


《睡虎地秦墓竹簡》(標記)舊版造字轉碼說明

中研院資訊所文獻處理實驗室
中央研究院語言所文獻語料小組
2009/10/22 丁致伶

轉碼主要工作是把檔案中的舊版造字轉換成 Windows XP 能支援的 Unicode 字形，Unicode 目前共收錄漢字 70194 個字，而 XP 只能支援 20902 個字（詳如表一），不支援之字將以構字式表達。例：造字編號 2461 的「葱」字，Unicode 編碼是 44E4，由於 XP 並不支援，仍需使用構字式「忽」。

表一、Unicode 的字數及編碼區段

Unicode	新增字數	新增編碼區段	總字數	WinXP
1.1 版	20902	4E00-9FFF	20902	支援
3.0 版	6582	3400-4DFF	27484	不支援
3.1 版	42710	20000-2A6D6	70194	不支援

一、舊版造字轉碼分析：

《睡虎地秦墓竹簡》(標記)使用舊版造字 59 個，字頻 669 次，這 59 個造字中，19 個可轉成 Windows XP 能顯示的字，字頻 103 次；另外 40 個字必須轉成構字式，字頻 566 次。

轉碼完成製作轉碼分析表，請參考附件一《睡虎地秦墓竹簡》(標記)轉碼分析表，欄位說明如下：

- (一) 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
- (二) 造字：舊版造字。
- (三) 頻次：舊版造字在文件的出現次數。
- (四) Big5：造字的 Big5 碼。
- (五) Unicode：造字所對應的 Unicode 碼。

- (六) WinXP：造字在 Windows XP 的對應字形。
- (七) 構字式：Windows XP 無法對應字形改用構字式。
- (八) 備註凡例：備註欄中記錄轉碼後字形及修改原因，凡例如下：
1. 部份使用不同字形，手動修改：檔案中為同一舊版造字，校對時卻發現原書使用兩個以上不同的字形，可能部份舊版造字為錯字或原書使用兩種異體字形，導致同一舊版造字轉碼時對應到兩個以上的字形；將一部份手動修改，剩餘的頻次則用程式取代成對應的 Unicode 或構字式。例：造字編號 761 的「攵」字頻次為 9，1 字依原書以程式取代為「攵」，剩餘 8 字手動修改為「攵」。手動修改部份請參考附件二《睡虎地秦墓竹簡》(標記)手動取代表。
 2. 錯字，以程式全部取代：與原書字形不符，並查詢教育部異體字字典確認非異體字後，皆歸類為錯字，以程式取代為正確字形。例：檔案中使用造字編號 338 的「𠂇」字，而原書使用之字為「由」，「𠂇」為錯字，以程式取代為正確字形「由」。
 3. Unicode 字型呈現差異：Unicode 字型與舊版造字有些微差異，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 3978 的「猪」字，Unicode 字型呈現為「猪」，實際上仍為同一字。

附件一、《睡虎地秦墓竹簡》(標記)轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
265	𠂇	1	FBCD			𠂇	
296	𠂇	60	FBEC			𠂇	
297	𠂇	79	FBED			𠂇	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
298	厶	19	FBEE			厶	
299	𠂇	50	FBEF			𠂇	
300	𠂈	50	FBF0			𠂈	
317	𠂉	2	FC42			𠂉	中，錯字，以程式全部取代。
321	臣	1	FC46			臣	
327	𠂊	1	FC4C			𠂊	
338	𠂋	1	FC57			𠂋	由，錯字，以程式全部取代。
377	扁	1	FC7E			扁	
379	支	1	FCA2			支	
387	𠂌	1	FCAA			𠂌	
407	𠂍	12	FCBE			𠂍	
447	𠂎	13	FCE6			𠂎	
486	毆	132	FD4E			毆	
488	毘	1	FD50			毘	
529	𠂏	1	FD79			𠂏	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
556	商	1	FDB6	5547	商		
628	繚	1	FDFE			繚	
656	音	3	FE5B			音	
658	疒	5	FE5D			疒	
659	尢	1	FE5E			尢	
660	丂	2	FE5F			丂	
717	疒	7	FEBA			疒	
738	𠂇	18	FECF			𠂇	
745	𠂇	1	FED6			𠂇	
750	𠂇	3	FEDB			𠂇	
758	𠂇	1	FEE3			𠂇	
759	疒	6	FEE4			疒	
761	𠂇	9	FEE6			𠂇	部份使用不同字形「𠂇」，手動修改。
763	𠂇	3	FEE8			𠂇	
766	𠂇	9	FEEB			𠂇	
768	𠂇	7	FEED			𠂇	
770	𠂇	1	FEFF			𠂇	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
771	攵	2	FEF0			攵	
775	亠	10	FEF4			亠	
777	宀	7	FEF6			宀	
780	讠	14	FEF9			讠	
784	亻	5	FEFD			亻	
835	彳	3	8E71	5380	彳		
1077	稟	15	8FE8	7A3E	稟		
1088	暘	1	8FF3	757C	暘		
2461	葱	1	98CB	44E4		𠄎 葱忽	
3796	决	3	815B	51B3	决		
3846	壻	3	81AF	58FB	壻		
3978	猪	1	8274	732A	猪		Unicode 字型 呈現差異。
4035	辵	1	82CF	7A7D	辵		
4038	竈	1	82D2	7AC8	竈		
4184	迹	16	83C7	8FF9	迹		
4242	鬪	18	8442	9B2D	鬪		
4266	鼃	1	845A	9F02	鼃		
4496	段	30	85C5	53DA	段		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4727	𡗗	2	8750	58C4	𡗗		
4996	𡗗	2	88E2	65B5	𡗗		
5289	𡗗	1	8ACD	6164	𡗗		
5411	𡗗	2	8BAA	42AA		米𡗗萬	
5461	𡗗	1	8BDC	6B6B	𡗗		
5677	𡗗	25	8D58	4674		𡗗	

二、手動取代表說明：

《睡虎地秦墓竹簡》(標記)檔案中之舊版造字，因字形部份不同而無法全部以程式取代，皆手動修改成Windows XP能顯示的字或構字式，並製作手動取代表，請參考附件二。欄位說明如下：

1. 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
2. 造字：舊版造字字形。
3. 取代：實際使用字形與舊版造字不同時，於此欄列出。
4. 檔案原文摘錄：摘錄檔案《睡虎地秦墓竹簡》(標記)中的文句段落，其中紅字底線的部分，為需要手動取代的原文。
5. 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。
6. 備註：記錄修改相同辭句的次數或其他事項。

附件二、《睡虎地秦墓竹簡》(標記)手動取代表

編號	造字	取代	檔案原文摘錄	手動取代結果	備註
----	----	----	--------	--------	----

761	欠	欠	<u> </u> (VH1)[+jj]	<u>欠</u> (VH1)[+jj]	有 2 處
			<u> </u> (NI)	<u>欠</u> (NI)	有 2 處
			不(DC) <u> </u>	不(DC) <u>欠</u>	
			負 <u> </u> (VC1)	負 <u>欠</u> (VC1)	
			<u> </u> 以 (VM)[+others]	<u>欠</u> 以 (VM)[+others]	有 2 處