

《關漢卿戲曲集》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
2009/12/18 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：

《關漢卿戲曲集》使用舊版造字 88 個，字頻 246 次，這 88 個造字中，54 個可轉成 Windows XP 能顯示的字，字頻 126 次；另外 34 個字必須轉成構字式，字頻 120 次。

轉碼完成製作轉碼分析表，請參考附件一《關漢卿戲曲集》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 4012 的「𪗇」字，是「鼓」的異體字，以標準字「鼓」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 4953 的「𪗇」字，而原書使用之字為「𪗇」，查詢教育部異體字字典確認「𪗇」並非「𪗇」之異體字，因此歸為錯字，以程式取代為正確字形「𪗇」。
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 4764 的「𪗇」字，Unicode 字型呈現為「𪗇」，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 4512 的「口𪗇它」。

附件一、《關漢卿戲曲集》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
20	𪗇	2	FA53			𪗇 𪗇 𪗇	
296	𪗇	13	FBEC			𪗇	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
297	厶	35	FBED			厶	
298	厶	12	FBEE			厶	
299	𠂇	2	FBEF			𠂇	
300	𠂇	2	FBF0			𠂇	
304	𠂇	1	FBF4			𠂇	
533	𠂇	1	FD7D			𠂇	
573	𠂇	3	FDC7			𠂇	
595	𠂇	3	FDDD			𠂇	
632	𠂇	1	FE43			𠂇	
639	𠂇	1	FE4A			𠂇	
717	𠂇	1	FEBA			𠂇	
746	𠂇	1	FED7			𠂇	
748	𠂇	1	FED9			𠂇	
752	𠂇	2	FEDD			𠂇	
765	𠂇	1	FEEA			𠂇	
770	𠂇	9	FEEF			𠂇	
772	𠂇	2	FEF1			𠂇	
780	𠂇	1	FEF9			𠂇	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
784	亻	2	FEFD			亻	
972	犇	2	8F5D	7287	犇		
1014	玠	2	8FA9	73CE	玠		
1032	憊	2	8FBB	2039D		形亻 夂 心 心	
1310	蠊	2	9175	45EB		虫 厶 麻	
1394	啗	1	91EB	35D6		口 厶 音	
1482	窳	1	92A6	20816		形穴 頁 卩	
1491	擷	1	92AF	6527	擷		
1634	杼	2	93A1	5FEC	杼		
1642	脉	3	93A9	8109	脉		
1660	賺	1	93BB	8181	賺		
1691	髡	1	93DA	4BFC		髟 彡 狄	
1905	裊	1	9554	88E9	裊		
1981	讐	9	95C2	8B90	讐		
1999	贄	1	95D4			贄	
2021	趁	6	95EA	8D82	趁		
2022	漣	5	95EB	703D	漣		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2044	踈	2	9642	8E08	踈		
2244		3	976D			ノ厶虫	ノ厶虫，異體字問題。
2292		1	97BF			形足〇 火言火 ◻	躩，異體字問題。
2492	頽	1	98EA	983A	頽		
2592	贓	1	99B1	8D1C	贓		
2811		1	9AEF			亞厶田	畱，異體字問題。
2899		1	9BAA			亅厶粉 厶大	待造字。
3004	籊	2	9C54	7BD0	籊		
3019	腩	1	9C63	817C	腩		
3048	叅	7	9CA2	53C5	叅		
3777	俵	2	8148	5008	俵		
3789	冲	1	8154	51B2	冲		冲，異體字問題。
3796	决	1	815B	51B3	决		决，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
3805	勅	1	8164	52C5	勅		
3843	堦	1	81AC	5826	堦		
3870	帑	5	81C7	5E0B	帑		
3886	悞	3	81D7	609E	悞		
3928	糝	1	8242	685A	糝		
3963	炆	1	8265	7081	炆		
4012	鼓	2	82B8	76B7	鼓		鼓，異體字問題。
4038	竈	2	82D2	7AC8	竈		
4042	笋	2	82D6	7B0B	笋		
4067	綉	8	82EF	7D89	綉		
4069	綑	2	82F1	7DB3	綑		
4074	縑	1	82F6	7E27	縑		
4129	虬	1	836E	866C	虬		
4139	袴	2	8378	88B4	袴		
4173	駝	1	83BC	8EAD	駝		
4193	鈎	2	83D0	920E	鈎		鈎，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4226	凜	3	83F1	98E1	凜		
4242	鬪	1	8442	9B2D	鬪		
4267	鼈	1	845B	9F08	鼈		
4455	劔	7	857A			僉 ㄩ 刀	
4496	段	1	85C5	53DA	段		
4512		1	85D5			口 ㄩ 它	待造字。
4515	呪	2	85D8	546A	呪		
4519	咲	2	85DC	54B2	咲		
4764	摠	2	8775	64A7	摠		Unicode 字型 呈現差異。
4778	鞞	4	87A5	97B4	鞞		
4953	敝	2	88B7	22835		忄 ㄩ 敝	忄 ㄩ 敝，錯 字，以程式全 部取代。
5034	弃	3	8949	5F03	弃		棄，異體字問 題。
5101	馱	1	89AE	4B7E		馬 ㄩ 犬	
5120	虛	1	89C1	8657	虛		
5131	窻	2	89CC	7ABB	窻		Unicode 字型

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							呈現差異。
5147	恠	11	89DC	6060	恠		
5306	簡	1	8ADE			竹 ㄟ 間	
5371	冑	1	8B60	808E	冑		
5439	栢	1	8BC6	6822	栢		
5478	飴	2	8BED	4D77	飴		
5558	龕	1	8C7E	9E84	龕		
5583	殮	1	8CB9	98F1	殮		

二、手動取代表說明：

《關漢卿戲曲集》檔案中需手動修改的情況分為以下幾種：

(一)《關漢卿戲曲集》檔案中出現「？」字形，共計5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《關漢卿戲曲集》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、《關漢卿戲曲集》「？」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 274	是誰人??的	吁吁	

頁碼	檔案原文摘錄	原書字形	備註
p. 459	嫉妒的桑辛?	嬾	
p. 812	眼前面?率	堯	
p. 813	?的心疼	堯	
p. 274	是誰人??的	吡吡	

(二)《關漢卿戲曲集》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《關漢卿戲曲集》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、《關漢卿戲曲集》缺字外錯字修改

檔案原文摘錄	手動取代結果	備註
<u>沉</u> 惹得后代人	<u>枉</u> 惹得后代人	
我醜則醜刁 <u>起</u> 決	我醜則醜刁決	
(<u>妳</u>) (全部取代)	(<u>周</u>) (全部取代)	有 33 處
<u>妳</u> 公禱上	<u>周</u> 公禱上	
幾般兒 <u>妳</u> 周正正	幾般兒 <u>周</u> 周正正	
<u>妳</u> 德威 (全部取代)	<u>周</u> 德威 (全部取代)	有 8 處

檔案原文摘錄	手動取代結果	備註
某姓 <u>姝</u> 。名德威	某姓 <u>周</u> 。名德威	
<u>姝</u> 將軍（全部取代）	<u>周</u> 將軍（全部取代）	有 2 處
葛從 <u>姝</u> （全部取代）	葛從 <u>周</u> （全部取代）	有 3 處
非是閑 <u>姝</u> 次	非是閑 <u>周</u> 次	
廡 <u>姝</u> 全	廡 <u>周</u> 全	
記和 <u>姝</u> 易	記和 <u>周</u> 易	
不作 <u>姝</u> 方	不作 <u>周</u> 方	
<u>姝</u> 舍（全部取代）	<u>周</u> 舍（全部取代）	有 34 處
<u>姝</u> 同知	<u>周</u> 同知	
央告你。（ <u>姝</u>	央告你。（ <u>周</u>	
來的敢是 <u>姝</u> 郎	來的敢是 <u>周</u> 郎	
<u>容</u> 。 <u>易</u> 把這女子	<u>容易</u> 。把這女子	
早晚不見 <u>姝</u> 小哥來	早晚不見 <u>周</u> 小哥來	
<u>姝</u> 奪科	<u>周</u> 奪科	
<u>姝</u> 奪休書科	<u>周</u> 奪休書科	
<u>姝</u> 扯二旦云	<u>周</u> 扯二旦云	
得相公 <u>姝</u> 全	得相公 <u>周</u> 全	
<u>叨</u> （全部取代）	<u>仙</u> （全部取代）	有 24 處
<u>搆</u> （全部取代）	<u>感</u> （全部取代）	有 9 處
<u>況</u> 了（全部取代）	<u>枉</u> 了（全部取代）	有 10 處
西子 <u>況</u> 傾城	西子 <u>枉</u> 傾城	

檔案原文摘錄	手動取代結果	備註
<u>況</u> 送了存孝	<u>枉</u> 送了存孝	
階下 <u>況</u> 拴龍駒	階下 <u>枉</u> 拴龍駒	
<u>況</u> 讀了恭儉	<u>枉</u> 讀了恭儉	
<u>況</u> 教你坐黃堂	<u>枉</u> 教你坐黃堂	
冤 <u>況</u> (全部取代)	冤 <u>枉</u> (全部取代)	有 9 處
<u>況</u> 展污了你	<u>枉</u> 展污了你	
<u>況</u> 打壞了	<u>枉</u> 打壞了	
<u>況</u> 作的業	<u>枉</u> 作的業	
無 <u>況</u> 生受	無 <u>枉</u> 生受	
<u>況</u> 誤了良辰美景	<u>枉</u> 誤了良辰美景	
<u>況</u> 著別人	<u>枉</u> 著別人	
<u>況</u> 著你煩惱	<u>枉</u> 著你煩惱	
<u>況</u> 將他氣殺	<u>枉</u> 將他氣殺	
<u>裴</u> (全部取代)	<u>與</u> (全部取代)	有 275 處
<u>炮</u> (全部取代)	<u>烈</u> (全部取代)	有 10 處