SC22/WG20 N896
L2/01-476

Universal multiple-octet coded character set
International organization for standardization
Organisation internationale de normalisation

Title: Ordering rules for Khmer

Source: Kent Karlsson
Date: 2001-12-19
Status: Expert contribution
Document type: Working group document
Action: For consideration by the UTC and JTC 1/SC 22/WG 20

1 Introduction

The Khmer script in Unicode/10646 uses conjoining characters, just like the Indic scripts and Hangul. An alternative that is suitable (and much more elegant) for Khmer, but not for Indic scripts, would have been to have combining-below (and sometimes a bit to the side) consonants and combining-below independent vowels, similar to the combining-above Latin letters recently encoded, as well as how Tibetan is handled. However that is not the chosen solution, which instead uses a combining character (COENG) that makes characters conjoin (glue) like it is done for Indic (Brahmic) scripts and like Hangul Jamo, though the latter does not have a separate gluer character.

In the Khmer script, using the COENG based approach, the words are formed from orthographic syllables, where an orthographic syllable has the following structure [add ligation control?]:

Khmer-syllable ::= (K H)* K M*

where K is a Khmer consonant (most with an inherent vowel that is pronounced only if there is no consonant, independent vowel, or dependent vowel following it in the orthographic syllable) or a Khmer independent vowel, H is the invisible Khmer conjoint former COENG, M is a combining character (including COENG, though that would be a misspelling), in particular a combining Khmer vowel (noted A below) or modifier sign. The latter should come in a particular order, when present at all: ROBAT, consonant modifier, dependent vowel (including VIRIAM), vowel

modifier, and finally other combining characters. Not following this sequence, leads to a different ordering, and may have an effect on other things as well, like rendering and editing. Khmer orthographic syllable breaks are where the above syntax no longer matches. Thus, a syllable break is detectable by the occurrence of (spacing) punctuation, a digit, a non-Khmer spacing character, or a following (Khmer) consonant or independent vowel that is not preceded by a COENG. Note that not only can consonants be underscripted by consonants, but also by independent vowels (though that is rare), and independent vowels can be underscripted by consonants or independent vowels (both rare).

The COENG makes the adjacent Khmer consonant/independent vowel characters conjoining (like VIRAMAs make adjacent Indic scripts's consonants conjoining; unlike the VIRAMAs the COENG is never rendered, not even as a fallback). Thus, combining characters that follow the (K H)* K part apply to the entire (K H)* K part. In Khmer the non-first consonants/independent vowels in an orthographic syllable are typographically underscripted, with a modified glyph. Thus, in many cases, combining characters, with exception of combining-below characters, *glyphically* apply to the first consonant in the orthographic syllable rather than the last one; exceptions are when some part of the underscripted letter "sticks up" on the left or on the right. Note also that ligated glyph forms are common.

2 Questionable characters in the Khmer encoding

The BEYYAL character supposedly stands for "any abbreviation for the Khmer correspondence of 'et cetera'". This is an exceptionally bad idea, since it is no longer the author that decides which abbreviation to use, but it instead depends on the font used. Authors would no longer be in control over the abbreviation convention used. This character should simply not be used. That way a stable abbreviation convention can be maintained in a document also when using multiple fonts, or when changing fonts. It is collated below as if it was *one* of 'its' abbreviations.

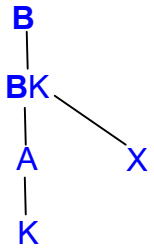
Originally, the intent was to handle lunar calendar symbols via compositions using the COENG (as well as one combining character). But now there seems to be unanimity in that they should be handled by different means.

Two "independent vowels" (U+17A3 and U+17A4) were encoded under the assumption that there was a need to distinguish the Khmer and 'transcribed from Sanskrit' uses at the character level. That appears unfounded.

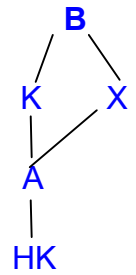
There are two explicit, but invisible, vowel signs that have INHERENT in their names. Note that by "inherent" we mean a vowel that is inherent in a consonant letter (or, indeed, an independent vowel), while by "INHERENT" we mean a Khmer character that has "INHERENT" in its name. The inclusion of the INHERENT characters appear to be a mistake, parallel to that of the two "independent vowels" U+17A3 and U+17A4.

The COENG vs. the combining-under encoding models is one of the major controversies about the encoding of Khmer. It is obvious that the combining-under (i.e. underscripts) model is much more elegant, as it is for Tibetan, and simplifies both the model and the computer processing of Khmer text. It also allows a more logical order for the ROBAT and consonant modifiers. However, the COENG (VIRAMA-like) model adopted appears to be adequate, though it is clumsy for Khmer, but it might on the other hand allow for in-syllable ligation control (ZWJ, ZWNJ).

3 Khmer ordering issues



Nominally, when ordering Khmer strings, they are grouped by orthographic syllables (regarded as clusters or segments), where the dependent vowels are collated before the consonants in the alphabetic ordering (ignoring the COENG that is used in the encoding). As always (see *WG20 N891R*), this ordering clustering can be achieved by inserting lightweight cluster end mark characters (**B**) by prehandling. This is similar to how Hangul can be handled (see *WG20 N891R*), which is grouped by consonant clusters and vowel clusters, (first level) word-by-word ordering, as well as numerical numeral ordering.



However, Khmer has a particular encoding where consonants (and independent vowels) that are not first in a syllable are preceded by a particular character, the COENG. By simply weighting the consonants lighter than the dependent vowels and the COENG heavier than the dependent vowels, the expected clustered ordering is achieved without inserting any cluster end marks. To make the collation keys a bit shorter, the COENG+consonant (or COENG+independent vowel) pairs can get contracted weightings. These would actually correspond one-to-one to underscripted consonants/independent vowels. Similarly, other related scripts that use VIRAMAs to achieve the coinjoinment, can weight the consonants (and independent vowels) before the dependent vowels and last (heaviest) the VIRAMA. This is already the general collation ordering of these characters in ISO/IEC 14651:2001 on a per script basis. Doing it on a per-script basis, however, may give problems with the ordering clustering at the end of the Khmer (Devanagari, etc.) string, unless a light-weighted character is, or is inserted (**B**) by prehandling, after the end of the Khmer (Devanagari, etc.) substring. Ignoring edge cases, this too can be handled without prehandling though: collate the dependent vowels (of Khmer, all Indic scripts, and Tibetan) as being last in the 'alphabetical' order, *after all scripts*, and the non-first consonants thereafter (i.e., COENG, VIRAMAs, combining consonants (Tibetan), and also COMBINING GRAPHEME JOINER, as well as Hangul trailing consonants and thereafter Hangul vowels).

One way of ordering most of the independent vowels are as if they were a glottal stop (there is a Khmer consonant letter for glottal stop) followed by the corresponding dependent vowel, even though they do not have a Unicode decomposition that way. This is how they are ordered in the Chuon Nath's dictionary. They are distinguished from an actual glottal stop followed by the corresponding dependent vowel at the second collation level via a <VRNTn> weight. This is the ordering adopted here. However, four of the independent vowels actually (phonetically) begin with a consonant (R for two of them, and L for the other two). They are ordered just after the corresponding consonant, just like in the Chuon Nath's dictionary.

The inherent vowel in a consonant is not explicitly weighted. Doing so only when it is not cancelled by a follow-on character would complicate the weighting too much, and would require prehandling. Not including a weight for an inherent vowel is justified by that the inherent vowel is the first vowel, and non-first consonants have collation weights after (heavier than) the vowels. So leaving out the weight for the inherent vowel does not disturb the ordering at all. The inherent vowel actually varies depending on circumstances, but that should not affect collation order. The INHERENT characters are invisible, and should be ignored in collation, as is done below.

The KHMER SIGN ROBAT, which is combining and cannot come first in an orthographic syllable, spells a syllable initial RO. To get the ordering proper, prehandling is needed to move the ROBAT to the initial position of the syllable for collation (but *only* for collation; it should not really combine with whatever is before the syllable it is in).

The nasalisation sign and other pseudo-vowels are ordered as the last few, in the 'alphabetical order', dependent vowels. In addition, two nasalised vowels are ordered as separate letters. Below the latter is done via contractions below.

Even though most applications of the consonant modifiers only have an effect on the second collation level, their applications to BA are ordered as separate letters. Below that is done via contractions.

The RY, RYY, LY, and LYY independent vowels are really consonants with different inherent vowels than usual. A phonetic alternative might be to order these as *variants* of RO or LO followed by Y or YY:

```
<U17AB> "<S179A><S17B9>";"<BASE><VRNT1><BASE>";"<MIN><MIN><MIN>";<U17AB> % KHMER INDEPENDENT VOWEL RY
<U17AC> "<S179A><S17BA>";"<BASE><VRNT1><BASE>";"<MIN><MIN><MIN>";<U17AC> % KHMER INDEPENDENT VOWEL RYY
<U17AD> "<S179B><S17B9>";"<BASE><VRNT1><BASE>";"<MIN><MIN><MIN>";<U17AD> % KHMER INDEPENDENT VOWEL LY
<U17AE> "<S179B><S17BA>";"<BASE><VRNT1><BASE>";"<MIN><MIN><MIN>";<U17AE> % KHMER INDEPENDENT VOWEL LYY
```

However, the glyphs for these letters are completely different from the corresponding consonant-vowel combinations, and even resemble unrelated letters. Further, the Chuon Nath's dictionary orders them as separate letters after RO and LO respectively. For these, we follow the Chuon Nath's dictionary in the ordering rules below.

The ordering rules below have not yet been thoroughly reviewed by experts in Khmer, so there may be changes.

4 Prehandling

As mentioned above, one can entirely avoid prehandling for the clustering part of the ordering of Khmer.

But to get the fully correct ordering when any ROBAT characters occur, the following prehandling is needed: For each Khmer orthographic syllable, move any KHMER SIGN ROBAT in it to the initial position of the syllable and insert a COENG after it. Note that the string after prehandling will not display properly; the KHMER SIGN ROBAT will then appear to combine with whatever is before the syllable it really is in. This prehandling is assumed in the weighting table below. (Note that in a combining underscript model, any ROBAT could just have been in the character string right after the initial consonant. It could then have been handled by contractions instead...)

Further, to get the correct placement of the weighting for consonant modifiers (MUUSIKATOAN, TRIISAP, TOANDAKHIAT), especially for the two contractions with BA (see below): For each Khmer orthographic syllable, move any consonant modifier in it to just after the initial consonant (but before any COENG applied to the initial consonant). Note that the string after prehandling is unlikely to display properly; it does not follow the orthographic syllable syntax anymore (in the COENG model; there would be no problem in the combining underscripts model), and the syllable

breaks (for the display engine) will be wrong. However, since any COENG after it is kept, the ordering will be ok. (Note that in a combining underscript model, any initial consonant modifier could just have been in the character string right after the initial consonant anyway. No need to move it by prehandling then...)

However, since the ROBAT never occurs first in a word, just (pronunciation-wise) first in orthographic syllables inside words, doing this prehandling is not a high priority, and may be omitted for non-dictionary purposes. The ROBAT should then be weighted as if it was a COENG followed by a RO, and with a distinction at the second level:

```
<U17CC> "<S17D2><S179A>";"<BASE><BASE><VRNT1>";"<MIN><MIN><MIN>";<U17CC> % KHMER SIGN ROBAT no prehandling
```

Further, if the two contractions of BA and a consonant modifier (see below) are removed, if that is acceptable, the consonant modifiers will only have effect on the second collation level. Then it may be felt appropriate to omit the prehandling movement of the consonant modifiers, since that change will not be noticeable in the ordering and enables the complete omission of any prehandling.

For some applications, it may instead be appropriate with *more* prehandling. Like when it is desired to make a collation distinction based on contextual differences, like different inherent vowels depending on whether a word is of Khmer or Sanskrit origin. (Insertion of the two invisible “INHERENT VOWEL” characters at places might be used for this, though that might still not be appropriate.)

5 Khmer collation weighting rules

```
% Declaration of weighting symbols.  Order in this first section is arbitrary.


% Consonant modifiers (level 2 weights):
collating_symbol <D17C9> % KHMER SIGN MUUSIKATOAN (muusekatoan)
collating_symbol <D17CA> % KHMER SIGN TRIISAP (treisap)
collating_symbol <D17CD> % KHMER SIGN TOANDAKHIAT % marks character not to be pronounced


% Vowel modifiers (level 2 weights):
collating_symbol <D17CB> % KHMER SIGN BANTOC (bantak) % shortens preceding vowel
collating_symbol <D17CE> % KHMER SIGN KAKABAT % sign used with some exclamations
collating_symbol <D17CF> % KHMER SIGN AHSDA % sign used for single-consonant words
collating_symbol <D17D0> % KHMER SIGN SAMYOK SANNYA % used to indicate shortened inherent vowel
```

```
% Consonants:
collating_symbol <S1780> % KHMER LETTER KA
collating_symbol <S1781> % KHMER LETTER KHA
collating_symbol <S1782> % KHMER LETTER KO
collating_symbol <S1783> % KHMER LETTER KHO
collating_symbol <S1784> % KHMER LETTER NGO
collating_symbol <S1785> % KHMER LETTER CA
collating_symbol <S1786> % KHMER LETTER CHA
collating_symbol <S1787> % KHMER LETTER CO
collating_symbol <S1788> % KHMER LETTER CHO
collating_symbol <S1789> % KHMER LETTER NYO
collating_symbol <S178A> % KHMER LETTER DA
collating_symbol <S178B> % KHMER LETTER TTHA
collating_symbol <S178C> % KHMER LETTER DO
collating_symbol <S178D> % KHMER LETTER TTHO
collating_symbol <S178E> % KHMER LETTER NNO (na)
collating_symbol <S178F> % KHMER LETTER TA
collating_symbol <S1790> % KHMER LETTER THA
collating_symbol <S1791> % KHMER LETTER TO
collating_symbol <S1792> % KHMER LETTER THO
collating_symbol <S1793> % KHMER LETTER NO
collating_symbol <S1794> % KHMER LETTER BA
collating_symbol <S1794_S17C9> % KHMER LETTER BA, KHMER SIGN MUUSIKATOAN
collating_symbol <S1794_S17CA> % KHMER LETTER BA, KHMER SIGN TRIISAP
collating_symbol <S1795> % KHMER LETTER PHA
collating_symbol <S1796> % KHMER LETTER PO
collating_symbol <S1797> % KHMER LETTER PHO
collating_symbol <S1798> % KHMER LETTER MO
collating_symbol <S1799> % KHMER LETTER YO
collating_symbol <S179A> % KHMER LETTER RO
collating_symbol <S17AB> % KHMER INDEPENDENT VOWEL RY % glyph based on glyph for 1794
collating_symbol <S17AC> % KHMER INDEPENDENT VOWEL RYY % glyph based on glyph for 1794
collating_symbol <S179B> % KHMER LETTER LO
collating_symbol <S17AD> % KHMER INDEPENDENT VOWEL LY % glyphs based on glyph for 1796
collating_symbol <S17AE> % KHMER INDEPENDENT VOWEL LYY % glyphs based on glyph for 1796
collating_symbol <S179C> % KHMER LETTER VO
collating_symbol <S179D> % KHMER LETTER SHA
collating_symbol <S179E> % KHMER LETTER SSO (ssa)
collating_symbol <S179F> % KHMER LETTER SA
collating_symbol <S17A0> % KHMER LETTER HA
collating_symbol <S17A1> % KHMER LETTER LA
collating_symbol <S17A2> % KHMER LETTER QA (glottal stop)
```

```
% Weights after (heavier than) all scripts:


% Dependent vowels:
collating_symbol <S17D1> % KHMER SIGN VIRIAM % works like the Tibetan HALANT; silent dependent vowel?
collating_symbol <S17B6> % KHMER VOWEL SIGN AA
collating_symbol <S17B7> % KHMER VOWEL SIGN I
collating_symbol <S17B8> % KHMER VOWEL SIGN II
collating_symbol <S17B9> % KHMER VOWEL SIGN Y
collating_symbol <S17BA> % KHMER VOWEL SIGN YY
collating_symbol <S17BB> % KHMER VOWEL SIGN U
collating_symbol <S17BC> % KHMER VOWEL SIGN UU
collating_symbol <S17BD> % KHMER VOWEL SIGN UA
collating_symbol <S17BE> % KHMER VOWEL SIGN OE
collating_symbol <S17BF> % KHMER VOWEL SIGN YA
collating_symbol <S17C0> % KHMER VOWEL SIGN IE
collating_symbol <S17C1> % KHMER VOWEL SIGN E
collating_symbol <S17C2> % KHMER VOWEL SIGN AE
collating_symbol <S17C3> % KHMER VOWEL SIGN AI
collating_symbol <S17C4> % KHMER VOWEL SIGN OO
collating_symbol <S17C5> % KHMER VOWEL SIGN AU


% Nasalised vowels and pseudo-vowels:
collating_symbol <S17BB_S17C6> % KHMER VOWEL SIGN U, KHMER SIGN NIKAHIT: nasalised U
collating_symbol <S17C6> % KHMER SIGN NIKAHIT
collating_symbol <S17B6_S17C6> % KHMER VOWEL SIGN AA, KHMER SIGN NIKAHIT: nasalised AA
collating_symbol <S17C7> % KHMER SIGN REAHMUK (used with (nearly) each of the dependent vowels and nasalised vowels)
collating_symbol <S17C8> % KHMER SIGN YUUKALEAPINTU (yukaleakpintu)



% The COENG, the consonant gluer:
collating_symbol <S17D2> % KHMER SIGN COENG (combining halant; AND makes adjacent Khmer consonant characters conjoining)



% Declararation of contractions
collating_element <U1794_U17C9> from "<U1794><U17C9>" % KHMER LETTER BA, KHMER SIGN MUUSIKATOAN (PA)
collating_element <U1794_U17CA> from "<U1794><U17CA>" % KHMER LETTER BA, KHMER SIGN TRIISAP
collating_element <U17BB_U17C6> from "<U17BB><U17C6>" % KHMER VOWEL SIGN U, KHMER SIGN NIKAHIT: nasalised U
collating_element <U17B6_U17C6> from "<U17B6><U17C6>" % KHMER VOWEL SIGN AA, KHMER SIGN NIKAHIT: nasalised AA
```

```
% Weighting of collation symbols.  Order in this second section is important.

% Consonant modifiers (level 2 weights):
<D17C9> % KHMER SIGN MUUSIKATOAN
<D17CA> % KHMER SIGN TRIISAP
<D17CD> % KHMER SIGN TOANDAKHIAT % marks character not to be pronounced

% Vowel modifiers (level 2 weights):
<D17CB> % KHMER SIGN BANTOC % shortens preceding vowel
<D17CE> % KHMER SIGN KAKABAT % sign used with some exclamations
<D17CF> % KHMER SIGN AHSDA % sign used for single-consonant words
<D17D0> % KHMER SIGN SAMYOK SANNYA % used to indicate shortened inherent vowel


% Consonants:
<S1780> % KHMER LETTER KA
<S1781> % KHMER LETTER KHA
<S1782> % KHMER LETTER KO
<S1783> % KHMER LETTER KHO
<S1784> % KHMER LETTER NGO
<S1785> % KHMER LETTER CA
<S1786> % KHMER LETTER CHA
<S1787> % KHMER LETTER CO
<S1788> % KHMER LETTER CHO
<S1789> % KHMER LETTER NYO
<S178A> % KHMER LETTER DA
<S178B> % KHMER LETTER TTHA
<S178C> % KHMER LETTER DO
<S178D> % KHMER LETTER TTHO
<S178E> % KHMER LETTER NNO
<S178F> % KHMER LETTER TA
<S1790> % KHMER LETTER THA
<S1791> % KHMER LETTER TO
<S1792> % KHMER LETTER THO
<S1793> % KHMER LETTER NO
<S1794> % KHMER LETTER BA
<S1794_S17C9> % KHMER LETTER BA, KHMER SIGN MUUSIKATOAN
<S1794_S17CA> % KHMER LETTER BA, KHMER SIGN TRIISAP
<S1795> % KHMER LETTER PHA
<S1796> % KHMER LETTER PO
<S1797> % KHMER LETTER PHO
<S1798> % KHMER LETTER MO
<S1799> % KHMER LETTER YO
<S179A> % KHMER LETTER RO
<S17AB> % KHMER INDEPENDENT VOWEL RY % glyph based on glyph for 1794
<S17AC> % KHMER INDEPENDENT VOWEL RYY % glyph based on glyph for 1794
```

```
<S179B> % KHMER LETTER LO
<S17AD> % KHMER INDEPENDENT VOWEL LY % glyphs based on glyph for 1796
<S17AE> % KHMER INDEPENDENT VOWEL LYY % glyphs based on glyph for 1796
<S179C> % KHMER LETTER VO
<S179D> % KHMER LETTER SHA
<S179E> % KHMER LETTER SSO
<S179F> % KHMER LETTER SA
<S17A0> % KHMER LETTER HA
<S17A1> % KHMER LETTER LA
<S17A2> % KHMER LETTER QA (glottal stop)


% Weights after all scripts:


% Dependent vowels:
<S17D1> % KHMER SIGN VIRIAM % works like the Tibetan HALANT; silent dependent vowel?
<S17B6> % KHMER VOWEL SIGN AA
<S17B7> % KHMER VOWEL SIGN I
<S17B8> % KHMER VOWEL SIGN II
<S17B9> % KHMER VOWEL SIGN Y
<S17BA> % KHMER VOWEL SIGN YY
<S17BB> % KHMER VOWEL SIGN U
<S17BC> % KHMER VOWEL SIGN UU
<S17BD> % KHMER VOWEL SIGN UA
<S17BE> % KHMER VOWEL SIGN OE
<S17BF> % KHMER VOWEL SIGN YA
<S17C0> % KHMER VOWEL SIGN IE
<S17C1> % KHMER VOWEL SIGN E
<S17C2> % KHMER VOWEL SIGN AE
<S17C3> % KHMER VOWEL SIGN AI
<S17C4> % KHMER VOWEL SIGN OO
<S17C5> % KHMER VOWEL SIGN AU

% Pseudo-vowels and nasalised vowels:
<S17BB_S17C6> % KHMER VOWEL SIGN U, KHMER SIGN NIKAHIT: nasalised U
<S17C6> % KHMER SIGN NIKAHIT
<S17B6_S17C6> % KHMER VOWEL SIGN AA, KHMER SIGN NIKAHIT: nasalised AA
<S17C7> % KHMER SIGN REAHMUK
<S17C8> % KHMER SIGN YUUKALEAPINTU


% The COENG, the consonant gluer:
<S17D2> % KHMER SIGN COENG (combining halant; AND makes adjacent Khmer consonant characters conjoining)
```

```
% Weighting table for Khmer.
% The order in this third section is arbitrary (except for the fourth level weight, which is unimportant),
% but the order used here is, for review purposes, the one implied by the weights as assigned above.


% Characters ignored (on levels 1-3) for collation:

<U17B5> IGNORE;IGNORE;IGNORE;<U17B5> % (glyphless) KHMER VOWEL INHERENT AA
<U17B4> IGNORE;IGNORE;IGNORE;<U17B4> % (glyphless) KHMER VOWEL INHERENT AQ

<U17D3> IGNORE;IGNORE;IGNORE;<U17D3> %  , KHMER SIGN BATHAMASAT
      % very rare sign used in historic lunar dates; MISTAKE IN THE ENCODING;
      % the real PATHAMASAT is not combining, looks different, and has a host of sibling characters.


<U17D4> IGNORE;IGNORE;IGNORE;<U17D4> %  , KHMER SIGN KHAN
      % functions as full stop, ellipsis, abbreviation (can be used to write one of the 'beyyal' abbreviations)

<U17D5> IGNORE;IGNORE;IGNORE;<U17D5> %  , KHMER SIGN BARIYOOSAN  % end of section

<U17D6> IGNORE;IGNORE;IGNORE;<U17D6> %  , KHMER SIGN CAMNUC PII KUUH  % functions as colon or semicolon

<U17D9> IGNORE;IGNORE;IGNORE;<U17D9> %  , KHMER SIGN PHNAEK MUAN  % a list bullet

<U17DA> IGNORE;IGNORE;IGNORE;<U17DA> %  , KHMER SIGN KOOMUUT  % indicates end of book or treatise

<U17DC> IGNORE;IGNORE;IGNORE;<U17DC> %  , KHMER SIGN AVAKRAHASANYA
      % rare, shows a deleted Sanskrit vowel, like an apostrophe


<U17D7> IGNORE;IGNORE;IGNORE;<U17D7> %  , KHMER SIGN LEK TOO % repetition sign


<U17DB> IGNORE;IGNORE;IGNORE;<U17DB> %  , KHMER CURRENCY SYMBOL RIEL
      % [RO with bar; CHANGE: order as other currency signs;
      % in CTT currency signs have primary weights before digits,
      % in EOR currency signs are ignored at levels 1-3]
```

% Consonant modifiers:

<U17C9> IGNORE;<D17C9>;<MIN>;<U17C9> % , KHMER SIGN MUUSIKATOAN

<U17CA> IGNORE;<D17CA>;<MIN>;<U17CA> % , KHMER SIGN TRIISAP

<U17CD> IGNORE;<D17CD>;<MIN>;<U17CD> % , KHMER SIGN TOANDAKHIAT % marks (consonant? independent vowel?) character
% not to be pronounced; is the inherent vowel still pronounced?

% Vowel modifiers:

<U17CB> IGNORE;<D17CB>;<MIN>;<U17CB> % , KHMER SIGN BANTOC % shortens preceding dependent vowel

<U17CE> IGNORE;<D17CE>;<MIN>;<U17CE> % , KHMER SIGN KAKABAT % sign used with some exclamations

<U17CF> IGNORE;<D17CF>;<MIN>;<U17CF> % , KHMER SIGN AHSDA % sign used for single-consonant words

<U17D0> IGNORE;<D17D0>;<MIN>;<U17D0> % , KHMER SIGN SAMYOK SANNYA % used to indicate shortened inherent vowel
% collate as vowel?

% Digits:

<U17E0> <S0030>;"<BASE><KHMER>";"<MIN><MIN>";<U17E0> % , KHMER DIGIT ZERO

<U17E1> <S0031>;"<BASE><KHMER>";"<MIN><MIN>";<U17E1> % , KHMER DIGIT ONE

<U17E2> <S0032>;"<BASE><KHMER>";"<MIN><MIN>";<U17E2> % , KHMER DIGIT TWO

<U17E3> <S0033>;"<BASE><KHMER>";"<MIN><MIN>";<U17E3> % , KHMER DIGIT THREE

<U17E4> <S0034>;"<BASE><KHMER>";"<MIN><MIN>";<U17E4> % , KHMER DIGIT FOUR

<U17E5> <S0035>;"<BASE><KHMER>";"<MIN><MIN>";<U17E5> % , KHMER DIGIT FIVE

<U17E6> <S0036>;"<BASE><KHMER>";"<MIN><MIN>";<U17E6> % , KHMER DIGIT SIX

<U17E7> <S0037>;"<BASE><KHMER>";"<MIN><MIN>";<U17E7> % , KHMER DIGIT SEVEN

<U17E8> <S0038>;"<BASE><KHMER>";"<MIN><MIN>";<U17E8> % , KHMER DIGIT EIGHT

<U17E9> <S0039>;"<BASE><KHMER>";"<MIN><MIN>";<U17E9> % , KHMER DIGIT NINE

% Consonants:

<U1780> <S1780>;<BASE>;<MIN>;<U1780> % , KHMER LETTER KA

<U1781> <S1781>;<BASE>;<MIN>;<U1781> % , KHMER LETTER KHA

<U1782> <S1782>;<BASE>;<MIN>;<U1782> % , KHMER LETTER KO

<U1783> <S1783>;<BASE>;<MIN>;<U1783> % , KHMER LETTER KHO

<U1784> <S1784>;<BASE>;<MIN>;<U1784> % , KHMER LETTER NGO

<U1785> <S1785>;<BASE>;<MIN>;<U1785> % , KHMER LETTER CA

<U1786> <S1786>;<BASE>;<MIN>;<U1786> % , KHMER LETTER CHA

<U1787> <S1787>;<BASE>;<MIN>;<U1787> % , KHMER LETTER CO

<U1788> <S1788>;<BASE>;<MIN>;<U1788> % , KHMER LETTER CHO

<U1789> <S1789>;<BASE>;<MIN>;<U1789> % , KHMER LETTER NYO

<U178A> <S178A>;<BASE>;<MIN>;<U178A> % , KHMER LETTER DA

<U178B> <S178B>;<BASE>;<MIN>;<U178B> % , KHMER LETTER TTHA

<U178C> <S178C>;<BASE>;<MIN>;<U178C> % , KHMER LETTER DO

<U178D> <S178D>;<BASE>;<MIN>;<U178D> % , KHMER LETTER TTHO

<U178E> <S178E>;<BASE>;<MIN>;<U178E> % , KHMER LETTER NNO

<U178F> <S178F>;<BASE>;<MIN>;<U178F> % , KHMER LETTER TA

<U1790> <S1790>;<BASE>;<MIN>;<U1790> % , KHMER LETTER THA

<U1791> <S1791>;<BASE>;<MIN>;<U1791> % , KHMER LETTER TO

<U1792> <S1792>;<BASE>;<MIN>;<U1792> % , KHMER LETTER THO

<U1793> <S1793>;<BASE>;<MIN>;<U1793> % , KHMER LETTER NO

<U1794> <S1794>;<BASE>;<MIN>;<U1794> % , KHMER LETTER BA; note the need for prehandling for the following two:

<U1794_U17C9> **<S1794_S17C9>**;<BASE>;<MIN>;<U1794_U17C9> % , **KHMER LETTER BA,** KHMER SIGN MUUSIKATOAN (PA)

<U1794_U17CA> **<S1794_S17CA>**;<BASE>;<MIN>;<U1794_U17CA> % , **KHMER LETTER BA,** KHMER SIGN TRIISAP

<U1795> <S1795>;<BASE>;<MIN>;<U1795> % , KHMER LETTER PHA

<U1796> <S1796>;<BASE>;<MIN>;<U1796> % , KHMER LETTER PO

<U1797> <S1797>;<BASE>;<MIN>;<U1797> % , KHMER LETTER PHO

<U1798> <S1798>;<BASE>;<MIN>;<U1798> % , KHMER LETTER MO

<U1799> <S1799>;<BASE>;<MIN>;<U1799> % , KHMER LETTER YO

<U179A> <S179A>;<BASE>;<MIN>;<U179A> % , KHMER LETTER RO (lacks inherent vowel)

<U17CC> <S179A>;"<BASE><VRNT1>";"<MIN><MIN>";<U17CC> % , KHMER SIGN ROBAT (combining); note the need for prehandling

% corresponds to [syllable, not word] initial r in Indian loan words, but treated as a diacritic

<U17AB> <S17AB>;<BASE>;<MIN>;<U17AB> % , KHMER INDEPENDENT VOWEL RY % glyph based on glyph for 1794

<U17AC> <S17AC>;<BASE>;<MIN>;<U17AC> % , KHMER INDEPENDENT VOWEL RYY % glyph based on glyph for 1794

<U179B> <S179B>;<BASE>;<MIN>;<U179B> % , KHMER LETTER LO

<U17D8> <S179B>;<BASE>;<MIN>;<U17D8> **% , KHMER SIGN BEYYAL**

% et cetera [ENCODING MISTAKE; don't use this character, spell out the beyyal in the desired (abbreviated) form]

<U17AD> <S17AD>;<BASE>;<MIN>;<U17AD> % , KHMER INDEPENDENT VOWEL LY % glyphs based on glyph for 1796

<U17AE> <S17AE>;<BASE>;<MIN>;<U17AE> % , KHMER INDEPENDENT VOWEL LYY % glyphs based on glyph for 1796

<U179C> <S179C>;<BASE>;<MIN>;<U179C> % , KHMER LETTER VO

<U179D> <S179D>;<BASE>;<MIN>;<U179D> % , KHMER LETTER SHA % used only for Pali/Sanskrit transliteration

<U179E> <S179E>;<BASE>;<MIN>;<U179E> % , KHMER LETTER SSO % used only for Pali/Sanskrit transliteration

<U179F> <S179F>;<BASE>;<MIN>;<U179F> % , KHMER LETTER SA

<U17A0> <S17A0>;<BASE>;<MIN>;<U17A0> % , KHMER LETTER HA

<U17A1> <S17A1>;<BASE>;<MIN>;<U17A1> % , KHMER LETTER LA

<U17A2> <S17A2>;<BASE>;<MIN>;<U17A2> % , KHMER LETTER QA (glottal stop)

% Independent vowels (glottal stop + dependent vowel).
% They are collated as variants of the glottal stop + vowel combination.

<U17A3> <S17A2>;<BASE>;<MIN>;<U17A3> % , KHMER INDEPENDENT VOWEL QAQ

% looks exactly like 17A2 [BOGUS CHARACTER; encoding mistake; use U+17A2 instead;
% differentiated collation should be done via higher level protocols if at all desired]

<U17A4> "<S17A2>**<S17B6>**";"<BASE><BASE>";"<MIN><MIN>";<U17A4> % , KHMER INDEPENDENT VOWEL QAA

% looks exactly like <17A2, 17B6> [BOGUS CHARACTER; encoding mistake; use <U+17A2, U+17B6> instead;
% differentiated collation should be done via higher level protocols if at all desired]

<U17A5> "<S17A2><S17B7>";"<BASE><VRNT1><BASE>";"<MIN><MIN><MIN>";<U17A5> % , KHMER INDEPENDENT VOWEL QI

<U17A6> "<S17A2><S17B8>";"<BASE><VRNT1><BASE>";"<MIN><MIN><MIN>";<U17A6> % , KHMER INDEPENDENT VOWEL QII

<U17A7> "<S17A2><S17BB>";"<BASE><VRNT1><BASE>";"<MIN><MIN><MIN>";<U17A7> % , KHMER INDEPENDENT VOWEL QU

<U17A8> "<S17A2><S17BB>";"<BASE><VRNT**2**><BASE>";"<MIN><MIN><MIN>";<U17A8> % , KHMER INDEPENDENT VOWEL QUK

<U17A9> "<S17A2><S17BC>";"<BASE><VRNT1><BASE>";"<MIN><MIN><MIN>";<U17A9> % , KHMER INDEPENDENT VOWEL QUU

<U17AA> "<S17A2><S17BC>";"<BASE><VRNT2><BASE>";"<MIN><MIN><MIN>";<U17AA> % , KHMER INDEPENDENT VOWEL QUUV

<U17AF> "<S17A2><S17C2>";"<BASE><VRNT1><BASE>";"<MIN><MIN><MIN>";<U17AF> % , KHMER INDEPENDENT VOWEL QE

<U17B0> "<S17A2><S17C3>";"<BASE><VRNT1><BASE>";"<MIN><MIN><MIN>";<U17B0> % , KHMER INDEPENDENT VOWEL QAI

<U17B1> "<S17A2><S17C4>";"<BASE><VRNT1><BASE>";"<MIN><MIN><MIN>";<U17B1> % , KHMER INDEPENDENT VOWEL QOO TYPE ONE

<U17B2> "<S17A2><S17C4>";"<BASE><VRNT2><BASE>";"<MIN><MIN><MIN>";<U17B2> % , KHMER INDEPENDENT VOWEL QOO TYPE TWO

<U17B3> "<S17A2><S17C5>";"<BASE><VRNT1><BASE>";"<MIN><MIN><MIN>";<U17B3> % , KHMER INDEPENDENT VOWEL QAU

% Dependent vowels (in collation order *after* all scripts):

<U17D1> <S17D1>;<BASE>;<MIN>;<U17D1> % , KHMER SIGN VIRIAM % seems to work like a silent dependent vowel

% mostly obsolete**?**
% But does work like the Tibetan HALANT (which is a bit virama-like): silences final inherent vowel.
% indicates the base character is part of the previous word **[indicates end of word]**
% If it's "mostly obsolete", what is nowadays used for spelling words that end in a consonant
% (like Chuon, Samyok, Kakabat, Triisap, ...)? Is it just context dependent if the inherent vowel
% is pronounced or not at the end of a word?

<U17B6> <S17B6>;<BASE>;<MIN>;<U17B6> % , KHMER VOWEL SIGN AA

% "<S17B6><S17C7>";"<BASE><BASE>";"<MIN><MIN>";<U17B6> % , KHMER VOWEL SIGN AA, KHMER SIGN REAHMUK

<U17B7> <S17B7>;<BASE>;<MIN>;<U17B7> % , KHMER VOWEL SIGN I

% "<S17B7><S17C7>";"<BASE><BASE>";"<MIN><MIN>" % , KHMER VOWEL SIGN I, KHMER SIGN REAHMUK

<U17B8> <S17B8>;<BASE>;<MIN>;<U17B8> % , KHMER VOWEL SIGN II

% <S17B8><S17C7>;<BASE><BASE>;<MIN><MIN> % , KHMER VOWEL SIGN II, KHMER SIGN REAHMUK

<U17B9> <S17B9>;<BASE>;<MIN>;<U17B9> % , KHMER VOWEL SIGN Y

% <S17B9><S17C7>;<BASE><BASE>;<MIN> % , KHMER VOWEL SIGN Y, KHMER SIGN REAHMUK

<U17BA> <S17BA>;<BASE>;<MIN>;<U17BA> % , KHMER VOWEL SIGN YY

% <S17BA><S17C7>;<BASE><BASE>;<MIN><MIN> % , KHMER VOWEL SIGN YY, KHMER SIGN REAHMUK

<U17BB> <S17BB>;<BASE>;<MIN>;<U17BB> % , KHMER VOWEL SIGN U

% <S17BB><S17C7>;<BASE><BASE>;<MIN><MIN> % , KHMER VOWEL SIGN U, KHMER SIGN REAHMUK

<U17BC> <S17BC>;<BASE>;<MIN>;<U17BC> % , KHMER VOWEL SIGN UU

% <S17BC><S17C7>;<BASE><BASE>;<MIN> % , KHMER VOWEL SIGN UU, KHMER SIGN REAHMUK

<U17BD> <S17BD>;<BASE>;<MIN>;<U17BD> % , KHMER VOWEL SIGN UA

% <S17BD><S17C7>;<BASE><BASE>;<MIN><MIN> % , KHMER VOWEL SIGN UA, KHMER SIGN REAHMUK

% (editorial note: the Khmer font for the remaining dependent vowels here is incorrect, the left side part is missing; temporary mock-up used)

<U17BE> <S17BE>;<BASE>;<MIN>;<U17BE> % , KHMER VOWEL SIGN OE

% <S17BE><S17C7>;<BASE><BASE>;<MIN><MIN> % , KHMER VOWEL SIGN OE, KHMER SIGN REAHMUK

<U17BF> <S17BF>;<BASE>;<MIN>;<U17BF> % , KHMER VOWEL SIGN YA

% <S17BF><S17C7>;<BASE><BASE>;<MIN> % , KHMER VOWEL SIGN YA, KHMER SIGN REAHMUK

<U17C0> <S17C0>;<BASE>;<MIN>;<U17C0> % , KHMER VOWEL SIGN IE

% <S17C0><S17C7>;<BASE><BASE>;<MIN><MIN> % , KHMER VOWEL SIGN IE, KHMER SIGN REAHMUK

<U17C1> <S17C1>;<BASE>;<MIN>;<U17C1> % , KHMER VOWEL SIGN E

% <S17C1><S17C7>;<BASE><BASE>;<MIN><MIN> % , KHMER VOWEL SIGN E, KHMER SIGN REAHMUK

<U17C2> <S17C2>;<BASE>;<MIN>;<U17C2> % , KHMER VOWEL SIGN AE

% <S17C2><S17C7>;<BASE><BASE>;<MIN><MIN> % , KHMER VOWEL SIGN AE, KHMER SIGN REAHMUK

<U17C3> <S17C3>;<BASE>;<MIN>;<U17C3> % , KHMER VOWEL SIGN AI

% <S17C3><S17C7>;<BASE><BASE>;<MIN><MIN> % , KHMER VOWEL SIGN AI, KHMER SIGN REAHMUK

<U17C4> <S17C4>;<BASE>;<MIN>;<U17C4> % , KHMER VOWEL SIGN OO

% <S17C4><S17C7>;<BASE><BASE>;<MIN><MIN> % , KHMER VOWEL SIGN OO, KHMER SIGN REAHMUK

<U17C5> <S17C5>;<BASE>;<MIN>;<U17C5> % , KHMER VOWEL SIGN AU

% <S17C5><S17C7>;<BASE><BASE>;<MIN><MIN> % , KHMER VOWEL SIGN AU, KHMER SIGN REAHMUK

% Nasalisation pseudo-vowel and reordered nasalisations:

<U17BB_U17C6> **<S17BB_S17C6>**;<BASE>;<MIN>;<U17BB_U17C6> % , KHMER VOWEL SIGN U, KHMER SIGN NIKAHIT: nasalised U

% **<S17BB_S17C6>**<S17C7>;<BASE><BASE>;<MIN><MIN> % , KHMER VOWEL SIGN U, KHMER SIGN NIKAHIT, KHMER SIGN REAHMUK

<U17C6> <S17C6>;<BASE>;<MIN>;<U17C6> % , KHMER SIGN NIKAHIT % , anusvara, final nasalization

<U17B6_U17C6> **<S17B6_S17C6>**;<BASE>;<MIN>;<U17B6_U17C6> % , KHMER VOWEL SIGN AA, KHMER SIGN NIKAHIT: nasalised AA

% **<S17B6_S17C6>**<S17C7>;<BASE><BASE>;<MIN><MIN> % , KHMER VOWEL SIGN AA, KHMER SIGN NIKAHIT, KHMER SIGN REAHMUK

% Pseudo-vowels:

<U17C7> <S17C7>;<BASE>;<MIN>;<U17C7> % , KHMER SIGN REAHMUK % visarga

<U17C8> <S17C8>;<BASE>;<MIN>;<U17C8> % , KHMER SIGN YUUKALEAPINTU

% **makes the inherent vowel short and** with an abrupt glottal stop

% Note that the dependent vowel + REAHMUK combinations need not get
% contraction weighting above, since the proper order results anyway (see comments above).

% The COENG, the consonant gluer:
<U17D2> <S17D2>;<BASE>;<MIN>;<U17D2> % (glyphless) KHMER SIGN COENG (combining; makes certain adjacent characters conjoining)

% functions as virama; note that the VIRIAM character works like the Tibetan HALANT.

% If a combining-under approach had been used instead (or also), the combining-under consonants and

```
% combining-under independent vowels would be ordered here, after the COENG, in the same order as the
% corresponding spacing ones are above. For illustration, the weighting of the underscripts are given here.

% Underscript consonants (for compatibility with the COENG model, a COENG weight is used; this does not affect
% the ordering compared to having separate weights for these, when looking at the encoding models *separately*):

% <Unn80> "<S17D2><S1780>";"<BASE><BASE>";"<MIN><MIN>";<Unn80> % KHMER LETTER COENG KA
% <Unn81> "<S17D2><S1781>";"<BASE><BASE>";"<MIN><MIN>";<Unn81> % KHMER LETTER COENG KHA
% <Unn82> "<S17D2><S1782>";"<BASE><BASE>";"<MIN><MIN>";<Unn82> % KHMER LETTER COENG KO
% <Unn83> "<S17D2><S1783>";"<BASE><BASE>";"<MIN><MIN>";<Unn83> % KHMER LETTER COENG KHO
% <Unn84> "<S17D2><S1784>";"<BASE><BASE>";"<MIN><MIN>";<Unn84> % KHMER LETTER COENG NGO
% <Unn85> "<S17D2><S1785>";"<BASE><BASE>";"<MIN><MIN>";<Unn85> % KHMER LETTER COENG CA
% <Unn86> "<S17D2><S1786>";"<BASE><BASE>";"<MIN><MIN>";<Unn86> % KHMER LETTER COENG CHA
% <Unn87> "<S17D2><S1787>";"<BASE><BASE>";"<MIN><MIN>";<Unn87> % KHMER LETTER COENG CO
% <Unn88> "<S17D2><S1788>";"<BASE><BASE>";"<MIN><MIN>";<Unn88> % KHMER LETTER COENG CHO
% <Unn89> "<S17D2><S1789>";"<BASE><BASE>";"<MIN><MIN>";<Unn89> % KHMER LETTER COENG NYO
% <Unn8A> "<S17D2><S178A>";"<BASE><BASE>";"<MIN><MIN>";<Unn8A> % KHMER LETTER COENG DA
% <Unn8B> "<S17D2><S178B>";"<BASE><BASE>";"<MIN><MIN>";<Unn8B> % KHMER LETTER COENG TTHA
% <Unn8C> "<S17D2><S178C>";"<BASE><BASE>";"<MIN><MIN>";<Unn8C> % KHMER LETTER COENG DO
% <Unn8D> "<S17D2><S178D>";"<BASE><BASE>";"<MIN><MIN>";<Unn8D> % KHMER LETTER COENG TTHO
% <Unn8E> "<S17D2><S178E>";"<BASE><BASE>";"<MIN><MIN>";<Unn8E> % KHMER LETTER COENG NNO
% <Unn8F> "<S17D2><S178F>";"<BASE><BASE>";"<MIN><MIN>";<Unn8F> % KHMER LETTER COENG TA
% <Unn90> "<S17D2><S1790>";"<BASE><BASE>";"<MIN><MIN>";<Unn90> % KHMER LETTER COENG THA
% <Unn91> "<S17D2><S1791>";"<BASE><BASE>";"<MIN><MIN>";<Unn91> % KHMER LETTER COENG TO
% <Unn92> "<S17D2><S1792>";"<BASE><BASE>";"<MIN><MIN>";<Unn92> % KHMER LETTER COENG THO
% <Unn93> "<S17D2><S1793>";"<BASE><BASE>";"<MIN><MIN>";<Unn93> % KHMER LETTER COENG NO
% <Unn94> "<S17D2><S1794>";"<BASE><BASE>";"<MIN><MIN>";<Unn94> % KHMER LETTER COENG BA
% <Unn95> "<S17D2><S1795>";"<BASE><BASE>";"<MIN><MIN>";<Unn95> % KHMER LETTER COENG PHA
% <Unn96> "<S17D2><S1796>";"<BASE><BASE>";"<MIN><MIN>";<Unn96> % KHMER LETTER COENG PO
% <Unn97> "<S17D2><S1797>";"<BASE><BASE>";"<MIN><MIN>";<Unn97> % KHMER LETTER COENG PHO
% <Unn98> "<S17D2><S1798>";"<BASE><BASE>";"<MIN><MIN>";<Unn98> % KHMER LETTER COENG MO
% <Unn99> "<S17D2><S1799>";"<BASE><BASE>";"<MIN><MIN>";<Unn99> % KHMER LETTER COENG YO
% <Unn9A> "<S17D2><S179A>";"<BASE><BASE>";"<MIN><MIN>";<Unn9A> % KHMER LETTER COENG RO (lacks inherent vowel)
% <UnnAB> "<S17D2><S17AB>";"<BASE><BASE>";"<MIN><MIN>";<UnnAB> % KHMER LETTER COENG RY
%%% <UnnAC> "<S17D2><S17AC>";"<BASE><BASE>";"<MIN><MIN>";<UnnAC> % KHMER LETTER COENG RYY
% <Unn9B> "<S17D2><S179B>";"<BASE><BASE>";"<MIN><MIN>";<Unn9B> % KHMER LETTER COENG LO
%%% <UnnAD> "<S17D2><S17AD>";"<BASE><BASE>";"<MIN><MIN>";<UnnAD> % KHMER LETTER COENG LY
%%% <UnnAE> "<S17D2><S17AE>";"<BASE><BASE>";"<MIN><MIN>";<UnnAE> % KHMER LETTER COENG LYY
% <Unn9C> "<S17D2><S179C>";"<BASE><BASE>";"<MIN><MIN>";<Unn9C> % KHMER LETTER COENG VO
% <Unn9D> "<S17D2><S179D>";"<BASE><BASE>";"<MIN><MIN>";<Unn9D> % KHMER LETTER COENG SHA % used only for Pali/Sanskrit
      transliteration
% <Unn9E> "<S17D2><S179E>";"<BASE><BASE>";"<MIN><MIN>";<Unn9E> % KHMER LETTER COENG SSO % used only for Pali/Sanskrit
      transliteration
% <Unn9F> "<S17D2><S179F>";"<BASE><BASE>";"<MIN><MIN>";<Unn9F> % KHMER LETTER COENG SA
% <UnnA0> "<S17D2><S17A0>";"<BASE><BASE>";"<MIN><MIN>";<UnnA0> % KHMER LETTER COENG HA
%%% <UnnA1> "<S17D2><S17A1>";"<BASE><BASE>";"<MIN><MIN>";<UnnA1> % KHMER LETTER COENG LA
```

```
%
% <UnnA2> "<S17D2><S17A2>";"<BASE><BASE>";"<MIN><MIN>";<UnnA2> % KHMER LETTER COENG QA (underscript glottal stop)
%
%
% Underscript independent vowels (ordered as underscript glottal stop + dependent vowel; Hmmm, is that appropriate?).
%%% <UnnA5> "<S17D2><S17A2><S17B7>";"<BASE><BASE><VRNT1><BASE>";"<MIN><MIN><MIN><MIN>";<UnnA5> % KHMER LETTER COENG QI
%%% <UnnA6> "<S17D2><S17A2><S17B8>";"<BASE><BASE><VRNT1><BASE>";"<MIN><MIN><MIN><MIN>";<UnnA6> % KHMER LETTER COENG QII
% <UnnA7> "<S17D2><S17A2><S17BB>";"<BASE><BASE><VRNT1><BASE>";"<MIN><MIN><MIN><MIN>";<UnnA7> % KHMER LETTER COENG QU
%%% <UnnA8> "<S17D2><S17A2><S17BB>";"<BASE><BASE><VRNT2><BASE>";"<MIN><MIN><MIN><MIN>";<UnnA8> % KHMER LETTER COENG QUK
%%% <UnnA9> "<S17D2><S17A2><S17BC>";"<BASE><BASE><VRNT1><BASE>";"<MIN><MIN><MIN><MIN>";<UnnA9> % KHMER LETTER COENG QUU
%%% <UnnAA> "<S17D2><S17A2><S17BC>";"<BASE><BASE><VRNT2><BASE>";"<MIN><MIN><MIN><MIN>";<UnnAA> % KHMER LETTER COENG QUUV
% <UnnAF> "<S17D2><S17A2><S17C2>";"<BASE><BASE><VRNT1><BASE>";"<MIN><MIN><MIN><MIN>";<UnnAF> % KHMER LETTER COENG QE
%%% <UnnB0> "<S17D2><S17A2><S17C3>";"<BASE><BASE><VRNT1><BASE>";"<MIN><MIN><MIN><MIN>";<UnnB0> % KHMER LETTER COENG QAI
%%% <UnnB1> "<S17D2><S17A2><S17C4>";"<BASE><BASE><VRNT1><BASE>";"<MIN><MIN><MIN><MIN>";<UnnB1> % KHMER LETTER COENG QOO
      TYPE ONE
%%% <UnnB2> "<S17D2><S17A2><S17C4>";"<BASE><BASE><VRNT2><BASE>";"<MIN><MIN><MIN><MIN>";<UnnB2> % KHMER LETTER COENG QOO
      TYPE TWO
%%% <UnnB3> "<S17D2><S17A2><S17C5>";"<BASE><BASE><VRNT1><BASE>";"<MIN><MIN><MIN><MIN>";<UnnB3> % KHMER LETTER COENG QAU
```

6 Acknowledgements

Thanks to Maurice Bauhahn for explaining the principles of Khmer collation. Any errors or shortcomings here are of course mine (especially since I've done some interpretations and changes).

7 References

ISO/IEC 10646-1:2000	Information Technology – Universal multiple-octet coded character set (UCS), Part 1, second edition.
Unicode 3.0	The Unicode standard, version 3.0.
UCD 3.1	Unicode character database, version 3.1.
ISO/IEC 14651:2001	International string ordering and comparison – Method for comparing character strings and description of the common template tailorable ordering.
UTS 10	Unicode technical standard 10, Unicode collation algorithm.

ISO/IEC JTC 1/SC22/WG20 N891R	Kent Karlsson, Hangul ordering rules. 2001-11-29.
ISO/IEC JTC 1/SC2/WG2 N2405R	Same as ISO/IEC JTC 1/SC22/WG20 N891R.
L2/01-469	Same as ISO/IEC JTC 1/SC22/WG20 N891R.
Choun Nath's	Chuon Nath's Khmer—Khmer dictionary.
ISO/IEC JTC 1/SC2/WG2 N2380R	Cambodian official objection to the existing Khmer block in UCS.
ISO/IEC JTC 1/SC2/WG2 N2385	Response to Cambodian official objection to the existing Khmer block in UCS (N2380).
ISO/IEC JTC 1/SC2/WG2 N2406	Response to WG2 document N2385.
ISO/IEC JTC 1/SC2/WG2 N2nnn	[response to that...]
	Khmer ordering analysis. Maurice Bauhahn. http://www.bauhahnm.clara.net/Khmer/KhmerSortingUnicodegamma.pdf.
	Segmental collation and the UCA. Mark Davis. "Very draft", 2001-11-16.

--- end ---