

《元刊雜劇三十種新校(上冊)》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
2009/12/09 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 1691 的「鬚」字，Unicode 編碼是 4BFC，由於 XP 並不支援，仍需使用構字式「鬚𠄎狄」。

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：

《元刊雜劇三十種新校(上冊)》使用舊版造字 33 個，字頻 94 次，這 33 個造字中，23 個可轉成 Windows XP 能顯示的字，字頻 44 次；另外 10 個字必須轉成構字式，字頻 50 次。

轉碼完成製作轉碼分析表，請參考附件一《元刊雜劇三十種新校(上冊)》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. 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 4764 的「擱」字，Unicode 字型呈現為「擱」，實際上仍為同一字。

附件一、《元刊雜劇三十種新校(上冊)》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
63	啐	1	FA7E	54F0	啐		
297	𠄎	36	FBED			𠄎	
298	𠄏	4	FBEE			𠄏	
299	𠄐	1	FBEF			𠄐	
300	𠄑	1	FBF0			𠄑	
746	𠄒	1	FED7			𠄒	
770	𠄓	1	FEEF			𠄓	
899	馳	1	8ED3	99DE	馳		
1341	豨	2	91B6	8C7C	豨		
1491	擱	1	92AF	6527	擱		
1691	鬚	1	93DA	4BFC		鬚彡彡	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2022	漣	5	95EB	703D	漣		
2044	踈	1	9642	8E08	踈		
2206	袂	3	9747	8884	袂		
2346	鐳	3	97F5	9427	鐳		
2432	鞞	1	98AE	29350		革厶登	
2722	吡	1	9A74	5421	吡		
3098	揼	1	9CD4	63F8	揼		Unicode 字型 呈現差異。
3777	侏	8	8148	5008	侏		
3851	姪	2	81B4	59F9	姪		
3902	拊	1	81E7	636C	拊		
4000	疴	1	82AC	75B4	疴		
4074	縑	1	82F6	7E27	縑		
4115	藁	1	8360	85C1	藁		
4129	虬	2	836E	866C	虬		
4195	鉢	1	83D2	9262	鉢		
4764	摠	3	8775	64A7	摠		Unicode 字型 呈現差異。
4778	鞞	1	87A5	97B4	鞞		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4953	𢗗	3	88B7	22835		忄 厶 𢗗	
5008	幪	2	88EE	5E5E	幪		
5074	𠂇	1	8971	5C2B	𠂇		
5075	𠂇	1	8972			𠂇	
5240	擗	1	8A7A	6481	擗		

二、手動取代表說明：

《元刊雜劇三十種新校(上冊)》檔案中需手動修改的情況分為以下幾種：

(一)《元刊雜劇三十種新校(上冊)》檔案中出現「？」字形，共計20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《元刊雜劇三十種新校(上冊)》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、《元刊雜劇三十種新校(上冊)》「？」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 102	這等人粧?處	么	
p. 137	折末你叫??	𠂇𠂇	

頁碼	檔案原文摘錄	原書字形	備註
p. 165	劃馬騎單鞭?	拞	
p. 166	不比?鋼槍槩	衡	
p. 176	相?掇、英布	擻	改為攏
p. 179	和?寬厚	衡	
p. 189	手內?錕鋳	拞	
p. 195	團?得更	拞	
p. 197	分明?志誠	衡	
p. 200	?竹分茶	擻	
p. 201	幾年間?僂	衡	
p. 201	兩只手?撈	衡	
p. 202	花攢柳?	衬	改為襯
p. 212	咱出酸?門	枣	改為棗
p. 214	似拳中?沙	拞	
p. 214	那家，叫??	吡吡	
p. 215	沒個把俺?	係	
p. 215	無人俦?	係	

(二)《元刊雜劇三十種新校(上冊)》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《元刊雜劇三十種新校(上冊)》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。文句後方標示（全部取代）者，表示執行全部取代的動作，不一一列舉

所有原文。

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。文句後方標示（全部取代）者，表示執行全部取代的動作，不一一列舉所有原文。
3. 備註：記錄修改相同詞句的次數或其他事項。

附件三、《元刊雜劇三十種新校(上冊)》缺字外錯字修改

檔案原文摘錄	手動取代結果	備註
<u>么</u> （全部取代）	<u>么</u> （全部取代）	有 32 處
<u>形</u> 口刀血 <u>。</u> 公房	<u>啗</u> 公房	