

Ordering rules for Khmer

Kent Karlsson  
2003-10-13

## 1 Introduction

The Khmer script in Unicode/10646 uses the virama model, like the Indic scripts. An alternative that would have been suitable, would have been to have combining-below (and sometimes a bit to the side) consonants and combining-below independent vowels (compare how Tibetan and Limbu are handled). However, that is not the chosen solution. The chosen solution is instead to use a combining character (COENG) that forms 'conjuncts' like it is done for Indic (Brahmic) scripts. The suitability of this is debatable, but is now not possible to change.

In the Khmer script, which does not use spaces between words, using the COENG based approach, the words are formed from orthographic syllables, where an orthographic syllable has the following structure:

Khmer-syllable ::= K N? (G K)\* A? 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, **G** is the invisible Khmer conjoint former COENG, **N** is a combining character that is a Khmer Sign Robot or a Khmer consonant modifier ("register shift"), **A** is a dependent Khmer vowel, VIRIAM, YUUKALEAPINTU, ATTHACAN, TOANDAKHIAT, or SAMYOK SANNA, and **M** is some other combining character, in particular a vowel modifier sign (which should be first, if any, among the **M** for a syllable). 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). In Khmer the non-first consonants/independent vowels in an orthographic syllable are typographically underscripted, with a modified glyph.

## 2 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).

However, Khmer has an 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. Indeed, the vowels and the COENG should be ordered after all scripts, so that the clustering is maintained also when other scripts may be involved in a string.

For instance, <KA, AA>, should come before <KA, COENG, KA>. So COENG must be heavier than any dependent vowel. In addition, e.g., <KA, U+3400> should come before <KA, AA, U+3400>. So AA, and other dependent vowel signs, must be heavier than any "independent" character, regardless of script.

To make the collation keys a bit shorter, the COENG+consonant (or COENG+independent vowel) pairs can get contracted weightings (these are not given below, since they are not necessary for getting the desired ordering). 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 (see example above).

One way of ordering most of the independent vowels in Khmer 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 at level 2. 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 the fact 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, as well as recommended not to be used, and should be ignored in collation, as they are in the rules 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, contractions are defined in the ordering rules below, for consonants and independent vowels directly followed, in the character stream, by a ROBAT. The rules for the contractions then put the weight for the ROBAT before the weight for the consonant or independent vowel it is applied to.

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

Even though most applications of the consonant modifiers only have an effect on the second collation level, their applications to BA are sometimes ordered as separate letters. That could be done via contractions, but is commented out below, since these would be too much of an exception to a general rule, and does not appear to be fully established.

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.

### 3 Khmer collation weighting rules

```
%%% Note that Khmer characters added after version 3.2 of the Unicode standard, apart from
%%% ATTHACAN, are not included below. A complete update for Khmer should of course include
%%% those characters too, but it is not the centre of this proposal.
```

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

**% Modifiers (level 2 weights):**

**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 <D17D1>** % **KHMER SIGN VIRIAM % works like the Tibetan HALANT and Thai YAMAKKAN**

**collating\_symbol <D17D0>** % KHMER SIGN SAMYOK SANNYA % used to indicate shortened inherent vowel (order as vowel?)

**collating\_symbol <D17C8>** % KHMER SIGN YUUKALEAPINTU (yukaleapintu)

**collating\_symbol <D17DD>** % KHMER SIGN ATTHACAN

**collating\_symbol <D17CB>** % KHMER SIGN BANTOC (bantak) % shortens preceding dependent vowel

**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 (cancelled)

**% 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**

*%% the following two are NOT weighted as separate letters, since it is NOT definite from Choun Nath's,*

*%% and in addition would complicate automatic generation of rules for Khmer (via the "sifter" program):*

**%% 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, nasalised vowels, and pseudo-vowels:**

```

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
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)

```

**% The COENG, the consonant gluer:**

```

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

```

**% Declaration 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 <U1780\_U17CC> from "<U1780><U17CC>" % , KHMER LETTER KA; , KHMER SIGN ROBAT

collating\_element <U1781\_U17CC> from "<U1781><U17CC>" % , KHMER LETTER KHA; , KHMER SIGN ROBAT

collating\_element <U1782\_U17CC> from "<U1782><U17CC>" % , KHMER LETTER KO; , KHMER SIGN ROBAT

collating\_element <U1783\_U17CC> from "<U1783><U17CC>" % , KHMER LETTER KHO; , KHMER SIGN ROBAT

collating\_element <U1784\_U17CC> from "<U1784><U17CC>" % , KHMER LETTER NGO; , KHMER SIGN ROBAT

collating\_element <U1785\_U17CC> from "<U1785><U17CC>" % , KHMER LETTER CA; , KHMER SIGN ROBAT

collating\_element <U1786\_U17CC> from "<U1786><U17CC>" % , KHMER LETTER CHA; , KHMER SIGN ROBAT

collating\_element <U1787\_U17CC> from "<U1787><U17CC>" % , KHMER LETTER CO; , KHMER SIGN ROBAT

collating\_element <U1788\_U17CC> from "<U1788><U17CC>" % , KHMER LETTER CHO; , KHMER SIGN ROBAT

collating\_element <U1789\_U17CC> from "<U1789><U17CC>" % , KHMER LETTER NYO; , KHMER SIGN ROBAT

collating\_element <U178A\_U17CC> from "<U178A><U17CC>" % , KHMER LETTER DA; , KHMER SIGN ROBAT

collating\_element <U178B\_U17CC> from "<U178B><U17CC>" % , KHMER LETTER TTHA; , KHMER SIGN ROBAT

collating\_element <U178C\_U17CC> from "<U178C><U17CC>" % , KHMER LETTER DO; , KHMER SIGN ROBAT

collating\_element <U178D\_U17CC> from "<U178D><U17CC>" % , KHMER LETTER TTHO; , KHMER SIGN ROBAT

collating\_element <U178E\_U17CC> from "<U178E><U17CC>" % , KHMER LETTER NNO; , KHMER SIGN ROBAT

collating\_element <U178F\_U17CC> from "<U178F><U17CC>" % , KHMER LETTER TA; , KHMER SIGN ROBAT

collating\_element <U1790\_U17CC> from "<U1790><U17CC>" % , KHMER LETTER THA; , KHMER SIGN ROBAT

collating\_element <U1791\_U17CC> from "<U1791><U17CC>" % , KHMER LETTER TO; , KHMER SIGN ROBAT

collating\_element <U1792\_U17CC> from "<U1792><U17CC>" % , KHMER LETTER THO; , KHMER SIGN ROBAT

collating\_element <U1793\_U17CC> from "<U1793><U17CC>" % , KHMER LETTER NO; , KHMER SIGN ROBAT

collating\_element <U1794\_U17CC> from "<U1794><U17CC>" % , KHMER LETTER BA; , KHMER SIGN ROBAT

collating\_element <U1795\_U17CC> from "<U1795><U17CC>" % , KHMER LETTER PHA; , KHMER SIGN ROBAT

collating\_element <U1796\_U17CC> from "<U1796><U17CC>" % , KHMER LETTER PO; , KHMER SIGN ROBAT

collating\_element <U1797\_U17CC> from "<U1797><U17CC>" % , KHMER LETTER PHO; , KHMER SIGN ROBAT

collating\_element <U1798\_U17CC> from "<U1798><U17CC>" % , KHMER LETTER MO; , KHMER SIGN ROBAT

collating\_element <U1799\_U17CC> from "<U1799><U17CC>" % , KHMER LETTER YO; , KHMER SIGN ROBAT

collating\_element <U179A\_U17CC> from "<S179A><U17CC>" % , KHMER LETTER RO; , KHMER SIGN ROBAT

collating\_element <U17AB\_U17CC> from "<U17AB><U17CC>" % , KHMER INDEPENDENT VOWEL RY; , KHMER SIGN ROBAT

collating\_element <U17AC\_U17CC> from "<U17AC><U17CC>" % , KHMER INDEPENDENT VOWEL RYY; , KHMER SIGN ROBAT

collating\_element <U179B\_U17CC> from "<U179B><U17CC>" % , KHMER LETTER LO; , KHMER SIGN ROBAT

collating\_element <U17AD\_U17CC> from "<U17AD><U17CC>" % , KHMER INDEPENDENT VOWEL LY; , KHMER SIGN ROBAT

collating\_element <U17AE\_U17CC> from "<U17AE><U17CC>" % , KHMER INDEPENDENT VOWEL LYY; , KHMER SIGN ROBAT

collating\_element <U179C\_U17CC> from "<U179C><U17CC>" % , KHMER LETTER VO; , KHMER SIGN ROBAT

collating\_element <U179D\_U17CC> from "<U179D><U17CC>" % , KHMER LETTER SHA; , KHMER SIGN ROBAT

collating\_element <U179E\_U17CC> from "<U179E><U17CC>" % , KHMER LETTER SSO; , KHMER SIGN ROBAT

```

collating_element <U179F_U17CC> from "<U179F><U17CC>" % , KHMER LETTER SA; , KHMER SIGN ROBAT
collating_element <U17A0_U17CC> from "<U17A0><U17CC>" % , KHMER LETTER HA; , KHMER SIGN ROBAT
collating_element <U17A1_U17CC> from "<U17A1><U17CC>" % , KHMER LETTER LA; , KHMER SIGN ROBAT
collating_element <U17A2_U17CC> from "<U17A2><U17CC>" % , KHMER LETTER QA (glottal stop); , KHMER SIGN ROBAT

```

```

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

```

```

collating_element <U17A5_U17CC> from "<U17A5><U17CC>" % , KHMER INDEPENDENT VOWEL QI; , KHMER SIGN ROBAT
collating_element <U17A6_U17CC> from "<U17A6><U17CC>" % , KHMER INDEPENDENT VOWEL QII; , KHMER SIGN ROBAT
collating_element <U17A7_U17CC> from "<U17A7><U17CC>" % , KHMER INDEPENDENT VOWEL QU; , KHMER SIGN ROBAT
collating_element <U17A8_U17CC> from "<U17A8><U17CC>" % , KHMER INDEPENDENT VOWEL QUK; , KHMER SIGN ROBAT
collating_element <U17A9_U17CC> from "<U17A9><U17CC>" % , KHMER INDEPENDENT VOWEL QUU; , KHMER SIGN ROBAT
collating_element <U17AA_U17CC> from "<U17AA><U17CC>" % , KHMER INDEPENDENT VOWEL QUUV; , KHMER SIGN ROBAT
collating_element <U17AF_U17CC> from "<U17AF><U17CC>" % , KHMER INDEPENDENT VOWEL QE; , KHMER SIGN ROBAT
collating_element <U17B0_U17CC> from "<U17B0><U17CC>" % , KHMER INDEPENDENT VOWEL QAI; , KHMER SIGN ROBAT
collating_element <U17B1_U17CC> from "<U17B1><U17CC>" % , KHMER INDEPENDENT VOWEL QOO TYPE ONE; , KHMER SIGN ROBAT
collating_element <U17B2_U17CC> from "<U17B2><U17CC>" % , KHMER INDEPENDENT VOWEL QOO TYPE TWO; , KHMER SIGN ROBAT
collating_element <U17B3_U17CC> from "<U17B3><U17CC>" % , KHMER INDEPENDENT VOWEL QAU; , KHMER SIGN ROBAT

collating_element <U17C6_U17BB> from "<U17BB><U17C6>" % KHMER VOWEL SIGN U, KHMER SIGN NIKAHIT: nasalised U
collating_element <U17BB_U17C6> from "<U17C6><U17BB>" % KHMER SIGN NIKAHIT, KHMER VOWEL SIGN U: nasalised U
collating_element <U17C6_U17B6> from "<U17B6><U17C6>" % KHMER VOWEL SIGN AA, KHMER SIGN NIKAHIT: nasalised AA
collating_element <U17B6_U17C6> from "<U17C6><U17B6>" % KHMER SIGN NIKAHIT, KHMER VOWEL SIGN AA: nasalised AA

```

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

**% Modifiers (level 2 weights):**

<D17CE> % KHMER SIGN KAKABAT % sign used with some exclamations

<D17CF> % KHMER SIGN AHSDA % sign used for single-consonant words

<D17D1> % **KHMER SIGN VIRIAM % works a bit like the Thai YAMAKKAN**

<D17D0> % KHMER SIGN SAMYOK SANNYA % used to indicate shortened inherent vowel (order as vowel?)

<D17C8> % KHMER SIGN YUUKALEAPINTU

<D17DD> % KHMER SIGN ATTHACAN

<D17CB> % KHMER SIGN BANTOC % shortens preceding dependent vowel

<D17C9> % KHMER SIGN MUUSIKATOAN

<D17CA> % KHMER SIGN TRIISAP

<D17CD> % KHMER SIGN TOANDAKHIAT % marks character not to be pronounced

**% 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, pseudo-vowels and nasalised vowels:**

```

<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
<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

```

% The COENG, the consonant gluer (should be weighted among viramas):  
<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; these three characters are MISTAKES IN THE ENCODING;  
% the real PATHAMASAT is not combining, looks different, and has a host of sibling characters.

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

<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

<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]

**% Modifiers:**

<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

<U17D1> IGNORE;<D17D1>;<MIN>;<U17D1> % , KHMER SIGN VIRIAM

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

<U17C8> IGNORE;<D17C8>;<MIN>;<U17C8> % , KHMER SIGN YUUKALEAPINTU

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

<U17DD> IGNORE;<D17DD>;<MIN>;<U17DD> % KHMER SIGN ATTHACAN

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

<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 character not to be pronounced (cancelled)

**% 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**  
 %% <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)  
 % corresponds to [syllable, not word] initial r in Indic loan words, but treated as a diacritic  
 <U1780\_U17CC> --<S1780>;<BASE><VRNT1><BASE><BASE>;<MIN><MIN><MIN><MIN>;<U1780\_U17CC> % , KHMER LETTER KA; , KHMER  
 SIGN ROBAT

<U1781\_U17CC> "<S179A><S17D2><S1781>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U1781\_U17CC> % , KHMER LETTER  
KHA; , KHMER SIGN ROBAT

<U1782\_U17CC> "<S179A><S17D2><S1782>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U1782\_U17CC> % , KHMER LETTER  
KO; , KHMER SIGN ROBAT

<U1783\_U17CC> "<S179A><S17D2><S1783>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U1783\_U17CC> % , KHMER  
LETTER KHO; , KHMER SIGN ROBAT

<U1784\_U17CC> "<S179A><S17D2><S1784>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U1784\_U17CC> % , KHMER LETTER  
NGO; , KHMER SIGN ROBAT

<U1785\_U17CC> "<S179A><S17D2><S1785>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U1785\_U17CC> % , KHMER LETTER  
CA; , KHMER SIGN ROBAT

<U1786\_U17CC> "<S179A><S17D2><S1786>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U1786\_U17CC> % , KHMER LETTER  
CHA; , KHMER SIGN ROBAT

<U1787\_U17CC> "<S179A><S17D2><S1787>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U1787\_U17CC> % , KHMER LETTER  
CO; , KHMER SIGN ROBAT

<U1788\_U17CC> "<S179A><S17D2><S1788>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U1788\_U17CC> % , KHMER  
LETTER CHO; , KHMER SIGN ROBAT

<U1789\_U17CC> "<S179A><S17D2><S1789>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U1789\_U17CC> % , KHMER  
LETTER NYO; , KHMER SIGN ROBAT

<U178A\_U17CC> "<S179A><S17D2><S178A>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U178A\_U17CC> % , KHMER LETTER  
DA; , KHMER SIGN ROBAT

<U178B\_U17CC> "<S179A><S17D2><S178B>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U178B\_U17CC> % , KHMER LETTER  
TTHA; , KHMER SIGN ROBAT

<U178C\_U17CC> "<S179A><S17D2><S178C>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U178C\_U17CC> % , KHMER LETTER  
DO; , KHMER SIGN ROBAT

<U178D\_U17CC> "<S179A><S17D2><S178D>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U178D\_U17CC> % , KHMER  
LETTER TTHO; , KHMER SIGN ROBAT

<U178E\_U17CC> "<S179A><S17D2><S178E>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U178E\_U17CC> % , KHMER  
LETTER NNO; , KHMER SIGN ROBAT

<U178F\_U17CC> "<S179A><S17D2><S178F>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U178F\_U17CC> % , KHMER LETTER  
TA; , KHMER SIGN ROBAT

<U1790\_U17CC> "<S179A><S17D2><S1790>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U1790\_U17CC> % , KHMER LETTER  
THA; , KHMER SIGN ROBAT

<U1791\_U17CC> "<S179A><S17D2><S1791>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U1791\_U17CC> % , KHMER LETTER  
TO; , KHMER SIGN ROBAT

<U1792\_U17CC> "<S179A><S17D2><S1792>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U1792\_U17CC> % , KHMER LETTER  
THO; , KHMER SIGN ROBAT

<U1793\_U17CC> "<S179A><S17D2><S1793>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U1793\_U17CC> % , KHMER LETTER  
NO; , KHMER SIGN ROBAT

<U1794\_U17CC> "<S179A><S17D2><S1794>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U1794\_U17CC> % , KHMER LETTER  
BA; , KHMER SIGN ROBAT

%% <U1794\_U17C9\_U17CC> "<S179A><S17D2><S1794\_S17C9>;"; "<BASE><VRNT1><BASE>;"; "<MIN><MIN><MIN><MIN>;"; <U1794\_U17C9\_U17CC>  
 % , KHMER LETTER BA, KHMER SIGN MUUSIKATOAN (PA); , KHMER SIGN ROBAT

%% <U1794\_U17CA\_U17CC> "<S179A><S17D2><S1794\_S17CA>;"; "<BASE><VRNT1><BASE>;"; "<MIN><MIN><MIN><MIN>;"; <U1794\_U17CA\_U17CC>  
 % , KHMER LETTER BA, KHMER SIGN TRIISAP; , KHMER SIGN ROBAT

<U1795\_U17CC> "<S179A><S17D2><S1795>;"; "<BASE><VRNT1><BASE><BASE>;"; "<MIN><MIN><MIN><MIN>;"; <U1795\_U17CC> % , KHMER LETTER  
 PHA; , KHMER SIGN ROBAT

<U1796\_U17CC> "<S179A><S17D2><S1796>;"; "<BASE><VRNT1><BASE><BASE>;"; "<MIN><MIN><MIN><MIN>;"; <U1796\_U17CC> % , KHMER LETTER  
 PO; , KHMER SIGN ROBAT

<U1797\_U17CC> "<S179A><S17D2><S1797>;"; "<BASE><VRNT1><BASE><BASE>;"; "<MIN><MIN><MIN><MIN>;"; <U1797\_U17CC> % , KHMER LETTER  
 PHO; , KHMER SIGN ROBAT

<U1798\_U17CC> "<S179A><S17D2><S1798>;"; "<BASE><VRNT1><BASE><BASE>;"; "<MIN><MIN><MIN><MIN>;"; <U1798\_U17CC> % , KHMER LETTER  
 MO; , KHMER SIGN ROBAT

<U1799\_U17CC> "<S179A><S17D2><S1799>;"; "<BASE><VRNT1><BASE><BASE>;"; "<MIN><MIN><MIN><MIN>;"; <U1799\_U17CC> % , KHMER  
 LETTER YO; , KHMER SIGN ROBAT

<U179A\_U17CC> "<S179A><S17D2><S179A>;"; "<BASE><VRNT1><BASE><BASE>;"; "<MIN><MIN><MIN><MIN>;"; <U179A\_U17CC> % , KHMER LETTER  
 RO; , KHMER SIGN ROBAT

<U17AB\_U17CC> "<S179A><S17D2><S17AB>;"; "<BASE><VRNT1><BASE><BASE>;"; "<MIN><MIN><MIN><MIN>;"; <U17AB\_U17CC> % , KHMER  
 INDEPENDENT VOWEL RY; , KHMER SIGN ROBAT

<U17AC\_U17CC> "<S179A><S17D2><S17AC>;"; "<BASE><VRNT1><BASE><BASE>;"; "<MIN><MIN><MIN><MIN>;"; <U17AC\_U17CC> % , KHMER  
 INDEPENDENT VOWEL RYY; , KHMER SIGN ROBAT

<U179B\_U17CC> "<S179A><S17D2><S179B>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U179B\_U17CC> % , KHMER  
 LETTER LO; , KHMER SIGN ROBAT

<U17AD\_U17CC> "<S179A><S17D2><S17AD>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U17AD\_U17CC> % , KHMER  
 INDEPENDENT VOWEL LY; , KHMER SIGN ROBAT

<U17AE\_U17CC> "<S179A><S17D2><S17AE>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U17AE\_U17CC> % , KHMER  
 INDEPENDENT VOWEL LYY; , KHMER SIGN ROBAT

<U179C\_U17CC> "<S179A><S17D2><S179C>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U179C\_U17CC> % , KHMER LETTER  
 VO; , KHMER SIGN ROBAT

<U179D\_U17CC> "<S179A><S17D2><S179D>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U179D\_U17CC> % , KHMER LETTER  
 SHA; , KHMER SIGN ROBAT

<U179E\_U17CC> "<S179A><S17D2><S179E>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U179E\_U17CC> % , KHMER LETTER  
 SSO; , KHMER SIGN ROBAT

<U179F\_U17CC> "<S179A><S17D2><S179F>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U179F\_U17CC> % , KHMER  
 LETTER SA; , KHMER SIGN ROBAT

<U17A0\_U17CC> "<S179A><S17D2><S17A0>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U17A0\_U17CC> % , KHMER  
 LETTER HA; , KHMER SIGN ROBAT

<U17A1\_U17CC> "<S179A><S17D2><S17A1>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U17A1\_U17CC> % , KHMER  
 LETTER LA; , KHMER SIGN ROBAT

<U17A2\_U17CC> "<S179A><S17D2><S17A2>"; "<BASE><VRNT1><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U17A2\_U17CC> % , KHMER LETTER  
 QA (glottal stop); , KHMER SIGN ROBAT

% Independent vowels (glottal stop + dependent vowel).

% They are collated as variants of the glottal stop + vowel combination.

<U17A5\_U17CC>

"<S179A><S17D2><S17A2><S17B7>"; "<BASE><VRNT1><BASE><BASE><VRNT1><BASE>"; "<MIN><MIN><MIN><MIN><MIN><MIN>"; <U17A5\_U

17CC> % , KHMER INDEPENDENT VOWEL QI; , KHMER SIGN ROBAT

<U17A6\_U17CC>

"<S179A><S17D2><S17A2><S17B8>"; "<BASE><VRNT1><BASE><BASE><VRNT1><BASE>"; "<MIN><MIN><MIN><MIN><MIN><MIN>"; <U17A6\_U

17CC> % , KHMER INDEPENDENT VOWEL QII; , KHMER SIGN ROBAT

<U17A7\_U17CC>

"<S179A><S17D2><S17A2><S17BB>"; "<BASE><VRNT1><BASE><BASE><VRNT1><BASE>"; "<MIN><MIN><MIN><MIN><MIN><MIN>"; <U17A7\_U

17CC> % , KHMER INDEPENDENT VOWEL QU; , KHMER SIGN ROBAT

<U17A8\_U17CC>

"<S179A><S17D2><S17A2><S17BB>"; "<BASE><VRNT1><BASE><BASE><VRNT2><BASE>"; "<MIN><MIN><MIN><MIN><MIN><MIN>"; <U17A8\_U

17CC> % , KHMER INDEPENDENT VOWEL QUK; , KHMER SIGN ROBAT

<U17A9\_U17CC>

"<S179A><S17D2><S17A2><S17BC>"; "<BASE><VRNT1><BASE><BASE><VRNT1><BASE>"; "<MIN><MIN><MIN><MIN><MIN><MIN>"; <U17A9\_U

17CC> % , KHMER INDEPENDENT VOWEL QUU; , KHMER SIGN ROBAT

<U17AA\_U17CC>

"<S179A><S17D2><S17A2><S17BC>"; "<BASE><VRNT1><BASE><BASE><VRNT2><BASE>"; "<MIN><MIN><MIN><MIN><MIN><MIN>"; <U17AA\_U

17CC> % , KHMER INDEPENDENT VOWEL QUUV; , KHMER SIGN ROBAT

<U17AF\_U17CC>

"<S179A><S17D2><S17A2><S17C2>"; "<BASE><VRNT1><BASE><BASE><VRNT1><BASE>"; "<MIN><MIN><MIN><MIN><MIN><MIN>"; <U17AF\_U

17CC> % , KHMER INDEPENDENT VOWEL QE; , KHMER SIGN ROBAT

<U17B0\_U17CC>

"<S179A><S17D2><S17A2><S17C3>"; "<BASE><VRNT1><BASE><BASE><VRNT1><BASE>"; "<MIN><MIN><MIN><MIN><MIN><MIN>"; <U17B0\_U

17CC> % , KHMER INDEPENDENT VOWEL QAI; , KHMER SIGN ROBAT

<U17B1\_U17CC>

"<S179A><S17D2><S17A2><S17C4>"; "<BASE><VRNT1><BASE><BASE><VRNT1><BASE>"; "<MIN><MIN><MIN><MIN><MIN><MIN>"; <U17B1\_U

17CC> % , KHMER INDEPENDENT VOWEL QOO TYPE ONE; , KHMER SIGN ROBAT

<U17B2\_U17CC>

"<S179A><S17D2><S17A2><S17C4>"; "<BASE><VRNT1><BASE><BASE><VRNT2><BASE>"; "<MIN><MIN><MIN><MIN><MIN><MIN>"; <U17B2\_U

17CC> % , KHMER INDEPENDENT VOWEL QOO TYPE TWO; , KHMER SIGN ROBAT

```

<U17B3_u17cc>
  "<S179A><S17D2><S17A2><S17C5>";"<BASE><VRNT1><BASE><BASE><VRNT1><BASE>";"<MIN><MIN><MIN><MIN><MIN><MIN>";<U17B3_u
    17cc> % , KHMER INDEPENDENT VOWEL QAU; , KHMER SIGN ROBAT

<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>;<COMPAT>;<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>;<COMPAT>;<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>";"<COMPAT><COMPAT>";<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><S1780>"; "<BASE><VRNT2><BASE><BASE>"; "<MIN><MIN><MIN><MIN>"; <U17A8> % , KHMER INDEPENDENT VOWEL  
QUK (should that be "<S17A2><S17BB><S17D2><S1780>" instead?)**  
<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

%%  
After all scripts; among "Han heavy seconds", Hangul trail consonants, Hangul vowels, and Indic dependent vowels.  
%%

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

<U17B6> <S17B6>;<BASE>;<MIN>;<U17B6> % , KHMER VOWEL SIGN AA  
<U17B7> <S17B7>;<BASE>;<MIN>;<U17B7> % , KHMER VOWEL SIGN I  
<U17B8> <S17B8>;<BASE>;<MIN>;<U17B8> % , KHMER VOWEL SIGN II  
<U17B9> <S17B9>;<BASE>;<MIN>;<U17B9> % , KHMER VOWEL SIGN Y  
<U17BA> <S17BA>;<BASE>;<MIN>;<U17BA> % , KHMER VOWEL SIGN YY  
<U17BB> <S17BB>;<BASE>;<MIN>;<U17BB> % , KHMER VOWEL SIGN U  
<U17BC> <S17BC>;<BASE>;<MIN>;<U17BC> % , KHMER VOWEL SIGN UU  
<U17BD> <S17BD>;<BASE>;<MIN>;<U17BD> % , KHMER VOWEL SIGN UA

*% (editorial note: the Khmer font for the remaining dependent vowels here is incorrect, the left side part  
% is missing; temporary mock-up used, which will result in erroneous glyphs when using a correct Khmer font)*

<U17BE> <S17BE>;<BASE>;<MIN>;<U17BE> % , KHMER VOWEL SIGN OE  
<U17BF> <S17BF>;<BASE>;<MIN>;<U17BF> % , KHMER VOWEL SIGN YA  
<U17C0> <S17C0>;<BASE>;<MIN>;<U17C0> % , KHMER VOWEL SIGN IE  
<U17C1> <S17C1>;<BASE>;<MIN>;<U17C1> % , KHMER VOWEL SIGN E  
<U17C2> <S17C2>;<BASE>;<MIN>;<U17C2> % , KHMER VOWEL SIGN AE  
<U17C3> <S17C3>;<BASE>;<MIN>;<U17C3> % , KHMER VOWEL SIGN AI  
<U17C4> <S17C4>;<BASE>;<MIN>;<U17C4> % , KHMER VOWEL SIGN OO  
<U17C5> <S17C5>;<BASE>;<MIN>;<U17C5> % , KHMER VOWEL SIGN AU

**% Nasalisation pseudo-vowel and reordered nasalizations:**

```

<U17BB_U17C6> <S17BB_S17C6>;<BASE>;<MIN>;<U17BB_U17C6> % , KHMER VOWEL SIGN U, KHMER SIGN NIKAHIT: nasalised U
<U17C6_U17BB> <S17BB_S17C6>;<BASE>;<MIN>;<U17C6_U17BB> % , KHMER SIGN NIKAHIT, KHMER VOWEL SIGN U: nasalised U
<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
<U17C6_U17B6> <S17B6_S17C6>;<BASE>;<MIN>;<U17C6_U17B6> % , KHMER SIGN NIKAHIT, KHMER VOWEL SIGN AA: nasalised AA

% Pseudo-vowel:
<U17C7> <S17C7>;<BASE>;<MIN>;<U17C7> % , KHMER SIGN REAHMUK % visarga

% Note that the dependent vowel + REAHMUK combinations need not get contractions
% weightings above, since the proper order results from the given weighting anyway.

% The COENG, the consonant gluer (order among other viramas):
<U17D2> <S17D2>;<BASE>;<MIN>;<U17D2> % KHMER SIGN COENG (combining; makes certain adjacent characters conjoining)
% glyphless; functions as virama; note that the VIRIAM character seems to work like the Thai YAMAKKAN.

```

## 4 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).

## 5 References

- |                             |   |
|-----------------------------|---|
| <i>ISO/IEC 10646-1:2000</i> | Information Technology – Universal multiple-octet coded character set (UCS), Part 1, second edition.  |
| <i>Unicode 4.0</i>          | The Unicode standard, version 4.0.  |
| <i>UCD 4.0.0</i>            | Unicode character database, version 4.0.0.  |
| <i>ISO/IEC 14651:2001</i>   | International string ordering and comparison – Method for comparing character strings and description of the common template tailorable ordering. |
| <i>UTS 10</i>               | Unicode technical standard 10, Unicode collation algorithm.   |

*Choun Nath's*

Chuon Nath's Khmer—Khmer dictionary.

Khmer ordering analysis. Maurice Bauhahn.

<http://www.bauhahn.clara.net/Khmer/KhmerSortingUnicodegamma.pdf>.

--- end ---