Diff between UAX 14-17 (Unicode 4.1) and UAX 14-19 (Unicode 5.0)

Note: it completely clear whether the changes (rule 25) to allow PO at the start of a number an PR at the end of a number are independent of the changes (rule 30) to not break constructs like person(s).

2/7/2008 Andy Heninger

**Numbers:**

Do not break alphanumerics.

**LB23** Do not break within ‘a9’, ‘3a’, or ‘H%’.

\[
\begin{align*}
\text{ID} \times \text{PO} \\
\text{AL} \times \text{NU} \\
\text{NU} \times \text{AL}
\end{align*}
\]

**LB24** Do not break between prefix and letters or ideographs.

\[
\begin{align*}
\text{PR} \times \text{ID} \\
\text{PR} \times \text{AL} \\
\text{PO} \times \text{AL}
\end{align*}
\]

In general, it is recommended to not break lines inside numbers of the form described by the following regular expression:

\[
(\text{PR} | \text{PO})? (\text{OP} | \text{HY})? \text{NU} (\text{NU} | \text{SY} | \text{IS})^* \text{CL}^? (\text{PR} | \text{PO})?
\]

*Examples:* $(12.35)$ 2,1234 $(12)¢$ 12.54¢

The default line breaking algorithm approximates this with the following rule. Note that some cases have already been handled, such as ‘9’, ‘[9’. For a tailoring that supports the regular expression directly, as well as a key to the notation see Section 8.2, *Examples of Customization.*

**LB25** Do not break between the following pairs of classes relevant to numbers:

\[
\begin{align*}
\text{CL} \times \text{PO} \\
\text{CL} \times \text{PR} \\
\text{NU} \times \text{PO} \\
\text{NU} \times \text{PR} \\
\text{PO} \times \text{OP} \\
\text{PO} \times \text{NU} \\
\text{PR} \times \text{OP} \\
\text{PR} \times \text{NU} \\
\text{HY} \times \text{NU} \\
\text{IS} \times \text{NU} \\
\text{NU} \times \text{NU} \\
\text{SY} \times \text{NU}
\end{align*}
\]

Finally, join alphabetic letters into words and break everything else.

**LB28** Do not break between alphabets (“at”).

\[ \text{AL} \times \text{AL} \]

**LB29** Do not break between numeric punctuation and alphabets (“e.g.”).

\[ \text{IS} \times \text{AL} \]

**LB30** Do not break between letters, numbers, or ordinary symbols and opening or closing punctuation.

\[
(\text{AL} | \text{NU}) \times \text{OP} \\
\text{CL} \times (\text{AL} | \text{NU})
\]

The purpose of this rule is to prevent breaks in common cases where a part of a word appears between delimiters—for example, in “person(s)”.

**LB31** Break everywhere else.

\[ \text{ALL} \div \]

\[ \div \text{ALL} \]