

《新序》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
2010/05/05 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 964 的「𪗇」字，Unicode 編碼是 28EF3，由於 XP 並不支援，仍需使用構字式「卩𪗇焉」。

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：

《新序》使用舊版造字 69 個，字頻 161 次，這 69 個造字中，62 個可轉成 Windows XP 能顯示的字，字頻 145 次；另外 7 個字必須轉成構字式，字頻 16 次。

轉碼完成製作轉碼分析表，請參考附件一《新序》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. 異體字問題：為了使用者查詢和使用的方便，在處理異體字時最主要的方式是以標準字取代，除非是專有名詞或特殊情形，如：人名、地名等。例：造字編號 4134 的「衛」字，是「衛」的異體字，以標準字「衛」取代。
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 4512 的「口𠃉它」字，而原書使用之字為「咤」，查詢教育部異體字字典確認「咤」並非「口𠃉它」之異體字，因此歸為錯字，以程式取代為正確字形「咤」。
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 4773 的「姉」字，Unicode 字型呈現為「姉」，實際上仍為同一字。

附件一、《新序》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
297	𠃉	2	FBED			𠃉	
403	𠃉	2	FCBA			𠃉	
826	癱	1	8E68	7655	癱		
964	𠃉	1	8F55	28EF3		𠃉𠃉馬	
972	犇	2	8F5D	7287	犇		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
981	豺	1	8F66	72B2	豺		
1024	虵	8	8FB3	8675	虵		
1039	稟	2	8FC2	69C0	稟		
1067	碁	1	8FDE	68CA	碁		
1925	羈	1	9568	8989	羈		羈，異體字問題。
1981	讐	13	95C2	8B90	讐		
1999	贊	1	95D4	8CDB	贊		
2152	蕻	1	96D0	979F	蕻		
2162	醜	1	96DA	918E	醜		
2415	鬪	4	987B	95D8	鬪		鬪，異體字問題。
2651	鸚	1	99EC	9E1C	鸚		
2812	舉	2	9AF0	64E7	舉		
2948	聦	3	9BDB	43C2		耳厶叟	
2969	譖	2	9BF0	8B5B	譖		
2997	鹽	5	9C4D	25081		形土宀 口皿	
3789	冲	1	8154	51B2	冲		冲，異體字問題

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
3804	効	1	8163	52B9	効		
3813	邛	6	816C	536D	邛		
3840	坂	2	81A9	5742	坂		
3858	寃	1	81BB	5BC3	寃		寃，異體字問題。
3868	卮	2	81C5	5DF5	卮		
3883	徧	5	81D4	5FA7	徧		
3891	慙	3	81DC	6159	慙		
4012	鼓	1	82B8	76B7	鼓		鼓，異體字問題。
4038	竈	2	82D2	7AC8	竈		
4041	豎	3	82D5	7AEA	豎		
4055	粃	4	82E3	7C83	粃		
4070	縹	2	82F2	7DE4	縹		
4077	罇	1	82F9	7F47	罇		
4085	耻	1	8342	803B	耻		
4089	脇	2	8346	8107	脇		
4094	館	1	834B	8218	館		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4134	衛	1	8373	885E	衛		衛，異體字問題。
4139	袴	1	8378	88B4	袴		
4155	訶	1	83AA	8B0C	訶		
4156	諂	1	83AB	8B1F	諂		
4169	趙	1	83B8	8DA6	趙		
4184	迹	7	83C7	8FF9	迹		
4190	邳	4	83CD	90C4	邳		
4193	鈎	4	83D0	920E	鈎		
4237	髡	8	83FC	9AE0	髡		
4250	鱖	3	844A	9C0C	鱖		
4252	鰓	2	844C			魚𩺰眾	鰓，異體字問題。
4259	麇	1	8453	9E95	麇		
4267	鼈	2	845B	9F08	鼈		
4283	戲	1	846B	622F	戲		
4309	刼	3	84A7	523C	刼		
4502	俎	2	85CB	723C	俎		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4512		1	85D5			口厶它	咤，錯字，以程式全部取代。
4558	厩	3	8644	53A9	厩		厩，異體字問題。
4614	嫌	1	867C	3858		巾厶兼	
4768	妬	4	8779	59AC	妬		
4773	姉	1	877E	59C9	姉		Unicode 字型呈現差異。
4894		5	885A			宀厶心 厶冉	寤，異體字問題。
4904	尅	1	8864	5C05	尅		
4907	橐	1	8867	69D6	橐		
5034	弃	1	8949	5F03	弃		棄，異體字問題。
5289	慤	1	8ACD	6164	慤		敵，錯字，以程式全部取代。
5299	戟	1	8AD7	39B8		卓厶戈	戟，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5355	𠂇	2	8B50	20AE4		𠂇	
5439	栢	1	8BC6	6822	栢		
5541	蠹	1	8C6D	8827	蠹		
5591	杞	2	8CC1	233CC		木𠂇巳	杞，異體字問題。
5702	厨	1	8D71	53A8	厨		

二、手動取代表說明：

《新序》檔案中出現「●」字形，共計2字，比對原書後以Unicode字形或構字式手動取代，並製作「●」字手動取代表，請參考附件二。欄位說明如下：

1. 頁碼：「●」字形所在位置之原書頁碼。
2. 檔案原文摘錄：摘錄《新序》檔案有「●」字形的文句。
3. 原書字形：「●」在原書中之字形，以此字取代檔案原文的「●」。
4. 備註：記錄修改相同詞句的次數或其他事項。

附件二、《新序》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 64	先生老●歟	𠂇 𠂇 昏	
p. 66	麋麋●鹿	麋	