

# 《新書》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  
2010/05/05 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 4855 的「汨」字，Unicode 編碼是 3CC5，由於 XP 並不支援，仍需使用構字式「汨 厶不」。

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：

《新書》使用舊版造字 57 個，字頻 180 次，這 57 個造字中，47 個可轉成 Windows XP 能顯示的字，字頻 164 次；另外 10 個字必須轉成構字式，字頻 16 次。

轉碼完成製作轉碼分析表，請參考附件一《新書》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 4134 的「衛」字，是「衛」的異體字，以標準字「衛」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 4869 的「嬾」字，而原書使用之字為「𪛗 𪛗靡」，查詢教育部異體字字典確認「𪛗 𪛗靡」並非「嬾」之異體字，因此歸為錯字，以程式取代為正確字形「𪛗 𪛗靡」。
  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 4878 的「宄」字，Unicode 字型呈現為「宄」，實際上仍為同一字。

附件一、《新書》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
841	礩	1	8E77	7932	礩		
890	斐	1	8ECA	71AD	斐		
1024	虵	2	8FB3	8675	虵		
1040	聃	1	8FC3	803C	聃		
1981	讐	5	95C2	8B90	讐		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2036	礪	1	95F9			石 厶 萬	
2103		1	967D			形 之 口 亏 □	還，異體字問題。
2415	鬪	4	987B	95D8	鬪		鬪，異體字問題。
2817	窈	1	9AF5	3AC4		窈	
3791	况	4	8156	51B5	况		况，異體字問題。
3804	効	3	8163	52B9	効		
3835	噍	1	81A4	564D	噍		
3855	擘	4	81B8	5B7C	擘		
3860	尠	1	81BD	5C20	尠		
3883	徧	3	81D4	5FA7	徧		
3891	慙	1	81DC	6159	慙		
3910	擘	1	81EF	64E5	擘		
3924	朞	1	81FD	671E	朞		
3931	椀	1	8245	6900	椀		
3974	牀	1	8270	7240	牀		
3996	疎	5	82A8	758E	疎		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4015	眦	1	82BB	7726	眦		
4040	竝	12	82D4	7ADD	竝		
4045	箒	1	82D9	7B92	箒		
4055	粃	8	82E3	7C83	粃		
4072	縲	1	82F4	7DE5	縲		
4104	莅	1	8355	8385	莅		
4134	衛	27	8373	885E	衛		衛，異體字問題。
4146	覩	1	83A1	89A9	覩		
4155	訶	2	83AA	8B0C	訶		
4172	屨	1	83BB	8E77	屨		
4184	迹	12	83C7	8FF9	迹		
4190	邳	12	83CD	90C4	邳		
4250	鱒	2	844A	9C0C	鱒		
4267	鼈	3	845B	9F08	鼈		
4271	齧	1	845F	9F69	齧		
4371	倜	2	84E5	50F4	倜		
4455	劒	1	857A			僉  刃	
4533	唸	1	85EA	552B	唸		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4855	汩	4	87F2	3CC5		汩 ㄓ 不	
4869	癩	1	8841	5B4A	癩		癩 ㄓ 靡，錯字，以程式全部取代。
4878	宄	1	884A	5B82	宄		Unicode 字型呈現差異。
5034	弃	9	8949	5F03	弃		棄，異體字問題。
5048	婚	3	8957	36F0		女 ㄓ 昏	
5091	鄒	1	89A4	92E3	鄒		
5108	惛	3	89B5	60FD	惛		
5109	櫛	1	89B6	6A10	櫛		
5306	簡	6	8ADE			竹 ㄓ 閒	簡，異體字問題。
5325	踈	2	8AF1	28082		足 ㄓ 炙	
5335	蚤	1	8AFB	459F		亡 ㄓ 虫	
5355	𧈧	1	8B50	20AE4		𧈧	
5371	冃	9	8B60	808E	冃		
5398	滄	1	8B7B	6E7B	滄		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5651	難	1	8CFD	3E10		難 𠂇 𠂈 𠂉	
5682	蠶	1	8D5D	881A	蠶		
5739		2	8DB8			𠂇 𠂈 𠂉	𠂇，異體字問題。
5752	摠	1	8DC5	63D4	摠		

## 二、手動取代表說明：

《新書》檔案中出現「●」字形，共計41字，比對原書後以Unicode字形或構字式手動取代，並製作「●」字手動取代表，請參考附件二。欄位說明如下：

1. 頁碼：「●」字形所在位置之原書頁碼。
2. 檔案原文摘錄：摘錄《新書》檔案有「●」字形的文句。
3. 原書字形：「●」在原書中之字形，以此字取代檔案原文的「●」。
4. 備註：記錄修改相同詞句的次數或其他事項。

### 附件二、《新書》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 4	氓●之人	𠂇 上 示 隶 𠂈	
p. 20	韓王信●胡	𠂇 𠂈 奇	改為倚
p. 27	章而知其勢●	山 𠂇 人 𠂈 子	
p. 29	●過黑	慟	

頁碼	檔案原文摘錄	原書字形	備註
p. 29	●自完足矣	慟	
p. 39	更其名曰●	火△為	
p. 39	襄王出逃伯●	門△糾	
p. 41	以係、縛、●	木△芻	
p. 43	罵詈而●答之	木△芻	
p. 52	禽獸●焉耳	亻△勤	待造字
p. 64	然且吟●	齒△休	
p. 73	又孰敢●然	忄△分	
p. 75	●炙肉	𠃉月者火□	
p. 75	垂●而相告	彳△羨	
p. 78	適●拘也	彳△葦	
p. 80	啗臄●●	夕△肉嘽	
p. 80	●多飲酒	彳△瞽	
p. 84	王●不踏蹴	亻△勤	待造字
p. 95	●問小誦之	儻	
p. 103	●帷而入	扌△擐	
p. 106	過則●	言△間	
p. 113	●牲而	亻△貳	
p. 117	漻然●然	彳△愁	
p. 118	廢首●肘	𠃉亻彳工□	改為低
p. 119	如屋之●	彳△工	待造字

頁碼	檔案原文摘錄	原書字形	備註
p. 135	食●餽	豐△葬	
p. 139	●虵生	虫△兀	
p. 139	食●菹	形 <sup>艹</sup> 禾盧 <sup>口</sup>	待造字
p. 141	●，重裘而立	酉△旦	待造字
p. 149	不●於鄉曲	作△心	
p. 149	不●於朝廷	作△心	
p. 151	嘗試傳白●黑	騰	
p. 155	●·弟敬愛兄	夂△齒	待造字
p. 155	反端為●	跂	
p. 156	義為●	忄△夢	
p. 156	欣●可安	忄△熏	待造字
p. 157	反嚴為●	車△辰	
p. 159	●測之也	角△冏	待造字
p. 168	真德●也	角△冏	待造字
p. 175	猶景之●形	角△冏	待造字