



## Proposed Draft Unicode Technical Report #50

# UNICODE PROPERTIES FOR VERTICAL TEXT LAYOUT

Editor	Eric Muller ( <a href="mailto:emuller@adobe.com">emuller@adobe.com</a> )
Date	2011-10-04
This Version	<a href="http://www.unicode.org/reports/tr50/tr50-1.html">http://www.unicode.org/reports/tr50/tr50-1.html</a>
Previous Version	n/a
Latest Version	<a href="http://www.unicode.org/reports/tr50/">http://www.unicode.org/reports/tr50/</a>
Latest Proposed Update	<a href="http://www.unicode.org/reports/tr50/proposed.html">http://www.unicode.org/reports/tr50/proposed.html</a>
Revision	<u>1</u>

## Summary

The layout of Japanese text follows different conventions than the layout of Western texts. Many of the requirements are described in the W3C Working Group Note “Requirements for Japanese Text Layout” [[JLREQ](#)]. This report describes two Unicode character properties which can be used to implement those requirements.

## Status

This is a **draft** document which may be updated, replaced, or superseded by other documents at any time. Publication does not imply endorsement by the Unicode Consortium. This is not a stable document; it is inappropriate to cite this document as other than a work in progress.

A Unicode Technical Report (UTR) contains informative material. Conformance to the Unicode Standard does not imply conformance to any UTR. Other specifications, however, are free to make normative references to a UTR.

Please submit corrigenda and other comments with the online reporting form [[Feedback](#)]. Related information that is useful in understanding this document is found in [References](#). For the latest version of the Unicode Standard see [[Unicode](#)]. For a list of current Unicode Technical Reports see [[Reports](#)]. For more information about versions of the Unicode Standard, see [[Versions](#)].

## Contents

- 1 [Introduction](#)
  - 2 [Grapheme Clusters](#)
  - 3 [Plain Text and Markup](#)
  - 4 [The East Asian Class Property](#)
    - 4.1 [Presentation](#)
    - 4.2 [Property Values](#)
    - 4.3 [Guidelines for Assigning Property Values](#)
      - 4.3.1 [Graphic Characters](#)
      - 4.3.2 [Private-Use Code Points](#)
      - 4.3.3 [Reserved Code Points](#)
      - 4.3.4 [Other Code Points](#)
    - 4.4 [Using the Property](#)
  - 5 [The East Asian Orientation Property](#)
    - 5.1 [Presentation](#)
    - 5.2 [Property Values](#)
    - 5.3 [Relationship to the EAC Property](#)
    - 5.4 [Glyphs Changes for Vertical Orientation](#)
  - 6 [Data File](#)
  - [Acknowledgments](#)
  - [References](#)
  - [Modifications](#)
- 

## 1 Introduction

The JLREQ document uses terms such as “Japanese” and “Western”, and focuses on Japanese typography with Western text elements thrown in.

It is our belief that Japanese and Chinese typography have enough commonality that the goal of the properties should be to handle both cases. It seems that Chinese typography is both somewhat simpler (for one thing, it does not use Hiragana and Katakana) and less codified. We expect that it is possible to handle Chinese by simply “downscaling” what is done for Japanese. However, we do not have enough information to explain this downscaling. The same goes for Korean written using the CJK ideographs. As for Korean written using Hangul – including the occasional CJK ideographs seen today –, we think it follows essentially the Western typographic methods and is therefore out of

scope. Any information that would clarify the situation of the CJK locales besides Japan is very welcome.

As for the “Western” part, we similarly believe that it extends to other parts of the world as well and should really be interpreted in that way.

One difficulty for this document that needs a resolution is whether to use the same terms as JLREQ – in particular “Western” – and to gloss that we really mean something broader, or to use terms that are more easily understood at the cost of a disconnect with the JLREQ terminology.

The layout of Japanese text follows different convention than the layout of Western texts. Many of the requirements are described in the W3C Working Group Note “Requirements for Japanese Text Layout” [JLREQ]. In particular, this note explains the spacing of characters in lines: character occurrences are classified, with classes such as cl-07, commas and cl-19, ideographic, and tables indexed by the classes of adjacent occurrences determine the actual spacing between those occurrences.

This report describes two Unicode character properties which can be use to implement those requirements.

The first property, East Asian Class, is a classification of the characters which eventually contributes to the determination of inter-character spacing, both in unjustified and in justified lines. For example:

### Figure 1. Spacing of Characters

[Figure similar to the bottom part of [Figure 64](#) in JLREQ]

In the first line, the character U+3001 、 IDEOGRAPHIC COMMA is classified as cl-07, commas and the following ideograph is classified as cl-19, ideographic. This particular combination results in an 1/2 em space introduced by layout between the two characters. Similarly, the classification of the middle dot and the characters surrounding them in the last line cause a 1/4 em space around the middle dot. Furthermore, those spaces can shrink or expand as needed for justification.

The second property, East Asian Orientation, determines the orientation of the characters in vertical lines, which can differ from their orientation in horizontal

lines. For example:

## Figure 2. Orientation of Characters

[Figure similar to [Figure 25](#) in JLREQ, also showing the same text in horizontal lines]

The character U+306E の HIRAGANA LETTER NO stays upright in vertical lines, while the character U+0065 e LATIN SMALL LETTER E goes sideways in vertical lines.

The scope of application for the properties and algorithms described here is limited to the texts which are typeset according to Japanese rules. A Japanese word in the context of a English document would not obey the same spacing rules that it would obey in a Japanese document. Also, the issue of character orientation in vertical lines is entirely different when the text is set following, for example, English rules, and the algorithm described here is not relevant to that situation.

The title of this TR is obviously not well aligned with the scope and content. It is too broad given the paragraph above. It is also too narrow as the East Asian Class is not just for vertical text. The motivation for the inclusion of East Asian Class is that it is closely related to East Asian Orientation; one can effectively validate the proposed assignments for East Asian Orientation only in the context of the assignments of the East Asian Class property.

## 2 Grapheme Clusters

As in all matters of typography, the interesting unit of text is not the character, but something of the order of a grapheme cluster: it does not make sense to use a base character upright and a combining mark attached to it sideways.

It is expected that the client of the two properties defined here will select a notion of grapheme cluster, and is interested in obtaining a class and orientation for the clusters. This can be done by simply taking the property value of the first character in each cluster.

## 3 Plain Text and Markup

The properties and algorithms defined here are directly applicable to plain text. As a consequence, some classes listed in JLREQ do not appear here. For

example, the JLREQ class cl-28 warichu opening brackets applies to occurrences of characters which are used to bracket a warichu construct (a form of inline parenthetical); however, there is no way in plain text to express a warichu construct.

When the properties and algorithms are used in rich text, it is acceptable to resolve occurrences of characters to classes not listed here. It is also acceptable to support markup that override the algorithms presented here.

## 4 The East Asian Class Property

### 4.1 Presentation

The JLREQ document describes how character classes affect the spacing of character within lines, and its Appendix A seems to provide the classification of the characters. However, the situation is a bit more complicated.

First, Appendix A does not directly provide a method to classify the character occurrences, but rather provides the lists of classes in which an occurrence can be classified. For example, the character U+0041 A LATIN CAPITAL LETTER A appears in the list cl-27, Western characters, in the list cl-25, unit symbols, and in the list cl-19 ideographic character. The whole document just assumes that character occurrences have been somehow classified and is silent on how that is done.

Second, JLREQ ignores the existence and the common use in desktop computing of the fullwidth characters. The character U+FF21 A FULLWIDTH LATIN CAPITAL LETTER A is not listed anywhere in appendix A, yet in practice, it is used contrastively with U+0041 A LATIN CAPITAL LETTER A to distinguish occurrences which should be classified as cl-19 ideographic character rather than cl-27, Western characters.

Third, JLREQ restricts itself to the ISO/IEC 10646 [\[ISO 10646\]](#) collections 285 Basic Japanese and 286 Japanese Non Ideographics extension; of course, a solution that covers all of Unicode is needed.

To fill those gaps, this report defines a character property, and an algorithm using that property to classify character occurrences.

### 4.2 Property Values

The property values are a subset of the classes defined in JLREQ, where a few classes have been split to facilitate implementation.

Will need to come up with short names and long names, and respect whatever constraints apply to those names.

**Table 1. Property Values for the East Asian Class Property**

cl-01.1	opening bracket, corner
cl-01.2	opening bracket, round
cl-01.3	opening bracket, other
cl-02.1	closing bracket, corner
cl-02.2	closing bracket, round
cl-02.3	closing bracket, other
cl-03	hyphen
cl-04	dividing punctuation mark
cl-05	middle dot
cl-06	full-width stop
cl-07	comma
cl-08	inseparable character
cl-09	iteration mark
cl-10	prolonged sound mark
cl-11	small kana
cl-12	prefixed abbreviation
cl-13	postfixed abbreviation
cl-14	full-width ideographic space
cl-15	hiragana
cl-16	katakana
cl-19.1	ideographic character
cl-19.2	fullwidth number
cl-19.3	symbol
cl-26	western word space
cl-27	western character

The split of cl-01, opening bracket and cl-02, closing bracket is purely on a

graphical basis. The purpose of the split is that it is sometimes desirable to use slightly different amounts of space depending on the shape of the bracket. The split of cl-19, ideographic serves a similar purpose.

Do we need a class for fixed width Western spaces? if we make them cl-27, they will get non-zero aki (space introduced by layout) next to ideographs or in math formulas.

## 4.3 Guidelines for Assigning Property Values

### 4.3.1 Graphic Characters

If a character is routinely considered as an integral part of the Japanese writing system, it is assigned to one of the classes cl-01..cl-19. This is the case for characters in ISO/IEC 10646 collections 285 Basic Japanese and 286 Japanese Non Ideographics extension, except that Basic Latin characters are replaced by their companion character from the Halwidth and Fullwidth forms block. It is also the case for characters outside those collections which clearly are part of a set where a large part of the set is in the collections; for example, JLREQ includes U+2032 ' PRIME and U+2033 " DOUBLE PRIME in class cl-13; it is only natural to treat U+2034 "' TRIPLE PRIME and U+2057 "" QUADRUPLE PRIME in the same way.

Characters which are more symbolic than alphabetic are assigned to cl-19.3, because they can function typographically as ideographs.

Remaining characters are classified in cl-26 or cl-27.

### 4.3.2 Private-Use Code Points

The largest use of PUA code points in Japanese texts is for ideographs, therefore PUA code points are assigned the class cl-19.1.

### 4.3.3 Reserved Code Points

Generally, reserved code points should be assigned the class most likely for their intended future assignment, to maximize forward compatibility. For the ideographic block of the SMP, planes 2 and 3, the class is cl-19.1. For the blocks of symbols, the class is cl-19.3. Other reserved code points are classified cl-27.

### 4.3.4 Other Code Points

The property is not relevant to format, control, surrogate and noncharacter code points.

## 4.4 Using the Property

Because of the restriction of our algorithm to plain text, and given that actual usage of fullwidth characters, it turns out that one can simply use the property value of a character as the class to assign to all occurrences of that character. In other words, the algorithm is a simple lookup of the property value.

There is actually one character for which a contextual determination would be useful and reliable: U+00AE ® REGISTERED SIGN, which can occur both following terms in kanji/kana and following terms in Latin. An occurrence of ® should be assigned the same class as the character it follows. Others? Enough to warrant the complexity of contextual rules?

There are other cases where the character is used routinely in both Japanese and Western contexts: the quotes are a good example. While contextual determination would be useful, it's probably the case that it's not going to be reliable.

## 5 The East Asian Orientation Property

### 5.1 Presentation

In vertical texts, most characters are presented either upright or sideways.

A few characters have a different appearance altogether. This is for example the case for square katakana symbols, which have two to four component kanas arranged so that that they form two lines; the two lines are stacked horizontally in horizontal text, and vertically in vertical text. A complete chart showing representative glyphs for both orientations is provided below.

### 5.2 Property Values

The possible property values are given in table 2.

**Table 2. Property Values for the East Asian Orientation Property**

U characters that are displayed upright



S characters that are displayed sideways

SB brackets displayed sideways

T characters which are not just upright or sideways, but require a different glyph

TK small kanas, that are displayed upright and shifted

The property values SB and TK are meant to ensure some level of compatibility with existing fonts.

The SB property is conceptually a subclass of S. It captures the common practice in fonts to actually handle those characters as if they were transformed.

The TK property is conceptually a subclass of T, as the characters are transformed; it captures the fact the transformation is a small shift which, while desirable, is not critical to the reading of the text.

Do we need TK at all?

### 5.3 Relationship to the EAC Property

Characters with an East Asian Class cl-27 all have the orientation S.

Characters with the East Asian Class cl-11 are exactly those with the orientation TK.

Characters with the East Asian Classes cl-01 or cl-02 are exactly those with orientation SB.

Characters with the East Asian Class cl-19 all have orientation U.

### 5.4 Glyphs Changes for Vertical Orientation

Table 3 provides representative glyphs for the horizontal and vertical appearance of characters with the property value T.

Add glyphs for all the entries: 301F, 332C, FF61, FF64, 1F200, 1F201, halfwidth small kanas. Some glyphs (2018, 2019) may not be correct.

Table 3. Glyph Changes for Vertical Orientation

character	H	V	comment
-----------	---	---	---------

character	H	V	comment
U+2018 LEFT SINGLE QUOTATION MARK	‘	’	
U+2019 RIGHT SINGLE QUOTATION MARK	‚	ƒ	
U+3001 IDEOGRAPHIC COMMA	、	、	
U+3002 IDEOGRAPHIC STOP	。	。	
U+301C WAVE DASH	～	﹂	mirroring, not just rotation
U+301D REVERSED DOUBLE PRIME QUOTATION MARK	≡	≡	
U+301E DOUBLE PRIME QUOTATION MARK	≡	≡	
U+301F LOW DOUBLE PRIME QUOTATION MARK			
U+30FC KATAKANA-HIRAGANA PROLONGED SOUND MARK	ー	丨	mirroring, not just rotation
U+3041 HIRAGANA LETTER SMALL A	あ	あ	
U+3043 HIRAGANA LETTER SMALL I	い	い	
U+3045 HIRAGANA LETTER SMALL U	う	う	
U+3047 HIRAGANA LETTER SMALL E	え	え	
U+3049 HIRAGANA LETTER SMALL O	お	お	
U+3063 HIRAGANA LETTER SMALL TU	つ	つ	
U+3083 HIRAGANA LETTER SMALL YA	や	や	
U+3085 HIRAGANA LETTER SMALL YU	ゆ	ゆ	
U+3087 HIRAGANA LETTER SMALL YO	よ	よ	

character	H	V	comment
U+308E HIRAGANA LETTER SMALL WA	わ	わ	
U+3095 HIRAGANA LETTER SMALL KA	か	か	
U+3096 HIRAGANA LETTER SMALL KE	け	け	
U+30A1 KATAKANA LETTER SMALL A	ア	ア	
U+30A3 KATAKANA LETTER SMALL I	イ	イ	
U+30A5 KATAKANA LETTER SMALL U	ウ	ウ	
U+30A7 KATAKANA LETTER SMALL E	エ	エ	
U+30A9 KATAKANA LETTER SMALL O	オ	オ	
U+30C3 KATAKANA LETTER SMALL TU	ツ	ツ	
U+30E3 KATAKANA LETTER SMALL YA	ヤ	ヤ	
U+30E5 KATAKANA LETTER SMALL YU	ユ	ユ	
U+30E7 KATAKANA LETTER SMALL YO	ヨ	ヨ	
U+30EE KATAKANA LETTER SMALL WA	ワ	ワ	
U+30F5 KATAKANA LETTER SMALL KA	カ	カ	
U+30F6 KATAKANA LETTER SMALL KE	ケ	ケ	
U+31F0 KATAKANA LETTER SMALL KU	ク	ク	
U+31F1 KATAKANA LETTER SMALL SI	シ	シ	
U+31F2 KATAKANA LETTER SMALL SU	ス	ス	
U+31F3 KATAKANA LETTER SMALL TO	ト	ト	

character	H	V	comment
U+31F4 KATAKANA LETTER SMALL NU	ヌ	ヌ	
U+31F5 KATAKANA LETTER SMALL HA	ハ	ハ	
U+31F6 KATAKANA LETTER SMALL HI	ヒ	ヒ	
U+31F7 KATAKANA LETTER SMALL HU	フ	フ	
U+31F8 KATAKANA LETTER SMALL HE	ヘ	ヘ	
U+31F9 KATAKANA LETTER SMALL HO	ホ	ホ	
U+31FA KATAKANA LETTER SMALL MU	ム	ム	
U+31FB KATAKANA LETTER SMALL RA	ラ	ラ	
U+31FC KATAKANA LETTER SMALL RI	リ	リ	
U+31FD KATAKANA LETTER SMALL RU	ル	ル	
U+31FE KATAKANA LETTER SMALL RE	レ	レ	
U+31FF KATAKANA LETTER SMALL RO	ロ	ロ	
U+3300 SQUARE APAATO	アパ ート	丨ア トパ	
U+3301 SQUARE ARUHUA	アル ファ	フア アル	
U+3302 SQUARE ANPEA	アン ペア	ペア アン	
U+3303 SQUARE AARU	ア ール	ルア 丨	
U+3304 SQUARE ININGU	イニ ング	ンイ ゲニ	
U+3305 SQUARE INTI	イン チ	チイ ン	
U+3306 SQUARE UON	ウオ ン	ンウ オ	

character	H	V	comment
U+3307 SQUARE ESUKUUDO	エ ク ド	ク エ ド ス	
U+3308 SQUARE EEKAA	エ カ ー	カ エ ー ー	
U+3309 SQUARE ONSU	オ ン ス	ス オ ン	
U+330A SQUARE OOMU	オ ム	ム オ ー	
U+330B SQUARE KAIRI	カ イ リ	リ カ イ	
U+330C SQUARE KARATTO	カ ラ ッ ト	ツ カ ラ ト	
U+330D SQUARE KARORII	カ ロ リ ー	リ カ ー ロ	
U+330E SQUARE GARON	ガ ロ ン	ン ガ ロ	
U+330F SQUARE GANMA	ガ ン マ	マ ガ ン	
U+3310 SQUARE GIGA	ギ ガ	ギ ガ	
U+3311 SQUARE GINII	ギ ニ ー	ー ギ ニ	
U+3312 SQUARE KYURII	キ ュ リ ー	リ キ ー ユ	
U+3313 SQUARE GIRUDAA	ギ ル ダ ー	ダ ギ ー ル	
U+3314 SQUARE KIRO	キ ロ	キ ロ	
U+3315 SQUARE KIROGURAMU	キ ロ グ ラ ム	グ キ ー ラ ム ロ	
U+3316 SQUARE KIROMEETORU	キ ロ メ ー ト ル	ー キ ロ メ ト ル	
U+3317 SQUARE KIROWATTO	キ ロ ワ ッ ト	ワ ッ キ ロ ト	
U+3318 SQUARE GURAMU	グ ラ ム	ム グ ラ	
U+3319 SQUARE GURAMUTON	グ ラ ム ト ン	ト グ ラ ム	

character	H	V	comment
U+331A SQUARE KURUZEIRO	クル ゼロ	ゼク クル	
U+331B SQUARE KUROONE	クロ ーネ	ーク ネロ	
U+331C SQUARE KEESU	ケー ス	スケ ー	
U+331D SQUARE KORUNA	コル ナ	ナコ ル	
U+331E SQUARE KOOPO	コー ポ	ポコ ー	
U+331F SQUARE SAIKURU	サイ クル	クサ ルイ	
U+3320 SQUARE SANTIIMU	サン チム	チサ ムン	
U+3321 SQUARE SIRINGU	シリ ング	ンシ ゲリ	
U+3322 SQUARE SENTI	セン チ	チセ ン	
U+3323 SQUARE SENTO	セン ト	トセ ン	
U+3324 SQUARE DAASU	ダー ス	スダ ー	
U+3325 SQUARE DESI	デ シ	デ シ	
U+3326 SQUARE DORU	ドル ル	ドル ル	
U+3327 SQUARE TON	トン ン	トン ン	
U+3328 SQUARE NANO	ナ ノ	ナ ノ	
U+3329 SQUARE NOTTO	ノッ ト	トノ ッ	
U+332A SQUARE HAITU	ハイ ツ	ツハ イ	
U+332B SQUARE PAASENTO	パー セント	セパ ー	
U+332C SQUARE PAATU			

character	H	V	comment
U+332D SQUARE BAARERU	バー レル	レバ ルI	
U+332E SQUARE PIASUTORU	ピア ストル	スピ トルア	
U+332F SQUARE PIKURU	ピク ル	ルピ ク	
U+3330 SQUARE PIKO	ピ コ	ピ コ	
U+3331 SQUARE BIRU	ビ ル	ビ ル	
U+3332 SQUARE HUARADDO	ファ ラッド	ラフ ッドア	
U+3333 SQUARE HUIITO	フィ ート	Iフ トイ	
U+3334 SQUARE BUSSYERU	ブッ シェル	シブ ェルツ	
U+3335 SQUARE HURAN	フラ ン	ンフ ラ	
U+3336 SQUARE HEKUTAARU	ヘク タール	ター ルク	
U+3337 SQUARE PESO	ペ ソ	ペ ソ	
U+3338 SQUARE PENIHI	ペニ ヒ	ヒペ ニ	
U+3339 SQUARE HERUTU	ヘル ツ	ツヘ ル	
U+333A SQUARE PENSU	ペン ス	スペ ン	
U+333B SQUARE PEEZI	ペー ジ	ジペ I	
U+333C SQUARE BEETA	ベー タ	タバ I	
U+333D SQUARE POINTO	ポイ ント	ンポ トイ	
U+333E SQUARE BORUTO	ボル ト	トボ ル	
U+333F SQUARE HON	ホ ン	ホ ン	

character	H	V	comment
U+3340 SQUARE PONDO	ボン ド	ドボン	
U+3341 SQUARE HOORU	ホー ル	ルホー ー	
U+3342 SQUARE HOON	ホー ン	ンホー ー	
U+3343 SQUARE MAIKURO	マイ クロ	クマ ロイ	
U+3344 SQUARE MAIRU	マイ ル	ルマイ イ	
U+3345 SQUARE MAHHA	マッ ハ	ハマ ッ	
U+3346 SQUARE MARUKU	マル ク	クマル	
U+3347 SQUARE MANSYON	マン ション	シマ ション	
U+3348 SQUARE MIKURON	ミク ロン	ロンミ ク	
U+3349 SQUARE MIRI	ミ リ	ミ リ	
U+334A SQUARE MIRIBAARU	ミリ バル	バミ ルリ	
U+334B SQUARE MEGA	メ ガ	メ ガ	
U+334C SQUARE MEGATON	メガ トン	トメ ンガ	
U+334D SQUARE MEETORU	メー トル	トメ ル	
U+334E SQUARE YAADO	ヤー ド	ドヤ ー	
U+334F SQUARE YAARU	ヤー ル	ルヤ ー	
U+3350 SQUARE YUAN	ユア ン	ンユ ア	
U+3351 SQUARE RITTORU	リッ トル	トリ ル	
U+3352 SQUARE RIRA	リ ラ	リ ラ	



character	H	V	comment
U+3353 SQUARE RUPPII	ル ー	ル ピ	
U+3354 SQUARE RUUBURU	ル ブル	ブル ル	
U+3355 SQUARE REMU	レ ム	レ ム	
U+3356 SQUARE RENTOGEN	レ トゲン	ト レ ジン	
U+3357 SQUARE WATTO	ワ ット	ト ワ ット	
U+337B SQUARE ERA NAME HEISEI	平 成	平 成	
U+337C SQUARE ERA NAME SYOUWA	昭 和	昭 和	
U+337D SQUARE ERA NAME TAISYOU	大 正	大 正	
U+337E SQUARE ERA NAME MEIZI	明 治	明 治	
U+337F SQUARE CORPORATION	株 式 会 社	株 式 会 社	
U+FF61 HALFWIDTH IDEOGRAPHIC FULL STOP			
U+FF64 HALFWIDTH IDEOGRAPHIC COMMA			
U+FF67 HALFWIDTH KATAKANA LETTER SMALL A			
U+FF68 HALFWIDTH KATAKANA LETTER SMALL I			
U+FF69 HALFWIDTH KATAKANA LETTER SMALL U			
U+FF6A HALFWIDTH KATAKANA LETTER SMALL E			
U+FF6B HALFWIDTH KATAKANA LETTER SMALL O			
U+FF6C HALFWIDTH KATAKANA LETTER SMALL YA			

character	H	V	comment
U+FF6D HALFWIDTH KATAKANA LETTER SMALL YU			
U+FF6E HALFWIDTH KATAKANA LETTER SMALL YO			
U+FF6F HALFWIDTH KATAKANA LETTER SMALL TU			
U+1F200 SQUARE HIRAGANA HOKA			
U+1F201 SQUARE KATAKAN KOKO			

## 6 Data File

The data file, in UCD syntax: [ealayout.txt](#).

To help during the review, the following data files are available as well:

- [bycp.html](#), ordered by code point, with character names and block names.
- [byclass.html](#), grouped by class.
- [byorientation.html](#), grouped by orientation.

U+3030 ~ WAVY DASH; should it be rotated/mirrored like U+301C ~ WAVE DASH?

U+2016 || DOUBLE VERTICAL LINE; JRLEQ classifies this character as cl-19 ideographic; typically, this is a clue that it is upright; also, JIS 0213:2000 does not give a vertical variant. On the other hand, it seems that 'vert' often presents it sideways. Which is right? Could it be that font vendors have been influenced by U+30A0 = KATAKANA-HIRAGANA DOUBLE HYPHEN?

## Acknowledgments

Thanks to the reviewers: Julie Allen, Ken Lunde, Nat McCully, Ken Whistler, Taro Yamamoto.

## References

[JLREQ] [Requirements for Japanese Text layout](#), W3C Working Group Note 4 June 2009

- [Errata] Updates and Errata  
<http://www.unicode.org/errata>
- [Feedback] <http://www.unicode.org/reporting.html>  
For reporting errors and requesting information online.
- [ISO 10646] International Organization for Standardization. Information Technology—Universal Multiple-Octet Coded Character Set (UCS). (ISO/IEC 10646:2011).  
For availability, see: <http://www.iso.org>
- [Reports] Unicode Technical Reports  
<http://www.unicode.org/reports/>  
For information on the status and development process for technical reports, and for a list of technical reports.
- [Unicode] The Unicode Standard, Version 6.1.0, defined by: The Unicode Standard, Version 6.1.0 (Mountain View, CA: The Unicode Consortium, 2012. ISBN 978-1-936213-02-3)  
<http://www.unicode.org/versions/Unicode6.1.0>
- [Versions] Versions of the Unicode Standard  
<http://www.unicode.org/versions/>  
For details on the precise contents of each version of the Unicode Standard, and how to cite them.

## Modifications

This section indicates the changes introduced by each revision.

### Revision 1

- (draft 1) First working draft.

---

Copyright © 2011 Unicode, Inc. All Rights Reserved. The Unicode Consortium makes no expressed or implied warranty of any kind, and assumes no liability for errors or omissions. No liability is assumed for incidental and consequential damages in connection with or arising out of the use of the information or programs contained or accompanying this technical report. The Unicode [Terms of Use](#) apply.

Unicode and the Unicode logo are trademarks of Unicode, Inc., and are registered in some jurisdictions.