

《水經注校釋》舊版造字轉碼說明

中研院資訊所文獻處理實驗室 2007/12/20 陳建安 製作

一、《水經注校釋》一書(水經注校釋.xml)使用舊版造字 23 個，字頻 1065 次，詳如附件一。這 23 個造字中，18 個可轉成 Windows XP 能顯示的字，字頻 977 次；另外 5 個字必須轉成構字式，字頻 88 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，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 1565，舊漢籍造字為「𧯛」，比對 txt 檔及 xml 檔後，發現原先在 txt 檔使用舊漢籍造字「𧯛」者，在 xml 檔中部分已人工修改為「帛𧯛系」，但由於 Unicode 已有「𧯛」字，故將「帛𧯛系」改為「𧯛」，其餘未修改的部分亦皆手動改為「𧯛」。

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 5610，字碼為 8CD4，

對應的舊漢籍造字為「泠」，部件外字為「𠂔」，對原書發現書中皆取部件外字「𠂔」，故不需轉碼。但是在字碼 8DFE 這個碼位，對應舊漢籍造字為「踰」，而構字式對應的部件外字為「㊦」，此部件漢籍用來表達未定義的字形（可能表達任何字形），若直接採用，等於沒有解決此處的缺字問題。因此在此字碼重碼的情況下，無論是採用舊漢籍造字「踰」，或採用部件外字「㊦」，都要以人工的方式取代為原書字形，詳細取代結果請參考附件三。

- 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 4686 的「埤」字，Unicode 編碼為 3654 由於 Windows XP 並不支援，仍須使用構字式「土𠂔韋」。

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：實際使用字形與舊漢籍造字不同時，於此欄列出。
- 丙、檔案原文摘錄：摘錄檔案(水經注校釋.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。

- 五、 《水經注校釋》中未確定字形「㊦」共 32 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(水經注校釋.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。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。
- 丙、 備註：註明相同文句在檔案中出現的次數，未註明者表示只出現一次。

附件一、《水經注校釋》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
850	圪	16	2	8EA2			土  冗	土  冗，校對問題，新舊版缺字並存。
1565	繇	3	1	92F9	7DDC	繇		繇，校對問題，新舊版缺字並存，並且將帛  系改為繇。
3791	况	20	4	8156	51B5	况		况，校對問題，新舊版缺字並存，並且將  兄改為况。
4016	着	8	4	82BC	7740	着		着，校對問題，新舊版缺字並存，並且將𠂔  目改為着。
4040	竝	186	4	82D4	7ADD	竝		竝，校對問題，新舊版缺字並存，並且將〇〇立改為竝。
4068	綫	3	3	82F0	7DAB	綫		
4080	羣	67	3	82FC	7FA3	羣		羣，校對問題，新舊版缺字並存。
4132		123	2	8371			血  𠂔	眾，校對問題，新舊版缺字並存。
4184	迹	29	1	83C7	8FF9	迹		迹，校對問題，新舊版缺字並存，並且將  亦改為迹。
4188	邨	3	1	83CB	90A8	邨		邨，校對問題，新舊版缺字並存，並且將屯  𠂔改為邨。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4193	鈎	3	1	83D0	920E	鈎		鈎，校對問題，新舊版缺字並存，並且將金𠂔勾改為鈎。
4240	髯	8	1	8440	9AF4	髯		髯，校對問題，新舊版缺字並存，並且將髡𠂔弗改為髯。
4262	麪	2	2	8456	9EAA	麪		與昏重碼，重碼問題，採用部件外字，不轉碼。
4455	劒	30	1	857A			僉𠂔刃	僉𠂔刃，校對問題，新舊版缺字並存。
4686	埤	16	9	86E6	3654		土𠂔韋	土𠂔韋，校對問題，新舊版缺字並存。
4817	寶	58	2	87CC	5BF3	寶		寶，校對問題，新舊版缺字並存，並且將𠂔𠂔王尔貝𠂔改為寶。
4927	岝	2	1	887B	5C9E	岝		與與重碼，重碼問題，採用部件外字，不轉碼。
5610	泠	1	2	8CD4	6C75	泠		與𠂔重碼，重碼問題，採用部件外字，不轉碼。
5631	泯	19	1	8CE9	6CFF	泯		泯，校對問題，新舊版缺字並存，並且將泯𠂔艮改為泯。
5717	陘	1	1	8DA2	967F	陘		與支重碼，重碼問題，採用部件外字，不轉碼。
5797	灩	5	913	8DF2	7054	灩		與𠂔重碼，重碼問題，採用部件外字，不轉碼。
5798	虧	2	74	8DF3	48DC		虛𠂔卩	與𠂔重碼，重碼問題，採用部件外字，不轉碼。
5809	踰	2	32	8DFE	8E79	踰		與𠂔重碼，重碼問題，採用部件外字，人工修改，詳見附件三。

附件二、《水經注校釋》手動取代表

編號	取代	檔案原文摘錄	手動取代結果
850		<u> </u> (全部取代)	<u>土</u> <u>厶</u> <u>冗</u> (全部取代)
1565		但 <u> </u> 古芒昧	但 <u>繇</u> 古芒昧
		<u>帛</u> <u>厶</u> <u>系</u> (全部取代)	<u>繇</u> (全部取代)
3791		<u>牽</u> (全部取代)	<u>況</u> (全部取代)
		<u>讠</u> <u>厶</u> <u>兄</u> (全部取代)	<u>況</u> (全部取代)
4016		<u>進</u> (全部取代)	<u>着</u> (全部取代)
		<u>羊</u> <u>厶</u> <u>目</u> (全部取代)	<u>着</u> (全部取代)
4040		<u>艺</u> (全部取代)	<u>竝</u> (全部取代)
		<u>oo</u> <u>立</u> (全部取代)	<u>竝</u> (全部取代)
4080		<u>倉</u> (全部取代)	<u>羣</u> (全部取代)
4132	眾	<u>公</u> (全部取代)	<u>眾</u> (全部取代)
4184		兼考故城之 <u>壙</u>	兼考故城之 <u>迹</u>
		<u>讠</u> <u>厶</u> <u>亦</u> (全部取代)	<u>迹</u> (全部取代)
4188		東潛 <u>尊</u> 民識	東潛 <u>邨</u> 民識
		<u>屯</u> <u>厶</u> <u>卩</u> (全部取代)	<u>邨</u> (全部取代)
4193		與近本 <u>暴</u> 稽校勘	與近本 <u>鈞</u> 稽校勘
		<u>金</u> <u>厶</u> <u>勾</u> (全部取代)	<u>鈞</u> (全部取代)
4240		縱髻 <u>彡</u> 前聞	縱髻 <u>髻</u> 前聞
		<u>髟</u> <u>厶</u> <u>弗</u> (全部取代)	<u>髻</u> (全部取代)
4455		另一本作“ <u>𠂔</u> ”	另一本作“ <u>金</u> <u>厶</u> <u>刃</u> ”
4686		<u>匡</u> (全部取代)	<u>土</u> <u>厶</u> <u>韋</u> (全部取代)
4817		<u>丿</u> (全部取代)	<u>寶</u> (全部取代)
		<u>𠂔</u> <u>王</u> <u>尔</u> <u>貝</u> <u>𠂔</u> (全部取代)	<u>寶</u> (全部取代)
5631		根據《 <u>𠂔</u> 水注》	根據《 <u>退</u> 水注》
		<u>讠</u> <u>厶</u> <u>艮</u> (全部取代)	<u>退</u> (全部取代)

附件三、《水經注校釋》「𠂔」字形手動取代表

檔案原文摘錄	手動取代結果	備註
主 <u>𠂔</u> 訟兵陣	主 <u>鬪</u> 訟兵陣	
與河水 <u>𠂔</u> (全部取代)	與河水 <u>鬪</u> (全部取代)	2
偃王愛民不 <u>𠂔</u>	偃王愛民不 <u>鬪</u>	
<u>𠂔</u> ，南洑為榮澤	<u>鬪</u> ，南洑為榮澤	
七里有 <u>𠂔</u> 雞臺 (全部取代)	七里有 <u>鬪</u> 雞臺 (全部取代)	2
燕太子丹 <u>𠂔</u> 雞	燕太子丹 <u>鬪</u> 雞	

檔案原文摘錄	手動取代結果	備註
與河伯馮夷 ^②	與河伯馮夷 ^鬪	
所謂穀、洛 ^② 也	所謂穀、洛 ^鬪 也	
謂之穀、洛 ^②	謂之穀、洛 ^鬪	
穀、洛二水 ^②	穀、洛二水 ^鬪	
有似于 ^②	有似于 ^鬪	
西北有 ^② 雞臺	西北有 ^鬪 雞臺	
龍 ^② 于時門之外	龍 ^鬪 于時門之外	
與內蛇 ^②	與內蛇 ^鬪	
禍起 ^② 雞矣	禍起 ^鬪 雞矣	
布性不喜合 ^②	布性不喜合 ^鬪	
^② 。乃植戟于門	^鬪 。乃植戟于門	
可留決 ^②	可留決 ^鬪	
羣 ^② 于縣	羣 ^鬪 于縣	
白黑烏 ^②	白黑烏 ^鬪	
漢與江 ^②	漢與江 ^鬪	
令尹 ^② 祁	令尹 ^鬪 祁	
有兩牛 ^② 于江岸	有兩牛 ^鬪 于江岸	
謂官屬曰：吾 ^②	謂官屬曰：吾 ^鬪	
操刀入水與神 ^②	操刀入水與神 ^鬪	
^② ，便山習水	^鬪 ，便山習水	
便習戰 ^② 者	便習戰 ^鬪 者	
上岸 ^② ，角軟	上岸 ^鬪 ，角軟	
利距善 ^②	利距善 ^鬪	
世以家雞 ^② 之	世以家雞 ^鬪 之	