

《人天眼目》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
2010/03/09 丁玟伶

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 2534 的「堊」字，Unicode 編碼是 213BA，由於 XP 並不支援，仍需使用構字式「祝_厶土」。

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：

《人天眼目》使用舊版造字 80 個，字頻 361 次，這 80 個造字中，74 個可轉成 Windows XP 能顯示的字，字頻 341 次；另外 6 個字必須轉成構字式，字頻 20 次。

轉碼完成製作轉碼分析表，請參考附件一《人天眼目》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 4203 的「隣」字，是「鄰」的異體字，以標準字「鄰」取代。
2. 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 2752 的「木𣎵栗」。
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 3978 的「猪」字，Unicode 字型呈現為「猪」，實際上仍為同一字。

附件一、《人天眼目》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
814	墓	1	8E5C	87C7	墓		
841	礪	2	8E77	7932	礪		
887	鷲	1	8EC7	9E0E	鷲		
893	峩	1	8ECD	5CE9	峩		
1151	劄	1	9073	5273	劄		
1536	麤	10	92DC	9E81	麤		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1809	困	1	94B3	56E6	困		
1981	讐	2	95C2	8B90	讐		
1999	贊	1	95D4	8CDB	贊		
2325	睜	1	97E0	77A4	睜		
2386	瑯	5	985E	7458	瑯		
2420	垸	1	98A2	57A7	垸		
2454	鍤	1	98C4	28AFC		金 𠂔 敖	
2534	塋	1	9955	213BA		祝 宀 土	
2598	鄒	3	99B7	9381	鄒		
2719	函	1	9A71	51FE	函		
2752		3	9AB4	235D6		木 𠂔 栗	待造字。
3805	勅	3	8164	52C5	勅		
3814	却	51	816D	5374	却		
3816	厓	2	816F	5393	厓		
3826	叶	6	8179	53F6	叶		
3858	寃	6	81BB	5BC3	寃		寃，異體字問題。
3864	峯	50	81C1	5CEF	峯		
3909	擡	4	81EE	64E1	擡		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3940	鬱	2	824E	6B1D	鬱		
3944	氷	11	8252	6C37	氷		
3955	澗	6	825D	3D4E		澗	澗，異體字問題。
3966	烟	15	8268	70DF	烟		
3977	狗	1	8273	72E5	狗		
3978	猪	2	8274	732A	猪		Unicode 字型呈現差異。
3979	猫	3	8275	732B	猫		
3992	嘗	10	82A4	751E	嘗		
3996	疎	3	82A8	758E	疎		
4013	盃	11	82B9	76CC	盃		
4014	蓋	2	82BA	8462	蓋		
4016	着	2	82BC	7740	着		
4028	稟	6	82C8	7980	稟		
4041	豎	3	82D5	7AEA	豎		
4044	筋	1	82D8	7B6F	筋		
4067	綉	1	82EF	7D89	綉		
4089	脇	1	8346	8107	脇		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4090	脚	22	8347	811A	脚		
4118	藥	1	8363	8602	藥		
4139	袴	2	8378	88B4	袴		
4146	靚	5	83A1	89A9	靚		
4184	迹	4	83C7	8FF9	迹		
4193	鈎	1	83D0	920E	鈎		
4195	鉢	10	83D2	9262	鉢		
4198	鑛	1	83D5	945B	鑛		
4203	隣	2	83DA	96A3	隣		鄰，異體字問題。
4220	穎	8	83EB	9834	穎		
4256	鷄	2	8450	9DC4	鷄		
4267	鼈	1	845B	9F08	鼈		
4278	亘	1	8466	4E98	亘		
4432		1	8563			𠃉 纟 厂 墨 𠃉	纏，異體字問題。
4447	捏	2	8572	63D1	捏		
4515	呪	2	85D8	546A	呪		
4519	咲	2	85DC	54B2	咲		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4643	徃	1	86BB	5F83	徃		
4778	鞞	1	87A5	97B4	鞞		
4816	亻	1	87CB			亻	
4856	疊	1	87F3	3CB2		疊厶毛	
4891	類	13	8857	4AEB		婁厶頁	
4940	韵	1	88AA	97F5	韵		
4957	崑	1	88BB	5D53	崑		
5031	閔	1	8946	95B4	閔		
5072	扃	1	896F	6244	扃		
5163	憂	2	89EC	621E	憂		
5290	携	2	8ACE	643A	携		
5300	鬪	1	8AD8	9B2A	鬪		
5316	晉	1	8AE8	664B	晉		晉，異體字問題。
5360	鄜	1	8B55	913D	鄜		
5453	涅	6	8BD4	23D40		形: 日 工: 工	涅，異體字問題。
5472	瓦	1	8BE7		瓦		
5555	迥	3	8C7B	9008	迥		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5559	欸	4	8CA1	6B35	欸		
5590	朶	5	8CC0	6736	朶		
5742	澁	1	8DBB	6F81	澁		
5753	潛	5	8DC6	6F5C	潛		
5768	濶	3	8DD5	6FF6	濶		

二、手動取代表說明：

《人天眼目》檔案中出現「●」字形，共計13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《人天眼目》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、《人天眼目》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 300	●訛	諺	有 2 處
p. 303	不●無高下	恪	
p. 312	關●子	換	有 4 處
p. 312	眼●●	目𠃉𩺰	

頁碼	檔案原文摘錄	原書字形	備註
		目𩺰𩺱	
p. 313	杖子●跳	足𩺰𩺱	
p. 321	牯●以巧	羿	
p. 330	悲難摸●	扌𩺰𩺱	
p. 333	秦時轆●鑽	車𩺰𩺱	