Subject: Line break property value of U+22EF MIDLINE HORIZONTAL ELLIPSIS

Date: 2015-Feb-04
To: UTC
From: Peter Edberg

Proposal: For U+22EF MIDLINE HORIZONTAL ELLIPSIS, change the value of the Line_Break property from AL to IN.

Motivation: In some CJK text, U+22EF MIDLINE HORIZONTAL ELLIPSIS is used as an ellipsis that is intended to function in same way as U+2026 HORIZONTAL ELLIPSIS. The latter has Line_Break value IN and thus there is no break between it and a preceding ID (ideographic) character. However, U+22EF has Line_Break value AL which does permit a break between it and a preceding ID character, so that the MIDLINE HORIZONTAL ELLIPSIS can wrap to the beginning of a succeeding line.

Background:
• Both characters have been in Unicode since version 1.1.
• U+2026 HORIZONTAL ELLIPSIS has General Category Po, other punctuation.
• U+22EF MIDLINE HORIZONTAL ELLIPSIS has General Category Sm, math symbol; it is in the Mathematical Operators block 2200-22FF.
• The Line_Break value for both characters has been unchanged since the first version of LineBreak.txt in Unicode 3.2.
• Nearly all characters in the Mathematical Operators block have Line_Break property values of AL or AI, which is likely why U+22EF has that value.

However: There does not appear to be anything about the mathematical usage of U+22EF that particularly requires the AL behavior. UTR #25 "Unicode Support for Mathematics" does not mention U+22EF (and Murray Sargent indicated that Unicode Line_Break properties are largely ignored for such math layout anyway).

The examples of mathematical ellipsis usage in this Wikipedia article section show it being used in a manner very similar to the mathematical usage of U+2026 HORIZONTAL ELLIPSIS (so that different line break behavior would not seem to be a benefit).

Example of correct line wrap with U+2026:

正在設定系統語
言……

Example of poor line wrap with U+22EF:

正在設定系統語言
……