

《大戴禮記》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
2010/04/08 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 4087 的「鯁」字，Unicode 編碼是 4C47，由於 XP 並不支援，仍需使用構字式「魚𩺰旦」。

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：

《大戴禮記》使用舊版造字 55 個，字頻 122 次，這 55 個造字中，39 個可轉成 Windows XP 能顯示的字，字頻 73 次；另外 16 個字必須轉成構字式，字頻 49 次。

轉碼完成製作轉碼分析表，請參考附件一《大戴禮記》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. 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 4012 的「鼓」字，是「鼓」的異體字，以標準字「鼓」取代。

附件一、《大戴禮記》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
296	𠄎	8	FBEC			𠄎	
297	𠄏	18	FBED			𠄏	
298	𠄐	1	FBEE			𠄐	
299	𠄑	5	FBEF			𠄑	
300	𠄒	5	FBF0			𠄒	
436	𠄓	2	FCDB			𠄓	
452	𠄔	1	FCEB			𠄔	
675	𠄕	1	FE6E			𠄕	
707	𠄖	1	FEB0			𠄖	
739	𠄗	1	FED0			𠄗	
775	𠄘	1	FEF4			𠄘	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
972	犇	1	8F5D	7287	犇		
988	芊	1	8F6D	7F8B	芊		
1002	鳩	2	8F7B	9D02	鳩		
1024	虵	2	8FB3	8675	虵		
1053	傲	1	8FD0	6160	傲		
1088	暘	3	8FF3	757C	暘		
1565	絲	2	92F9	7DDC	絲		
1591	羗	4	9354	7F97	羗		
1982	調	2	95C3	8B87	調		
2168	疇	1	96E0	91BB	疇		
2472	雷	1	98D6	9741	雷		
2506	飭	1	98F8	991D	飭		
2610	鰈	1	99C3	9C53	鰈		
3826	叶	1	8179	53F6	叶		
3855	孽	2	81B8	5B7C	孽		
3883	徧	3	81D4	5FA7	徧		
3915	斃	5	8459	6585	斃		
3937	檝	1	824B	6A9D	檝		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3980	獾	2	8276	7328	獾		
4012	鼓	1	82B8	76B7	鼓		鼓，異體字問題。
4040	竝	2	82D4	7ADD	竝		
4087	鯰	1	8344	4C47		魚𩺰旦	
4104	莅	6	8355	8385	莅		
4113	簡	1	835E			艹 叕 間	
4119		2	8364			𠃉 ^艹 正 頁 巳 久 ◻	夔，異體字問題。
4140	衽	2	8379	88B5	衽		
4172	屨	1	83BB	8E77	屨		
4184	迹	4	83C7	8FF9	迹		
4212	鞞	2	83E3	97B8	鞞		
4222	頰	1	83ED	9847	頰		
4242	鬪	3	8442	9B2D	鬪		
4250	鱈	2	844A	9C0C	鱈		
4267	鼈	1	845B	9F08	鼈		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4342	倮	3	84C8	502E	倮		
4455	劒	1	857A			僉△刃	
4816	亻	1	87CB			亻	
4996	斲	1	88E2	65B5	斲		
5108	惛	1	89B5	60FD	惛		
5360	鄺	1	8B55	913D	鄺		
5525	豸	1	8C5D	8C83	豸		
5591	杞	1	8CC1	233CC		木△巳	杞，異體字問題。
5592	桌	1	8CC2	3B9A		桌	
5709	蟀	1	8D78	872F	蟀		
5778		1	8DDF			𠂇△田 △黽	鼉，異體字問題。