

《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

中研院史語所漢籍電子文獻工作小組

2008/1/7 陳建安 製作

一、《大戴禮記》一書(大戴禮記.xml)使用舊版造字 55 個，字頻 122 次，詳如附件一。這 55 個造字中，36 個可轉成 Windows XP 能顯示的字，字頻 69 次；另外 19 個字必須轉成構字式，字頻 53 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、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 5591 的「木𠂔巳」字，原字為「杞」。

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 4119 的「𠂔⁺⁺正頁已久□」係「夔」字之異體，故以「夔」字取代。

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。所以造字編號 4087 的「𩺰」字，Unicode 編碼為 4C47，由於 Windows XP 並不支援，仍須使用構字式「魚𩺰旦」。

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、《大戴禮記》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
296	𡗗	8	8	FBEC			𡗗	
297	𡗘	18	18	FBED			𡗘	
298	𡗙	1	1	FBEE			𡗙	
299	𡗚	5	5	FBEF			𡗚	
300	𡗛	5	5	FBF0			𡗛	
436	𡗜	2	2	FCDB			𡗜	
452	𡗝	1	1	FCEB			𡗝	
675	𡗞	1	1	FE6E			𡗞	
707	𡗟	1	1	FEB0			𡗟	
739	𡗠	1	1	FED0			𡗠	
775	𡗡	1	1	FEF4			𡗡	
972	𡗢	1	1	8F5D	7287	𡗢		
988	𡗣	1	1	8F6D	7F8B	𡗣		
1002	𡗤	2	2	8F7B	9D02	𡗤		

1024	虵	2	2	8FB3	8675	虵		
1053	傲	1	1	8FD0	6160	傲		
1088	暘	3	3	8FF3	757C	暘		
1565	繇	2	2	92F9	7DDC	繇		
1591	羗	4	4	9354	7F97	羗		
1982	調	2	2	95C3	8B87	調		
2168	醕	1	1	96E0	91BB	醕		
2472	雷	1	1	98D6	9741	雷		
2506	飭	1	1	98F8	991D	飭		
2610	鰾	1	1	99C3	9C53	鰾		
3826	叶	1	1	8179	53F6	叶		
3855	擘	2	2	81B8	5B7C	擘		
3883	徧	3	3	81D4	5FA7	徧		
3915	斆	5	5	8459	6585	斆		
3937	檝	1	1	824B	6A9D	檝		
3980	獮	2	2	8276	7328	獮		
4012	鼓	1	1	82B8	76B7	鼓		
4040	竝	2	2	82D4	7ADD	竝		
4087	鰓	1	1	8344	4C47		魚𩺰旦	

4104	莅	6	6	8355	8385	莅		
4113	蒯	1	1	835E			𠂔 𠂔 𠂔 間	
4119		2	2	8364			𠂔 𠂔 正頁已久 𠂔	變，異體字問題。
4140	衽	2	2	8379	88B5	衽		
4172	屨	1	1	83BB	8E77	屨		
4184	迹	4	4	83C7	8FF9	迹		
4212	鞞	2	2	83E3	97B8	鞞		
4222	頤	1	1	83ED	9847	頤		
4242	鬪	3	3	8442	9B2D	鬪		
4250	鰭	2	2	844A	9C0C	鰭		
4267	鼈	1	1	845B	9F08	鼈		
4342	倮	3	3	84C8	502E	倮		
4455	劒	1	1	857A			僉 𠂔 刀	
4816	𠂔	1	1	87CB			𠂔	
4996	斲	1	1	88E2	65B5	斲		
5108	惛	1	1	89B5	60FD	惛		
5360	廩	1	1	8B55	913D	廩		
5525	貊	1	1	8C5D	8C83	貊		
5591	杞	1	1	8CC1	233CC		木 𠂔 巳	杞，校對問題，舊版漢籍錯字，可用

								程式全部取代。
5592	𩑦	1	1	8CC2	3B9A		𩑦	
5709	𧈧	1	1	8D78	872F	𧈧		
5778		1	1	8DDF			𠂇𠂇田𠂇𠂇	𧈧，校對問題，舊版漢籍錯字，可用程式全部取代。