Title: Ordering rules for Khmer, Thai, and Lao;
suggestions for annex C.2 changes and a new annex B.5

Source: Kent Karlsson
Date: 2004-03-09
Status: Expert contribution
Document Type: Working group document, regarding 14651 amd 2
Action: For consideration by the UTC and JTC 1/SC 2

1 Suggested changes to annex C.2

The collation rules for Thai in the 14651 CTT (and the UTS 10 data files) do not follow the collation rules given in annex C.2, Thai string ordering, of ISO/IEC 14651. Firstly, the informal rules need some corrections. Secondly, the CTT should follow the updates to the order given in annex C.2 (see companion paper "Ordering rules for Khmer, Thai, and Lao; CTT suggestion"). Thirdly, add a new annex B.5 with a Khmer/Thai/Lao tailoring (detailed below).

The following text is a _corrected_ quote (in seriffed font) from ISO/IEC 14651 beginning of annex C.2 (new annex B.5). The actual text of ISO/IEC 14651 should incorporate the corrections below.

The widely accepted standard for Thai lexicographical ordering is defined in the Royal Institute Dictionary 2525 B.E. Edition (1982 A.D.), the official standard Thai dictionary. The ordering principles are:

- Words are ordered alphabetically, not phonetically. Consonants order is:

  ก ข ฃ ค ฅ ฆ ง จ ฉ ช ซ ฌ ญ ฎ ฏ ฐ ฑ ฒ ณ ด ต ถ ท
  ธ น บ ป ผ ฝ พ ฟ ภ ม ย ร ฤ ล ฦ ว ศ ษ ส ห ฬ อ ฮ

- Vowels, and nikhahit, are also ordered by written forms, not by sounds. Vowels and nikhahit order is:

  ◌ั ◌ะ ◌ั ◌า ◌ๅ ◌ำ ◌ิ ◌ี ◌ึ ◌ื ◌ุ ◌ู เ◌ แ◌ โ◌ ใ◌ ไ◌ ◌ฺ

(อ ว ย are always ordered as consonants, although they sometimes act as vowels.)

- Tones and diacritics are ignored at level 1, at level 2 the order is:

  $\overset{\varepsilon}{-}\quad \overset{\sigma}{-}\quad \overset{\text{๛}}{-}\quad \overset{!}{-}\quad \overset{\text{๛}}{-}\quad \overset{\text{๛}}{-}\quad \overset{+}{-}$

- String comparison is performed from left to right, but considering initial consonants before vowels in the same syllable.

LAKKHANGYAO (ๅ) is a vowel letter, a variant of SARA AA (า). LAKKHANGYAO should be ordered as a compatibility variant of SARA AA, and no collating elements (contractions) should be defined for combinations using LAKKHANGYAO.

Annex C.2 later should continue:

- Leading vowels (เ- แ- โ- ใ- ไ-, corresponding to characters U+0E40-U+0E44), which are written before consonants, must be considered after the initial consonant. Therefore, the rearrangement is needed before comparison.
- Diacritics and tone marks ($\overset{\varepsilon}{-}\ \overset{\sigma}{-}\ \overset{\text{๛}}{-}\ \overset{!}{-}\ \overset{\text{๛}}{-}\ \overset{\text{๛}}{-}\ \overset{+}{-}$) must be ignored in the first pass, and be considered at later pass if the first pass yields equality.

Even later:

When punctuation marks (ๆ ฯ ๚ ๐ ๚ะ) are concerned, another level of weights is required for them.

Dependent vowels (and would-be dependent vowels) are ordered like dependent vowels for Indic scripts, at the very end of the weighting scale for level 1, along with Hangul vowels.

Note also that PAIYANNOI (ฯ) and repeat mark (ๆ) are regarded as punctuation marks.

Below, formal ordering rules (for the suggested new annex B.5), as a tailoring of the suggested CTT part for Khmer, Thai, and Lao that attempt to follow the informal description (as amended above, after discussions) for Thai in annex C.2 are given.

The tailoring rules below contain a number of contractions. Using these contractions, special prehandling for Thai and Lao (which is now specified in UTS 10) is not needed. This method works very well with ISO/IEC 14651, which does not require that there is any prehandling at all.

2 Suggestion for a new annex B.5, a Khmer, Thai, and Lao, tailoring of the CTT

There is a separate suggestion for updating the CTT regarding Khmer, Thai, and Lao (see the companion paper "Ordering rules for Khmer, Thai, and Lao; CTT suggestion"). However, in order to keep that suggestion relatively simple, and also avoid bringing some optional features into the common CTT, a tailoring (of the new CTT) is here suggested as an informative annex.

Since these scripts are mainly used in neighbouring countries, all of these three scripts should be covered by a single tailoring (rather than one tailoring per script).

```
reorder-after <MAX>


% Thai:

collating-element <U0E5A_0E30> from "<U0E5A><U0E30>" % special end symbol

              % Thai pre-vowels:
collating-element <U0E40_0E01>..<U0E40_0E2E> from "<U0E40><U0E01>..<U0E2E>"
collating-element <U0E40_0E40_0E01>..<U0E40_0E40_0E2E> from "<U0E40><U0E40><U0E01>..<U0E2E>"
collating-element <U0E41_0E01>..<U0E41_0E2E> from "<U0E41><U0E01>..<U0E2E>"
collating-element <U0E42_0E01>..<U0E42_0E2E> from "<U0E42><U0E01>..<U0E2E>"
collating-element <U0E43_0E01>..<U0E43_0E2E> from "<U0E43><U0E01>..<U0E2E>"
collating-element <U0E44_0E01>..<U0E44_0E2E> from "<U0E44><U0E01>..<U0E2E>"

collating-element <U0E4D_0E32> from "<S0E4D><S0E32>" % AM




% Lao:
              % Lao pre-vowels:
collating-element <U0EC0_0E81>..<U0EC0_0EAE> from "<U0EC0><U0E81>..<U0EAE>"
collating-element <U0EC0_0EC0_0E81>..<U0EC0_0EC0_0EAE> from "<U0EC0><U0EC0><U0E81>..<U0EAE>"
collating-element <U0EC1_0E81>..<U0EC1_0EAE> from "<U0EC1><U0E81>..<U0EAE>"
collating-element <U0EC2_0E81>..<U0EC2_0EAE> from "<U0EC2><U0E81>..<U0EAE>"
collating-element <U0EC3_0E81>..<U0EC3_0EAE> from "<U0EC3><U0E81>..<U0EAE>"
collating-element <U0EC4_0E81>..<U0EC4_0EAE> from "<U0EC4><U0E81>..<U0EAE>"

              % LAO HO NO
collating-element <U0EC0_0EDC> from "<U0EC0><U0EDC>"
collating-element <U0EC0_0EC0_0EDC> from "<U0EC0><U0EC0><U0EDC>"
collating-element <U0EC1_0EDC> from "<U0EC1><U0EDC>"
collating-element <U0EC2_0EDC> from "<U0EC2><U0EDC>"
collating-element <U0EC3_0EDC> from "<U0EC3><U0EDC>"
collating-element <U0EC4_0EDC> from "<U0EC4><U0EDC>"

              % LAO HO MO
collating-element <U0EC0_0EDD> from "<U0EC0><U0EDD>"
collating-element <U0EC0_0EC0_0EDD> from "<U0EC0><U0EC0><U0EDD>"
collating-element <U0EC1_0EDD> from "<U0EC1><U0EDD>"
collating-element <U0EC2_0EDD> from "<U0EC2><U0EDD>"
collating-element <U0EC3_0EDD> from "<U0EC3><U0EDD>"
collating-element <U0EC4_0EDD> from "<U0EC4><U0EDD>"

              % U+0EAB LAO LETTER HO SUNG, U+200D ZERO WIDTH JOINER, U+0EA5 LAO LETTER LO LOOT
collating-element <U0EC0_0EAB_200D_0EA5> from "<U0EC0><U0EAB><U200D><U0EA5>"
collating-element <U0EC0_0EC0_0EAB_200D_0EA5> from "<U0EC0><U0EC0><U0EAB><U200D><U0EA5>"
collating-element <U0EC1_0EAB_200D_0EA5> from "<U0EC1><U0EAB><U200D><U0EA5>"
collating-element <U0EC2_0EAB_200D_0EA5> from "<U0EC2><U0EAB><U200D><U0EA5>"
collating-element <U0EC3_0EAB_200D_0EA5> from "<U0EC3><U0EAB><U200D><U0EA5>"
collating-element <U0EC4_0EAB_200D_0EA5> from "<U0EC4><U0EAB><U200D><U0EA5>"

collating-element <U0ECD_0EB2> from "<S0ECD><S0EB2>" % AM
collating-element <U0EB2_0ECD> from "<S0EB2><S0ECD>" % AM
```

```
% Khmer:

collating-symbol <S1794_S17C9> % KHMER LETTER BA, KHMER SIGN MUUSIKATOAN
collating-symbol <S1794_S17CA> % KHMER LETTER BA, KHMER SIGN TRIISAP

collating-symbol <S17BB_S17C6> % KHMER VOWEL SIGN U, KHMER SIGN NIKAHIT: nasalised U
collating-symbol <S17B6_S17C6> % KHMER VOWEL SIGN AA, KHMER SIGN NIKAHIT: nasalised AA

collating-symbol <C1780>..<C179C>

        % Declaration of Khmer contractions

collating-element <U1794_17C9> from "<U1794><U17C9>" %  , KHMER LETTER BA, KHMER SIGN MUUSIKATOAN
        (PA)
collating-element <U1794_17CA> from "<U1794><U17CA>" %  , KHMER LETTER BA, KHMER SIGN TRIISAP

collating-element <SW_17CC_1780>..<SW_17CC_17A2> from "<U1780>..<U17A2><U17CC>"
        % KHMER LETTER KA, KHMER SIGN ROBAT..KHMER LETTER QA, KHMER SIGN ROBAT
collating-element <SW_17CC_17A5>..<SW_17CC_17B3> from "<U17A5>..<U17B3><U17CC>"
        % KHMER INDEPENDENT VOWEL QI, KHMER SIGN ROBAT..KHMER INDEPENDENT VOWEL QAU, KHMER SIGN
        ROBAT

collating-element <U17C6_17BB> from "<U17BB><U17C6>"
        % KHMER VOWEL SIGN U, KHMER SIGN NIKAHIT: nasalised U
collating-element <U17BB_17C6> from "<U17C6><U17BB>"
        % KHMER SIGN NIKAHIT, KHMER VOWEL SIGN U: nasalised U
collating-element <U17C6_17B6> from "<U17B6><U17C6>"
        % KHMER VOWEL SIGN AA, KHMER SIGN NIKAHIT: nasalised AA
collating-element <U17B6_17C6> from "<U17C6><U17B6>"
        % KHMER SIGN NIKAHIT, KHMER VOWEL SIGN AA: nasalised AA

collating-element <U17D2_1780>..<U17D2_179C> from "<U17D2><U1780>..<U179C>"
        % COENG, KHMER LETTER KA..COENG, KHMER LETTER QA
collating-element <U17D2_17A5>..<U17D2_17B3> from "<U17D2><U17A5>..<U17B3>"
        % COENG, KHMER INDEPENDENT VOWEL QI..COENG, KHMER INDEPENDENT VOWEL QAU




reorder-after <S1794> % KHMER LETTER BA
<S1794_17C9> % KHMER LETTER BA, KHMER SIGN MUUSIKATOAN
<S1794_17CA> % KHMER LETTER BA, KHMER SIGN TRIISAP



reorder-after <S0ECD> % LAO NIGGAHITA
<S0EB3> % LAO VOWEL SIGN AM

reorder-after <S17C5> KHMER VOWEL SIGN AU
<S17BB_17C6> % KHMER VOWEL SIGN U, KHMER SIGN NIKAHIT: nasalised U (OM)
reorder-after <S17C6> KHMER SIGN NIKAHIT
<S17B6_17C6> % KHMER VOWEL SIGN AA, KHMER SIGN NIKAHIT: nasalised AA (AM)



reorder-after <S17D2>
<C1780>..<C1794> % COENG, KHMER LETTER KA..COENG, KHMER LETTER BA
<C1795>..<C179A> % COENG, KHMER LETTER PHA..COENG, KHMER LETTER RO
<C17AB> % COENG, KHMER INDEPENDENT VOWEL RY
<C17AC> % COENG, KHMER INDEPENDENT VOWEL RYY
<C179B> % COENG, KHMER LETTER LO
<C17AD> % COENG, KHMER INDEPENDENT VOWEL LY
<C17AE> % COENG, KHMER INDEPENDENT VOWEL LYY
<C179C>..<C17A2> % COENG, KHMER LETTER VO..COENG, KHMER LETTER QA
```

4

```
reorder-after <SFFFF>

order_start forward;forward;forward;forward



<U0E5A_0E30> IGNORE;IGNORE;IGNORE;<U0E5A_0E30> % THAI CHARACTER ANGKHANKHU, THAI CHARACTER SARA A



<U1794_17C9> <S1794_17C9>;<BASE>;<MIN>;<U1794_17C9>
        %  , KHMER LETTER BA, KHMER SIGN MUUSIKATOAN (PA)

<U1794_17CA> <S1794_17CA>;<BASE>;<MIN>;<U1794_17CA>
        %  , KHMER LETTER BA, KHMER SIGN TRIISAP



%% The Robat contractions should be used only in an "advanced" tailoring for
%% Khmer/Thai/Lao, since ROBAT is rather rarely used, and these contractions
%% may impact on the efficiency of the key computation even if ROBAT does not
%% occur, since these contractions begin with commonly used letters.

<SW_17CC_1780>..<SW_17CC_17A2> "<S179A><S17D2><S1780>..<S17A2>";
        "<BASE><VRNT1><BASE><BASE>";"<MIN><MIN><MIN><MIN>"; <SW_17CC_1780>..<SW_17CC_17A2>
        % KHMER LETTER KA, KHMER SIGN ROBAT..KHMER LETTER QA, KHMER SIGN ROBAT

<SW_17CC_17A5>..<SW_17CC_17A6> "<S179A><S17D2><S17A2><S17B7>..<S17B8>";
        "<BASE><VRNT1><BASE><BASE><VRNT1><BASE>";"<MIN><MIN><MIN><MIN><MIN><MIN>";
        <SW_17CC_17A5>..<SW_17CC_17A6> % KHMER INDEPENDENT VOWEL QI, KHMER SIGN ROBAT..KHMER
        INDEPENDENT VOWEL QII, KHMER SIGN ROBAT

<SW_17CC_17A7> "<S179A><S17D2><S17A2><S17BB>";"<BASE><VRNT1><BASE><BASE><VRNT1><BASE>";
        "<MIN><MIN><MIN><MIN><MIN><MIN>";<SW_17CC_17A7>
        % KHMER INDEPENDENT VOWEL QU, KHMER SIGN ROBAT

<SW_17CC_17A8> "<S179A><S17D2><S17A2><S17BB>";"<BASE><VRNT1><BASE><BASE><VRNT2><BASE>";
        "<MIN><MIN><MIN><MIN><MIN><MIN>";<SW_17CC_17A8>
        %  , KHMER INDEPENDENT VOWEL QUK;   , KHMER SIGN ROBAT

<SW_17CC_17A9> "<S179A><S17D2><S17A2><S17BC>";"<BASE><VRNT1><BASE><BASE><VRNT1><BASE>";
        "<MIN><MIN><MIN><MIN><MIN><MIN>";<SW_17CC_17A9>
        %  , KHMER INDEPENDENT VOWEL QUU;   , KHMER SIGN ROBAT

<SW_17CC_17AA> "<S179A><S17D2><S17A2><S17BC>";"<BASE><VRNT1><BASE><BASE><VRNT2><BASE>";
        "<MIN><MIN><MIN><MIN><MIN><MIN>";<SW_17CC_17AA>
        %  , KHMER INDEPENDENT VOWEL QUUV;   , KHMER SIGN ROBAT

<SW_17CC_17AF>..<SW_17CC_17B1> "<S179A><S17D2><S17A2><S17C2>..<S17C4>";
        "<BASE><VRNT1><BASE><BASE><VRNT1><BASE>";"<MIN><MIN><MIN><MIN><MIN><MIN>";
        <SW_17CC_17AF>..<SW_17CC_17B1> % KHMER INDEPENDENT VOWEL QE, KHMER SIGN ROBAT..KHMER
        INDEPENDENT VOWEL QOO TYPE ONE, KHMER SIGN ROBAT

<SW_17CC_17B2> "<S179A><S17D2><S17A2><S17C4>";"<BASE><VRNT1><BASE><BASE><VRNT2><BASE>";
        "<MIN><MIN><MIN><MIN><MIN><MIN>";<SW_17CC_17B2>
        %  , KHMER INDEPENDENT VOWEL QOO TYPE TWO;   , KHMER SIGN ROBAT

<SW_17CC_17B3> "<S179A><S17D2><S17A2><S17C5>";"<BASE><VRNT1><BASE><BASE><VRNT1><BASE>";
        "<MIN><MIN><MIN><MIN><MIN><MIN>";<SW_17CC_17B3>
        %  , KHMER INDEPENDENT VOWEL QAU;   , KHMER SIGN ROBAT




%%% Thai consonants's swap for pre-vowels (note that SARA AE is ordered as <SARA E, SARA E>):
```

```
<U0E40_0E01>..<U0E40_0E2E> "<S0E01>..<S0E2E><S0E40>";"<BASE><BASE>";"<MIN><MIN>";
                <U0E40_0E01>..<U0E40_0E2E> % E, KO KAI..E, HO NOHUK
<U0E40_0E40_0E01>..<U0E40_0E40_0E2E> "<S0E01>..<S0E2E><S0E40><S0E40>";"<BASE><BASE><BASE>";
                "<MIN><MIN><MIN>";<U0E40_0E40_0E01>..<U0E40_0E40_0E2E> % E, E, KO KAI.. E, E, HO
                NOHUK
<U0E41_0E01>..<U0E41_0E2E> "<S0E01>..<S0E2E><S0E40><S0E40>";"<BASE><BASE><BASE>";
                "<MIN><MIN><MIN>";<U0E41_0E01>..<U0E41_0E2E> % AE, KO KAI..AE, HO NOHUK
<U0E42_0E01>..<U0E42_0E2E> "<S0E01>..<S0E2E><S0E42>";"<BASE><BASE>";"<MIN><MIN>";
                <U0E42_0E01>..<U0E42_0E2E> % O, KO KAI..O, HO NOHUK
<U0E43_0E01>..<U0E43_0E2E> "<S0E01>..<S0E2E><S0E43>";"<BASE><BASE>";"<MIN><MIN>";
                <U0E43_0E01>..<U0E43_0E2E> % AI MAIMUAN, KO KAI..AI MAIMUAN, HO NOHUK
<U0E44_0E01>..<U0E44_0E2E> "<S0E01>..<S0E2E><S0E44>";"<BASE><BASE>";"<MIN><MIN>";
                <U0E44_0E01>..<U0E44_0E2E> % AI MAIMALAI, KO KAI..AI MAIMALAI, HO NOHUK


%%% Lao consonants's swap for pre-vowels (note that VOWEL EI is ordered as <VOWEL E, VOWEL E>):
%%% (The code point ranges contain some code points which are not assigned to any character.
%%% However, the (level 1) weights should be declared and approriately weighted.)

<U0EC0_0E81>..<U0EC0_0EAE> "<S0E81>..<S0EAE><S0EC0>";"<BASE><BASE>";"<MIN><MIN>";
                <U0EC0_0E81>..<U0EC0_0EAE> % E, KO..E, HO TAM
<U0EC0_0EC0_0E81>..<U0EC0_0EC0_0EAE> "<S0E81>..<S0EAE><S0EC0><S0EC0>";"<BASE><BASE><BASE>";
                "<MIN><MIN><MIN>";<U0EC0_0EC0_0E81>..<U0EC0_0EC0_0EAE> % E, E, KO..E, E, HO TAM
<U0EC1_0E81>..<U0EC1_0EAE> "<S0E81>..<S0EAE><S0EC0><S0EC0>";"<BASE><BASE><BASE>";
                "<MIN><MIN><MIN>";<U0EC1_0E81>..<U0EC1_0EAE> % EI, KO, EI, HO TAM
<U0EC2_0E81>..<U0EC2_0EAE> "<S0E81>..<S0EAE><S0EC2>";"<BASE><BASE>";"<MIN><MIN>";
                <U0EC2_0E81>..<U0EC2_0EAE> % O, KO..O, HO TAM
<U0EC3_0E81>..<U0EC3_0EAE> "<S0E81>..<S0EAE><S0EC3>";"<BASE><BASE>";"<MIN><MIN>";
                <U0EC3_0E81>..<U0EC3_0EAE> % AY, KO..AY, HO TAM
<U0EC4_0E81>..<U0EC4_0EAE> "<S0E81>..<S0EAE><S0EC4>";"<BASE><BASE>";"<MIN><MIN>";
                <U0EC4_0E81>..<U0EC4_0EAE> % AI, KO..AI, HO TAM


<U0EC0_0EDC> "<S0EAB><S0E99><S0EC0>";"<BASE><BASE><BASE>";"<COMPAT><COMPAT><COMPAT>";
                <U0EC0_0EDC> % E, HO NO
<U0EC0_0EC0_0EDC> "<S0EAB><S0E99><S0EC0><S0EC0>";"<BASE><BASE><BASE><BASE>";
                "<COMPAT><COMPAT><COMPAT><COMPAT>";<U0EC0_0EC0_0EDC> % E, E, HO NO
<U0EC1_0EDC> "<S0EAB><S0E99><S0EC0><S0EC0>";"<BASE><BASE><BASE><BASE>";
                "<COMPAT><COMPAT><COMPAT><COMPAT>"; <U0EC1_0EDC> % EI, HO NO
<U0EC2_0EDC> "<S0EAB><S0E99><S0EC2>";"<BASE><BASE><BASE>";"<COMPAT><COMPAT><COMPAT>";
                <U0EC2_0EDC> % O, HO NO
<U0EC3_0EDC> "<S0EAB><S0E99><S0EC3>";"<BASE><BASE><BASE>";"<COMPAT><COMPAT><COMPAT>";
                <U0EC3_0EDC> % AY, HO NO
<U0EC4_0EDC> "<S0EAB><S0E99><S0EC4>";"<BASE><BASE><BASE>";"<COMPAT><COMPAT><COMPAT>";
                <U0EC3_0EDC> % AI, HO NO


<U0EC0_0EDD> "<S0EAB><S0EA1><S0EC0>";"<BASE><BASE><BASE>";"<COMPAT><COMPAT><COMPAT>";
                <U0EC0_0EDD> % E, HO MO
<U0EC0_0EC0_0EDD> "<S0EAB><S0EA1><S0EC0><S0EC0>";"<BASE><BASE><BASE><BASE>";
                "<COMPAT><COMPAT><MIN><MIN>";<U0EC0_0EC0_0EDD> % E, E, HO MO
<U0EC1_0EDD> "<S0EAB><S0EA1><S0EC0><S0EC0>";"<BASE><BASE><BASE><BASE>";
                "<COMPAT><COMPAT><COMPAT><COMPAT>";<U0EC1_0EDD> % EI, HO MO
<U0EC2_0EDD> "<S0EAB><S0EA1><S0EC2>";"<BASE><BASE><BASE>";"<COMPAT><COMPAT><COMPAT>";
                <U0EC2_0EDD> % O, HO MO
<U0EC3_0EDD> "<S0EAB><S0EA1><S0EC3>";"<BASE><BASE><BASE>";"<COMPAT><COMPAT><COMPAT>";
                <U0EC3_0EDD> % AY, HO MO
<U0EC4_0EDD> "<S0EAB><S0EA1><S0EC4>";"<BASE><BASE><BASE>";"<COMPAT><COMPAT><COMPAT>";
                <U0EC4_0EDD> % AI, HO MO


%% HO LO is a ligature which does not have its own code point, so ZWJ is used.
<U0EC0_0EAB_200D_0EA5> "<S0EAB><S0EA5><S0EC0>";"<BASE><BASE><BASE>";"<MIN><MIN><MIN>";
                <U0EC0_0EAB_200D_0EA5> % E, HO LO,
<U0EC0_0EC0_0EAB_200D_0EA5> "<S0EAB><S0EA5><S0EC0><S0EC0>";"<BASE><BASE><BASE><BASE>";
                "<MIN><MIN><MIN><MIN>";<U0EC0_0EC0_0EAB_200D_0EA5> % E, E, HO LO
<U0EC1_0EAB_200D_0EA5> "<S0EAB><S0EA5><S0EC0><S0EC0>";"<BASE><BASE><BASE><BASE>";
                "<MIN><MIN><COMPAT><COMPAT>";<U0EC1_0EAB_200D_0EA5> % EI, HO LO
<U0EC2_0EAB_200D_0EA5> "<S0EAB><S0EA5><S0EC2>";"<BASE><BASE><BASE>";"<MIN><MIN><MIN>";
                <U0EC2_0EAB_200D_0EA5> % O, HO LO
<U0EC3_0EAB_200D_0EA5> "<S0EAB><S0EA5><S0EC3>";"<BASE><BASE><BASE>";"<MIN><MIN><MIN>";
                <U0EC3_0EAB_200D_0EA5> % AY, HO LO
<U0EC4_0EAB_200D_0EA5> "<S0EAB><S0EA5><S0EC4>";"<BASE><BASE><BASE>";"<MIN><MIN><MIN>";
                <U0EC4_0EAB_200D_0EA5> % AI, HO LO
```

```
%%% Thai SARA AM is ordered as <SARA AA, NIKHAHIT>, but can be written the other way around:

<U0E4D_0E32> "<S0E32><S0E4D>";"<BASE><BASE>";"<MIN><MIN>";<U0E4D_0E32>
                % THAI CHARACTER NIKHAHIT, THAI CHARACTER SARA AA; reorder characters (sort of)



%%% Lao VOWEL AM is ordered as <VOWEL AA, NIGGAHITA>, but can be written the other way around:

<U0EB3> "<S0EB3>";"<BASE>";"<COMPAT>";<U0EB3> % LAO VOWEL SIGN AM
<U0EB2_0ECD> "<S0EB3>";"<BASE>";"<MIN>";<U0EB2_0ECD> % LAO VOWEL SIGN AA, LAO NIGGAHITA
<U0ECD_0EB2> "<S0EB3>";"<BASE>";"<MIN>";<U0ECD_0EB2>
                % LAO NIGGAHITA, LAO VOWEL SIGN AA; reorder characters (sort of)



%%% Khmer OM and AAM:

<U17BB_17C6> <S17BB_17C6>;<BASE>;<MIN>;<U17BB_17C6>
                % , KHMER VOWEL SIGN U, KHMER SIGN NIKAHIT: nasalised U
<U17C6_17BB> <S17BB_17C6>;<BASE>;<MIN>;<U17C6_17BB>
                % , KHMER SIGN NIKAHIT, KHMER VOWEL SIGN U: nasalised U

<U17B6_17C6> <S17B6_17C6>;<BASE>;<MIN>;<U17B6_17C6>
                % , KHMER VOWEL SIGN AA, KHMER SIGN NIKAHIT: nasalised AA
<U17C6_17B6> <S17B6_17C6>;<BASE>;<MIN>;<U17C6_17B6>
                % , KHMER SIGN NIKAHIT, KHMER VOWEL SIGN AA: nasalised AA



%%% The following may be included to shorten the Khmer collation keys somewhat.
%%% Adding these contractions will not affect efficiency very much, since they
%%% begin with the COENG, which occurs in only a fraction of the syllables.

<U17D2_1780>..<U17D2_1794> <C1780>..<C1794>;<BASE>;<MIN>;<U17D2_1780>..<U17D2_1794>
        % COENG, KHMER LETTER KA..COENG, KHMER LETTER BA
<U17D2_1795>..<U17D2_179A> <C1795>..<C179A>;<BASE>;<MIN>;<U17D2_1795>..<U17D2_179A>
        % COENG, KHMER LETTER PHA..COENG, KHMER LETTER RO
<U17D2_17AB> <C17AB>;<BASE>;<MIN>;<U17D2_17AB> % COENG, KHMER INDEPENDENT VOWEL RY
<U17D2_17AC> <C17AC>;<BASE>;<MIN>;<U17D2_17AC> % COENG, KHMER INDEPENDENT VOWEL RYY
<U17D2_179B> <C179B>;<BASE>;<MIN>;<U17D2_179B> % COENG, KHMER LETTER LO
<U17D2_17AD> <C17AD>;<BASE>;<MIN>;<U17D2_17AD> % COENG, KHMER INDEPENDENT VOWEL LY
<U17D2_17AE> <C17AE>;<BASE>;<MIN>;<U17D2_17AE> % COENG, KHMER INDEPENDENT VOWEL LYY
<U17D2_179C>..<U17D2_17A2> <C179C>..<C17A2>;<BASE>;<MIN>;<U17D2_179C>..<U17D2_17A2>
        % COENG, KHMER LETTER VO..COENG, KHMER LETTER QA

<U17D2_17A7> "<C17A2><S17BB>";"<BASE><VRNT1><BASE>";"<MIN><MIN><MIN>";<U17D2_17A7>
        % COENG, KHMER INDEPENDENT VOWEL QU
<U17D2_17AF> "<C17A2><S17C2>";"<BASE><VRNT1><BASE>";"<MIN><MIN><MIN>";<U17D2_17AF>
        % COENG, KHMER INDEPENDENT VOWEL QE

reorder-end
```

Acknowledgements

Thanks to Maurice Bauhahn for explaining the principles of Khmer collation. Thanks to Theppitak Karoonboonyanan for helping me with the Thai rules and Anousak Souphavanh for helping me with the Lao rules. Any errors or shortcomings here are of course mine (especially since I've done some interpretations and changes).