

# Proposal for Line Breaking of Halfwidth Forms

To: UTC

From: Koji Ishii (Google Inc.)

Date: 2015-10-25

## Overview

Currently letters in Halfwidth Forms have Unicode Line Breaking property value AL. This document proposes to change them to match to established practices for the scripts, and also to existing implementations.

## Motivations

Historically, Halfwidth forms were originally invented to use a minimum set of CJK characters to fit in 8-bits encoding systems, and were designed to have compatible width as ASCII characters in monospaced fonts to work well in terminal-like applications.

From this purpose, Halfwidth forms are expected to have the same typographic characteristics as their compatibility mapped characters.

This is unlike Fullwidth forms, where their purpose is to typeset foreign characters to fit with native writing systems and thus different typographic characteristics are desired. For instance, the [section 3.2.4 of Requirements for Japanese Text Layout](#) defines how differently Fullwidth Latin letters are typeset from narrow Latin letters, but it does not define different typographic rules for Halfwidth forms.

## Compatibility Considerations

The rules for the line breaking property was proposed in [L2/99-179](#) and then drafted as [L2/99-180](#). Since Halfwidth forms are Lo(h), they resolve to Lo(n) and are assigned AL.

On the other hand, products shipping in CJK in 1999 and today, including Microsoft Windows, Internet Explorer, and Office all handle Halfwidth Letters as ID.

Makoto Kato at Mozilla originally raised this issue. Chromium currently follows the current values, but is willing to change.

Since most popular products have already tailored or willing to change, this proposed changes do not pose any negative compatibility risks.

## Proposed Changes

Code Point	Current	Proposed	<narrow>	Name
FF61	CL		CL 3002	HALFWIDTH IDEOGRAPHIC FULL STOP
FF62	OP		OP 300C	HALFWIDTH LEFT CORNER BRACKET
FF63	CL		CL 300D	HALFWIDTH RIGHT CORNER BRACKET
FF64	CL		CL 3001	HALFWIDTH IDEOGRAPHIC COMMA
FF65	NS		NS 30FB	HALFWIDTH KATAKANA MIDDLE DOT
FF66	AL	ID	ID 30F2	HALFWIDTH KATAKANA LETTER WO
FF67..FF6F	CJ		CJ 30A1..30A9, 30E3..30E7, 30C3	HALFWIDTH KATAKANA LETTER SMALL A..HALFWIDTH KATAKANA LETTER SMALL TU
FF70	CJ		CJ 30FC	HALFWIDTH KATAKANA-HIRAGANA PROLONGED SOUND MARK
FF71..FF9D	AL	ID	ID 30A2..30F3	HALFWIDTH KATAKANA LETTER A..HALFWIDTH KATAKANA LETTER N
FF9E..FF9F	NS	(*1)	CM 3099..309A	HALFWIDTH KATAKANA VOICED SOUND MARK..HALFWIDTH KATAKANA SEMI-VOICED SOUND MARK
FFA0..FFBE	AL	ID	ID 3164, 3131..314A	HALFWIDTH HANGUL FILLER..HALFWIDTH HANGUL LETTER HIEUH
FFC2..FFC7	AL	ID	ID 314F..3154	HALFWIDTH HANGUL LETTER A..HALFWIDTH HANGUL LETTER E
FFCA..FFCF	AL	ID	ID 3155..315A	HALFWIDTH HANGUL LETTER YEO..HALFWIDTH HANGUL LETTER OE
FFD2..FFD7	AL	ID	ID 315B..3160	HALFWIDTH HANGUL LETTER YO..HALFWIDTH HANGUL LETTER YU
FFDA..FFDC	AL	ID	ID 3161..3163	HALFWIDTH HANGUL LETTER EU..HALFWIDTH HANGUL LETTER I

FFE8	AL	ID(*2)	AI 2502	HALFWIDTH FORMS LIGHT VERTICAL
FFE9..FFEC	AL	ID(*2)	AI 2190..2193	HALFWIDTH LEFTWARDS ARROW..HALFWIDTH DOWNWARDS ARROW
FFED..FFEE	AL	ID(*2)	AI 25A0, 25CB	HALFWIDTH BLACK SQUARE..HALFWIDTH WHITE CIRCLE

1. NS is appropriate for this specific case given the rules in [L2/99-179](#) and because these characters are Lm (spacing) while their compatibility equivalents are Mn (non-spacing.) But this proposal does not oppose changing to CM if preferred.
2. Symbols in U+FFE8-FFEE are included for consistency and because they are ID in implementations such as Microsoft Internet Explorer and Office. But this proposal does not oppose keeping the current value or changing to other values if preferred.