

《韓詩外傳》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
2007/12/26 黃榮順 製作

一、《韓詩外傳》一書(韓詩外傳.xml)使用舊版造字 50 個，字頻 118 次，詳如附件一。這 50 個造字中，42 個可轉成 Windows XP 可顯示的字，字頻 97 次；另外 8 個字必須轉成構字式，字頻 21 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、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 4442 的「潛」字，Unicode 字型呈現為「潛」，實際上仍為同一字。

2、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 2244 的「ノ亾虫」。

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 961 的「枴」字，Unicode 編碼為 3B4A，由於 Windows XP 並不支援，仍須改用構字式「木 𠂇 片」。

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

附件一、「韓詩外傳.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
296	𠂇	3	3	FBEC			𠂇	
297	𠂉	9	9	FBED			𠂉	
298	𠂊	1	1	FBEE			𠂊	
658	𣎵	1	1	FE5D			𣎵	
675	食	1	1	FE6E			食	
826	癰	1	1	8E68	7655	癰		
910	𡇃	1	1	8EDE	5D78	𡇃		
961	枴	4	4	8F52	3B4A		木𠂉片	
982	胷	1	1	8F67	80F7	胷		
1024	虵	1	1	8FB3	8675	虵		
1871	𧔻	1	1	94F1	87C1	𧔻		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1981	讐	1	1	95C2	8B90	讐		
2044	蹠	2	2	9642	8E08	蹠		
2053	蹣	2	2	964B	8EA7	蹣		
2152	鞚	1	1	96D0	979F	鞚		
2244		1	1	976D			ノ会虫	待造字。
2572	銜	1	1	997B	8858	銜		
3804	効	3	3	8163	52B9	効		
3813	卽	2	2	816C	536D	卽		
3850	姪	1	1	81B3	59D9	姪		
3883	徧	3	3	81D4	5FA7	徧		
3891	慙	6	6	81DC	6159	慙		
3929	犂	1	1	8243	68C3	犂		
3930	樟	1	1	8244	6901	樟		
3994	畊	1	1	82A6	754A	畊		
3996	疎	5	5	82A8	758E	疎		
4010	臯	5	5	82B6	7690	臯		Unicode 字型差異。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4035	筭	1	1	82CF	7A7D	筭		
4038	竈	1	1	82D2	7AC8	竈		
4103	蒼	1	1	8354	831D	蒼		
4109	薦	2	2	835A	84AD	薦		
4130	衅	1	1	836F	8845	衅		
4134	衛	20	20	8373	885E	衛		
4140	袵	3	3	8379	88B5	袵		
4167	賚	2	2	83B6	8CEB	賚		
4184	迹	2	2	83C7	8FF9	迹		
4193	鈎	3	3	83D0	920E	鈎		
4242	鬪	3	3	8442	9B2D	鬪		
4267	鼈	4	4	845B	9F08	鼈		
4342	倮	1	1	84C8	502E	倮		
4442	潛	1	1	856D	6F98	潛		Unicode 字型差異。
4519	咷	1	1	85DC	54B2	咷		
4768	妬	1	1	8779	59AC	妬		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4996	斬	1	1	88E2	65B5	斬		
5147	恠	1	1	89DC	6060	恠		
5308		1	1	8AE0			棹戈	待造字。
5482	璽	1	1	8BF1	8812	璽		
5680	憝	1	1	8D5B	6164	憝		
5684	渢	2	2	8D5F	6E12	渢		
5722	朶	4	4	8DA7	671E	朶		