

《儀禮》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
2011/02/18 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 3060 的「揆」字，Unicode 編碼是 3A0E，由於 XP 並不支援，仍需使用構字式「扌𠂔𠂔」。

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：

《儀禮》使用舊版造字 34 個，字頻 372 次，這 34 個造字中，29 個可轉成 Windows XP 能顯示的字，字頻 322 次；另外 5 個字必須轉成構字式，字頻 50 次。

轉碼完成製作轉碼分析表，請參考附件一《儀禮》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 4132 的「血𠂇𠂇」字，是「眾」的異體字，以標準字「眾」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 2981 的「金𠂇盈」字，而原書使用之字為「楹」，查詢教育部異體字字典確認「金𠂇盈」並非「楹」之異體字，因此歸為錯字，以程式取代為正確字形「楹」。
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 4091 的「臯」字，Unicode 字型呈現為「臯」，實際上仍為同一字。

附件一、《儀禮》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
296		32	FBEC			𠂇	
297		1	FBED			𠂇	
777		1	FEF6			𠂇	
848	髻	2	8E7E	9AE4	髻		
1024	虵	1	8FB3	8675	虵		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1077	稟	1	8FE8	7A3E	稟		
1381	穉	1	91DE	7A3A	穉		
2981		1	9BFC			金△盈	楹，錯字，以程式全部取代。
3060	換	15	9CAE	3A0E		扌△奕	
3770	个	43	8141	4E2A	个		
3787	冢	2	8152	51A2	冢		
3815	登	2	816E	537A	登		
3846	墻	13	81AF	58FB	墻		
3883	徧	2	81D4	5FA7	徧		
3918	擅	8	81F7	65DC	擅		
3924	菴	2	81FD	671E	菴		
3930	憵	3	8244	6901	憵		
3949	涖	7	8257	6D96	涖		
3974	牀	13	8270	7240	牀		
3992	嘗	1	82A4	751E	嘗		
4080	羣	6	82FC	7FA3	羣		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4091	臯	1	8348	81EF	臯		Unicode 字型呈現差異。
4104	菘	2	8355	8385	菘		
4132		152	8371			血𠄎𠄎	眾，異體字問題。
4134	衛	2	8373	885E	衛		衛，異體字問題。
4217	韡	9	83E8	97E0	韡		
4268	叢	1	845C	9F17	叢		
4298	廡	36	847A	5EBF	廡		
4804	冪	6	87BF	5E42	冪		
5274	經	1	8ABE	4793		赤𠄎𠄎	
5374	盼	1	8B63	670C	盼		
5456	韎	2	8BD7			韋𠄎未	韎，異體字問題。
5483	眡	1	8BF2	7721	眡		
5505	耆	1	8C49	264BE		老𠄎占	耆，錯字，以程式全部取代。