L2/01-347 ISO/IEC JTC1/SC2/WG2 N2366R 2001-09-20

Universal Multiple-Octet Coded Character Set International Organization for Standardization Organisation Internationale de Normalisation Международная организация по стандартизации

Doc Type:Working Group DocumentTitle:Proposal to add six phonetic characters to the UCSSource:Richard S. Cook, Jr., and Michael EversonStatus:Expert ContributionDate:2001-09-20

A. Administrative

Title
 Proposal to add six phonetic characters to the UCS.
 2. Requester's name
 Richard S. Cook, Jr., and Michael Everson.
 3. Requester type
 Expert contribution.
 4. Submission date
 2001-09-20
 5. Requester's reference
 6a. Completion
 This is a complete proposal.
 6b. More information to be provided?
 No.
 No.
 A. Submission to the provided of the properties of the properties of the provided o

B. Technical – General

1a. New script? Name? No. 1b. Addition of characters to existing block? Name? Yes. IPA and Latin Extended-B (proposed code points: U+0221, U+0234, U+0235, U+0236, U+02AE, U+02AF) 2. Number of characters 6. 3. Proposed category Category A. 4. Proposed level of implementation and rationale Level 1. Base character with no diacritics. 5a. Character names included in proposal? Yes. 5b. Character names in accordance with guidelines? Yes. 5c. Character shapes reviewable? Yes 6a. Who will provide computerized font? Michael Everson. **6b.** Font currently available? Yes. 6c. Font format? TrueType.

7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided? Yes.

7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?

Yes.

8. Does the proposal address other aspects of character data processing? No.

C. Technical – Justification

1. Contact with the user community? Yes. The Sino-Tibetan Etymological Dictionary and Thesaurus (STEDT) Project. 2. Information on the user community? Chinese and Sino-Tibetanist linguists. 3a. The context of use for the proposed characters? Phonetic transcription. **3b. Reference** 4a. Proposed characters in current use? Yes. 4b. Where? At least in China and North America. 5a. Characters should be encoded entirely in BMP? Yes. **5b.** Rationale Keeping them with other Latin characters. 6. Should characters be kept in a continuous range? No. 7a. Can the characters be considered a presentation form of an existing character or character sequence? No. 7b. Where? 7c. Reference 8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character? No. 8b. Where? **8c. Reference** 9a. Combining characters or use of composite sequences included? No.

9b. List of composite sequences and their corresponding glyph images provided? No.

10. Characters with any special properties such as control function, etc. included? No.

D. Proposal

At present although [c, z, η , η] used in Sino-Tibetanist linguistics are to be found in the Unicode Standard 3.0, at U+0255, U+0291, U+027F, and U+0285 respectively, the six characters [d, l, η , t, η , η , η] are not.

1. Alveolo-palatal consonants. As Pullum and Ladusaw (*Phonetic Symbol Guide*, second edition, 1996, ISBN 0-226-68536-5) report under their entry for [c] (p. 33):

IPA USAGE

Voiceless "alveolo-palatal" median laminal fricative. Articulated further forward than $[\varsigma]$ (true palatal) but not as far forward as $[\int]$ (palato-alveolar), and articulated laminally (with the flat blade of the tongue) rather than apically (with the tip of the tongue, as in retroflex [§]).

This character [c] is used to represent the Hanyu Pinyin *x*- initial of Mainland Standard Chinese (MSC, as in the word *xiàn* 'now'). To this it may be added that, according to Wu Zongji (1992:77) key differences between the [\int] as in English and MSC [c] are that, whereas the English sound involves a certain degree of lip-rounding and troughing (grooving) of the tongue, the Chinese sound does not: medial tongue closure is rather tight with [c], and the primary point of frication is rather back in comparison with [\int]. Note also that this [c] initial is always found with a following high front vowel, and so is in complementary distribution with MSC's retroflex [s].

What is being distinguished for [c, z] is a place of articulation, such that the following four places are distinguished in the continuum from alveolum to palate:

ALVEOLAR, PALATO-ALVEOLAR, ALVEOLO-PALATAL, PALATAL.

Although IPA [c, z] are apparently marginalized in the "Other Symbols" part of the IPA chart, the "curly-tail" manner of writing in these consonants does, in fact, lend itself to the same logical extension as that seen in the writings of symbols for other places of articulation, e. g. as seen in the IPA symbols for the retroflex series [d, l, n, t, s, t, z]. Pullum and Ladusaw call this "right tail" a "diacritic for retroflexion" (1996:177; "right tail" seems to us to be a "modification" rather than a "diacritic" per se). By the same reasoning the "curly-tail" mark itself could be viewed as a modification (or "diacritic"). Just as the alveolar [d, l, n, t] symbols may receive the "diacritic for retroflexion" to become apico-postalveolar retroflex [d, l, n, t], so too [d, l, n, t] are generated by the same principle, after the pattern of [c, z] to indicate laminal alveolo-palatals. This appears, at any rate, to have been the reasoning employed by linguists in their creation and usage of these [d, l, n, t] curly-tailed symbols. These symbols are employed by linguists working with Chinese and Sino-Tibetan linguistics, in apparent conformity with IPA principles. It is proposed here to encode:

U+0221	[ț] LATIN SMALL LETTER T WITH CURL
U+0234	$[\mathfrak{q}_{ m c}]$ latin small letter D with curl
U+0235	[],] LATIN SMALL LETTER L WITH CURL

U+0236 $[n_{\nu}]$ Latin small letter N with Curl

2. Rounded apical vowels. While [1, 1] are encoded (U+027F LATIN SMALL LETTER REVERSED R WITH FISHHOOK and U+0285 LATIN SMALL LETTER SQUAT REVERSED ESH), and are noted in Unicode 3.0 for Sinological use, two things may be observed.

1 These two characters $[\eta, \chi]$, which represent *unrounded* apical vowels have *rounded* counterparts $[\eta, \chi]$ which are as yet unencoded in the UCS. The glyphs are related to, but are not identical with, $[\eta]$ U+0265 LATIN SMALL LETTER TURNED H. They can be seen on page 5 below in the chart from Wu Zongji (1992), and are also described in Pullum and Ladusaw (1996:81-81). It is proposed here to encode:

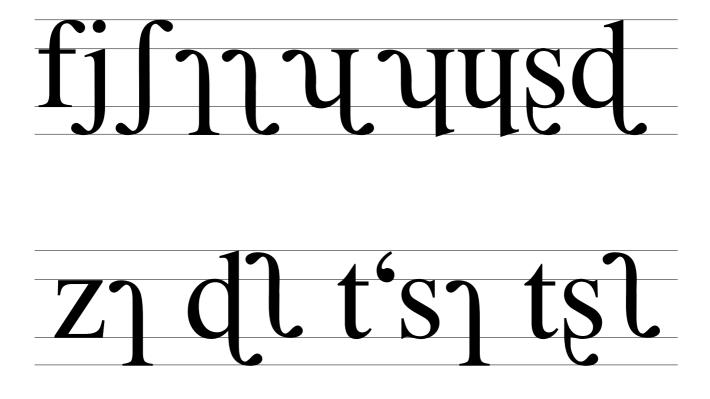
U+02AE	$[\gamma]$ latin small letter turned H with fishhook
U+02AF	$\left[\boldsymbol{\chi} ight]$ latin small letter turned H with Fishhook and Tail

2 The glyphs for $[1, \chi]$ shown in the Unicode Standard 3.0 are inconsistent with Sino-Tibetanist use. These letters *should* descend below the baseline, to the depth of the letter $\langle j \rangle$. It may be possible that U+027F was derived from U+027E LATIN SMALL LETTER R WITH FISHHOOK – though it can be noted that Pullum and Ladusaw identify this typographically as a turned small capital J (1996:98) as well as describing it as an *r* with its top left serif removed (1996:161). In any case, in current Sino-Tibetanist use it has grown to the same depth so that all four vowels form a set: [1, 1, 1, 4, 4]. Typographically there are some interesting variations to be found; for instance in *Yi wen lun* (1993) we find that both [1, 1] have the same length, though the second one, unusually, sits on the baseline rather than hanging from the x-height line (not to be recommended as this could easily be confused with $\int U+0283$ LATIN SMALL LETTER ESH).

凌 (21⁵⁵) "身", "中" (ku²¹) 是 "体", 意, "回"
(bu³³) 是 "形"意, 二字相加以人之形体表示"身"。
窗 (21⁵⁵) "庙", ">" (tgⁱ³³) 是 "陶" 意, "回"
(bu³⁵) 是 "(庙)"; "意, 二字相加以山脚有庙宇表示"庙"。
飞 (teⁱ³³) "毒", "沙" (sl³³) 是 "草" 意, "1"
(fu¹³) 是 "妖"意, 二字相加以妖草有毒表示"毒"。
ベ (fi¹³) "晒", "乙" (dze²¹) 是 "砍" 意, "⁴"
(t⁰³³) "是 "松"意, 二字相加以砍伐松太需要匣于表云 "匾"。
ỡ (fi⁰²¹) " (g) 烧", "→" (nu²) 是 "依" 意,

"U" (me²¹) 是"有"意,二字相加以有病发热表示"烧"。 该字指有病发烧发热的"烧"、"热",即非烧火的"烧",也 不是天热的"热"。

* (d1**) "断", "+" (ts'v²¹) 是"须"意为二"<u>**</u>" 相加表示须断的"断"。



			\backslash								
	268			\backslash	Chapitre III						
	n° (WFS)			9	7		74		70	5
		A (WFS)		*(11	*[h L	*l *1	,	****	ún –
		A (HP) làyĭpíng		*(*1	<u>n</u>	<u>•1</u> /		<u> </u>	
		yǎnghāo			[- -	λ_1	 _h-	- L		ç	- h-
		dànánshã	n		-		-	Į-			sh-
		dàshānji	iăo	-		-	··			-	_
		yejīpō			ζ.	1	-	l-			- 1!
		fēngxiān shíménk			к- dl-	1	-	l- 1-		Ç	
		jiǎdìng		t-		i	-	1-			sh h-
	[18]	băituō			-	i	-	1-		S	-
		tiěshí			-	ļ	-		l-		sh-
		gézhèng jiǎotuó			;- 8-	1	-	1- 1-		_	— V
$\overline{\ }$	[44]	Juono		6		b		[-		8	<u>y-</u>
		667									
	$\overline{}$	287		大			烧	焼山		轻	年轻
	养	蒿,h		`Ļhə ₍ '	1)			⁻			<u></u>
		L (†)								_	
		进1]0(1)			le	u (3)		Ĩŭ	a 33
		ן <u>ר</u>		lo(1)				Y(3)			41 (7)
		,。 岩】		lau.				u ¹⁸			
											(1) (1)
	高坡 lh 宗地 l		$h_{(1)}^{24}$ $l_{(1)}^{22}$		[hə(13)			/]ha(*)			
									[a(1b)		
		员1]u (1)				-			
			lhau, ³³ 40(1)		·						
	瑶里lhj 1 文界4j 4										
				1hju:	3(1)						
				4j0(**)]jeu(**)尚!		<u>. </u>					
						-					
	75.	û8母 養蒿 ↓ ↓h	臟乙坪 (lh) *	大南山	石門坎	擺托 ↓	甲定 Jh	紋坨 し	野鷄坡 ↓	楓香 lh	韻類號
	大病了	lhə‰ Ibər≫#		10:11	joan Joan M	lau ^{ss}	lhə ²¹	lo 👬	lu.3.	[hau∰	(7)
	帶子	(臍帶)]aŋ.f]	law 🐕	ວ ກ ິລິ	lhoŋ *	lua 🟗	_	lhoŋ.º	(24)
	燒山 年輕			leu.5	ley 55	lau. ¹³	lhə:	(ə 35)			(13)
		回反映形式是	由 <u>鎌</u> 母的	lua新 反映形式	la ^出 推測出來:	lo∄ 的。	¦ha \$	la 📩		_	(5)
	76.		職乙坪 L	大南山 1	石門坎 1(lfi)	擺托 1	甲定 1	紋坨 し	野鶴坡 1	楓 香 1	齦頬號
	鐮刀			la盘调 :	lie举調!	len 🗄	ιε <u>**</u>	t [æin∰		len ²⁴	(18)
	⊞ ≜⊮	Li29		la 31	lfiie 🦉	len 🐉	lε 20	læin [™]	len M	len 24	(18)
	畫 米 • 重			lo ²¹	lfiu∛	lon,≵ li∛a	ləŋ ½ li ¾	loŋ3	laŋ <u>.</u>]. li語韻!	 li 13	(22) (2)
	业 埋人		kaŋ%	lau ²¹	<u> </u>	log ²⁸	loŋ."	long	loŋ <u>s</u> ,	laŋ"	(21)
	喫憤	7 lat	La 42			lo 🏭	lu 5	ləa 🖁	la:c	la ll	(15)
	閃電		La 🖁	lai 👬	lfiai ¦	lai%	le 🍇	le l	le.3.	le 🕯	(10)

BIBLIOGRAPHIC REFERENCES (SELECTED)

In addition to the above statistics for data in the STEDT *Main Lexicon*, the following references seek to document attested usages of and comments on curly-tail alveolo-palatal place series symbols. Comments on usage of specific symbols are appended to some entries below.

• CHAO Yuen Ren 趙元任

- 1928 《現代吳語的研究》 <Xiandai Wuyu de yanjiu> [Research on Modern Wu Dialects.] Pei-p'ing: Ch'ing-hua hsueh hsiao yen chiu yuan. 2nd edition. Series: Ch'ing-hua hsueh hsiao yen chiu yuan ts'ung shu; ti 4 chung. {UCB East Asian 5155.4012.}
- **1934** "The non-uniqueness of phonemic solutions of phonetic systems." Nanking: Extract from v. 4 of the Bulletin of the National research Institute of History and Philology, Academia Sinica. (363-398 p. 28 cm.) {UCB Main P221; .C5.}

• International Phonetic Association (IPA)

- **2000** Current Charts Available at http://www.arts.gla.ac.uk/IPA/ipa.html. See specifically, http://www2.arts.gla.ac.uk/IPA/symbols.html for the Chart of OTHER SYMBOLS.
- **1999** Handbook of the International Phonetic Association. Cambridge: Cambridge University Press. {ISBN: 0 521 63751 1.}
- LI Zhenhua 李珍華 and ZHOU Changji 周長楫
- 1993 《漢字古今音表》,〔美〕李珍華,周長楫編撰。北京:中華書局。<Han Zi Gu Jin Yin Biao> [Historical Chinese Phonologic Tables]. LI Zhenhua and ZHOU Changji. Beijing: Zhonghua Shu Ju, 1993. {Harvard Yenching Library Number: 5120 4414; ISBN 7-101-01198-5/H.103.} 8,865 char. readings in 3 historical periods and 7 modern dialects. This work employs [t, d, n, c, z] in reconstructions of Old Chinese (上古); [t, d, c, z] in transcriptions of Middle Chinese (中古); [n, c, z] for the modern Wu dialect (吳語); [n, c] for the modern Gan dialect (贛語); [n,] for the modern Kejia (客家話) dialect. Many of LI and ZHOU's historical forms were also cited in RSC-ECC (COOK 1995).
- LIANG Min 梁民 (sp?) and ZHANG Junru
- 1996 《 侗台語族概論》 <Dong-Tai yuzu gailun> [Intro to Kam-Tai languages]. Beijing: Zhongguo shehui kexue chubanshe. Cf. pp. 972-973. {UCB TEMP98 3754 EAST.}
- MA Xueliang 马学良
- 1981 《语言学概要》<Yuyanxue gailun> [Elements of Linguistics]. Beijing: Huazhong gongxueyuan chubanshe. Cf. pp. 32,51. {UCB P121 .M3 EAST.}

• NIEDERER, Barbara

- 1998 Les langues Hmong-Mjen (Miao-Yao):phonologie historique. Munchen : Lincom Europa. {ISBN: 3-89586-211-8; UCB PL 4070 N543 1998 MAIN.} This text uses serif forms of [1].
- PULLUM, Geoffrey K. and William A. LADUSAW.
- **1996** *Phonetic Symbol Guide*, Second Edition. Chicago, London: University of Chicago Press. {ISBN 0-226-68536-5.}
- QIAN Nairong 錢乃榮
- 1992 《當代吳語研究》<Dangdai Wuyu yanjiu> [Contemporary Wu Dialect Studies]. Shanghai: Shanghai jiaoyu chubanshe. 1121 pp. {UCB PL1931 EAST.C45 1992.} Cf. p. 60.

- Unknown Committee 中國大百科全書編輯委員
- **1988** 《中国大百科全书》 <Zhongguo dabaike quanshu> [Great Chinese Encyclopedia]. Beijing: Zhongguo dabaike quanshu chubanshe. Cf. Yuyan-Wenzi vol., p.118 (IPA).
- WANG Fushi 王輔世 and MAO Zongwu 毛宗武
- **1995** 《苗瑤語古音構擬》, 王輔世,毛宗武 著。北京: 中國社會科學出版社. <Miao-yao Yu Gu Yin Gou Ni> [Reconstruction of Proto-Miao-Yao language]. Beijing: Chinese Social Sciences Press. {ISBN 7-5004-1383-1; UCB PL 4070 W36 1995 EAST.} This text uses serif forms of [1].
- WANG Fushi 王輔世
- 1985 苗語古音構擬/王輔世 著。東京: ILCAA. *Reconstruction of Proto-Miao [Proto-Hmong] language*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA), [1994]. {UCB PL3311.M5 W35 1994 EAST.} This text uses sans-serif forms of [し].
- Wu Zongji 吴宗济
- 1992 《现代汉语语音概要》, 吴宗济 主编。北京: 华语教学出版社。<Xiandai Hanyu -Yuyin Gaiyao> [Elements of Modern Chinese Phonetics]. Beijing: Sinolingua, 1992. {ISBN 7-80052-137-0.} A general handbook of Modern Standard (Beijing) Chinese phonetics, this work contains an IPA-style "Phonetic Symbol Table" (apparently derived derived from YUAN 1960) on page 196 (see attached copy) which includes the [t, d, n, c,

z] curly-tail symbols.

- YUAN Jiahua 袁家骅
- **1960** 《汉语方言概要》,袁家骅等著。北京:文字改革出版社。< Hanyu Fangyan Gaiyao> [Outline of Chinese Dialects]. Beijing: Wenzi Gaige Chubanshe. 2nd edition, 1989. (See the IPA-style chart on page 7, whence derives the chart in WU Zongji 1992).
- Yu Nae-wing 余迺永 (Yu Naiyong)
- 1993 《新校互註·宋本廣韻》香港:香港中文大學。 <Xin Jiao Hu Zhu Song Ben Guang Yun> [A New Revision of the Sung Edition of the Kuang-yun Rhyming Dictionary]. Hong Kong: Chinese University of Hong Kong. 2nd Ed. (3rd Ed., 2000). 1vol., ~900pp., hardcover, Chinese, indices and English appendices. {ISBN: 962-201-413-5.} Cf. the table, p. 82.

STEDT Source Bibliography ABBREVIATIONS

The Sino-Tibetan Etymological Dictionary and Thesuarus (STEDT) Project *Source Bibliography* abbreviations cited above are as follows (numbers at the end of each entry indicate the total number of records in the STEDT databases from each source):

• AW-TBT	= WEIDERT,1987.	• RSC-ECC	= COOK.
• CSL-YIzd	= CHEN, 1979.	 SHK-Sulung 	= SUN, 1993.
• DHFRL	= DAI, 1991.	• SLZO-MLD	= SUN 1980.
• JP-Idu	= PULU, 1978.	• TBL	= DAI, 1992.
• LTBA	= MATISOFF 1974	• ZMYYC	= SUN 1991.

CHEN Shilin

1979 《彝漢字典》<Yihan zidian> [Yi-Chinese dictionary]. CHEN Shilin, LI Min, et al., eds. Sichuan: Yi Language Work Unit, People's Committee of Sichuan.

Соок, Richard S. 曲理察

1995 *The Etymology of Chinese* 辰 *Chén*. Monograph volume of LTBA, Vol. 18.2/Fall 1995. Berkeley: University of California. 1vol., 278pp.《 '辰'字的原始義》。藏緬語言學 報·希18.2。貝克萊市:美國加利福尼亞州大學在貝克萊市語言學部, 1995.

DAI Qingxia

- 1992 《藏緬語族語言詞彙》<Zang Mian Yu Zu Yu Yan Ci Hui> [A Tibeto-Burman lexicon]. DAI Qingxia et al. Beijing: Central Institute of Minorities. {ISBN 7-81001-347-5/H}
- 1991 《藏緬語十五種》<Zangmianyu shiwuzhong> [Fifteen Tibeto-Burman languages]. DAI Qingxia et al., eds. Beijing: Yanshan Chubanshe. DHFRL, authors: DAI, HUANG, FU, RENZENG, LIU.

MATISOFF, James A.

1974- The Biannual Journal *Linguistics of the Tibeto-Burman Area*. The University of California at Berkeley, James A. MATISOFF, editor. {LOC 82-640813, ISSN 0731-3500, OCLC: 4790670.} http://stedt.berkeley.edu/ltba/

PULU, Jatan.

1978 *Idu phrase book*. Shillong: The Director of Information and Public Relations, Arunachal Pradesh.

SUN Hongkai

- **1993** Questionnaire (Sulong). Unpublished.
- **1991** <Zangmianyu yuyin he cihui> [Tibeto-Burman phonology and lexicon]. SUN Hongkai et al. Beijing: Chinese Social Sciences Press. {ISBN 7-5004-0797-1/H.}
- **1980** <Menba, Luoba, Dengren de yuyan> [The languages of the Menba, Luoba and Deng peoples]. SUN Hongkai, LU Shaozun, ZHANG Jichuan and OUYANG Jueya, eds. Beijing: Social Sciences Press.

WEIDERT, Alfons K.

1987 *Tibeto-Burman tonology: A comparative account.* (Current Issues in Linguistic Theory, Vol. 54.) Amsterdam and Philadelphia: John Benjamins Publishing Co. (5426)

ACKNOWLEDGEMENTS

This document was prepared by Richard S. COOK <rscook@socrates.berkeley.edu> of the STEDT Project <http://stedt.berkeley.edu/>, in association with Michael Everson <everson@indigo.ie> and Kenneth WHISTLER <kenw@sybase.com>.

STEDT Project research is supported in part by grants from:

• The National Science Foundation (NSF), Division of Behavioral & Cognitive Sciences, Linguistics, Grant Nos. BNS-86-17726, BNS-90-11918, DBS-92 09481, FD-95-11034, SBR-9808952 and BCS-9904950;

• The National Endowment for the Humanities (NEH), Preservation and Access, Grant Nos. RT-20789-87, RT-21203-90, RT-21420-92, PA-22843 96 and PA-23353-99.

Thanks are due to: Martin HEIJDRA <mheijdra@princeton.edu> of the Gest Oriental Library at Princeton University, for his kind assistance in researching several of the bibliographic references; David MORTENSEN <dmort@socrates.berkeley.edu> of the UC Berkeley Linguistics Dept., for his suggestions and assistance with the sources and scans relating to [1].