

Title:	Comments accompanying the US negative vote on the PDAM2 to ISO/IEC 10646-1:2000
Source:	NCITS/L2
Action:	Forward to NCITS

The US National body is voting NO on the PDAM2 to ISO/IEC 10646-1:2000. The accommodation of comment T9 will change its vote to YES. In addition, the US has the following comments:

### **Technical Comments:**

#### **T.1 Limbu**

After further study on the Limbu script described by WG2 N2410, The US National body requests the following changes to the proposed Limbu script.

The following characters should be added:

1909 LIMBU LETTER JHA  
 190A LIMBU LETTER YAN  
 191A LIMBU LETTER SSA  
 1932 LIMBU SMALL LETTER ANUSVARA  
 1940 LIMBU SIGN LOO  
 1944 LIMBU EXCLAMATION MARK

The Limbu vowel signs should be re-arranged as follows (former name and location in parenthesis)

1920 LIMBU VOWEL SIGN A (1920 LIMBU VOWEL SIGN A)  
 1921 LIMBU VOWEL SIGN I (1922 LIMBU VOWEL SIGN I)  
 1922 LIMBU VOWEL SIGN U (1923 LIMBU VOWEL SIGN U)  
 1923 LIMBU VOWEL SIGN EE (1924 LIMBU VOWEL SIGN E)  
 1924 LIMBU VOWEL SIGN AI (1925 LIMBU VOWEL SIGN AI)  
 1925 LIMBU VOWEL SIGN OO (1926 LIMBU VOWEL SIGN O)  
 1926 LIMBU VOWEL SIGN AU (1927 LIMBU VOWEL SIGN AU)  
 1927 LIMBU VOWEL SIGN E (1929 LIMBU VOWEL SIGN EE)  
 1928 LIMBU VOWEL SIGN O (1928 LIMBU VOWEL SIGN OO)  
 1929 LIMBU SUBJOINED LETTER YA (192A LIMBU SUBJOINED LETTER Y)  
 192A LIMBU SUBJOINED LETTER RA (192B LIMBU SUBJOINED LETTER R)  
 192B LIMBU SUBJOINED LETTER WA (192C LIMBU SUBJOINED LETTER W)

The existing characters between 1930 and 1938 should be renamed as follows ('A' appended to their name):

1930 LIMBU SMALL LETTER KA  
 1931 LIMBU SMALL LETTER NGA  
 1933 LIMBU SMALL LETTER TA  
 1934 LIMBU SMALL LETTER NA  
 1935 LIMBU SMALL LETTER PA  
 1936 LIMBU SMALL LETTER MA  
 1937 LIMBU SMALL LETTER RA  
 1938 LIMBU SMALL LETTER LA

The character in code position 193F should be moved to 1945 (by similarity with the addition in position 1944) with the slightly modified name as follows:

1945 LIMBU QUESTION MARK

## **T.2 Normalization reference**

At the last WG2 meeting in Singapore, it was proposed to add a normative reference to normalization into the standard (document WG2 N2398). The US national body is asking to add the text proposed by N2398 into this amendment. The additions and changes are as follows (quoted from N2398)

<<

In clause '3 Normative reference', add the following reference:

Unicode Standard Annex, UAX#15, Unicode Normalization Forms, Version 3.1., 2001-03-23.

Insert a new clause 24 between the current clause 23 and 24 (the following clauses are incremented by one):

### **24 Normalization forms**

Normalization forms are the mechanisms allowing the selection of a unique coded representation among alternative, but equivalent coded text representations of the same text. Normalization forms for use with ISO/IEC 10646 are specified in the Unicode Standard UAX#15.

NOTE 1 – By definition, the result of applying any of these normalization forms is stable over time. It means that a normalized representation of text remains normalized even when the standard is amended.

NOTE 2 – Some normalizations forms favor composite sequences over shorter representations of text, other favor the shorter representations. The backward compatibility requirement is provided by establishing ISO/IEC 10646-1:2000 and ISO/IEC 10646-2:2001 as the reference versions for the definition of the shorter representation of text.

In the current clause 24 (which becomes 25), insert a new clause 25.3 (after the current clause 24.2):

### **25.3 Alternate coded representations**

Alternate coded representations of text are generated by using multiple combining characters in different orders, or using various equivalent combinations of characters and composite sequences. These alternate coded representations result in multiple representation of the same text. Normalizing these coded representations creates a unique representation.

NOTE – For example, in implementation level 3 the French word "là" may be represented by the characters LATIN SMALL LETTER L followed by LATIN SMALL LETTER A WITH GRAVE, or may be represented by the characters LATIN SMALL LETTER L followed by LATIN SMALL LETTER A followed by COMBINING GRAVE ACCENT. When the normalizations forms are applied on those alternate coded representations, only one representation remains. The form of the remaining representation depends on the normalization form used.

In the current clause 24.3 (new 25.4), remove the Note (replaced by note mentioned above)

>>

## **T.3 Arabic characters addition**

The US I asking for the addition for the following Arabic symbols with the proposed code location and name (reference documents L2/01-427 and L2/02-061 with WG2 documents to be provided):

0600	ARABIC NUMBER SIGN
0601	ARABIC YEAR SIGN
0602	ARABIC FOOTNOTE MARKER
060D	ARABIC POETIC VERSE SIGN
060E	ARABIC DATE SEPARATOR
0610	ARABIC SIGN SALLALLAHOU ALAYHE WASALLAM
0611	ARABIC SIGN ALAYHE ASSALAM
0612	ARABIC SIGN RAHMATULLAH ALAYHE
0613	ARABIC SIGN RADI ALLAHOU ANHU
0614	ARABIC SIGN NOM DE PLUME
0656	ARABIC SUBSCRIPT ALEF
0657	ARABIC TURNED DAMMA
0658	ARABIC BREVE
06EE	ARABIC LETTER DAL WITH INVERTED V
06EF	ARABIC LETTER REH WITH INVERTED V
06FF	ARABIC LETTER HEH WITH INVERTED V

Three of the new characters (0600-602) should be added to a new category of format characters (see T.4). The characters 0610-0614, 0656-0658 needs to be added to annex B.1.

#### **T.4 New format characters (general format characters and head marker)**

There should be a brief description of the Combining Grapheme Joiner introduced by 10646-1:2001 Amd 1. This could be satisfied by adding a new entry in sub-clause F.1.1 as follows:

**COMBINING GRAPHEME JOINER (034F):** The Combining Grapheme Joiner is used to indicate that adjacent characters belong to the same grapheme cluster. Grapheme clusters are sequences of one or more coded characters that correspond to what users think of as characters. They include, but are not limited to, composite sequences such as (g + °), digraphs such as Slovak “ch”, or sequences with letter modifiers such as k<sup>w</sup>.

In addition, there is a need to describe a type of alternate format character represented by the SYRIAC ABBREVIATION MARK (070F) as well as the ARABIC END OF AYAH (06DD), the latter currently classified as a combining character. The US national body is asking for the addition of a new clause in Annex F as follows:

#### **F.5 Subtending format characters**

The following characters are used to subtend a sequence of subsequent characters:

ARABIC NUMBER SIGN (0600)  
ARABIC YEAR SIGN (0601)  
ARABIC FOOTNOTE MARKER (0602)  
ARABIC END OF AYAH (06DD),  
SYRIAC ABBREVIATION MARK (070F)

The scope of these characters is the subsequent sequence of digits (plus certain other characters), with the exact specification as defined in Unicode 3.2 for ARABIC END OF AYAH.

In addition the ARABIC END OF AYAH (06DD) should be removed from Annex B – List of combining characters and its glyph should be updated in the chart to remove the dotted circle and use a dotted rectangular frame to be consistent with other format characters.

#### **T5. Korean CJK compatibility character mapping**

The CJK COMPATIBILITY IDEOGRAPH F951 is mapped by the new clause 22.2 (Amd.1) to the CJK UNIFIED IDEOGRAPH 96FB, it should be mapped to 964B instead. The corresponding entry in the file CJKC0SR.txt needs to be updated.

#### **T6. Variation selectors**

In sub-clause 24.5 (Amd.1), the following table row should be fixed as follows:

<2269, FE00>	GREATER-THAN BUT NOT EQUAL TO with vertical stroke
--------------	--

#### **T.7 Addition of symbols**

The US National Body recommends adding the character NEGATIVE CIRCLED DIGIT ZERO to complete the series of negative circled digits U+2776-U+277F, U+24EB-24F4. This makes it similar to U+24EA, U+2460-U+2473. The proposed code point and name are as follows:

24FF NEGATIVE CIRCLED DIGIT ZERO

This makes the collection 43 ENCLOSED ALPHANUMERICS fixed.

The US also recommends adding the EJECT SYMBOL as a required user interface symbol for IT usage with code position and name as follows (document L2/01-414 to be submitted as a WG2 document):

23CF EJECT SYMBOL

## **T.8 Khmer**

Following discussions with representatives of the Khmer Standardization national body, the US national body is asking for the addition of the following Khmer characters. An appropriate WG2 document will be provided.

17EA	KRUNG VOICING OR PREGLOTTALIZATION SIGN
17EB	KRUNG VOICING AND PREGLOTTALIZATION SIGN
17DD	KHMER SIGN ATTHACAN
17F0	KHMER SYMBOL LEK ATTAK SON
17F1	KHMER SYMBOL LEK ATTAK MUOY
17F2	KHMER SYMBOL LEK ATTAK PII
17F3	KHMER SYMBOL LEK ATTAK BEI
17F4	KHMER SYMBOL LEK ATTAK BUON
17F5	KHMER SYMBOL LEK ATTAK PRAM
17F6	KHMER SYMBOL LEK ATTAK PRAM-MUOY
17F7	KHMER SYMBOL LEK ATTAK PRAM-PII
17F8	KHMER SYMBOL LEK ATTAK PRAM-BEI
17F9	KHMER SYMBOL LEK ATTAK PRAM-BUON
19E0	KHMER SYMBOL PATHAMASAT
19E1	KHMER SYMBOL MUOY KOET
19E2	KHMER SYMBOL PII KOET
19E3	KHMER SYMBOL BEI KOET
19E4	KHMER SYMBOL BUON KOET
19E5	KHMER SYMBOL PRAM KOET
19E6	KHMER SYMBOL PRAM-MUOY KOET
19E7	KHMER SYMBOL PRAM-PII KOET
19E8	KHMER SYMBOL PRAM-BEI KOET
19E9	KHMER SYMBOL PRAM-BUON KOET
19EA	KHMER SYMBOL DAP KOET
19EB	KHMER SYMBOL DAP-MUOY KOET
19EC	KHMER SYMBOL DAP-PII KOET
19ED	KHMER SYMBOL DAP-BEI KOET
19EE	KHMER SYMBOL DAP-BUON KOET
19EF	KHMER SYMBOL DAP-PRAM KOET
19F0	KHMER SYMBOL TUTEYASAT
19F1	KHMER SYMBOL MUOY ROC
19F2	KHMER SYMBOL PII ROC
19F3	KHMER SYMBOL BEI BOC
19F4	KHMER SYMBOL BUON ROC
19F5	KHMER SYMBOL PRAM ROC
19F6	KHMER SYMBOL PRAM-MUOY ROC
19F7	KHMER SYMBOL PRAM-PII ROC
19F8	KHMER SYMBOL PRAM-BEI ROC
19F9	KHMER SYMBOL PRAM-BUON ROC
19FA	KHMER SYMBOL DAP ROC
19FB	KHMER SYMBOL DAP-MUOY ROC
19FC	KHMER SYMBOL DAP-PII ROC
19FD	KHMER SYMBOL DAP-BEI ROC
19FE	KHMER SYMBOL DAP-BUON ROC
19FF	KHMER SYMBOL DAP-PRAM ROC

## **T.9 Symbols to be removed**

The characters 267E DO NOT LITTER SIGN and 267F RECYCLABLE PACKAGING must be removed. These characters have created many issues concerning usage and identity, so the US doesn't consider them mature for encoding.

The character 31D0-31D4 (VULGAR FRACTION ... with horizontal bars) are duplicates of characters already encoded in the LATIN-1 SUPPLEMENT and NUMBER FORMS blocks and must not be encoded again as character. The existing characters have always been explicitly intended to unify both forms of the fraction.

## **T.10 DPRK symbols**

The name of the character 2614 SHOWER WEATHER should be changed to ‘UMBRELLA WITH RAIN DROPS’.

In accordance with the glyph used for 2614, the graphic representation of 2602 UMBRELLA should be updated accordingly.

The name of the character 2615 TEA SYMBOL should be changed to ‘HOT BEVERAGE SYMBOL’.

The character 2618 CIRCLED UPWARD INDICATION seems to be a variation of 261D WHITE UP POINTING INDEX and may not be needed. At minimum, the name should be changed to reflect the similarity with 261D.

The character 2700 LEFTWARDS SCISSORS seems a variation of 2702 SCISSORS and should be removed accordingly.

The characters a position 23CF WHITE UP-POINTING TRIANGLE WITH CAUTION MARK and 23D0 HIGH VOLTAGE CAUTION MARK should be moved and renamed as follows:

26A0	WARNING SIGN
26A1	HIGH VOLTAGE SIGN

The block Supplemental Arrows-C (in row 2B) should be renamed ‘Miscellaneous Symbols and Arrows’.

The first 6 characters of the newly proposed Supplemental CJK Compatibility should be moved either in holes of existing CJK blocks (for the first five characters: 31C0-31C4) or in the Letterlike Symbols block (for the fifth: 31C5) with the new location as follows:

32CC	SQUARE HG
32CD	SQUARE ERG
32CE	SQUARE EV
32CF	LIMITED LIABILITY SIGN
3250	PARTNERSHIP SIGN
213B	FACSIMILE SIGN

Therefore, there is no need yet for the Supplemental CJK Compatibility block and it should be removed.

## **T.11 Indic scripts character addition**

The US is asking for the addition of the following characters with the suggested code position and names:

0904	DEVANAGARI SHORT LETTER A
09BD	BENGALI SIGN AVAGRAHA
0A01	GURMUKHI SIGN ADAK BINDI
0A03	GURMUKHI SIGN VISARGA
0A8C	GUJARATI LETTER VOCALIC L
0AE1	GUJARATI LETTER VOCALIC LL
0AE2	GUJARATI VOWEL SIGN VOCALIC L
0AE3	GUJARATI VOWEL SIGN VOCALIC LL
0AF1	GUJARATI RUPEE SIGN
0B35	ORIYA LETTER VA
0CBC	KANNADA SIGN NUKTA
0CBD	KANNADA SIGN AVAGRAHA

These additions are the results of several communications from the Indian government, in response to their review of ISO 10646 (reference L2/01-430R and L2/01-431R, WG2 document to be provided)

----

End of US comments