

Unicode request for modifier capital letters (b)

L2/21-089

Kirk Miller, kirkmiller@gmail.com

2021 February 11

This is a supplement to L2/20-251, which added modifier capitals <^{C F Q}> to the Latin Extended-D block. When those were accepted by the UTC, I received the following comments from James Kirby, a member of the Alphabet, Charts & Fonts subcommittee of the IPA Council:

I was delighted to see the proposal accepted, however, I was confused by the inclusion of superscript capital C, F, Q but not S, X, Y, Z. I think it is not hard to find examples of these being used in the literature: for example, superscript capital S is used to differentiate different stopped tones in Tai linguistics (DL vs. DS); X is sometimes used to indicate extra-high tone/register; etc.

James sent the references below, and I am submitting <^{S X}> for consideration.

Characters

- ^S A7F0 MODIFIER LETTER CAPITAL S.
- ^X A7F1 MODIFIER LETTER CAPITAL X.

Chart

During the discussion over L2/20-251, Michael Everson suggested all modifier capital letters of the basic Latin alphabet should be added to Latin Extended-D.

	...0	...1	...2	...3	...4	...5	...6	...7	...8	...9	...A	...B	...C	...D	...E	...F
Latin Extended-D																
U+A7Fx	^S	^X														

Properties

A7F0;MODIFIER LETTER CAPITAL S;Lm;0;L;<super> 0053; ; ; ; ;N; ; ; ; ;

A7F1;MODIFIER LETTER CAPITAL X;Lm;0;L;<super> 0058; ; ; ; ;N; ; ; ; ;

References

Alexis Michaud & Bonny Sands (2020) *Tonogenesis*. In Aronoff (ed.) *Oxford Research Encyclopedia of Linguistics*. Oxford University Press,

Pittayawat Pittayaporn (2009) *The Phonology of Proto-Tai*. PhD dissertation, Cornell University.

Figures

Gloss	Li	Siamese	Lungchow	Po-ai
‘male’	*p ^h -	p ^h u: ^{C1}	p ^h u: ^{C1}	pu: ^{C2}
‘to carry’	*t ^h -	t ^h u: ^{A1}	t ^h u: ^{A1}	tu: ^{A2}
‘to ride’	*k ^h -	k ^h i: ^{B1}	k ^h wi: ^{B1}	kuj ^{B2}
‘ripe, cooked’	*s-	suk ^{DS1} ▲	luk ^{DS1} ▲	εuk ^{DS2} ▲

Figure 1. Pittayaporn (2009: 78, table 3-4). The traditional tone categories of the Tai languages are *A *B *C with sonorant finals, and *D with obstruent finals. These split historically into two sets, 1 and 2, based on the voicing of the initial consonant. *D further split into *DS and *DL based on the length of the vowel. As with the H M L tone-letter convention, Tai tones are typically written as full letters in the abstract but as superscripts when modifying a syllable.

²⁶ The etymology is my own. For CT, see Western Nung /sɔk^{DS1}/, Ningming /sak^{DS2}/, and Qinzhou /εak^{DS2}/, etc. For NT, see Yay /sak^{DS2}/, Wuming /εak^{DS2}/, Yongbei /tsak^{DS2}/, etc. The forms in Ningming, Qinzhou, Wuming, Yongbei, and other NT

Figure 2. Pittayaporn (2009: 82, fn 26).

Table 8. Comparative data bringing out the origin of extra-high tone in Moba. Data and analysis from Rialland (2001, p. 317). XH = extra-high.

gloss	Gulmancema	Moba
he walked on (sthg)	õ kándì (^{LM} o ^H kan ^L di)	ũ kánt (^{LM} u ^{XH} kant)
he walks on (sthg)	õ kándí (^{LM} o ^H kan ^H di)	ũ kánt (^{LM} u ^H kant)

Figure 3. Michaud & Sands (2020: 17, table 8) The H-M-L tone convention is extended to five tone levels with X.

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹.**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html for guidelines and details before filling this form.

Please ensure you are using the latest Form from std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest Roadmaps.

A. Administrative

1. Title:	<i>Modifier capital letters (b)</i>
2. Requester's name:	<i>Kirk Miller</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>individual</i>
4. Submission date:	<i>2021 February 11</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<i>yes</i>
(or) More information will be provided later:	

B. Technical - General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	<i>yes</i>	
Proposed name of script:	<i>Latin Extended-D</i>	
b. The proposal is for addition of character(s) to an existing block:	<i>no</i>	
Name of the existing block:		
2. Number of characters in proposal:	<i>2</i>	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input checked="" type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?	<i>yes</i>	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<i>yes</i>	
b. Are the character shapes attached in a legible form suitable for review?	<i>yes</i>	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Kirk Miller</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>SIL (Gentium release)</i>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>yes</i>	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>yes</i>	
7. Special encoding issues:		
Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?	<i>no</i>	

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database (<http://www.unicode.org/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	<i>no</i>
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, with whom? If YES, available relevant documents:	<i>no</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>phonetic</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>yes</i> <i>See references</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	<i>preferably</i> <i>preference of Michael Everson</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<i>yes</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>no</i>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>no</i>
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>no</i>
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	<i>no</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	<i>no</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	<i>no</i>