

《說苑》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
2010/05/07 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「𠂔」字，Unicode 編碼是 28EF3，由於 XP 並不支援，仍需使用構字式「𠂔𠂔𠂔」。

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：

《說苑》使用舊版造字 82 個，字頻 255 次，這 82 個造字中，71 個可轉成 Windows XP 能顯示的字，字頻 234 次；另外 11 個字必須轉成構字式，字頻 21 次。

轉碼完成製作轉碼分析表，請參考附件一《說苑》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 5308 的「棹 Δ 戈」字，而原書使用之字為「檝」，查詢教育部異體字字典確認「檝」並非「棹 Δ 戈」之異體字，因此歸為錯字，以程式取代為正確字形「檝」。
 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 3935 的「槩」字，Unicode 字型呈現為「槩」，實際上仍為同一字。
 4. 待造字：Unicode 及漢字構形資料庫皆未收錄的舊漢籍造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 4684 的「禘 Δ 受」。

附件一、《說苑》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
910	嵒	1	8EDE	5D78	嵒		
936	鷓	2	8EF8	4D0A		義 Δ 鳥	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
959	揄	2	8F50	3A49		扌 ㄩ	
964	隅	7	8F55	28EF3		阝 厶 馬	
972	犇	1	8F5D	7287	犇		
982	曾	1	8F67	80F7	曾		
988	芊	1	8F6D	7F8B	芊		
1024	虵	11	8FB3	8675	虵		
1032	憊	1	8FBB	2039D		形 彳 扶 心 心	
1039	稟	2	8FC2	69C0	稟		
1040	聃	2	8FC3	803C	聃		
1053	傲	1	8FD0	6160	傲		
1056	阤	4	8FD3	9641	阤		
1565	繇	6	92F9	7DDC	繇		
1673	芮	1	93C8	34BC	芮		
1857	蚩	2	94E3	86A0	蚩		
1900	桂	1	954F	88BF	桂		
1981	讐	3	95C2	8B90	讐		
2053	躑	1	964B	8EA7	躑		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2449	雒	3	98BF			名 隹	
2598	鄒	1	99B7	9381	鄒		
2610	鰓	2	99C3	9C53	鰓		
2638	鷓	2	99DF	9D55	鷓		
3787	冢	2	8152	51A2	冢		
3801	剗	1	8160	5257	剗		
3804	効	1	8163	52B9	効		
3813	邛	1	816C	536D	邛		
3835	噍	1	81A4	564D	噍		
3846	壻	1	81AF	58FB	壻		
3855	擘	2	81B8	5B7C	擘		
3883	徧	4	81D4	5FA7	徧		
3891	慙	7	81DC	6159	慙		
3894	筋	1	81DF	4225		竹 会 勐	
3924	菴	5	81FD	671E	菴		
3930	挀	1	8244	6901	挀		
3935	槩	1	8249	3BA3	槩		Unicode 字型 呈現差異。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3977	狗	1	8273	72E5	狗		
3996	疎	4	82A8	758E	疎		
4012	鼓	1	82B8	76B7	鼓		鼓，異體字問題。
4031	杭	1	82CB	79D4	杭		
4034	穉	1	82CE	7A49	穉		
4038	竈	3	82D2	7AC8	竈		
4041	豎	1	82D5	7AEA	豎		
4052	簾	1	82E0	7C34	簾		Unicode 字型呈現差異。
4070	縹	2	82F2	7DE4	縹		
4077	罇	1	82F9	7F47	罇		
4089	脇	8	8346	8107	脇		
4109	葛	2	835A	84AD	葛		
4114	藥	1	835F	8616	藥		
4134	衛	66	8373	885E	衛		衛，異體字問題。
4140	衽	4	8379	88B5	衽		
4145	羈	1	837E	898A	羈		Unicode 字型

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							呈現差異。
4146	覩	4	83A1	89A9	覩		
4184	迹	10	83C7	8FF9	迹		
4190	邳	1	83CD	90C4	邳		
4209	鞞	1	83E0	9789	鞞		
4217	鞞	1	83E8	97E0	鞞		
4226	凜	3	83F1	98E1	凜		
4242	鬪	2	8442	9B2D	鬪		
4250	鰩	7	844A	9C0C	鰩		
4256	鷄	1	8450	9DC4	鷄		雞，異體字問題。
4259	麇	1	8453	9E95	麇		
4265	鼃	1	8459	9F03	鼃		
4267	鼃	7	845B	9F08	鼃		
4342	倮	5	84C8	502E	倮		
4460	勅	2	85A1	52D1	勅		
4493	特	1	85C2	3E40		牛 厶 寺	
4519	咲	2	85DC	54B2	咲		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4545	踮	1	85F6	28105		足厶區	
4684		1	86E4			衤厶受	待造字。
4768	妬	5	8779	59AC	妬		
4799	顛	1	87BA	9856	顛		恩厶頁，異體字問題。
4805	瀆	1	87C0	51DF	瀆		
4996	斲	2	88E2	65B5	斲		
5034	弃	1	8949	5F03	弃		棄，異體字問題。
5093	鸛	1	89A6	9D2A	鸛		
5111	旆	1	89B8	65BE	旆		
5308		1	8AE0			棹厶戈	檝，錯字，以程式全部取代。
5381	脆	3	8B6A	8103	脆		脆，異體字問題。
5588	恠	1	8CBE	6061	恠		
5591	杞	1	8CC1	233CC		木厶巳	杞，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5722	朥	2	8DA7	671E	朥		

二、手動取代表說明：

《說苑》檔案中出現「●」字形，共計26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《說苑》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、《說苑》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 98	僕●首於軫	車𠃉頁	
p. 280	幸郎嫪●	毒	
p. 280	●專國事	毒	
p. 280	●懼誅	毒	
p. 280	●敗，始皇	毒	
p. 280	乃取●四肢車	毒	
p. 365	昌州州●州	食𠃉甚	
p. 411	可以為其●矣	木𠃉府	待造字
p. 418	澤鳴、犢●	雝𠃉十	有3處

頁碼	檔案原文摘錄	原書字形	備註
			待造字
p. 436	直●，傍有	堦	
p. 436	曲其●	堦	
p. 436	錄言曲●者	堦	
p. 436	曲●之策	堦	
p. 480	貪欲無●	馱	
p. 504	●穆無窮	目厶勿	
p. 569	●而知之	竹厶選	
p. 586	鳴蜩●●	噤噤	
p. 589	●於言者	賁	
p. 601	巖嵒●嶧	山厶累	
p. 641	鵲食●	狷	
p. 641	●食鵲	狷	
p. 679	●●憧憧	忄厶禺 忄厶禺	