Proposal to Update Properties for Two Khmer Characters

To: UTC via Script Ad Hoc
Status: Individual Contribution
Date: 2023-04-28

Authors: Steven R. LOOMIS <srl295@codehivetx.us> (Individual Unicode Member)

1. Summary

This is a request for modifications to the sentence break properties of two Khmer Characters.

- U+17D4 ']' KHMER SIGN KHAN
- U+17D5 ']' KHMER SIGN BARIYOOSAN

2. Proposal

Add to PropList.txt

```
# PropList.txt
17D4..17D5 ; Sentence_Terminal # Po [2] KHMER SIGN KHAN..KHMER SIGN BARIYOOSAN
```

Which will cause an update to SentenceBreakProperty.txt

```
# SentenceBreakProperty.txt
17D4 ; STerm # Po KHMER SIGN KHAN
17D5 ; STerm # Po KHMER SIGN BARIYOOSAN
```

3. Background

Note: This issue was originally filed as <https://unicode-org.atlassian.net/browse/CLDR-16220> by Mr. Makara SOK of SIL.
3.1 NamesList

The NamesList already indicates the functions of these characters.

- **17D4** KHMER SIGN KHAN
  * functions as a full stop, period
- **17D5** KHMER SIGN BARIYOOSAN
  * indicates the end of a section or a text

3.2 Example Text

This example text contains multiple sentences, with a U+17D4 (។) at the end of each sentence, and a U+17D5 (៖) at the end of the chapter. There are three variations for ៖:

1. ៖
2. ៖ mour
3. ៖ 0
   ។
Example 1  https://www.bible.com/bible/2287/MAT.1.GKHB

Example 2 shows U+17D5 at the end of a chapter, plus ៚ showing the end of a book.
Proposal to Update Properties for Two Khmer Characters

YouVersion

Search Bible.com

Get the app

GKHB

Example 1

https://www.bible.com/bible/2287/3JN.1.GKHB

Example 2

https://www.bible.com/bible/2287/3JN.1.GKHB
Example 3, a scanned Khmer book showing end of paragraph marks
4. Other Consideration

4.1: ATerm was considered, but ATerm is a fixed list in UAX#29 and seems to only interact differently than STerm for numeric and upper/lower letters.

4.2: The codepoint U+17D8 ឈ  discourage in the names list. But, the sequence ឈ  (‘i+ŋ+i’) means "etcetera" and so break rules may need to be evaluated. ឈ  is already handled as complex text.

Sometimes written -ឈ- but with ឈ at the end of a sentence. The document below shows that there are two variations of ឈ. It can be either ឈ or ឈ.

4.3: The question was asked, are there other contexts where ឈ is used as an abbreviation?

Makara SOK writes: "Not that I know of. Period (.) is used in abbreviation if a symbol is needed in between letters."

The conclusion is that there are no known contexts where ឈ is used as an abbreviation.