

《高僧傳》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
2009/06/17 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 888 的「捺」字，Unicode 編碼是 3B88，由於 XP 並不支援，仍需使用構字式「木𠂔奈」。

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：

《高僧傳》使用舊版造字 127 個，字頻 1351 次，這 127 個造字中，112 個可轉成 Windows XP 能顯示的字，字頻 1309 次；另外 15 個字必須轉成構字式，字頻 42 次。

轉碼完成製作轉碼分析表，請參考附件一《高僧傳》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 4561 的「𠃉〃厂曲比[⊙]」字，而原書使用之字為「鹿」，查詢教育部異體字字典確認「𠃉〃厂曲比[⊙]」並非「鹿」之異體字，因此歸為錯字，以程式取代為正確字形「鹿」。
 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。
 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 5637 的「保会言」。

附件一、《高僧傳》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
39	𠃉	2	FA66	27E5E		赤𠃉皮	
888	𠃉	1	8EC8	3B88		木𠃉奈	
972	𠃉	1	8F5D	7287	𠃉		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1067	綦	2	8FDE	68CA	綦		
1133	瓌	3	9061	74CC	瓌		
1238	戾	3	90EC	77E6	戾		
1400	窓	2	91F1	7A93	窓		
1513	糧	1	92C5	7CAE	糧		糧，異體字問題。
1530	紉	1	92D6	7CFA	紉		
1536	麤	3	92DC	9E81	麤		
1544	絇	2	92E4	7D59	絇		
1591	羗	2	9354	7F97	羗		
1708	棉	2	93EB	3B7F		木𠂇昂	
1801		1	94AB			𠂇卜一 八人日 又土𠂇	壑，錯字，以程式全部取代。
1886	榦	1	9541			𠂇車人 木𠂇	
1890	苾	1	9545	44D7		艹𠂇念	
1999	贊	60	95D4	8CDB	贊		贊，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2278	鎗	2	97B1	936E	鎗		
2298	殼	1	97C5	3C89		殼△卵	
2386	瑯	26	985E	7458	瑯		
2524	饒	1	994B	994D	饒		
2631	鷄	2	99D8	9D44	鷄		
2821	悞	1	9AF9	394F		悞△典	
2969	譖	2	9BF0	8B5B	譖		
3153	筓	2	9D4C	7BFC	筓		
3789	冲	4	8154	51B2	冲		沖，異體字問題。
3804	効	10	8163	52B9	効		
3805	勅	110	8164	52C5	勅		
3806	勳	1	8165	52E0	勳		
3813	邛	2	816C	536D	邛		
3814	却	12	816D	5374	却		
3817	廁	2	8170	53A0	廁		廁，異體字問題。
3824	叙	23	8177	53D9	叙		
3829	咏	1	817C	548F	咏		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3840	坂	4	81A9	5742	坂		
3843	堦	1	81AC	5826	堦		
3860	尠	1	81BD	5C20	尠		
3864	峯	15	81C1	5CEF	峯		
3866	崐	1	81C3	5D10	崐		
3877	廼	3	81CE	5EFC	廼		
3900	挖	1	81E5	62D5	挖		
3904	墟	1	81E9	588C	墟		
3920	皓	1	81F9	6667	皓		
3924	菴	1	81FD	671E	菴		
3940	鬱	7	824E	6B1D	鬱		
3944	氷	13	8252	6C37	氷		
3955	澗	19	825D	3D4E		汜 㗎 間	澗，異體字問題。
3966	烟	16	8268	70DF	烟		
3978	猪	3	8274	732A	猪		Unicode 字型呈現差異。
3980	獾	2	8276	7328	獾		
3981	珉	2	8277	73C9	珉		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3992	嘗	4	82A4	751E	嘗		
3996	疎	8	82A8	758E	疎		
4023	礎	2	82C3	78D3	礎		
4028	稟	46	82C8	7980	稟		
4041	豎	5	82D5	7AEA	豎		
4055	粃	1	82E3	7C83	粃		
4089	脇	6	8346	8107	脇		
4090	脚	18	8347	811A	脚		
4104	莅	1	8355	8385	莅		
4107	葱	19	8358	8471	葱		
4129	虬	4	836E	866C	虬		
4131	劔	1	8370	8842	劔		
4134	衛	41	8373	885E	衛		衛，異體字問題。
4138	裘	6	8377	88A0	裘		
4145	羈	1	837E	898A	羈		Unicode 字型呈現差異。
4146	覩	33	83A1	89A9	覩		睹，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4160	適	2	83AF	8B81	適		
4178	輓	1	83C1	8F2D	輓		
4184	迹	86	83C7	8FF9	迹		
4190	邳	10	83CD	90C4	邳		
4195	鉢	44	83D2	9262	鉢		
4203	隣	12	83DA	96A3	隣		
4213	鞞	1	83E4	97BE	鞞		
4214	鞞	1	83E5	97C0	鞞		
4220	穎	25	83EB	9834	穎		
4226	飡	6	83F1	98E1	飡		
4240	鬣	4	8440	9AF4	鬣		
4256	鷄	5	8450	9DC4	鷄		雞，異體字問題。
4262	麪	1	8456	9EAA	麪		Unicode 字型呈現差異。
4278	亘	5	8466	4E98	亘		
4296	堰	1	8478	5878	堰		
4321	嵇	1	84B3	5D47	嵇		
4366	邲	6	84E0	48BC		共𠄎𠄎	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4428	蹶	2	855F	8E70	蹶		
4436	碧	20	8567	40AE		勅 彳 石	
4448		4	8573			𠃉 纟 厂 墨 口	纏，異體字問題。
4458	勸	9	857D	52CC	勸		
4515	呪	69	85D8	546A	呪		
4561		1	8647			𠃉 乚 厂 曲 比 口	麤，錯字，以程式全部取代。
4566	礪	3	864C	78F5	礪		
4575	繼	2	8655	7E7F	繼		
4603	憇	3	8671	6187	憇		
4607	狫	1	8675	3E60		豸 厶 屯	
4625	岐	1	86A9	3A7B		危 厶 支	
4720		1	8749			土 厶 已	圮，異體字問題。
4735	壘	1	8758	58E0	壘		
4739	鞞	1	875C	97DB	鞞		
4856	疊	1	87F3	3CB2		疊 厶 毛	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4961	盼	1	88BF	80B9	盼		
4996	斲	1	88E2	65B5	斲		
5065	忼	1	8968	5FFC	忼		
5074	𠂔	1	8971	5C2B	𠂔		
5111	旆	3	89B8	65BE	旆		
5133	憇	6	89CE	6199	憇		
5188	𠂔	1	8A46	62EA	𠂔		
5225	邨	2	8A6B	460F		血𠂔卩	
5255	涇	3	8AAB	6DB6	涇		
5289	慤	1	8ACD	6164	慤		
5290	携	18	8ACE	643A	携		
5300	鬪	4	8AD8	9B2A	鬪		
5305	晒	1	8ADD	661E	晒		
5316	晋	260	8AE8	664B	晋		晉，異體字問題。
5439	栢	9	8BC6	6822	栢		
5453	涅	65	8BD4	23D40		形: 日 工: 工	涅，異體字問題。
5541	蠹	1	8C6D	8827	蠹		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5555	迥	1	8C7B	9008	迥		
5559	欸	4	8CA1	6B35	欸		
5583	殮	1	8CB9	98F1	殮		
5591	杞	1	8CC1	233CC		木𠂇巳	杞，異體字問題。
5637		1	8CEF			保會言	待造字。
5666	汚	1	8D4D	6C5A	汚		汚，異體字問題。
5678	𡗗	4	8D59	757A	𡗗		
5703	𡗗	1	8D72	9F53	𡗗		
5734		1	8DB3			●	𠂇 ^艹 臣 ^口 ，錯字，以程式全部取代。
5753	潜	50	8DC6	6F5C	潜		
5786	剗	1	8DE7	34F7		臬𠂇𠂇	晡，錯字，以程式全部取代。

二、手動取代表說明：

《高僧傳》檔案中需手動修改的情況分為以下幾種：

(一)《高僧傳》檔案中出現「●」字形，共計 63 字，比對原書後以

Unicod 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《高僧傳》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、《高僧傳》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 327	●下督富整	漲	
p. 331	[二七]●其節	虛△予	待造字
p. 332	眾心●伏莫	快	
p. 333	情無所●	恪	
p. 334	興●耶舍	貝△親	
p. 334	皆重●施	貝△親	
p. 335	隨母傭織●	毘	
p. 335	●為業	毛△登	
p. 336	[一九]●識不	恪	
p. 341	香薪成●	艹△積	
p. 341	●脹蟲爛壞	縫	
p. 345	●施相續	貝△親	
p. 348	●澗飲浪	喰	
p. 350	●錢絹及步輿	貝△親	
p. 350	披●[一九]頓	衿	

頁碼	檔案原文摘錄	原書字形	備註
p. 352	[一九]●倚像	𠂇 𠂇伯	待造字
p. 353	未能[四]●伏	扌 𠂇允	待造字
p. 353	●威蓄銳	扌 𠂇允	待造字
p. 354	月月修書·●遺	貝 𠂇親	
p. 358	●若隱而顯	𠂇火 ㄣ 冂 冂 目 女 冂	待造字
p. 363	授戒·神●	貝 𠂇親	
p. 363	不●過遂多矣	亥 𠂇羽	
p. 364	●流含吐彬蔚	𠂇 𠂇注	
p. 369	敬●遺相續	貝 𠂇親	
p. 372	●會有限名簿素	貝 𠂇親	
p. 379	[五]●幽凝	勑	
p. 380	瑯琊之●山	山 𠂇聶	待造字
p. 380	●山廟巫	山 𠂇聶	待造字
p. 382	以為記[五]●云	翦	
p. 385	而貪●未已	恪	
p. 386	薪●而焚之	𠂇 𠂇積	
p. 390	白[二]土[三]●	土 𠂇泉	
p. 390	賃●·謳唱	土 𠂇泉	
p. 390	引●以自	糸 𠂇乍	

頁碼	檔案原文摘錄	原書字形	備註
p. 390	[一四]●上	土厶象	
p. 390	白土●送與史宗	土厶象	
p. 390	白土●遇	土厶象	
p. 390	及至白土●	土厶象	
p. 391	所得●施迴	貝厶親	
p. 392	進即以為●	貝厶親	
p. 392	●主楊難當	彡	
p. 393	赴●施重疊	貝厶親	
p. 393	西長史劉景●	形 ^彡 麦玉 ^口	
p. 393	數日·●果	形 ^彡 麦玉 ^口	
p. 395	●固無以校也	^彡 厶朋	待造字
p. 396	●錢一萬蜜	貝厶親	
p. 396	秦隱居麥●山	^彡 厶積	
p. 397	羅●苦何由	扌	
p. 397	讒言●疑	扌	
p. 398	王親自●道	扌	
p. 399	薪●始然便有	^彡 厶積	
p. 403	名曇無●	辵	
p. 403	●於東山石室	^彡 厶積	
p. 405	●已[一]恫然	^彡 厶積	
p. 405	●獨何云遠	足厶育厶女	

頁碼	檔案原文摘錄	原書字形	備註
p. 405	見一物如龍從●	𠄎 積	
p. 405	密往●中以火	𠄎 積	
p. 406	則●而弗為	恪	
p. 410	禮拜[一]●塔	𠄎 貝 卩 𠄎	待造字
p. 415	別[三四]●三萬	貝 𠄎 親	
p. 416	●一萬後	貝 𠄎 親	
p. 420	齊瑯琊●山	山 𠄎 聶	待造字
p. 423	●·今以所著	𠄎 墨	待造字

(二)《高僧傳》檔案中，對於舊版造字以外發現的錯字，以手動修改，並製作錯字修改表，請參考附件三。欄位說明如下：

1. 檔案原文摘錄：摘錄檔案《高僧傳》中的文句段落，其中紅字底線的部分，為需要手動取代的原文。
2. 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。
3. 備註：記錄修改相同詞句的次數或其他事項。

附件三、《高僧傳》缺字外錯字修改

檔案原文摘錄	手動取代結果	備註
<u>晡</u> 當降	<u>時</u> 當降	