

**Title: Follow-up on Action Item 120-A58**

Source: Laurențiu Iancu, Microsoft Corporation

Date: 2010-02-01

Action: For consideration by the Unicode Technical Committee

This document addresses Action Item [120-A58](#):[UTC 120 / L2 217](#) Joint Meeting – Redmond, WA – August 10–14, 2009B.10.6.1.1. General feedback [[L2/09-230](#)][\[120-M1\]](#) Motion:

Add the property value “CP” Close\_Parenthesis with the contents U+0029 and U+005D, and make the rule changes for CP and CL as per [L2/09-128R \(UAX #14-23, draft 2\)](#) in Unicode 5.2.

[\[120-A58\]](#) Action Item for Laurențiu Iancu (updated in the [Open Action Items](#) register):

Supply additional test cases for CP and CL to Mark Davis for inclusion in LineBreakTest.txt, for Unicode 6.0 (shifted from Unicode 5.2).

The Unicode 5.2 revision of the Line Breaking Algorithm introduced a new Line\_Break property value, Close\_Parenthesis (CP), and revised a few rules accordingly. Two characters were assigned this property value, U+0029 RIGHT PARENTHESIS ‘)’ and U+005D RIGHT SQUARE BRACKET ‘]’. Prior to Unicode 5.2, these characters had the Line\_Break property value Close\_Punctuation (CL), similar to U+007D RIGHT CURLY BRACKET ‘}’.

The following table contains 70 test sequences that contain closing parentheses and braces. They are based on real-world samples such as “(s)he” or “'sI1əb(ə)l”, in combinations with characters of various line-breaking classes. Where relevant, those classes and certain code points are explicitly mentioned for clarity. A UTF-8 text file containing only the test cases will be supplied to Mark Davis for inclusion in LineBreakTest.txt.

When regenerating LineBreakTest.txt, the following typos should also be fixed:

- The mechanically added rule “0.2) sot ÷” documented in LineBreakTest.html should be replaced with “0.2) sot ×” in accordance with rule [LB2 of UAX #14](#). As a result of this change, all test cases in LineBreakTest.txt will start with a prohibited break instead of a break opportunity.
- The line count at the end of LineBreakTest.txt should be corrected. Currently it reads “# Lines: 5184” although there are 5202 test cases. The line count should be updated to account for the proposed test cases when LineBreakTest.txt is regenerated.

No	Test case	Relevant characters and/or Line_Break values	Comments
1	(s)he	AL	At beginning of word
2	{s}he		
3	'sIləb(ə)l		In medial position
4	'sIləb{ə}l		
5	code(s).	IS	With Infix_Numeric (IS)
6	code(s.)	IS	
7	code(s)!	EX	With Exclamation (EX)
8	code(s!)	EX	
9	code\(s\)	PR	Escaping (as in regular expressions)
10	code( s )	SP	Certain function call style
11	code{s}		See “bug(s)” in LinebreakTest.txt
12	code{s}.	IS	
13	code{s}!	EX	
14	code{s\}	PR	
15	code{ s }	SP	
16	cod(e)...(s)	IN	Combinations with Inseparable (IN)
17	(cod(e)...s	IN	
18	cod{e}...{s}	IN	
19	{cod{e}...}s	IN	
20	(con-)lang	HY U+002D	Hyphen-minus
21	(con-)lang	BA U+00AD	Soft hyphen
22	(con-)lang	GL U+2011	Non-breaking hyphen
23	(con)-lang	HY U+002D	Similar to the above cases
24	(con)-lang	BA U+00AD	
25	(con)-lang	GL U+2011	
26	{con-}lang	HY U+002D	Hyphen-minus
27	{con-}lang	BA U+00AD	Soft hyphen
28	{con-}lang	GL U+2011	Non-breaking hyphen
29	{con}-lang	HY U+002D	Similar to the above cases
30	{con}-lang	BA U+00AD	
31	{con}-lang	GL U+2011	
32	cré(é)(e)	CM U+0301	Adjacent parentheses
33	cré[er é(e)(s)]	CM U+0301	Nested parentheses
34	cré{er é(e)(s)}	CM U+0301	
35	ambigu(¨)(ë)	CM U+0308	Baseless combining mark
36	ambigu(«¨»)(ë)	...QU-CM-QU-CP...	Guillemets
37	ambigu(« ¨ »)(ë)	...CM-SP-QU-CP...	
38	ambigu« ( ¨ ) »(ë)	U+0020 U+0029 U+0020	Flanked by spaces
39	ambigu« ( ¨ ) »(ë)	U+00A0 U+0029 U+202F	Flanked by non-breaking spaces
40	ambigu{¨}(ë)		Cases with CL similar to the above
41	ambigu{«¨»}(ë)		
42	ambigu{« ¨ »}(ë)		
43	ambigu« { ¨ } »(ë)		
44	ambigu« { ¨ } »(ë)	U+00A0 U+007D U+202F	
45	(czerwono--)niebieska	BA-GL U+00AD U+2011	Combinations with SHY-NBHY

46	(czerwono-)-niebieska	BA-CP-GL	Combinations with SHY-NBHY
47	(czerwono)--niebieska	CP-BA-GL	
48	{czerwono--}niebieska	BA-GL-CL	Cases with CL similar to the above
49	{czerwono-}-niebieska	BA-CL-GL	
50	{czerwono}--niebieska	CL-BA-GL	
51	operator[]( $\emptyset$ );	NU	Adjacent parentheses and digits
52	operator[](){}		Adjacent parentheses and braces
53	本(を)読む	ID	In medial position
54	本(「を」)読む	ID-CL-CP-ID	
55	本「(を)」読む	ID-CP-CL-ID	
56	本{を}読む		Cases with CL similar to the above
57	本{「を」}読む	ID-CL-CL-ID	
58	本[(を)]読む	ID-CP-CP-ID	
59	(ニュー・)ヨーク	NS U+30FB	Combinations with Nonstarter (NS)
60	(ニュー)・ヨーク	NS-CP-NS	
61	{ニュー・}ヨーク		Cases with CL similar to the above
62	{ニュー}・ヨーク		
63	(འགོ་བོ་མཚོ་)གི་རྒྱུ	BB U+1806	Combinations with Todo SHY
64	(འགོ་བོ་མཚོ་)གི་རྒྱུ	CP-BB	
65	{འགོ་བོ་མཚོ་}གི་རྒྱུ	BB-CL	Cases with CL similar to the above
66	{འགོ་བོ་མཚོ་}གི་རྒྱུ	CL-BB	
67	(http://)xn--a	SY	With Break_Symbols (SY)
68	{http://}xn--a	SY	
69	( $\emptyset, 1$ )+(2,3) $\oplus$ (-4,5) $\ominus$ (6,7)	NU, PR, AI, AL	Digits (NU) and symbols (PR, AI, AL)
70	{ $\emptyset, 1$ }{+{2,3} $\oplus$ {-4,5} $\ominus$ {6,7}}	NU, PR, AI, AL	Case with CL similar to the above