

《大金國志校證》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

2008/1/3 陳建安 製作

一、《大金國志校證》一書(大金國志校證.xml)使用舊版造字 54 個，字頻 410 次，詳如附件一。這 54 個造字中，49 個可轉成 Windows XP 能顯示的字，字頻 391 次；另外 5 個字必須轉成構字式，字頻 19 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 4027 的「裨」字，原字為「裨」。
- 2、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 800，舊漢籍造字為「逕」，比對 txt 檔及 xml 檔後，發現原先在 txt 檔使用舊漢籍造字「逕」者，在 xml 檔中部分已人工修改為「𠂔眾」，但由於 Unicode 已有「逕」字，故將「𠂔眾」改為「逕」，其餘未修改的部分亦皆手動改為「逕」。
- 3、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 4132 的「血𡗗𡗗」係「眾」字之異體，故以「眾」字取代。

- 4、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 5338，字碼為 8AFE，對應的舊漢籍造字為「暎」，部件外字為「𠂇」，對原書發現書中皆取部件外字「𠂇」，故不需轉碼。
- 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 4158 的「譌」字，Unicode 字型呈現為「譌」，實際上仍為同一字。

三、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說明如下：

- 甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
- 乙、取代：實際使用字形與漢籍舊版造字不同時，於此欄列出。
- 丙、檔案原文摘錄：摘錄檔案(大金國志校證.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。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。
- 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。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。

四、《大金國志校證》中未確定字形「❓」共 12 處，詳細處理內容可參考附件三的未確定字形手動取代表。附件三的手動取代表說明如下：

- 甲、檔案原文摘錄：摘錄檔案(大金國志校證.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。

附件一、《大金國志校證》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
297	𡗗	21	3	FBED			𡗗	𡗗，校對問題，新舊版缺字並存。
766	𡗗	1	1	FEED			𡗗	
800	還	2	1	8E4E	905D	還		還，校對問題，新舊版缺字並存。
972	犇	23	1	8F5D	7287	犇		犇，校對問題，新舊版缺字並存。
1416	埃	6	2	9242	7AE2	埃		埃，校對問題，新舊版缺字並存。
1565	絲	1	1	92F9	7DDC	絲		
1603	倂	4	2	9360	5034	倂		倂，校對問題，新舊版缺字並存。
1893	袞	7	4	9548	886E	袞		袞，校對問題，新舊版缺字並存。
3799	刼	2	1	815E	5226	刼		刼，校對問題，新舊版缺字並存。
3804	効	9	1	8163	52B9	効		効，校對問題，新舊版缺字並存。
3814	却	34	10	816D	5374	却		却，校對問題，新舊版缺字並存。
3824	叙	7	1	8177	53D9	叙		叙，校對問題，新舊版缺字並存。
3835	噍	3	1	81A4	564D	噍		噍，校對問題，新舊版缺字並存。
3846	壻	20	7	81AF	58FB	壻		壻，校對問題，新舊版缺字並存。
3864	峯	16	2	81C1	5CEF	峯		峯，校對問題，新舊版缺字並存。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3978	猪	12	6	8274	732A	猪		猪，校對問題，新舊版缺字並存。 Unicode 呈現差異。
3979	猫	3	1	8275	732B	猫		猫，校對問題，新舊版缺字並存。
4017	蟬	1	1	82BD	880F	蟬		
4027	裨	1	1	82C7	7986	裨		裨，校對問題，舊版漢籍錯字，可用程式全部取代。
4043	筭	2	2	82D7	7B53	筭		Unicode 呈現差異。
4044	筋	2	1	82D8	7B6F	筋		筋，校對問題，新舊版缺字並存。
4080	羣	57	5	82FC	7FA3	羣		羣，校對問題，新舊版缺字並存。
4090	脚	4	4	8347	811A	脚		
4107	葱	1	1	8358	8471	葱		
4132		219	12	8371			血𠂇𠂇	眾，異體字問題。 眾，校對問題，新舊版缺字並存。
4134	衛	40	2	8373	885E	衛		衛，校對問題，新舊版缺字並存。
4139	袴	3	2	8378	88B4	袴		袴，校對問題，新舊版缺字並存。
4140	衽	3	3	8379	88B5	衽		
4145	羈	1	1	837E	898A	羈		Unicode 呈現差異。
4146	覩	64	1	83A1	89A9	覩		覩，校對問題，新舊版缺字並存。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4158	譌	12	1	83AD	8B4C	譌		譌，校對問題，新舊版缺字並存。 Unicode 呈現差異。
4176	輒	4	1	83BF	8F19	輒		輒，校對問題，新舊版缺字並存。
4184	迹	86	7	83C7	8FF9	迹		迹，校對問題，新舊版缺字並存。
4193	鈎	3	1	83D0	920E	鈎		鈎，校對問題，新舊版缺字並存。
4203	隣	4	1	83DA	96A3	隣		隣，校對問題，新舊版缺字並存。
4242	鬪	16	5	8442	9B2D	鬪		鬪，校對問題，新舊版缺字並存。詳見附件三。
4256	鷄	3	2	8450	9DC4	鷄		鷄，校對問題，新舊版缺字並存。
4278	亘	6	4	8466	4E98	亘		亘，校對問題，新舊版缺字並存。
4309	刼	1	1	84A7	523C	刼		
4507	吴	150	13	85D0	5434	吴		吴，校對問題，新舊版缺字並存。
4904	尅	4	1	8864	5C05	尅		尅，校對問題，新舊版缺字並存。
5008	幪	8	3	88EE	5E5E	幪		幪，校對問題，新舊版缺字並存。
5176		5	1	89F9			； 厶 冗	沉，異體字問題。 沉，校對問題，新舊版缺字並存。
5218		23	2	8A64			形 宀 覓、 覓	寬，異體字問題。 寬，校對問題，新舊版缺字並存。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5290	携	4	2	8ACE	643A	携		携，校對問題，新舊版缺字並存。
5338	暎	1	7	8AFE	668E	暎		與 ^彡 重碼，重碼問題，採用部件外字，不轉碼。
5439	栢	4	4	8BC6	6822	栢		
5542	鑠	4	2	8C6E	93C1	鑠		鑠，校對問題，新舊版缺字並存。
5550	纪	6	207	8C76	4E63	纪		與 ^亼 重碼，重碼問題，採用部件外字，不轉碼。
5552	弃	1	1	8C78	5F06	弃		
5559	欸	8	2	8CA1	6B35	欸		欸，校對問題，新舊版缺字並存。
5704	羶	10	2	8D73	6C0A	羶		羶，校對問題，新舊版缺字並存。
5706	敍	59	2	8D75	654D	敍		敍，校對問題，新舊版缺字並存。
5717	陘	1	57	8DA2	967F	陘		與 ^支 重碼，重碼問題，採用部件外字，不轉碼。

附件二、《大金國志校證》手動取代表

編號	取代	檔案原文摘錄	手動取代結果
297		___ (全部取代)	<u>厶</u> (全部取代)
800		車馬雜___	車馬雜 <u>逕</u>
		車馬雜 <u>𠂔</u> <u>厶</u> <u>眾</u>	車馬雜 <u>逕</u>
972		六月___蔡	六月 <u>𠂔</u> 蔡
		<u>品</u> <u>牛</u> (全部取代)	<u>𠂔</u> (全部取代)
1416		___ (全部取代)	<u>𠂔</u> (全部取代)
		<u>立</u> <u>厶</u> <u>矣</u> (全部取代)	<u>𠂔</u> (全部取代)

編號	取代	檔案原文摘錄	手動取代結果
1603		— (全部取代)	儕 (全部取代)
		1 𠂔 奔 (全部取代)	儕 (全部取代)
1893		— (全部取代)	袞 (全部取代)
		衣 𠂔 公 (全部取代)	袞 (全部取代)
3799		殺人剽 𠂔	殺人剽 刦
		殺人剽 去 𠂔 𠂔 者	殺人剽 刦 者
3804		非人人 𠂔 死戰	非人人 効 死戰
		交 𠂔 力 (全部取代)	効 (全部取代)
3814		屨 (全部取代)	却 (全部取代)
		去 𠂔 尸 (全部取代)	却 (全部取代)
3824		然其 𠂔 次簡質	然其 叙 次簡質
		余 𠂔 又 (全部取代)	叙 (全部取代)
3835		殺戮無 𠂔 類	殺戮無 𠂔 類
		口 𠂔 焦 (全部取代)	𠂔 (全部取代)
3846		鳥 (全部取代)	𠂔 (全部取代)
		士 𠂔 胥 (全部取代)	𠂔 (全部取代)
3864		𠂔 (全部取代)	𠂔 (全部取代)
		山 𠂔 峯 (全部取代)	𠂔 (全部取代)
3978		𠂔 (全部取代)	𠂔 (全部取代)
		𠂔 𠂔 者 (全部取代)	𠂔 (全部取代)
3979		猪羊、𠂔 犬	猪羊、𠂔 犬
		𠂔 𠂔 苗 (全部取代)	𠂔 (全部取代)
4044		無碗 𠂔	無碗 𠂔
		忠獻以 𠂔 𠂔 助 條三策	忠獻以 𠂔 條三策
4080		𠂔 (全部取代)	𠂔 (全部取代)
		君 𠂔 羊 (全部取代)	𠂔 (全部取代)
4132	眾	𠂔 (全部取代)	𠂔 (全部取代)
		血 𠂔 𠂔 (全部取代)	𠂔 (全部取代)
4134		𠂔 (全部取代)	𠂔 (全部取代)
		𠂔 行 五 口 𠂔 (全部取代)	𠂔 (全部取代)
4139		𠂔 (全部取代)	𠂔 (全部取代)
		𠂔 𠂔 夸 襪皆以皮	𠂔 襪皆以皮
4146		<qa>𠂔</qa>漢官威儀	<qa>𠂔</qa>漢官威儀
		者 𠂔 見 (全部取代)	𠂔 (全部取代)
4158		他本 𠂔 為興平府	他本 𠂔 為興平府
		言 𠂔 為 (全部取代)	𠂔 (全部取代)

編號	取代	檔案原文摘錄	手動取代結果
4176		入 <u>灾</u> 不可出	入 <u>輒</u> 不可出
		<u>車</u> △ <u>取</u> (全部取代)	<u>輒</u> (全部取代)
4184		<u>墮</u> (全部取代)	<u>迹</u> (全部取代)
		<u>之</u> △ <u>亦</u> (全部取代)	<u>迹</u> (全部取代)
4193		天祚混同江 <u>暴</u> 魚	天祚混同江 <u>鈎</u> 魚
		<u>金</u> △ <u>勾</u> (全部取代)	<u>鈎</u> (全部取代)
4203		先併 <u>墨</u> 國	先併 <u>隣</u> 國
		<u>卩</u> △ <u>薺</u> (全部取代)	<u>隣</u> (全部取代)
4242		<u>显</u> (全部取代)	<u>鬪</u> (全部取代)
		詳見附件三	
4256		<u>鷄</u> (全部取代)	<u>鷄</u> (全部取代)
		璘又遣別將至寶 <u>奚</u> △ <u>鳥</u>	璘又遣別將至寶 <u>鷄</u>
4278		<u>扶</u> (全部取代)	<u>亘</u> (全部取代)
		<u>一</u> △ <u>旦</u> (全部取代)	<u>亘</u> (全部取代)
4507		<u>垂</u> (全部取代)	<u>吳</u> (全部取代)
		<u>口</u> △ <u>天</u> (全部取代)	<u>吳</u> (全部取代)
4904		以奉安玄祖 <u>三</u> 者	以奉安玄祖 <u>尅</u> 者
		<u>克</u> △ <u>寸</u> (全部取代)	<u>尅</u> (全部取代)
5008		<u>𠂔</u> (全部取代)	<u>幪</u> (全部取代)
		<u>巾</u> △ <u>業</u> (全部取代)	<u>幪</u> (全部取代)
5176	沉	瓜玉斧 <u>专</u> 煙和	瓜玉斧 <u>沉</u> 煙和
		<u>汙</u> △ <u>冗</u> (全部取代)	<u>沉</u> (全部取代)
5218	寬	<u>𠂔</u> △ <u>覓</u> 、 <u>𠂔</u> (全部取代)	<u>寬</u> (全部取代)
		<u>𠂔</u> (全部取代)	<u>寬</u> (全部取代)
5290		<u>𠂔</u> (全部取代)	<u>携</u> (全部取代)
		<u>𠂔</u> △ <u>隹</u> 乃 <u>𠂔</u> (全部取代)	<u>携</u> (全部取代)
5542		<u>𠂔</u> (全部取代)	<u>鑠</u> (全部取代)
		<u>金</u> △ <u>巢</u> (全部取代)	<u>鑠</u> (全部取代)
5559		<u>𠂔</u> (全部取代)	<u>欸</u> (全部取代)
		<u>矣</u> △ <u>欠</u> (全部取代)	<u>欸</u> (全部取代)
5704		<u>𠂔</u> (全部取代)	<u>𠂔</u> (全部取代)
		<u>毛</u> △ <u>亘</u> (全部取代)	<u>𠂔</u> (全部取代)
5706		<u>𠂔</u> (全部取代)	<u>𠂔</u> (全部取代)
		<u>余</u> △ <u>支</u> (全部取代)	<u>𠂔</u> (全部取代)

附件三、《大金國志校證》「」字形手動取代表

檔案原文摘錄	手動取代結果
城破，士無  志	城破，士無  志
<qa>  </qa>之士	<qa>  </qa>之士
舍戰  之士	舍戰  之士
「  」原作「門」	「  」原作「門」
按南遷錄作「  」	按南遷錄作「  」
由是人無<qd>  </qd>	由是人無<qd>  </qd>
「固」原作「  」	「固」原作「  」
<qa>  </qa>	<qa>  </qa>
喜戰  「  」原作「門」	喜戰  「  」原作「門」
<qa>  </qa>。</s>	<qa>  </qa>。</s>
不若  事趙天子	不若  事趙天子