

《元代白話碑集錄》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
2009/11/24 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 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：

《元代白話碑集錄》使用舊版造字 39 個，字頻 307 次，這 39 個造字中，36 個可轉成 Windows XP 能顯示的字，字頻 301 次；另外 3 個字必須轉成構字式，字頻 6 次。

轉碼完成製作轉碼分析表，請參考附件一《元代白話碑集錄》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 4134 的「衛」字，是「衛」的異體字，以標準字「衛」取代。
 2. 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、《元代白話碑集錄》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
64	颺	1	FAA1	98B7	颺		
1079	總	12	8FEA	7DEB	總		
1513	糧	26	92C5	7CAE	糧		
1520	嚟	1	92CC	3B20		日 𠂔 煞	
1760	洩	1	9460	6DAD	洩		
1800	菌	22	94AA	8597	菌		
2252	乱	1	9775	4E71	乱		
3048	叅	2	9CA2	53C5	叅		
3770	个	3	8141	4E2A	个		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3789	冲	5	8154	51B2	冲		
3814	却	13	816D	5374	却		
3864	峯	7	81C1	5CEF	峯		
3883	徧	1	81D4	5FA7	徧		
3978	猪	7	8274	732A	猪		Unicode 字型 呈現差異。
4016	着	110	82BC	7740	着		
4036	窰	1	82D0	7AB0	窰		
4090	脚	1	8347	811A	脚		
4104	莅	1	8355	8385	莅		
4105	菓	5	8356	83D3	菓		
4134	衛	2	8373	885E	衛		衛，異體字問 題。
4146	覩	1	83A1	89A9	覩		
4167	賈	6	83B6	8CEB	賈		
4184	迹	2	83C7	8FF9	迹		
4195	鉢	1	83D2	9262	鉢		
4256	鷄	7	8450	9DC4	鷄		
4263	麩	3	8457	9EAF	麩		Unicode 字型

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							呈現差異。
4309	刼	1	84A7	523C	刼		
4419	璦	2	8556	24A8C		王𠂇璦	
4447	捏	5	8572	63D1	捏		
4817	寶	6	87CC	5BF3	寶		
5034	弃	1	8949	5F03	弃		
5104	礼	1	89B1	793C	礼		
5437	并	31	8BC4	5E77	并		
5439	栢	6	8BC6	6822	栢		
5472	瓦	3	8BE7		瓦		
5565	斛	3	8CA7	469F		角𠂇寻	
5569	壇	2	8CAB	74AE	壇		
5591	杞	2	8CC1	233CC		木𠂇巳	杞，異體字問題。
5707	棧	2	8D76	6830	棧		

二、手動取代表說明：

《元代白話碑集錄》檔案中出現「●」字形，共計 97 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《元代白話碑集錄》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、《元代白話碑集錄》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 9	和尚住持●峪山寺	山厶共	待造字
p. 9	俺每交●皇帝	与	
p. 13	瓦匠●人	久厶三	待造字
p. 13	來歷不明底●人	酉厶寻	待造字
p. 14	每却補●數	与	有 2 處
p. 15	道●朱哥那衍	与	
p. 15	仰依理應副●者	与	
p. 29	頭先生每●了退	与	
p. 29	別了在先●例	身厶本	待造字
p. 29	在前三教●例	身厶本	
p. 29	前●例安置	身厶本	有 2 處
p. 30	做沒●例勾	身厶本	
p. 31	東至古●直北灣	疒厶土	待造字
p. 31	充常住●寺地土	貝厶扇	待造字
p. 31	太子●秋無窮者	迂	

頁碼	檔案原文摘錄	原書字形	備註
p. 32	官●哥玉都	𠂔	
p. 32	並所屬●田	疒△土	
p. 33	大殿內總制院官●哥	𠂔	
p. 33	依奉●哥	𠂔	
p. 35	聖旨●例	身△𠂔	有 19 處
p. 35	聖旨●了	与	有 2 處
p. 35	稅糧休●者	与	
p. 35	無●例勾當	身△𠂔	有 8 處
p. 37	旨●例	身△𠂔	
p. 37	●者。但屬寺家	与	有 2 處
p. 39	沒●例的事	身△𠂔	
p. 40	沒●例的勾當	身△𠂔	
p. 42	演胡都●幹魯	疒△刁	
p. 43	沒●例勾當	身△𠂔	
p. 44	吉祥●把行	执	
p. 46	無●例的勾當	身△𠂔	有 5 處
p. 47	稅休●者	与	
p. 47	沒●例勾當	身△𠂔	
p. 49	依●例坐地	身△𠂔	
p. 49	法旨●來	与	
p. 49	無●例做呵	身△𠂔	

頁碼	檔案原文摘錄	原書字形	備註
p. 49	不依●例勾當	身△本	
p. 51	差發休交●	形、ノ目□	有 2 處
p. 51	差發休●	形、ノ目□	有 2 處
p. 51	每根底，●	与	
p. 51	稅休●者	形、ノ目□	
p. 51	觀裏的●田	疒△土	有 2 處
p. 52	●例裏	身△本	
p. 53	●伐樹	石△斤	
p. 54	子壹拾●疋	叁	
p. 55	經文●例	身△本	
p. 56	商稅休●者	形、ノ目□	
p. 56	觀裏的●田	疒△土	
p. 57	懿旨●了也	与	
p. 58	大●例裏	身△本	
p. 63	●把着行的聖旨	执	
p. 70	休●者	与	
p. 74	聖旨體例裏●者	与	
p. 74	印令旨●了	与	
p. 76	好生的●擾	形扌文見□	待造字
p. 89	聖旨●了也	与	
p. 90	用●錢物脩補	形ノ冊卩□	

頁碼	檔案原文摘錄	原書字形	備註
p. 93	交用●	𠄎ノ冊卩𠄎	
p. 94	遣●出寺者	赶	
p. 97	係宮觀●田	疒△土	
p. 121	郝德溫 郝●	𠄎身ㄣ王尔貝𠄎	改為贛