部取代)者，表示執行全部取代的動作，不一一列舉所有原文。
- 乙、 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。文句後方標示(全部取代)者，表示執行全部取代的動作，不一一列舉所有修改後結果。

附件一、《四庫全書總目提要》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
615	丂	1	1	FDF1			丂	
4262	麪	1	1	8456	9EAA	麪		與昏重碼，重碼問題，採用部件外字，不轉碼。
4350	侶	4	10	84D0	5058	侶		與虫重碼，重碼問題，採用部件外字，不轉碼。
4366	邾	2	1	84E0	48BC		共𠂔𠂔	與𠂔重碼，重碼問題，採用部件外字，不轉碼。
4369	晫	38	1	84E3	666B	晫		與𠂔重碼，重碼問題，採用部件外字，不轉碼。
5043	曠	1	1	8952	5F4D	曠		與自重碼，重碼問題，採用部件外字，不轉碼。
5098	寤	1	1	89AB	7AB9	寤		與川重碼，重碼問題，採用部件外字，不轉碼。
5233	覽	1	155	8A73	3F49		臨 𠂔 瓦	與月重碼，重碼問題，採用部件外字，不轉碼。
5239	擎	1	20	8A79	6486	擎		與丹重碼，重碼問題，採用部件外字，不轉碼。
5302	昏	35	3	8ADA	662C	昏		與羸重碼，重碼問題，採用部件外字，不轉碼。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5334	暉	1	14	8AFA	6692	暉		與扌重碼，重碼問題，採用部件外字，不轉碼。
5336	曙	1	14	8AFC	668F	曙		與亼重碼，重碼問題，採用部件外字，不轉碼。
5338	暎	8	172	8AFE	668E	暎		與艹重碼，重碼問題，採用部件外字，不轉碼。
5351	肫	1	20	8B4C	80AC	肫		與竹重碼，重碼問題，採用部件外字，不轉碼。
5371	冃	86	1	8B60	808E	冃		與业重碼，重碼問題，採用部件外字，不轉碼。
5388	黠	1	10	8B71	4D63		黑 𠃉 旦	與彡重碼，重碼問題，採用部件外字，不轉碼。
5392	息	2	2	8B75	60A4	息		與羊重碼，重碼問題，採用部件外字，不轉碼。
5423	悤	3	2	8BB6	6810	悤		與冂重碼，重碼問題，採用部件外字，不轉碼。
5445	覘	1	21	8BCC	6898	覘		與疒重碼，重碼問題，採用部件外字，不轉碼。
5470	椽	32	11	8BE5	68C5	椽		與畝重碼，重碼問題，採用部件外字，不轉碼。
5472	瓦	7	4	8BE7			瓦	與繼重碼，重碼問題，採用部件外字，不轉碼。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5478	飴	1	2	8BED	98F0	飴		與儿重碼，重碼問題，採用部件外字，不轉碼。
5479	龜	1	28	8BEE	9F03	龜		與丿重碼，重碼問題，採用部件外字，不轉碼。
5491	楨	4	1	8BFA	6967	楨		與关重碼，重碼問題，採用部件外字，不轉碼。
5493	迨	4	15	8BFC			迨	與彡重碼，重碼問題，採用部件外字，不轉碼。
5537	离	6	5	8C69	79BC	离		與冫重碼，重碼問題，採用部件外字，不轉碼。
5539	鬮	2	1	8C6B	29C5B		鬮 𠂔 𠂕	與商重碼，重碼問題，採用部件外字，不轉碼。
5555	迨	7	4	8C7B	9008	迨		與乚重碼，重碼問題，採用部件外字，不轉碼。
5559	欸	10	5	8CA1	6B35	欸		與亾重碼，重碼問題，採用部件外字，不轉碼。
5571	撫	16	2	8CAD	6A45	撫		與棊重碼，重碼問題，採用部件外字，不轉碼。
5580	离	2	7	8CB6	79BC	离		與攴重碼，重碼問題，採用部件外字，不轉碼。
5583	殮	5	21	8CB9	98F1	殮		與葬重碼，重碼問題，採用部件外字，不轉碼。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5592	桌	3	3	8CC2	3B9A		桌	與百重碼，重碼問題，採用部件外字，不轉碼。
5602	枱	2	2	8CCC	67D7	枱		Unicode 呈現差異。
5616	𦉳	1	1	8CDA	3849		巾 𦉳 弦	與𦉳重碼，重碼問題，採用部件外字，不轉碼。
5669	𧈧	1	7	8D50	2722B		工 𧈧 虫	與𧈧重碼，重碼問題，採用部件外字，不轉碼。
5684	淖	1	2	8D5F	6E12	淖		與𧈧重碼，重碼問題，採用部件外字，不轉碼。
5689	𧈧	1	5	8D64	20087		𧈧	與𧈧重碼，重碼問題，採用部件外字，不轉碼。
5695	𧈧	1	3	8D6A	3B3A		𧈧 𧈧 巾	與𧈧重碼，重碼問題，採用部件外字，不轉碼。
5844		1	1	C6C3				冂，校對問題，舊版漢籍錯字，可用程式全部取代。
5845		1	1	C6C4				冂，校對問題，舊版漢籍錯字，可用程式全部取代。
5847	勺	1	1	C6C6	52F9	勺		
5850		1	1	C6C9				厶，校對問題，舊版漢籍錯字，可用程式全

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
								部取代。
5852		3	3	C6CB				宀，校對問題，舊版漢籍錯字，可用程式全部取代。
5853		3	3	C6CC				《，校對問題，舊版漢籍錯字，可用程式全部取代。
5854	么	9	9	C6CD	5E7A	么		
5855		1	1	C6CE				广，校對問題，舊版漢籍錯字，可用程式全部取代。
5858	彡	2	2	C6D1			彡	
5861		1	1	C6D4				广，校對問題，舊版漢籍錯字，可用程式全部取代。

附件二、《四庫全書總目提要》「𠄎」字形手動取代表

檔案原文摘錄	手動取代結果
宋龔士𠄎編。士𠄎爵里無考	宋龔士𠄎編。士𠄎爵里無考
酉字用古體作𠄎	酉字用古體作𠄎
是以𠄎解𠄎矣	是以𠄎解𠄎矣
是仍以𠄎解𠄎	是仍以𠄎解𠄎
轉益其𠄎而已矣	轉益其𠄎而已矣
手部。篆作𠄎。隸省作𠄎	手部。篆作𠄎 ^𠄎 。隸省作𠄎 ^𠄎
佐𠄎。遂併毛鄭而棄之	佐𠄎。遂併毛鄭而棄之
亞誤𠄎。弓誤已	亞誤𠄎。弓誤已
𠄎甸不出車馬一條	𠄎甸不出車馬一條

檔案原文摘錄	手動取代結果
毛奇齡二家之 ^②	毛奇齡二家之 ^鬪
若小司徒四 ^② 為甸以下	若小司徒四 ^丘 為甸以下
而隨聲佐 ^② 者	而隨聲佐 ^鬪 者
而 ^② 為啟亂端	而 ^鬪 為啟亂端
罵詈 ^② 毆賭博諸律	罵詈 ^鬪 毆賭博諸律
引 ^② 辛以駁伍員	引 ^鬪 辛以駁伍員
楚婦人 ^② 伯比妻	楚婦人 ^鬪 伯比妻
誤以 ^② 𠂔 ^② 儀父為邾命卿	誤以 ^𠂔 𠂔 ^② 儀父為邾命卿
^② 。 ^② 古庚字。為陽金。 ^② 古鑄字	^𠂔 庸 ^體 說文 ^𠂔 。 ^𠂔 庚 ^體 說文 ^𠂔 古庚字。為陽金。 ^𠂔 用 ^體 說文 ^𠂔 古鑄字
親字下既云古尚書作 ^②	親字下既云古尚書作 ^𠂔 𠂔 ^親 體說文 ^𠂔
^② 即古瞿字也	^𠂔 目 ^體 說文 ^𠂔 即古瞿字也
乃竝云古老子作 ^②	乃竝云古老子作 ^𠂔 𠂔 ^體 說文 ^𠂔
^② ^② ^②	^𠂔 弋 ^一 體說文古文 ^𠂔 𠂔 ^式 體說文古文 ^𠂔 𠂔 ^弋 三體說文古文 ^𠂔
乃惟云 ^② 字出說文	乃惟云 ^𠂔 式 ^體 說文古文 ^𠂔 字出說文
^② ^② 字	^𠂔 弋 ^一 體說文古文 ^𠂔 𠂔 ^弋 三體說文古文 ^𠂔 字
簣字說文本作 ^②	簣字說文本作 ^𠂔 史 ^體 說文古文 ^𠂔
文本作 ^②	文本作 ^𠂔 寢 ^體 說文 ^𠂔
燒字說文本作 ^②	燒字說文本作 ^𠂔 燒 ^體 說文 ^𠂔
^② 鍤	^𠂔 鍤 ^體 說文古文 ^𠂔 鍤
如 ^② 本音畏	如 ^𠂔 本音畏
飛几 ^② 凶鹵矛辛	飛几 ^𠂔 𠂔 ^𠂔 𠂔 ^𠂔 凶鹵矛辛
蓐勃酋 ^② 斤	蓐勃酋 ^𠂔 斤
^② 之與禾	^𠂔 之與禾
其餘如 ^② 字	其餘如 ^𠂔 字
污汗幬 ^𠂔 巾 ^② ... ^𠂔 之類	污汗幬 ^𠂔 巾 ^𠂔 ... ^𠂔 之類
如丑字从 ^② 从丨	如丑字从 ^𠂔 又 ^體 說文 ^𠂔 从丨
之為鄰。 ^② 之	之為鄰。 ^𠂔 之
纍既 ^② 夫	纍既 ^𠂔 夫
^② 古攀字	^𠂔 古攀字
不知篆文作 ^②	不知篆文作 ^𠂔 冊 ^體 說文 ^𠂔
謂小篆日月作 ^② ^②	謂小篆日月作 ^𠂔 日 ^體 說文 ^𠂔 𠂔 ^月 體說文 ^𠂔
已不知古文之作 ^② ^②	已不知古文之作 ^𠂔 日 ^體 集成 4269 ^𠂔 𠂔 ^月 體包 2.214 ^𠂔
而古文作 ^②	而古文作 ^𠂔

檔案原文摘錄	手動取代結果
既而轉 [㊟] 干盤	既而轉 [㊟] 干盤
習於戰 [㊟]	習於戰 [㊟]
誇多 [㊟] 靡	誇多 [㊟] 靡
暹羅鬼與鄭和 [㊟] 法	暹羅鬼與鄭和 [㊟] 法
濶字笨字 [㊟]	濶字笨字 [㊟] ^一 [㊟] ^八 [㊟] ^士
試馬 [㊟] 羊	試馬 [㊟] 羊
八 [㊟] 訟	八 [㊟] 訟
[㊟] 訟為 [㊟] 毆	[㊟] 訟為 [㊟] 毆
唐律並入 [㊟] 訟	唐律並入 [㊟] 訟
分入 [㊟] 毆罵詈	分入 [㊟] 毆罵詈
至爭 [㊟] 賦役諸事	至爭 [㊟] 賦役諸事
固宜與循聲佐 [㊟] 者	固宜與循聲佐 [㊟] 者
格 [㊟] 而不休者	格 [㊟] 而不休者
是率天下而 [㊟] 也	是率天下而 [㊟] 也
淫蕩酗酒 [㊟] 很罔利	淫蕩酗酒 [㊟] 很罔利
習格 [㊟] 法	習格 [㊟] 法
銜奇 [㊟] 異	銜奇 [㊟] 異
以陵抵 [㊟] 食之說	以陵抵 [㊟] 食之說
譜 [㊟] 新角異	譜 [㊟] 新角異
以誇多 [㊟] 巧為事	以誇多 [㊟] 巧為事
左傳 [㊟] 宜申	左傳 [㊟] 宜申
不知志字上本从 [㊟]	不知志字上本从 [㊟]
謂 [㊟] 百草始於周	謂 [㊟] 百草始於周
與子路 [㊟] 事為證	與子路 [㊟] 事為證
謂 [㊟] 閱一事	謂 [㊟] 閱一事
官舍 [㊟] 鶴鷄	官舍 [㊟] 鶴鷄
方競以駢偶 [㊟] 工巧	方競以駢偶 [㊟] 工巧
史嵩之趙與 [㊟]	史嵩之趙與 [㊟] ^形 ^竹 ^焉 ^心 ^口
以誇多 [㊟] 靡為工	以誇多 [㊟] 靡為工
又有隼 [㊟] 集	又有隼 [㊟] ^矢 [△] ^吉 集
非徒 [㊟] 靡誇多	非徒 [㊟] 靡誇多
平心解 [㊟] 之論	平心解 [㊟] 之論
即戰 [㊟] 於流寇之	即戰 [㊟] 於流寇之
他如池平初 [㊟] 蛤	他如池平初 [㊟] 蛤
堡兵爭 [㊟] 之由	堡兵爭 [㊟] 之由
多 [㊟] 靡	多 [㊟] 靡

檔案原文摘錄	手動取代結果
鷄篇	鬪鷄篇
梁簡文但云	梁簡文但云鬪
意欲夸多捷	意欲夸多鬪捷
昇字原本作曰	昇字原本作
用差字	用身差字
曰。又詩人玉屑	。又詩人玉屑
亦作黃曰	亦作黃
實作曰字	實作字
昇字篆文作曰	昇字篆文作
故作曰字	故作字

附件三、《四庫全書總目提要》校出錯字表

檔案原文摘錄	手動取代結果
故貞不及日者	故貞不及睹
瓦（全部取代）	瓦（全部取代）