《馬可波羅行紀》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
2008/1/7 陳建安 製作

一、 《馬可波羅行紀》一書(馬可波羅行紀.xml)使用舊版造字 1 個，字頻 1 次，詳如附件一，但由於用重碼部件，故不需轉碼。

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、重碼問題，採用部件外字，不轉碼：舊漢籍造字與構字式部件外字，皆是建立在 Big5 的使用者造字區中，兩者所使用的碼區是重疊的，在同時使用構字式和舊漢籍造字的新版漢籍中，這些重疊的字碼可能被用以表達兩個不同字形，我們稱之為「重碼問題」。然而漢籍在舊轉新的過程中，若已將原舊漢籍造字的部分全數改為 Unicode 或構字式，則在 xml 出現的造字頻次皆採用部件外字，此時便不需轉碼。如編號 4278，字碼為 8466，對應的舊漢籍造字為「亘」，部件外字為「扶」，比對原書發現書中皆取部件外字「扶」，故不需轉碼。

三、 《馬可波羅行紀》中未確定字形「❷」共 1 處，詳細處理內容可參考附件二的未確定字形手動取代表。附件二的手動取代表說明如下：

甲、 檔案原文摘錄：摘錄檔案(馬可波羅行紀.xml)中包含未確定字形「❷」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、《馬可波羅行紀》造字分析表

編號	造字	頻次(txt)	頻次(xml)	Big5	Unicode	WinXP	構字式	備註
4278	亘	20	1	8466	4E98	亘		與扶重碼，重碼問題，採用部件外字，不轉碼。

附件二、《馬可波羅行紀》未確定字形手動取代表

檔案原文摘錄	手動取代結果
雙方戰❓	雙方戰鬬