

Author: Kenneth Whistler

Title: Character Properties for repetition marks

Kent Karlsson reported the following anomaly regarding the general category for repetition marks.

The proposed changes are not for new characters for 3.2, but for characters that already existed as of 3.0. Therefore I don't think these are merely errata fixes for the 3.2 beta files, but a request to formally correct properties that have already been established -- something that the UTC has to consider and approve or disapprove.

I agree that Lm is the intended general category for length marks and for most repetition marks, to the extent that they "stand in" for letters and are letterlike in form and/or function.

As for these two particular examples, my assessment is:

U+17D7 KHMER SIGN LEK TOO

This sign must be cognate with the Thai sign, U+0E46 THAI CHARACTER MAIYAMOK, which has similar form and function. Since U+0E46 is Lm, I think it would be consistent to make U+17D7 Lm, rather than Po.

U+0F34 TIBETAN MARK BSDUS RTAGS

This is also a repetition mark of some sort. Formally, it is a TSHEG sitting on what is probably two strokes representing the repetition (like a "ditto" mark). Currently it is So. I suspect that Lm might be more appropriate for it, but it would be best if someone who knows Tibetan could check further on its usage.

--Ken

Whistler: Character Properties for repetition marks

----- Begin Included Message -----

```
> From: Kent Karlsson [mailto:kentk@md.chalmers.se]
> Sent: Sunday, January 13, 2002 12:06 PM
> Subject: Two characters that are(?) modifier letters but have general
> category other than Lm
>
> Most non-combining repetition and lenght marking characters have
> general category Lm:
>
> 0E46          THAI CHARACTER MAIYAMOK (mai yamok)
>      = mai yamok
>      * repetition
> 0E46;THAI CHARACTER MAIYAMOK;Lm;0;L;;;;;N;THAI MAI YAMOK;mai yamok;;;
>
> 0EC6          LAO KO LA
>      * repetition
> 0EC6;LAO KO LA;Lm;0;L;;;;;N;;;;;
>
>
> 02D0;MODIFIER LETTER TRIANGULAR COLON;Lm;0;L;;;;;N;;;;;
> 02D1;MODIFIER LETTER HALF TRIANGULAR COLON;Lm;0;L;;;;;N;;;;;
> 3031;VERTICAL KANA REPEAT MARK;Lm;0;L;;;;;N;;;;;
> 3032;VERTICAL KANA REPEAT WITH VOICED SOUND MARK;Lm;0;L;;;;;N;;;;;
> 3033;VERTICAL KANA REPEAT MARK UPPER HALF;Lm;0;L;;;;;N;;;;;
> 3034;VERTICAL KANA REPEAT WITH VOICED SOUND MARK UPPER
> HALF;Lm;0;L;;;;;N;;;;;
> 3035;VERTICAL KANA REPEAT MARK LOWER HALF;Lm;0;L;;;;;N;;;;;
> 30FC;KATAKANA-HIRAGANA PROLONGED SOUND MARK;Lm;0;L;;;;;N;;;;;
> 30FD;KATAKANA ITERATION MARK;Lm;0;L;;;;;N;;;;;
> 30FE;KATAKANA VOICED ITERATION MARK;Lm;0;L;30FD 3099;;;;;N;;;;;
>
>
> Then it is a bit strange that there are two such characters that have
> general category So or Po:
>
> 0F34          TIBETAN MARK BSDUS RTAGS (du ta)
>      * repetition
> 0F34;TIBETAN MARK BSDUS RTAGS;So;0;L;;;;;N;du ta;;;
> So    ---> Lm?
>
> 17D7          KHMER SIGN LEK TOO
>      * repetition sign
> 17D7;KHMER SIGN LEK TOO;Po;0;L;;;;;N;;;;;
> Po    ---> Lm?
>
> Should not these two also have general category Lm? Or are they
> somehow
> really different?
>
```

----- End Included Message -----