

# 《劉知遠諸宮調校注》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  
2009/12/08 丁玟伶

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 1520 的「𠄎」字，Unicode 編碼是 3B20，由於 XP 並不支援，仍需使用構字式「日𠄎𠄎」。

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：

《劉知遠諸宮調校注》使用舊版造字 83 個，字頻 376 次，這 83 個造字中，63 個可轉成 Windows XP 能顯示的字，字頻 298 次；另外 20 個字必須轉成構字式，字頻 78 次。

轉碼完成製作轉碼分析表，請參考附件一《劉知遠諸宮調校注》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 4132 的「血𠂇𠂇」字，是「眾」的異體字，以標準字「眾」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 761 的「久」字，而原書使用之字為「攵」，查詢教育部異體字字典確認「攵」並非「久」之異體字，因此歸為錯字，以程式取代為正確字形「攵」。
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。

附件一、《劉知遠諸宮調校注》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
296	𠂇	12	FBEC			𠂇	
297	𠂇	32	FBED			𠂇	
298	𠂇	3	FBEE			𠂇	
299	𠂇	1	FBEF			𠂇	
300	𠂇	1	FBF0			𠂇	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
379	支	1	FCA2			支	
387	𠂇	1	FCAA			𠂇	
407	𠂈	1	FCBE			𠂈	
527	𠂉	1	FD77			𠂉	
639	𠂊	4	FE4A			𠂊	
658	𠂋	1	FE5D			𠂋	
675	𠂌	1	FE6E			𠂌	
746	𠂍	2	FED7			𠂍	
761	𠂎	1	FEE6			𠂎	𠂎，錯字，以程式全部取代。
766	𠂏	1	FEEB			𠂏	
770	𠂐	5	FEEF			𠂐	
899	馳	1	8ED3	99DE	馳		
972	犇	1	8F5D	7287	犇		
1032	𠂑	1	8FBB	2039D		𠂑 心	
1170	瞳	4	90A8	7583	瞳		
1226	罰	2	90E0	7F78	罰		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1418	𠂇	1	9244	4FF2	𠂇		
1520	𠂇	7	92CC	3B20		日 𠂇 𠂇	
1536	𠂇	3	92DC	9E81	𠂇		
1800	菌	1	94AA	8597	菌		
1823	鬚	1	94C1	9B00	鬚		
1905	𠂇	3	9554	88E9	𠂇		
2408	𠂇	4	9874	9599	𠂇		𠂇，異體字問題。
2505	𠂇	1	98F7	98DC	𠂇		
2520	𠂇	1	9947	993B	𠂇		
3066	𠂇	1	9CB4	519D	𠂇		
3770	个	1	8141	4E2A	个		
3789	冲	3	8154	51B2	冲		冲，異體字問題。
3792	凉	1	8157	51C9	凉		凉，異體字問題。
3797	凭	1	815C	51ED	凭		
3801	𠂇	1	8160	5257	𠂇		
3814	却	31	816D	5374	却		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3820	厦	1	8173	53A6	厦		
3825	叠	2	8178	53E0	叠		
3843	堦	11	81AC	5826	堦		
3846	堦	4	81AF	58FB	堦		
3858	冤	13	81BB	5BC3	冤		冤，異體字問題。
3864	峯	1	81C1	5CEF	峯		
3869	币	2	81C6	5E00	币		
3872	鵜	1	81C9	4CF5		勅△鳥	
3876	廻	8	81CD	5EFB	廻		
3884	忽	5	81D5	6031	忽		
3886	悞	2	81D7	609E	悞		
3899	担	1	81E4	62C5	担		
3909	擡	8	81EE	64E1	擡		
3931	椀	3	8245	6900	椀		
3978	猪	1	8274	732A	猪		Unicode 字型呈現差異。
3991	攜	3	82A3	64D5	攜		
3996	疎	3	82A8	758E	疎		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4016	着	64	82BC	7740	着		
4028	稟	4	82C8	7980	稟		
4038	竈	1	82D2	7AC8	竈		
4042	笋	1	82D6	7B0B	笋		
4065	紮	3	82ED	7D25	紮		
4067	綉	9	82EF	7D89	綉		
4080	羣	3	82FC	7FA3	羣		
4090	脚	7	8347	811A	脚		
4132		11	8371			血𠂇𠂇	眾，異體字問題。
4139	袴	1	8378	88B4	袴		
4146	覩	3	83A1	89A9	覩		
4184	迹	6	83C7	8FF9	迹		
4195	鉢	1	83D2	9262	鉢		
4242	鬪	9	8442	9B2D	鬪		
4256	鷄	1	8450	9DC4	鷄		
4307	罵	8	84A5	99E1	罵		罵，異體字問題。
4507	吳	4	85D0	5434	吳		吳，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
4513	昫	1	85D6	7717	昫		
4557	醣	1	8643	288A9		酉 𪛗 易	
4716	湜	2	8745	57FF	湜		
4738	睺	1	875B	777A	睺		
4768	妬	1	8779	59AC	妬		
5034	弃	2	8949	5F03	弃		
5176		8	89F9			氵 𪛗 冗	沉，異體字問題。
5290	携	1	8ACE	643A	携		
5482	璽	1	8BF1	8812	璽		
5558	龕	1	8C7E	9E84	龕		
5590	朶	9	8CC0	6736	朶		
5621	泪	5	8CDF	6CEA	泪		