

《上海博物館藏戰國楚竹書(七)》出土造字轉碼說明

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
2009/10/05 丁致伶

轉碼主要工作是把檔案中的語言所出土造字轉換成 Windows XP 能支援的 Unicode 字形和構字式，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 1 的「洵」字，Unicode 編碼是 23CE0，由於 XP 並不支援，需使用構字式「; 𠄎母」。

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、出土造字轉碼分析：

《上海博物館藏戰國楚竹書(七)》使用出土造字 180 個，字頻 751 次，這 180 個造字中，3 個可轉成 Windows XP 能顯示的字，字頻 7 次；另外 175 個字必須轉成構字式，字頻 741 次，2 個字必須轉成「&+造字內碼」，字頻 3 次。

轉碼完成製作轉碼分析表，請參考附件一《上海博物館藏戰國楚竹書(七)》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：出土造字。
- (三) 頻次：出土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**：檔案內的出土造字與 Unicode 或漢字構形資料庫所收字形，僅僅只有部件位置或異寫的差異，並無改變該字字義，以 Unicode 字形或漢字構形資料庫所收字形的構字式取代。例：造字編號 920 的「陳△土」字，是漢字構形資料庫所收字形「墜」的異體字，僅僅只有部件位置或異寫的差異，並無改變該字字義，因此以漢字構形資料庫所收字形的構字式「陳△土」取代。
2. **部份使用不同字形，手動修改**：檔案中為同一出土造字，校對原書時卻發現對應到兩個以上不同的字形，依書將一部份手動修改，剩餘的頻次則用程式取代成對應的 Unicode 或構字式。例：出土造字編號 4283 的「隸」字頻次為 2，原書卻對應到「隸」和「征△示」兩種異體字字形，依書將 1 字手動修改為「征△示」，另 1 字用程式轉為「隸」的構字式「正示」。手動修改部份請參考附件二：《上海博物館藏戰國楚竹書(七)》手動取代表。
3. **待造字**：Unicode 及漢字構形資料庫皆未收錄的出土造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 2365 的「知于日」。
4. **古漢字缺字**：檔案內的出土造字為漢字構形資料庫未收錄的古漢字，因不確定其對應的現代楷書字形，所以無法使用風格碼表達，暫轉為「&」符號加該字在 Big5 使用者造字區的內碼表示。例：編號 4309 為出土造字中漢字構形資料庫未收錄的古漢字，而該字在 Big5 使用者造字區的內碼為 84A7，暫轉為「&84A7」表示。
5. **符號使用問題，以統一符號取代**：出土造字字形中的「」符號，代表該部件難以辨識，與之前出土文獻用來代表部件不清或部件

缺符號「□」相異，為求統一，以「□」符號取代。例：造字編號 4355 的「𠂔△□」，字形中的「□」符號代表該部件難以辨識，與之前出土文獻代表部件不清符號「□」相異，為求統一，以「□」符號取代，改構字式為「𠂔△□」。

附件一、《上海博物館藏戰國楚竹書(七)》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1	洵	5	FA40	23CE0		洵△母	
12	息	6	FA4B			身△心	
19	熨	6	FA52			既△火	
47	悚	2	FA6E	3939		忄△求	
48	壘	6	FA6F			來△止	
53	訶	4	FA74			言△訶	
79	遊	10	FAB0			形△从 羊△	
87	𠂔	29	FAB8			𠂔△壬	
95	陞	10	FAC0			陀△土	
136	𠂔	1	FAE9			形△立自 欠白△	
138	𠂔	2	FAEB	65F9	𠂔		
141	𠂔	17	FAEE	2659E		昏△耳	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
156	酒	28	FAFD	287FB		酒	
182	会	3	FB58	4F8C	会		
195	复	4	FB65			复	
214	筭	19	FB78	421E		竹 会 冂	
224	迨	9	FBA4			辶 厶 己	
232	惹	2	FBAC			艹 会 女 会 心	
257	穿	1	FBC5			穿	
259	勑	14	FBC7			勑	
265	恁	1	FBCD			化 会 心	
278	戒	2	FBDA			戒	
288	装	2	FBE4			装	
300	忍	2	FBF0			川 会 心	
316	募	7	FC41			亠 会 分	
368	瀉	1	FC75	23FD5		氵 厶 廌	
387	审	2	FCAA			宀 会 中	
404	豉	1	FCBB	4734		豆 厶 支	
453	昇	2	FCEC			日 会 丌	
478	經	1	FD46	2601A		纟 厶 呈	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
479	笈	4	FD47			竹 厶 石	
500	糶	2	FD5C	23A4D		歹 厶 麥	
515	所	2	FD6B			所 厶 人	
570	梟	2	FDC4			白 厶 木	
589	紕	2	FDD7	25FC7		糸 厶 疋	
605	散	2	FDE7	22F38		散	
805	尿	4	8E53			尿	
862	會	8	8EAE	2332D		會	
865	槩	4	8EB1			明 厶 示	
920		3	8EE8			陳 厶 土	陳 厶 土，異體字問題。
930	逡	9	8EF2			辵 厶 復	
937	兪	1	8EF9			兪 厶 止	
997	𠄎	10	8F76			𠄎	
1174	吕	60	90AC			吕	
1180	苜	2	90B2			艹 厶 囧	
1186	虞	5	90B8			盧 厶 又	
1189	閑	1	90BB	28CFF		閑	
1197	聿	2	90C3	26614		聿	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1227	稟	2	90E1			高厶京	
1230	忒	1	90E4	225B6		工厶心	
1236	=	42	90EA			=	
1244	巜	1	90F2	386D		巜	
1246	戩	2	90F4	210E3		罍厶戈	
1334	袷	1	91AF	2566D		衤厶奎	
1352	禹	1	91C1			五厶禹	
1353	易	2	91C2	661C	易		
1461	尋	38	926F			貝厶寸	
1475	痲	1	927D			疒厶因	
1600	霤	2	935D	290A9		霤	
1646	軛	6	93AD	282A0		軛	
1781	伎	1	9475	36A2		亻厶女	
1805	虐	5	94AF			虍厶口	
1855	舍	1	94E1			余厶口	
1859	悟	4	94E5			忄厶舌	
1890	聾	5	9545			聾厶昏 耳	
1960	系	60	95AD			心厶系	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1964	瞳	1	95B1			貝厶童	
2019	赫	5	95E8	233DF		赫	
2268	象	19	97A7			心厶家	
2270	逡	8	97A9	284A5		辵厶夂	
2277	余	1	97B0			余厶ノ	
2279	懋	2	97B2	22613		母厶心	
2287	逵	1	97BA			辵厶圭	
2298	逵	2	97C5			辵厶真	
2324	宀	1	97DF	219CA		宀	
2326	慝	1	97E1	3951		再厶心	
2365		2	9849			形知于 日	待造字。
2367	司	6	984B			台厶司	
2389	墜	1	9861			阝厶土	
2395	嚙	2	9867	21139		形品受 幺	
2462	希	1	98CC			采厶巾	
2507	惹	3	98F9	2276C		者厶心	
2533	囡	1	9954	361D		囡	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2579	槐	12	99A4	29CA1		衤 ㄩ 鬼	
2581	昇	1	99A6	22331		与 厶 升	
2602	𧨧	1	99BB			𠂇 舟 頁 止 口	
2621	𠂇	4	99CE			凡 厶 口	
2622	𠂇	8	99CF			止 厶 所	
2631	𠂇	1	99D8			木 ㄩ 堂	
2669	堂	1	99FE			尚 厶 立	
2682	𠂇	1	9A4C			允 厶 火 厶 土	
2740	隆	3	9AA8			降 厶 止	
2773	羿	1	9AC9			羽 厶 廾	
3841		4	81AA			育 ㄩ 女	待造字。
3923	𠂇	1	81FC	41C3		立 ㄩ 巳	
3946	𠂇	2	8254	21858		𠂇 幸 卂 女 口	
3979	穀	2	8275	3745		穀 ㄩ 子	
4271	註	1	845F			言 ㄩ 土	
4273	𠂇	1	8461			𠂇 ㄩ 見	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
						目隹◻	
4274	莖	2	8462			形祈口 立◻	
4275	託	1	8463			冒厶毛	
4276	柚	1	8464			木厶曲	
4277	箎	2	8465			竹厶軍	
4278	覷	1	8466			形目心 見◻	
4279	宿	1	8467			形宀亻 日◻	
4281	訢	1	8469			形土斤 言斤◻	
4282	冢	1	846A	3CBE		宀厶水	
4283	隼	2	846B			形亻正 示◻	部份使用不同字形「征厶示」，手動修改。
4284		1	846C			征厶示	待造字。
4285	梃	1	846D			形木口 七◻	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4286	辱	2	846E			厂△母	
4287	呈	1	846F			尸△土	
4288	送	2	8470	284AB		辶△羊	
4289	輦	1	8471			外△車	
4290	覲	4	8472			堇△見	
4291	怠	2	8473			佻△心	
4293	忍	2	8475	3902		及△心	
4294	保	2	8476	20203		亻△禾	
4295	憊	1	8477			忄△龠	
4296	慵	1	8478	22931		忄△龍	
4297	憾	2	8479			臧△心	
4300	眷	2	847C			关△旨	
4301	緝	2	847D			纟△共	
4302	數	2	847E			形△炎土 女△	
4303	呈	2	84A1	2125D		丂△土	
4304	沅	2	84A2			〇〇元	
4306	仄	1	84A4	201A3		一△人	
4308	戩	2	84A6			囂△戈	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4309		2	84A7			&84A7	古漢字缺字。
4310	𠂔	2	84A8			七会十	
4311		2	84A9			𠂔 丿 丿 早 口	待造字。
4312		2	84AA			𠂔 冫 么 大 口	待造字。
4313	愬	2	84AB			𠂔 言 台 习 心 口	
4315	𠂔	2	84AD			角 会 大	
4316	誨	2	84AE			言 厶 母	
4317	𠂔	2	84AF			〇 矢	
4318	童	2	84B0			帝 会 止	
4319	窆	2	84B1			穴 会 圣	
4320	进	3	84B2			辶 会 升	
4321	𠂔	2	84B3			卯 会 飮	
4322	餐	1	84B4	296C1		卯 会 食	
4323	𠂔	2	84B5			𠂔 羽 厂 占 口	
4324	豸	1	84B6			豸 会 口	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4325	颺	2	84B7	295BD		風△皮	
4326	戡	24	84B8			形少音 戈□	
4327	曼	2	84B9			民△ △又	
4328	練	2	84BA			糸△東	
4329	敷	2	84BB			形白大 女□	
4330	捺	6	84BC	22B91		扌△彖	
4331	減	1	84BD			彳△戍	
4332		1	84BE			&84BE	古漢字缺字。
4333	憇	2	84BF			佻△心	
4334	斲	3	84C0			形臣口 又力□	
4335	省	2	84C1			少△口	
4336	均	2	84C2			公△勻	
4337	具	2	84C3			口△具	
4339	慙	1	84C5			形十言 斤心□	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4340		1	84C6			𠄎羊一 企◻	待造字。
4341	窳	1	84C7			𠄎穴告 火◻	
4342	慙	1	84C8			新㊦心	
4343	刑	1	84C9			刑㊦日	
4344		1	84CA			車㊦冕	待造字。
4346	戡	1	84CC			齊㊦戈	
4348	窞	1	84CE			宀㊦圭	
4349	啟	1	84CF			皮㊦攴	
4350	晷	1	84D0			日㊦日	
4351	戡	1	84D1			𠄎日 日 戈◻	
4355		1	84D5			𠄎◻	𠄎◻，符號使用問題，以統一符號取代。 待造字。
4356	整	1	84D6			爻㊦止	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4359	徠	1	84D9			彳厶袁	
4360	鑿	1	84DA			金厶盤	
4361		1	84DB			形彳丰 土□	待造字。
4362	靈	1	84DC	290D2		雨厶垂	
4363		1	84DD			畏厶心	待造字。
4364		1	84DE			來厶女	待造字。
4365		1	84DF			爰厶攵	待造字。
4366		1	84E0			形彳保 貝□	待造字。
5733	恙	2	8DB2			羊厶心	
5746	音	1	8DBF			立厶中 厶日	
5930	鄔	2	C75A			鄔	

二、手動取代表說明：

《上海博物館藏戰國楚竹書(七)》檔案中之出土造字，因字形部份不同而無法全部以程式取代，皆手動修改成Windows XP能顯示的字或構字式，並製作手動取代表，請參考附件二。欄位說明如下：

1. 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
2. 造字：出土造字字形。

3. 取代：實際使用字形與出土造字不同時，於此欄列出。
4. 檔案原文摘錄：摘錄檔案《上海博物館藏戰國楚竹書(七)》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。
5. 備註：記錄修改相同辭句的次數或其他事項。

附件二、《上海博物館藏戰國楚竹書(七)》手動取代表

編號	造字	取代	檔案原文摘錄	手動取代結果	備註
4283	隸	征 _△ 示	<u>形</u> 正示 _□ (懲) 酒 (將) 大	<u>征</u> _△ 示 (懲) 酒 (將) 大	