

**DATE:** 2002-02-13

**DOC TYPE:** Expert contribution

**TITLE:** **Proposal to encode Khmer subscript characters**

**SOURCE:** CHEA Sok Huor, LAO Kim Leang, HARADA Shiro, Norbert KLEIN

**PROJECT:**

**STATUS:** Proposal

**ACTION ID:**

**DUE DATE:** --

**DISTRIBUTION:** Worldwide

**MEDIUM:** Paper

**NO. OF PAGES:**

<b>A. Administrative</b>	
1. Title	Proposal to encode Khmer subscript characters
2. Requester's name	CHEA Sok Huor, LAO Kim Leang, HARADA Shiro, Norbert KLEIN
3. Requester type	Expert request.
4. Submission date	2002-02-13
5. Requester's reference	ISO/IETC JTC 1/SC 2/WG 2 N2380R
6a. Completion	This is a complete proposal.
6b. More information to be provided?	Only as required for clarification.

<b>B. Technical – General</b>	
1a. New script? Name?	No.
1b. Addition of characters to existing block? Name?	No. Extended Khmer
2. Number of characters	37
3. Proposed category	
4. Proposed level of implementation and rationale	
5a. Character names included in proposal?	Yes
5b. Character names in accordance with	Yes

guidelines?	
5c. Character shapes reviewable?	Yes.
6a. Who will provide computerized font?	
6b. Font currently available?	
6c. Font format?	TrueType.
7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?	
7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?	
8. Does the proposal address other aspects of character data processing?	Character classifications are provided.

## C. Technical – Justification

1. Contact with the user community?	Yes. National Body Contribution
2. Information on the user community?	Native.
3a. The context of use for the proposed characters?	
3b. Reference	
4a. Proposed characters in current use?	Yes.
4b. Where?	Cambodia.
5a. Characters should be encoded entirely in BMP?	Yes.
5b. Rationale	Daily use of characters in Cambodia.
6. Should characters be kept in a continuous range?	Yes.
7a. Can the characters be considered a presentation form of an existing character or character sequence?	Yes. All subscript forms are currently specified to be formed using U+17D2 followed by a letter to form the subscript.
7b. Where?	
7c. Reference	
8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?	

8b. Where?	
8c. Reference	
9a. Combining characters or use of composite sequences included?	
9b. List of composite sequences and their corresponding glyph images provided?	
10. Characters with any special properties such as control function, etc. included?	











<b>D. SC2/WG2 Administrative</b>	
To be completed by SC2/WG2	
1. Relevant SC 2/WG 2 document numbers:	
2. Status (list of meeting number and corresponding action or disposition)	
3. Additional contact to user communities, liaison organizations etc.	
4. Assigned category and assigned priority/time frame	
Other Comments	












The Cambodian body proposes to encode subscript forms of characters. The subscript form has been used for many centuries. All existing Cambodian computer systems currently represent subscripts as a single character. Compatibility with existing systems would be enhanced by encoding the subscript characters.













While it is acknowledged that the current Unicode encoding is capable of forming all subscript forms required, the use of U+17D2 to form subscripts has deficiencies, especially in the area of conversion between character encodings. Unlike the Indic virama, the U+17D2 does not function as a killer of the inherent vowel. Subscript characters can be placed on vowels. Furthermore, unlike the Myanmar language, there is no visual killer in Khmer.





The Cambodian body proposes encoding subscript forms in an Extended Khmer block (with the Lunar Dates) as a canonical equivalent of the current subscript model specified in the Unicode standard.

The following chart shows the Unicode 3.1 encoding of subscript characters that are known to be valid for modern and historical use. These are the characters that should be encoded in an Extended Khmer block.

Unicode 3.1 code points	Khmer representation	Subscript name	Unicode properties
U+17D2 U+1780		KHMER CONSONANT SIGN COENG KA	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+1781		KHMER CONSONANT SIGN COENG KHA	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+1782		KHMER CONSONANT SIGN COENG KO	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+1783		KHMER CONSONANT SIGN COENG KHO	Mc;0;L;;;;;N;;;;;
U+17D2 U+1784		KHMER CONSONANT SIGN COENG NGO	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+1785		KHMER CONSONANT SIGN COENG CA	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+1786		KHMER CONSONANT SIGN COENG CHA	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+1787		KHMER CONSONANT SIGN COENG CO	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+1788		KHMER CONSONANT SIGN COENG CHO	Mc;0;L;;;;;N;;;;;
U+17D2 U+1789		KHMER CONSONANT SIGN COENG NYO	Mn;0;NSM;;;;;N;;;;;

U+17D2 U+178A		KHMER CONSONANT SIGN COENG DA	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+178B		KHMER CONSONANT SIGN COENG TTHA	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+178C		KHMER CONSONANT SIGN COENG DO	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+178D		KHMER CONSONANT SIGN COENG TTHO	Mc;0;L;;;;;N;;;;;
U+17D2 U+178E		KHMER CONSONANT SIGN COENG NA	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+178F		KHMER CONSONANT SIGN COENG TA	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+1790		KHMER CONSONANT SIGN COENG THA	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+1791		KHMER CONSONANT SIGN COENG TO	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+1792		KHMER CONSONANT SIGN COENG THO	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+1793		KHMER CONSONANT SIGN COENG NO	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+1794		KHMER CONSONANT SIGN COENG BA	Mc;0;L;;;;;N;;;;;

U+17D2 U+1795		KHMER CONSONANT SIGN COENG PHA	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+1796		KHMER CONSONANT SIGN COENG PO	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+1797		KHMER CONSONANT SIGN COENG PHO	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+1798		KHMER CONSONANT SIGN COENG MO	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+1799		KHMER CONSONANT SIGN COENG YO	Mc;0;L;;;;;N;;;;;
U+17D2 U+179A		KHMER CONSONANT SIGN COENG RO	Mc;0;L;;;;;N;;;;;
U+17D2 U+179B		KHMER CONSONANT SIGN COENG LO	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+179C		KHMER CONSONANT SIGN COENG VO	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+179D		KHMER CONSONANT SIGN COENG SHA	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+179E		KHMER CONSONANT SIGN COENG SSA	Mc;0;L;;;;;N;;;;;
U+17D2 U+179F		KHMER CONSONANT SIGN COENG SA	Mc;0;L;;;;;N;;;;;
U+17D2 U+17A0		KHMER CONSONANT SIGN COENG HA	Mn;0;NSM;;;;;N;;;;;

U+17D2 U+17A2		KHMER CONSONANT SIGN COENG QA	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+17A7		KHMER INDEPENDENT VOWEL SIGN COENG QU	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+17AB		KHMER INDEPENDENT VOWEL SIGN COENG RY	Mn;0;NSM;;;;;N;;;;;
U+17D2 U+17AF		KHMER INDEPENDENT VOWEL SIGN COENG QE	Mn;0;NSM;;;;;N;;;;;