

《中論》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

2008/2/21 陳建安 製作

一、《中論》一書(中論.xml)使用舊版造字 66 個，字頻 170 次，詳如附件一。這 66 個造字中，59 個可轉成 Windows XP 能顯示的字，字頻 130 次；另外 7 個字必須轉成構字式，字頻 40 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 2191 的「金 Δ 戊」字，原字為「鉞」。
- 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 4132 的「血 Δ 𩇑」係「眾」字之異體，故以「眾」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 3935 的「槩」字，Unicode 字型呈現為「槩」，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 3097 的「彳 Δ 幸」。

三、《中論》中未確定字形「●」共 6 處，詳細處理內容可參考附件二的未確定字形「●」手動取代表。附件二《中論》未確定字形

「●」手動取代表欄位說明如下：

- 甲、 檔案原文摘錄：摘錄檔案(中論.xml)中包含未確定字形「●」的文句段落，其中紅字底線的部分，為需要手動取代的字形。
- 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。

附件一、《中論》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
915	遲	2	2	8EE3	905F	遲		
1067	綦	1	1	8FDE	68CA	綦		
1187	盖	5	5	90B9	76D6	盖		蓋，異體字問題。
1238	戾	1	1	90EC	77E6	戾		
1981	讐	7	7	95C2	8B90	讐		
1999	贊	1	1	95D4	8CDB	贊		
2044	踈	2	2	9642	8E08	踈		
2191	鉞	2	2	96F7			金𠩺戍	鉞，校對問題，舊版漢籍錯字，可用程式全部取代。
2434		1	1	98B0			𠩺𠩺迨	隨，異體字問題。
2516	餽	2	2	9943	9919	餽		
2826	巡	2	2	9AFE	5EF5	巡		
2883	涂	1	1	9B78	51C3	涂		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
2947	聰	1	1	9BDA	8066	聰		
2955	節	1	1	9BE2	83AD	節		
2972	賢	2	2	9BF3	8D12	賢		
3035	賓	7	7	9C73	8CD4	賓		
3096	歲	4	4	9CD2	5D57	歲		
3097		1	1	9CD3			彳 Δ 幸	待造字。
3791	況	17	17	8156	51B5	況		
3796	決	4	4	815B	51B3	決		
3804	効	1	1	8163	52B9	効		
3852	嫵	1	1	81B5	5A63	嫵		
3883	徧	1	1	81D4	5FA7	徧		
3889	慤	1	1	81DA	6164	慤		
3935	槩	1	1	8249	69E9	槩		Unicode 呈現差異。
3944	氷	1	1	8252	6C37	氷		
3961	灾	1	1	8263	707E	灾		
3977	狗	2	2	8273	72E5	狗		
3992	嘗	2	2	82A4	751E	嘗		
3996	疎	1	1	82A8	758E	疎		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4011	皤	1	1	82B7	76A1	皤		Unicode 呈現差異。
4040	竝	2	2	82D4	7ADD	竝		
4080	羣	3	3	82FC	7FA3	羣		Unicode 呈現差異。
4081	翱	1	1	82FD	7FFA	翱		Unicode 呈現差異。
4085	耻	6	6	8342	803B	耻		
4132		30	30	8371			血𠂇𠂇	眾，異體字問題。
4136	衰	1	1	8375	887A	衰		
4140	衽	1	1	8379	88B5	衽		
4144	霸	1	1	837D	8987	霸		Unicode 呈現差異。
4145	羈	2	2	837E	898A	羈		Unicode 呈現差異。
4155	訶	1	1	83AA	8B0C	訶		
4156	諂	1	1	83AB	8B1F	諂		諂，校對問題，舊版漢籍錯字，可用程式全部取代。
4173	𧢲	1	1	83BC	8EAD	𧢲		
4184	迹	3	3	83C7	8FF9	迹		
4193	鈎	1	1	83D0	920E	鈎		
4203	隣	1	1	83DA	96A3	隣		
4220	穎	1	1	83EB	9834	穎		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4250	鱸	1	1	844A	9C0C	鱸		
4400	堯	1	1	8543	5160	堯		
4502	俎	1	1	85CB	723C	俎		
4643	徃	9	9	86BB	5F83	徃		
4723	塲	1	1	874C	58C3	塲		
4827	𦍋	1	1	87D6	3DC0		𠂇火𠂇𠂇	𦍋，異體字問題。
4910		4	4	886A			𠂇 𠂇 𠂇	𦍋，異體字問題。
4996	斲	1	1	88E2	65B5	斲		
5015	甯	2	2	88F5	5BD7	甯		
5034	弃	1	1	8949	5F03	弃		Unicode 呈現差異。
5046	彝	1	1	8955	5F5C	彝		
5119	博	3	3	89C0	613D	博		
5220	𠂇	1	1	8A66	6BCC	𠂇		𠂇，校對問題，舊版漢籍錯字，可用程式全部取代。
5371	冂	3	3	8B60	808E	冂		
5395	規	1	1	8B78	4093		𠂇 𠂇 見	規，異體字問題。
5558	龕	1	1	8C7E	9E84	龕		
5753	潜	1	1	8DC6	6F5C	潜		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5768	潤	3	3	8DD5	6FF6	潤		
5792	籀	1	1	8DED	7C55	籀		

附件二、《中論》未確定字形「●」手動取代表

檔案原文摘錄	手動取代結果
●而成之者	ノ ^ㄉ 地而成之者
孺子之所以●膝	孺子之所以 ^ㄉ 困 ^ㄌ 膝
無名貪●之	無名貪 ^ㄌ 之
相●是非易位	相 ^ㄌ 是非易位
惟顏回●	惟顏回 ^ㄌ 處
大道●矣	大道 ^ㄌ 矣