

《說苑》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
2008/2/18 黃榮順 製作

一、《說苑》一書(說苑.xml)使用舊版造字 82 個，字頻 255 次，詳如附件一。這 82 個造字中，69 個可轉成 Windows XP 可顯示的字，字頻 232 次；另外 13 個字必須轉成構字式，字頻 23 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、構字式：Windows XP 若無對應字形，則改採用構字式

壬、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 5308 的「棹𣎵戈」字，原字應為「木𣎵戢」。
- 2、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 4134 的「衛」係「衛」字之異體，故以「衛」字取代。
- 3、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 4145 的「羈」字，Unicode 字型呈現為「羈」，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「𣎵𣎵受」。

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。所以如造字編號 959 的「掄」字，Unicode 編碼為 3A49，由於 Windows XP 並不支援，仍須改用構字式「扌厶翕」。

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、「說苑.xml」發現之未確定字形「●」共 26 處，修正情形詳列附件二，附件二「說苑.xml」「●」修改表體例說明如下：

- 甲、檔案原文摘錄：摘錄檔案(說苑.xml)中包含「●」的文句段落，其中紅字底線的部分，為需要手動取代的原文。
- 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。
- 丙、備註：修改後的字形若屬於「待造字」，則於此欄標示。

附件一、「說苑.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
910	嵒	1	1	8EDE	5D78	嵒		
936	鷁	2	2	8EF8	4D0A		義厶鳥	
959	掄	2	2	8F50	3A49		扌厶翕	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
964	隅	7	7	8F55	28EF3		阝厶焉	
972	犇	1	1	8F5D	7287	犇		
982	胃	1	1	8F67	80F7	胃		
988	𦍋	1	1	8F6D	7F8B	𦍋		
1024	虵	11	11	8FB3	8675	虵		
1032	僣	1	1	8FBB	2039D		𠂇𠂇𠂇心	
1039	𦍋	2	2	8FC2	69C0	𦍋		
1040	𦍋	2	2	8FC3	803C	𦍋		
1053	傲	1	1	8FD0	6160	傲		
1056	阤	4	4	8FD3	9641	阤		
1565	𦍋	6	6	92F9	7DDC	𦍋		
1673	𦍋	1	1	93C8	34BC		𦍋	
1857	𦍋	2	2	94E3	86A0	𦍋		
1900	𦍋	1	1	954F	88BF	𦍋		
1981	𦍋	3	3	95C2	8B90	𦍋		
2053	𦍋	1	1	964B	8EA7	𦍋		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
2449	雒	3	3	98BF			名𪛗隹	雒，異體字問題。
2598	鄧	1	1	99B7	9381	鄧		
2610	鰱	2	2	99C3	9C53	鰱		
2638	鷓	2	2	99DF	9D55	鷓		
3787	豕	2	2	8152	51A2	豕		
3801	剗	1	1	8160	5257	剗		
3804	効	1	1	8163	52B9	効		
3813	邛	1	1	816C	536D	邛		
3835	噍	1	1	81A4	564D	噍		
3846	壻	1	1	81AF	58FB	壻		Unicode 字型差異。
3855	擘	2	2	81B8	5B7C	擘		Unicode 字型差異。
3883	徧	4	4	81D4	5FA7	徧		
3891	慙	7	7	81DC	6159	慙		
3894	筋	1	1	81DF	4225		𦍋𠂔𠂔	
3924	朥	5	5	81FD	671E	朥		
3930	攄	1	1	8244	6901	攄		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3935	槩	1	1	8249	69E9	槩		Unicode 字型差異。
3977	狗	1	1	8273	72E5	狗		
3996	疎	4	4	82A8	758E	疎		
4012	鼓	1	1	82B8	76B7	鼓		
4031	杭	1	1	82CB	79D4	杭		
4034	穉	1	1	82CE	7A49	穉		
4038	竈	3	3	82D2	7AC8	竈		
4041	豎	1	1	82D5	7AEA	豎		
4052	簾	1	1	82E0			形竹虛入	簾，異體字問題。
4070	縹	2	2	82F2	7DE4	縹		
4077	罇	1	1	82F9	7F47	罇		
4089	脇	8	8	8346	8107	脇		
4109	葛	2	2	835A	84AD	葛		
4114	蘖	1	1	835F	8616	蘖		
4134	衛	66	66	8373	885E	衛		衛，異體字問題。
4140	衽	4	4	8379	88B5	衽		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4145	羈	1	1	837E	898A	羈		Unicode 字型差異。
4146	覩	4	4	83A1	89A9	覩		
4184	迹	10	10	83C7	8FF9	迹		
4190	邳	1	1	83CD	90C4	邳		
4209	鞞	1	1	83E0	9789	鞞		
4217	韡	1	1	83E8	97E0	韡		
4226	飡	3	3	83F1	98E1	飡		
4242	鬪	2	2	8442	9B2D	鬪		
4250	鰭	7	7	844A	9C0C	鰭		
4256	鷄	1	1	8450	9DC4	鷄		
4259	麇	1	1	8453	9E95	麇		Unicode 字型差異。
4265	鼃	1	1	8459	9F03	鼃		
4267	鼈	7	7	845B	9F08	鼈		
4342	倮	5	5	84C8	502E	倮		
4460	勅	2	2	85A1	52D1	勅		
4493	特	1	1	85C2	3E40		牛 𠂇 字	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4519	咲	2	2	85DC	54B2	咲		
4545	跣	1	1	85F6	28105		足厶區	
4684		1	1	86E4			衤厶受	待造字。
4768	妬	5	5	8779	59AC	妬		
4799	顙	1	1	87BA	9856	顙		
4805	瀆	1	1	87C0	51DF	瀆		
4996	斲	2	2	88E2	65B5	斲		
5034	弃	1	1	8949	5F03	弃		Unicode 字型差異。
5093	鸛	1	1	89A6	9D2A	鸛		
5111	旆	1	1	89B8	65BE	旆		
5308		1	1	8AE0			棹厶戈	木厶戠，舊漢籍錯字， 可用程式全部取代。
5381	脛	3	3	8B6A	8103	脛		Unicode 字型差異。
5588	悽	1	1	8CBE	6061	悽		
5591	杞	1	1	8CC1	233CC		木厶已	杞，異體字問題。
5722	暮	2	2	8DA7	671E	暮		

附件二、「說苑.xml」「●」修改表

檔案原文摘錄	手動取代結果	備註
僕●首於軫曰	僕 <u>車</u> △ <u>頁</u> 首於軫曰	
幸郎嫪●	幸郎嫪 <u>毒</u>	
●專國事	<u>毒</u> 專國事	
●懼誅，因作亂	<u>毒</u> 懼誅，因作亂	
●敗，始皇	<u>毒</u> 敗，始皇	
乃取●四肢車裂之	乃取 <u>毒</u> 四肢車裂之	
昌州州●州焉	昌州州 <u>食</u> △ <u>甚</u> 州焉	
梧可以為其●矣	梧可以為其 <u>木</u> △ <u>府</u> 矣	待造字
晉有澤鳴、犢●	晉有澤鳴、犢 <u>雥</u> △ <u>士</u>	待造字
召澤鳴、犢●	召澤鳴、犢 <u>雥</u> △ <u>士</u>	待造字
夫澤鳴、犢●	夫澤鳴、犢 <u>雥</u> △ <u>士</u>	待造字
見竈直●，傍有積薪	見竈直 <u>土</u> △ <u>突</u> ，傍有積薪	
曲其●，遠其積薪	曲其 <u>土</u> △ <u>突</u> ，遠其積薪	
反不錄言曲●者	反不錄言曲 <u>土</u> △ <u>突</u> 者	
陛下察客徙薪曲●	陛下察客徙薪曲 <u>土</u> △ <u>突</u>	
尸祿素飡，貪欲無●	尸祿素飡，貪欲無 <u>獸</u>	
感則能動，●穆無窮	感則能動， <u>目</u> △ <u>勿</u> 穆無窮	
智伯廚人亡炙●而知	智伯廚人亡炙 <u>𠂔</u> △ <u>選</u> 而知	
苑彼柳斯，鳴蜩●●	苑彼柳斯，鳴蜩 <u>口</u> △ <u>慧</u> <u>口</u> △ <u>慧</u>	
●於言者，華也	<u>貴</u> 於言者，華也	
夫山龍從●嶧	夫山龍從 <u>山</u> △ <u>累</u> 嶧	
鵲食●	鵲食 <u>犛</u> △ <u>胃</u>	
●食鵲義△鳥	<u>犛</u> △ <u>胃</u> 食鵲義△鳥	
方興未登，●●憧憧	方興未登， <u>忄</u> △ <u>禺</u> <u>忄</u> △ <u>禺</u> 憧憧	