

《瑞州洞山良价禪師語錄》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
2010/03/11 丁玫伶

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 5315 的「泐」字，Unicode 編碼是 3CC2，由於 XP 並不支援，仍需使用構字式「泐 厶公」。

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：

《瑞州洞山良价禪師語錄》使用舊版造字 23 個，字頻 110 次，這 23 個造字中，22 個可轉成 Windows XP 能顯示的字，字頻 107 次；另外 1 個字必須轉成構字式，字頻 3 次。

轉碼完成製作轉碼分析表，請參考附件一《瑞州洞山良价禪師語錄》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 3955 的「澗」字，是「澗」的異體字，以標準字「澗」取代。

附件一、《瑞州洞山良价禪師語錄》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
814	墓	2	8E5C	87C7	墓		
1400	窓	1	91F1	7A93	窓		
1536	麤	2	92DC	9E81	麤		
2598	鄒	1	99B7	9381	鄒		
3801	剗	1	8160	5257	剗		
3814	却	18	816D	5374	却		
3826	叶	1	8179	53F6	叶		
3864	峯	35	81C1	5CEF	峯		
3944	氷	1	8252	6C37	氷		
3955	澗	1	825D	3D4E		澗	澗，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3966	烟	2	8268	70DF	烟		
3984	瑠	1	827A	7460	瑠		
3992	嘗	1	82A4	751E	嘗		
3996	疎	14	82A8	758E	疎		
4041	豎	2	82D5	7AEA	豎		
4042	笋	1	82D6	7B0B	笋		
4090	脚	7	8347	811A	脚		
4105	菓	6	8356	83D3	菓		
4146	覩	1	83A1	89A9	覩		
4195	鉢	7	83D2	9262	鉢		
4643	徃	1	86BB	5F83	徃		
5300	鬪	1	8AD8	9B2A	鬪		
5315	公	3	8AE7	3CC2		汜 厶公	

二、手動取代表說明：

《瑞州洞山良价禪師語錄》檔案中出現「●」字形，共計2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《瑞州洞山良价禪師語錄》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、《瑞州洞山良价禪師語錄》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 522	這箇●·雲	灑	
p. 522	●·龍牙云	口△力	