

《晏子春秋》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
2010/04/13 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 4855 的「汜」字，Unicode 編碼是 3CC5，由於 XP 並不支援，仍需使用構字式「汆 厶不」。

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：

《晏子春秋》使用舊版造字 46 個，字頻 156 次，這 46 個造字中，43 個可轉成 Windows XP 能顯示的字，字頻 152 次；另外 3 個字必須轉成構字式，字頻 4 次。

轉碼完成製作轉碼分析表，請參考附件一《晏子春秋》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 3858 的「寃」字，是「冤」的異體字，以標準字「冤」取代。

附件一、《晏子春秋》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
842	𣪠	1	8E78	3AB7		𣪠	
852	貉	1	8EA4	72E2	貉		
957	遨	1	8F4E	906C	遨		
972	犇	5	8F5D	7287	犇		
994	楚	7	8F73	8E75	楚		
1039	稟	3	8FC2	69C0	稟		
1668	臍	1	93C3	81C8	臍		
1871	𧈧	2	94F1	87C1	𧈧		
1981	讐	2	95C2	8B90	讐		
3858	寃	3	81BB	5BC3	寃		冤，異體字問題。
3883	徧	3	81D4	5FA7	徧		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3891	慙	9	81DC	6159	慙		
3921	皙	1	81FA	6673	皙		
3924	菴	1	81FD	671E	菴		
3930	攄	1	8244	6901	攄		
3994	畊	5	82A6	754A	畊		
4014	蓋	1	82BA	8462	蓋		
4029	稷	1	82C9	799D	稷		
4040	竝	6	82D4	7ADD	竝		
4043	筭	1	82D7	7B53	筭		
4077	罇	4	82F9	7F47	罇		
4080	羣	20	82FC	7FA3	羣		
4104	莅	13	8355	8385	莅		
4132		28	8371			血𠂇丞	眾，異體字問題。
4139	袴	1	8378	88B4	袴		
4146	覩	1	83A1	89A9	覩		
4156	諂	2	83AB	8B1F	諂		
4172	屨	1	83BB	8E77	屨		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4227	鈇	1	83F2	98E6	鈇		
4239	髻	2	83FE	9AE5	髻		
4242	鬪	6	8442	9B2D	鬪		
4252	鰓	2	844C			魚厶眾	鰓，異體字問題。
4267	鼈	1	845B	9F08	鼈		
4342	倮	1	84C8	502E	倮		
4455	劒	1	857A			僉厶刃	
4460	勅	4	85A1	52D1	勅		
4532		1	85E9			口厶占	囧，異體字問題。
4855	汩	2	87F2	3CC5		汨厶不	
5169	危	1	89F2	6239	危		
5310	烏	1	8AE2	8204	烏		
5437	并	2	8BC4	5E77	并		
5447	躡	1	8BCE	8E83	躡		
5572	埠	1	8CAE	57BE	埠		
5588	悽	1	8CBE	6061	悽		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5666	汚	1	8D4D	6C5A	汚		汚，異體字問題。
5739		2	8DB8			冗𠂇古	𠂇，異體字問題。

二、手動取代表說明：

《晏子春秋》檔案中出現「●」字形，共計21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《晏子春秋》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、《晏子春秋》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 63	若何●●去	汙 𠂇 汙 𠂇	
p. 77	孽星在●	𠂇	
p. 128	隅●之削	𠂇	
p. 150	涕●而去	洩	
p. 165	接一搏●	豕 𠂇 肩	

頁碼	檔案原文摘錄	原書字形	備註
p. 268	其●久乎	竟	
p. 353	睹●冠反裘	敝△大	
p. 388	臣受命●邑	敝△大	
p. 397	●車	敝△大	有 5 處
p. 407	力●矣	敝△大	
p. 407	死近矣·●其	敝△大	
p. 413	衣食●薄	敝△大	有 2 處
p. 457	景公●于路	宿	
p. 458	身老，子●	子△禹	
p. 458	門外宇●之下	溜	