

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%’.

ID × PO

AL × NU

NU × AL

LB24 Do not break between prefix and letters or ideographs.

PR × ID

PR × AL

~~PO × AL~~

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:

CL × PO

~~CL × PR~~

NU × PO

~~NU × PR~~

~~PO × OP~~

~~PO × NU~~

PR × OP

PR × NU

HY × NU

IS × NU

NU × NU

SY × NU

Example pairs: ‘\$9’, ‘\$[’, ‘\$-’, ‘-9’, ‘/9’, ‘99’, ‘,9’, ‘9%’, ‘]’%

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

LB28 Do not break between alphabetics (“at”).

$AL \times AL$

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

$IS \times AL$

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

~~$(AL|NU) \times OP$~~

~~$CL \times (AL|NU)$~~

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.

$ALL \div$

$\div ALL$