《世說新語箋疏》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 2009/01/17 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 1032 的「恁」字,Unicode 編碼是 2039D,由於 XP 並不支援,仍需使用構字式「厖/ 扶心[·]。

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

表一、Unicode 的字數及編碼區段

一、舊版造字轉碼分析:

《世說新語箋疏》使用舊版造字 154 個,字頻 1197 次,這 154 個造字中,126 個可轉成 Windows XP 能顯示的字,字頻 1119 次;另外 28 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《世說新語箋疏》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:
- 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「衞」字,是「衛」的異體字,以標準字「決」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 3011 的「土▲空」字,而原書使用之字為「巾▲軍」,因此「土▲空」為錯字,以程式取代為正確字形「巾▲軍」。
- 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 3935 的「槩」字,Unicode 字型呈現為「槩」,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 1723的「**会泣」。

附件一、《世說新語箋疏》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
792	帬	2	8E46	5E2C	帬		
835	厀	10	8E71	5380	厀		
852	狢	1	8EA4	72E2	狢		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
893	峩	2	8ECD	5CE9	峩		
977	麐	1	8F62	9E8F	麐		
982	匈月	6	8F67	80F7	匈月		胸,異體字問題。
1032	倦	1	8FBB	2039D		厖(扶 心①	
1048	頟	1	8FCB	981F	額		
1066	頳	1	8FDD	4AC7		須▲冄	
1067	棊	30	8FDE	68CA	棊		
1097	峄	2	8FFC	37ED		山∭昬	
1133	瓌	3	9061	74CC	瓌		
1334	眇	1	91AF	26548		耳∭少	
1350	褚	1	91BF	7987	褚		褚,異體字問題。
1383	穅	2	91E0	7A45	穅		
1463	攞	1	9271	651E	攞		
1530	糺	4	92D6	7CFA	糺		
1544	絙	1	92E4	7D59	絙		
1565	緜	2	92F9	7DDC	緜		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1642	脉	1	93A9	8109	脉		
1660	膁	1	93BB	8181	膁		
1723		1	93FA			**会泣	待造字。
1886	榦	2	9541			形卓 △ 木①	
1906	裦	2	9555	88E6	裦		
1917	襥	2	9560	8965	襥		
1981	雙言	8	95C2	8B90	雙言		雠, 異體字問 題。
2044	踈	5	9642	8E08	疎		
2159	新	3	96D7	4D35		麥▲并	
2192	毹	1	96F8	23C05		翕▲毛	
2212	糴	2	974D	277CC		衤▲羅	
2485	艖	3	98E3	4C79		魚▲差	
2524	饍	1	994B	994D	饍		
2741		4	9AA9			麻▲心	待造字。
2811		8	9AEF			亞 会田	雷,異體字問題。
3011		1	9C5B			土瓜穵	巾▲軍,錯

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							字,以程式全部取代。
3780	儁	54	814B	5101	儁		
3787	冢	9	8152	51A2	冢		
3796	决	1	815B	51B3	决		決,異體字問題。
3805	勅	3	8164	52C5	勅		
3806	勠	1	8165	52E0	勠		
3807	騐	1	8166	9A10	騐		
3809	匳	1	8168	5333	匳		
3813	邛	1	816C	536D	卭		
3817	厠	1	8170	53A0	厠		廁,異體字問 題。
3819	斯	1	8172	53AE	厮		廝,異體字問 題。
3824	叙	4	8177	53D9	叙		
3840	坂	2	81A9	5742	坂		
3846	壻	21	81AF	58FB	壻		
3855	孼	1	81B8	5B7C			

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3858	寃	1	81BB	5BC3	寃		
3864	峯	7	81C1	5CEF	峯		
3868	巵	1	81C5	5DF5	巵		
3869	币	2	81C6	5E00	币		
3883	徧	1	81D4	5FA7	徧		
3891	慙	11	81DC	6159	慙		
3915	斅	1	8459	6585	斅		
3921	哲	5	81FA	6673	哲		
3924	朞	4	81FD	671E	朞		
3935	既六	3	8249	3BA3	旣		Unicode 字型 呈現差異。
3955	澗	1	825D	3D4E		氵▲閒	
3974	牀	48	8270	7240	牀		
3981	珉	8	8277	73C9	珉		
3984	瑠	3	827A	7460	瑠		
3996	疎	17	82A8	758E	疎		
4013	盌	14	82B9	76CC	盌		
4028	禀	2	82C8	7980	京		
4034	穉	18	82CE	7A49	稺		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4035	穽	2	82CF	7A7D	穽		
4040	竝	49	82D4	7ADD	竝		
4041	竪	2	82D5	7AEA	竪		
4044	筯	1	82D8	7B6F	筯		
4063	蘇	1	82EB	7CF5	薛来		
4066	絝	3	82EE	7D5D	絝		
4067	綉	1	82EF	7D89	綉		
4080	君	95	82FC	7FA3	羣		
4085	耻	2	8342	803B	耻		
4090	脚	1	8347	811A	脚		
4095	舩	1	834C	8229	舩		
4104	莅	1	8355	8385	莅		
4190		101	0071			, A 4.	眾,異體字問
4132		101	8371			血会瓜	題。
4134	衞	87	8373	885E	衞		衛,異體字問
1101	1111			0001	1451		題。
4139	袴	1	8378	88B4	袴		
4140	袵	1	8379	88B5	袵		
4146	覩	12	83A1	89A9	覩		睹,異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
4158	譌	1	83AD	8B4C	譌		Unicode 字型 呈現差異。
4173	躭	2	83BC	8EAD	躭		
4176	輙	2	83BF	8F19	輙		
4184	迹	20	83C7	8FF9	迹		
4185	逷	1	83C8	9037	逷		
4186	遡	2	83C9	9061	遡		
4190	郄	2	83CD	90C4	郄		
4195	鉢	2	83D2	9262	鉢		
4203	隣	2	83DA	96A3	隣		鄰,異體字問題。
4206	橡	1	83DD	3BD3		木▲翁	
4222	顇	4	83ED	9847	顇		
4224	頿	1	83EF	983F	頿		
4225	顦	1	83F0	9866	顦		
4226	飡	1	83F1	98E1	飡		
4238	髢	1	83FD	9AE2	髢		
4242	鬭	9	8442	9B2D	斷		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4256	鶏	2	8450	9DC4	鶏		雞,異體字問題。
4260	麞	2	8454	9E9E	麞		
4307	黑	1	84A5	99E1	馬		罵,異體字問題。
4365	制魚	1	84DF	4C65		制会魚	
4369	晫	5	84E3	666B	晫		
4436	䂮	1	8567	40AE		勃会 石	
4455	劒	9	857A			会 ∭刃	
4460	勑	3	85A1	52D1	勑		
4507	吴	214	85D0	5434	吴		吳,異體字問 題。
4546	鵶	1	85F7	9D76	鵶		
4567	紿	1	864D	25E4B		米∭台	
4579	儯	1	8659	347A		イ 孤隽	
4607	蓰	4	8675	3E60		犭▲屯	
4642	惲	12	86BA	3853		巾盔軍	
4643	徃	1	86BB	5F83	徃		
4699	暅	1	86F3	6685	暅		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4702	場	1	86F6	5872	場		
4752	懕	1	8769	61D5	懕		Unicode 字型 呈現差異。
4768	妬	3	8779	59AC	妬		妒 ,異體字問題。
4870	旅	2	8842	97F2	亦		
4878	宂	1	884A	5B82	宂		Unicode 字型 呈現差異。
4889	発	1	8855	3CAA		登▲毛	
4902	兖	3	8862	5156	兖		
4961	肹	1	88BF	80B9	肹		
4996	斵	1	88E2	65B5	斵		
5048	婚	7	8957	36F0		女∭昏	
5065	忼	7	8968	5FFC	忼		
5074	尫	1	8971	5C2B	尫		
5116	伷	3	89BD	4F37	伷		
5119	博	1	89C0	613D	博		
5131	窻	2	89CC	7ABB	窗		
5162	崼	1	89EB	37EA		山▲農	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5164	肈	2	89ED	8088	肈		
5192	隟	10	8A4A	969F	隟		
5241	我鳥	6	8A7B	9D5E	我爲		
5251	左	7	8AA7	53B7	左		
5289	慤	1	8ACD	6164	慤		
5347	慽	3	8B48	617D	慽		
5392	忽	4	8B75	60A4	愈		
5398	湻	2	8B7B	6E7B	湻		
5413	雝	2	8BAC	26265		雝会缶	
5434	湼	3	8BC1	6E7C	湼		
5439	栢	1	8BC6	6822	栢		
5455	厀	4	8BD6	48DB		泰▲阝	
5464	煙	5	8BDF	7157	煗		
5585	寓	5	8CBB	3762		☆会禹	
5587	歘	2	8CBD	6B58	歘		
5697	煑	1	8D6C	7151	煑		
5701		1	8D70				編,錯字,以 程式全部取 代。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5705	廐	1	8D74	5ED0	廐		廄,異體字問 題。
5706	敍	53	8D75	654D	敍		叙 ,異體字問 題。
5711	騪	5	8D7A	2998E		馬▲整	
5799	剡	1	8DF4	262FA		^四 会 剡	
5860	无	5	C6D3	65E0	无		

二、手動取代表說明:

《世說新語箋疏》檔案中出現「●」字形,共計 26 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《世說新語箋疏》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、《世說新語箋疏》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 1	以綿漬酒中,●乾	暴	
p. 32	二百里●山中	山瓜印	

頁碼	檔案原文摘錄	原書字形	備註
p. 64	躡●腳足	馬▲殳	待造字
p. 86	其水●渫而揚波	氵 ▲ 甲	
p. 136	住剡東●山	山瓜卯	
p. 203	擣●噉鐵杵	- 一	
p. 257	改「望」為「●」	厖扌隹冏⊙	
p. 374	與從弟●等選精	王瓜炎	
p. 485	人問王長史江●兄弟	桃▲虍	
p. 485	●及弟淳	桃▲虍	
p. 536	但恐狐狸●狢噉盡	猯	
p. 580	外孫●臼	軟会韭	
p. 580	字為好・●白	軟会韭	
p. 616	約疑亮●遺詔	囮穴亻 愛⊙	待造字
p. 620	鬢如反●皮	猬	
p. 626	便自有●處山澤	郵 穴 € •	待造字
p. 636	以粲貌●通脫	®穴亻 曼•	待造字
p. 648	聞上●然有聲	口瓜酋	待造字
p. 654	常獨●	郵 穴亻曼∙	待造字
p. 696	傾筐倒●	广▲支	
p. 760	王●已見	☆会 會	
p. 837	輕在角●	角▲弱	
p. 841	●布單衣		

頁碼	檔案原文摘錄	原書字形	備註
p. 880	恒冬天得韭蓱●	囮牙次韭∙	改為1887 中央非 ①
p. 880	韭蓱●是搗韭根	囮 中次韭∙	
p. 915	魚蝦●未可致	魚瓜羌	待造字