N3448 L2/08-175 Page 1

ISO/IEC JTC 1/SC 2/WG 2 **N3448**

DOC TYPE: Working Group Document

TITLE: Comments on N3297: Proposal to encode Tangut characters in UCS Plane 1 and Charts

SOURCE: UK National Body STATUS: Member contribution ACTION: For consideration by WG2 DISTRIBUTION: ISO/IEC JTC1/SC2/WG2

DATE: 2008-04-19

The UK welcomes the proposal to encode the Tangut script, and wishes to express its appreciation to Dr. Richard Cook for a detailed and well thought out proposal.

However, it is important that a proposal of this magnitude is carefully reviewed by experts in the field of Tangut studies, and that there is broad consensus on the character repertoire and ordering. We have been in contact with a number of Tangut scholars, and whilst we have had some positive feedback, we have been told that more time is required to thoroughly review the proposal and check all of the nearly 6,000 proposed characters. We would therefore not wish to see the proposal added to an amendment at this stage.

One issue in particular that has been raised is the question of Tangut radicals. Almost all the modern Tangut dictionaries include a radical-based index of Tangut characters. Some examples are appended to this document:

- Appendix A: Table of 357 radicals used in Li Fanwen's Xia-Han Zidian 夏漢字典 (Beijing, 1997)
- Appendix B: Table of 444 radicals used in Wenhai Yanjiu 文海研究 (Beijing, 1983)
- Appendix C: Table of 329 radicals used in Tatsuo Nishida's Seikabun Kojiten 西夏文小字典 (Tokyo, 1964)

We believe that it is essential that the Tangut proposal includes a set of radicals, so that radical indices such as these can be represented in electronic text. Note that unlike Han radicals, the majority of Tangut radicals defined in these dictionaries do not exist as free-form characters and so cannot be represented unless they are separately encoded. The actual radical scheme used varies from author to author, so we strongly recommend that the set of Tangut radical characters encoded is a superset of all the radicals used in the various dictionaries.

Furthermore, in addition to the actual radicals used in dictionaries there are a relatively small number of other recurring elements found in Tangut characters that it would be advisable to encode (these mostly occur as the right hand side component of characters, but have not been used as radicals in dictionaries as most Tangut radicals are artificially chosen from the topmost or left side component of a character). It would be very useful to encode all component elements used in Tangut characters, not just those used as radicals, for use in discussion of the composition of Tangut characters (something that is very important to scholars of the Tangut script) and for use in component-based input methods for the Tangut script.

Finally, we would like the proposal to address Tangut character strokes. Most Tangut strokes are the same as strokes used for Han ideographs, but there may be some strokes peculiar to Tangut that it would be beneficial to encode (we suggest adding any such strokes to the space at the end of the CJK Strokes block, 31E4..31EF). For example, Li Fanwen's *Tangut-Chinese Dictionary* shows 24 basic strokes:

护		和代號:			محم	bril	説明
性筆	名	號碼	筆	E 形	字	例	説 明
單	横	1		با -	双 双	。益。他。	横和横上下鈎
	垂	2	1	1 18	馒 飯	漁服	直和撇
1.第二	- 點	3	,	\ \\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\	屏談	继°乘。	點和捺
	叉	4	×	1 ++	3猜°崩	。赏飲	
復	+++	5	1	‡ ***	浅 "拼	。解。溝	一筆横穿三筆,直穿二至三筆
	4111	6	1	##	藤。世		一筆横、直穿四筆
	角	7	F	7 \$	。 麗。		一筆轉折和兩筆相接而成的角
筆	元	8	~/	^	% 新		類似漢字八和八的變形
	小	9		III	°h °稍		類似漢字小和小的變形

Appendix A: Table of radicals used in Li Fanwen's *Xia-Han Zidian* 夏漢字典 (Beijing, 1997)

		(一) 筆畫	部首	
筆		部	首	
畫	、點	一横	豎	〕撒
一畫		-171	11	1343
一畫	上474	=5=5√6 ₹6 另7	月7月9119	<u>3</u> 9 111

Ī			T	
	十11	〒16 〒16 〒16 〒16	F20 月20 中20 早21	第18月21月21分21
三	715	孝17剂17月17月17		久21冬22廿22
		ま17三18尹19土19		
畫	·	Z ₁₉ ++20 Z ₂₀ = 20		
	_		4	
1		~		
-	菜 22 前 22 1 22	7 25 7 26 7 26 7 26		羊37长39们37介37
一四	Ĭ1 23 1 23 Ť 24 ¥ 24	□ 26 □ 27 甬 27 季 27	# 37	Ŷ38 ≯38 ※ 38 ※ 38
	× 24 第 24 ₹ 24 ₹ 25	年27	110,	130 ¥30 X30 X38 X38
	72.72.42.42.	マッランフランフ ス 29 ス 29 ス 29		X 44 X 44 X 45 41 45
畫		‡31平31±32 = 32	:	- -
-		₹31 + 31 − 32 ∃ 32 ₹33 + 33		
		24 33 111 33	·	
4	支 45 前 45 前 46 参 46	平49	F E F #	\$ 1/2 W 2
	第46 第46 第46 第 46		序55 尾55 序55 序55	灸55 爻57 ≪58 炙58
	1 1 1 1 1 1 1 1 1 2 2 3 4 5 6 6 7 8 8 9 8 9 8 9 9 9 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 <	1 '' '	Ø 55 F 55	乳58 削59 ₹ 59 ₹ 59
五	第47 第47 第47 第47	月50月50月50月50 月50早50早50 月 51		羊59 并59 羊59 ፋ59
	美47 ** 48 第 48 	市51 及51 录51 录51		•
	第 49	〒51 × 51 × 51 × 51 → 51 第51 章 51 × 51 前 51		
	7 49 	カ51 井 51 ダ 51 川 51		
		和51 和51 加51 对52 前52 年52 年52 第52		•
畫		刑 52 片 52 片 52 刊 52 升 52	·	
上				
		≨ 53 ₩53 ☎53 ₩53		
		₩53 ₩53 ₩54		
	+ + + +	H		& & 3 S
	章60年60年60年60	#65 #65 #65 7 65	库67年67前67见67	
	〒60 亥61 反 61 仟 61	F65 第65 第65 第65	Q 67	第68
六	 	₹65 ₹65 ₹65 ₹65 ₹ = # #		青68年68
ŀ	菜61菜61菜63菜64 生生生生	亥66 而66		
#	常64	五66月66月66月66		
畫	₹ 65	集研算研集研奏研		
		事的良的萧约册的		

京の前の前の前の 第四章の首の前の 第四章の第四章の 第四章の第四章の	弄71 弄71 异71 麦71 麦71 【71 甭71 麦71		第73
	麥71飞71甭71爱71		
并60 Kco 菜60 1 60			
HOND LOS CO	茂71 希71 7 71 7 71		
羊的羊的羊的羊的	弄71 弄71 季71 麥71		
前的黄的黄的年70	页72 薛72 奉72 斋72		
告70	葡 72 夏 72 辛 72 青 72		
	莽72 		
	清73		
		夏75 年75	
		-	
	岁75 	1.	
	•		
育74 麥74 麥74			
法 \	R 南 夏 夏		
	±		14.*).* :
贝76 萬76 聚76	发76 夏76 叕76 與76		
	青70 第70 第71 青74 夏74 第74 艾 74 支 74 第74 第74 爻 74	青70 第70 第71	青70 第70 第71

Appendix B: Table of radicals used in Wenhai Yanjiu 文海研究 (Beijing, 1983)

一部首表										
	ما	679	J.	681	4	685	¥	689	玄	691
₹	674 工	(70	भुं	681		685	半計	689	X Z	691
	674 1	019	爻	681	灵术	i	₹] •¥	600	全	691
1	674	679	泛犹礼		ドイ	685	多計	689	リス	691
)	674 T 675 F			681	们工	685	Í	689	多多	691
¥		679	有	681	到	6 85	计	690	英美丽和新闻	
3 ,	cac &	679	オ	681	有看夏麦猪	68 5	4	l	凡	691
٧	676	679	*	682	支	686	芝	690	并开	691
7	676	679	*	682		686	芨	690	Ħ	692
ے	676	680 680	\$	682	F	686	支	690	戶戶	692
	676		3	682	F	686	才	690	Ħ	692
13	676	680	Ť	682	F	686	茅	690	빵	692
ージャンフラーアライエ	676 676 676 676	680 680	书事丰秀而	682	及	686	ř	690	-9997	692
几	676	680	#	682	13	686	关支方方言并交为都都并交及多并而并开	690	刻	692
- 17	676 H	680		682	if	686	支	690	冬	692
)]	676	680	并而反	682	文	68 6	1/3	690	全	692
11	677	680	一员	682	冬	686	新	690	新	692
~ 1	677 冬	680	787	683	交参外在半有介	686	浙	690	到冬季新新新爱	692
-7	1	680	47	684	七	686	并	690	金	692
من لا	677 久 678 有	681	亥	684	半	686	多	690	多	692
کے	678 3	681		684	名	686	泛	691	1	692
ļ	678	681	全	684	俞	686	多	691	早	692
1	1 /		五	685	11	686	5 第	691	多	692
して	678	681 681	王	685	3	68 68	5 第	691	큐	692
す	678 兰 679 芝	681 681	1	685	灰		b	691	青	692
Ţ	017 2	681 681	3	685	14	68	6 4	691	甭	693
ツンナナイモモテるす	679 直679 首	681	及年本尹王九司	685		68	6 4	691	1	693

六七一

美	693	并	696	菱	698	反	699	溪	700	量	701
A A	693	并光	6 96	孩弄	698	看	699	新	700	是	701
天子子子而子	693		696	華	698	風	699	及刻本多年科	700	三月三月三天三开三川三三子子丁	701
H H	693	并	696	夏	698		699	1	700	孟	701
圣	693	并	696	乔	698	秉	699	益	700	THE STATE OF THE S	701
华	693	及	696	弄	698	备	699	***	700	星	701
7	693	孝	696	桑	698	Ž	699	泽	700	专	
*	693	3	697	\$	698	锋	699	7	700	圣	701 701
差	693	公主	697	新弄杂及	698	夏 青春奏後年	699	及	700	力	701
A	693	方	697	冬	698	有	699	名	700	本	701
亥本存存夜本月月	69 3	泛多手支并并下沒	697	受到	698	14	699	往次段及又多文音三方三年三十五十五八百十五人四人世一天一天文文	700 700	范凌凌凌春	701
Z	693	崖	697	新	698	4	699		700	圣	701
性	6 93	市	697	1	698	役祭弟	699	及艺	700	名	701
E	693	方	697	素	698	美	699	事	700	A	701
岸	694		697	產	698	并	6 9 9	圣	700	女	701
孙	694	泛麥麦	697	清	698	毛汗流光	699	3	700	老	701
	694	支	697	麦	69 8	半	699	区	700	开	702
4	694		697	言	698		699	英	700	芨	702
外	694	半	697	華	698	话	699	窗	700	秦	702
年并公会下父女	694	岁 光	697	辛多有三川 支三口三井二日	698	并沒產	699	**	700	不佳老为本开与开交行各交当交运机三年出	702
雀	694	并	697	垂	698	杀	699	奉	700	釜	702
9	694		697	古	698	計	699	序	700	量	702
3	694	谷	697	夜季	698	美	699	交	701	孟	702
各	695	存	697	奉	698	三美弄新美	699	夜	701	拔	702
Ž	695	有	697	再	698	淋	699	泵	701	并	702
刻	695	有	697	排	699	菱	700	甭	701	姜	
多	695	1/4	697	[3]	699	存	700	幸	701	多	702 702
X	695	3	698	泽	699	存	700	麦	701	1	702
N. A.	695	多	698	岸	699 699	岸	700	萧	701	差	702
年 美利 委 芳 市 新	6 96	何很多家家	698	奉存箱到月月 月	699	يد	700	夜都南手麦部哥	701	旅游景交交茶等有	702
***		!		•		ı		1		ŀ	

七二

星	702	背	702	煮	702	胃	703	夏素愛到	7,03	盘	704
并当	702	茅	702	各	703	莉	703	案	703		704
国	702	ツツ	702	艺	703	君	703	蒙	703	预	704
及		女	702	松当	703	温	703	桑	703	答	704
多之	702	多	702	竹炭	703	多	703	湖	703	夜	704
3	702 702	3	702	当	703	丹子	703	Á	703	4	704
3	702	2	702	满	703	女	703	一關百四次於門本年是所養一百八百五百五百五百五百五百五百五五百五五五五五五五五五五五五五五五五五五五五	703	版話	704
首立		의		姜	703	*	703	對	7 03	秀	704
肾	702 702	考	702	英	703	西	703	翟	703	*	704
Ţ		新	702 7 02	浅	703	交	703	孟	703	湖	704
伊	702	利	702	冷	703	并	703	文	703	龙	704
俊	702	有五	702	13	703	孟	703	玄	704	河	704
B	702	支	702 702	为	703	足	703	五	704	殿	704
H	702	九	702	が	703	女	703	季	704	級	704
3	702	4	702	督	703	世当	703	玄	704	流	704
多	702	名	702	子	703	3			704	郊冬	704
当并当的主义是美国文文文学的一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个	702	学校奏奏奏到手新到前安竟不要有交流员	702	罪各月法役員并後遇過過天教及後後都一般夏日多多常	703	·新夏八百日至至夏季为班交奉中奏五夏日报文世出图章·新开文世	703	養級愛	704	班版城自夜暴五版旗演者,曹震殿。殿。旅流就奏奏教	704
贵	702	孙		多世	703	對	703	級	704	一般	104
图	702	奚	702	新	100	益	703	菱		1	

N3448

Appendix C: Table of radicals used in Tatsuo Nishida's Seikabun Kojiten 西夏文小字典 (Tokyo, 1964)

	_	. ~	^	字小字典目》		
001	309	028	4	344	055	· 36
002	Γ310	029		346	056	∄ 36-
003	<i>-</i> 311	030	外	346	057	‡ ····· 369
004	Ī 311	031	1	346	058	¥ 36
005	ے۔ 312	032	久	347	059	4 369
006	<u> </u>	033	処	347	060	≨ 360
007	<i></i> 313	034	z.	348	061	详 360
800	ີ້ 319	035	丞	348	062	≇36
009	= 320	036	2	348	063	₹ 36°
010	4 322	037	٠.,	348	064	‡ 36
011	 324	038	ñ	349	065	- -
012	夫 324	039	J	349	066	₹ 36
013	<u>s</u> 324	040	J	353	067	≢ 36′
014	<u>e</u> 324	041	Ĭį	353	068	F 367
015	£ 324	042	ij,	359	069	∮ ····· 368
016	±	043	1	359	070	∮ 368
017	324	044	1	359	071	* ····· 369
810	₃₃₄	045	省	359	072	孝 369
019	* 337	046	7	360	073	‡ 369
020	338	047	1	360	074	‡ 370
021	339	048	7	361	074b	菲 370
022	340	049	看	361	075	片 370
023	± 340	050	有	362	076	并 371
024	<u>ж</u> 340	051	Ť	362	077	₮ 371
025	4 340	052	1	363	078	# 371
026	<i></i>	053	ħ	363	079	₮ 372
027	.a 343	054	耆	363	080	# 372

N3448 Page 10

西夏文字小字典目次

081	f 373	111 ff	*
		_	140 1
082	# ······ 374	112 #	141 #
083	# ······ 374	113 🖣390	142 🕴 993
084	∤ 374	114 # 390	143 # 993
085	<i>₹</i> 374	115 # 390	144 🖣 399
086	≱ 375	116 # 390	145 🕯 401
087	₹ ······ 375	117 # 391	146 🕯401
088	≱ 376	118 # 391	147 - 🐐 401
089	₹ 376	119 # 391	147b # 402
090	# 377	120 # 391	148 🐐 402
091	2 ······ 377	121 🖁391	149 🕯 402
092	₹ ······ 377	122 🖡 392	150 着 402
093	ቹ ······ 377	123 # 392	151 1 402
094	† 377	123b 🛊 394	152 # 402
095	Ť 377	124 # 394	153 🕯 403
096	f	125 🛊 394	154 🖣403
097	7 378	126 🛉 394	155 🖷 403
098	† ····· 378	127 # 394	156 🖷 404
099	f 379	128 # 394	157 着 404
100	# ····· 380	129 # 395	158
101	fi 380	130 🕯395	159
102	ii 380	131 青 395	159b \$ 406
103	∄ 382	132 🕴 395	160
104	ā · 383	133 /395	161 着
105	ii 383	134 ‡ 396	162 🀐 407
106	fi 384	135 ቹ 396	163
107	∄ ······ 385	136 🕽	164 🎄 407
108	F 386	137	165 🍇 408
109	ii 387	138 \$	166 🗱 408
110	fi 388	139 fi	167 東
	300		450

西夏文字小字典目次

168	表 408	198 🐧 428	228 🐧 465
169	清408	199 🕴 438	229 💈 466
170	萧408	200 🕻 428	230 🕏466
171	ñ ······ 408	201 🤌 428	231 🔻468
172	₹ ····· 409	202 🛪429	232
173	∄ 409	203 🔅 430	233 🕏469
174	∄ ······ 409	204 🕉 431	234 💈469
175	第 ·······409	205 刻 449	235 💈 469
176	fi ······ 409	206 \$ 451	235b 🌡 470
177	f410	207 🕺 451	236 💈 471
178	煮 ····································	208 💈451	237 🎉
179	着410	209 🕉 452	238 💈 476
180	4 410	210 🕱 452	239 🐧 477
181	; 412	211 🕺 455	240 麦 477
182	₹418	212 💈 459	241 🖟
183	\$ 421	213 🕻 459	242 🕯 478
184	₹421	214 🕉 459	243 🛱 479
185	∄ 422	215 💈 459	244 🕻 480
186	₹ ······ 423	216 🛊 459	245 🖁 480
187	ž 424	217 🛊 460	246 🖟 480
188	3 424	218 🕺 460	247 🛱 481
189	3 425	219 🕻 461	247b 莨 ······· 481
190	§ 425	220 \$ 461	248 🕻 481
191	ž 425	221 \$ 462	249 🛱 481
192	3 425	222 🔻 463	250 🕈 481
193	§ 426	223 🖟 464	251 🕴 482
194	3 426	224 🐧 464	252 * 482
195	§427	225 🐧 464	253 🔻 482
196	3 427	226 🐧 465	254 🔻 482
197	\$ 427	227 💈	255 🖟

西夏文字小字典目次

256	# 485	279	302 贅 499
257	Ø 486	280 L 493	303 🏂 499
258	团486	281 🐔 493	304 \$ 500
259	1 486	282 🗓 493	305 🍹 500
260	¾ 487	283	306 4 501
261	Ñ 488	284 🖺 493	307 4 501
262	演488	285 🗓 493	308 🕴 501
263	₫ ······ 489	286 瓦 494	309 § 501
264	ß 489	287	310 🕈 501
265	ii 489	288494	311 🛊 501
266	§ 489	289 🕇494	312 . 眷 502
267	麗	290 🛊 495	313 🐧 502
268	L 489	291 實 495	314 🛊 502
269	礼 ······· 489	292 夏 495	315 🛊 502
270	il 491	293 🛊 495	316
271	心491	294 🐧 495	317 🥞 503
272	Æ 491	295 畫495	318 💈 503
273	₹ 491	296 🕴 496	319 ** 503
274	風 491	297 🐧 496	017b 504
275	纟491	298 景 498	017c # 504
276	£ 492	299 🕏 498	017d # 504
277		300 备 499	017e 🕱 505
278	克 ······ 492	301 🕻 499	