



Proposed Update Unicode® Standard Annex Technical Report #50

UNICODE VERTICAL TEXT LAYOUT

Version	Unicode 10.0.0 (draft 4)
Editors	Koji Ishii 石井宏治 (Google Inc.), Laurentiu Iancu (Microsoft Corporation)
Date	2017-03-03
This Version	http://www.unicode.org/reports/tr50/tr50-18.html
Previous Version	http://www.unicode.org/reports/tr50/tr50-17.html
Latest Version	http://www.unicode.org/reports/tr50/
Latest Proposed Update	http://www.unicode.org/reports/tr50/proposed.html
Data	http://www.unicode.org/Public/vertical/revision-18/
Revision	18

Summary

The Unicode code charts generally show characters oriented for horizontal presentation. However, some of the glyphs are actually oriented for vertical presentation. A few characters change shape or orientation when the text is rotated from horizontal to vertical.

When text is presented, there are various conventions for the orientation of the characters with respect to the line. In most cases, characters are oriented in an upright manner similar to their presentation in the Unicode code charts. In a few cases, when presented in vertical lines, the characters will appear rotated or transformed in various ways. For example, in East Asia, Han ideographs, Kana syllables, Hangul syllables, and Latin letters in acronyms are upright, while words and sentences in the Latin script are typically sideways. This report describes a Unicode character property which can serve as a stable default orientation of characters for reliable document interchange.

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This is a **draft** document which may be updated, replaced, or superseded by other documents at any time.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.~~

A Unicode Standard Annex (UAX) forms an integral part of the Unicode Standard, but is published online as a separate document. The Unicode Standard may require conformance to normative content in a Unicode Standard Annex, if so specified in the Conformance chapter of that version of the Unicode Standard. The version number of a UAX document corresponds to the version of the Unicode Standard of which it forms a part.

Please submit corrigenda and other comments with the online reporting form [\[Feedback\]](#). Related information that is useful in understanding this annex is found in Unicode Standard Annex #41, “[Common References for Unicode Standard Annexes](#).” 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 [\[Versions\]](#). For any errata which may apply to this annex, see [\[Errata\]](#).

Contents

- 1 [Overview and Scope](#)
- 2 [Conformance](#)
- 3 [The Vertical Orientation Property \(vo\)](#)
 - 3.1 [Property Values](#)
 - 3.2 [Grapheme Clusters](#)
 - 3.3 [Vertical Glyphs in the Code Charts](#)
- 4 [Tailorings](#)
 - 4.1 [The Brackets](#)
 - 4.2 [The Arrows](#)
- 5 [Glyphs Changes for Vertical Orientation](#)
- 6 [Data File](#)
- [Acknowledgments](#)
- [References](#)
- [Modifications](#)

1 Overview and Scope

When text is displayed in vertical lines, there are various conventions for the orientation of the characters with respect to the line. In East Asia, Han ideographs, Kana syllables, and Hangul syllables, along with Latin letters of acronyms, remain upright, meaning that they appear with the same orientation as in the code charts, but words and sentences that are composed of characters of the Latin script are typically oriented sideways, as can be seen in [Figure 1](#).

Figure 1. Japanese [Vertical Text](#)

平気で2時間近くに及ぶというパフォーマンスの凄さでも知られるティナリウエンだが、そのステージに向けて格好の予習テキストとなりそうなDVDも出る。ジュールズ・ホランドがホストを務め、10年以上にわたり人気を誇っている英のTV音楽番組「Later... With Jools Holland」に出演したワイルド・ミュージック系の33組のパフォーマンスをまとめた『レイタ

In many parts of the world, most characters are upright, as can be seen in **F**igure 2.

Figure 2. Western **Vertical **T**ext**



Most languages and scripts are written horizontally and vertical presentation is a special case, usually used for short runs of text (as in **F**igure 2). Some languages, however, have publishing traditions that provide for long-format vertical text presentation, notably East Asian languages such as Japanese. In those languages, the orientation in which

characters are laid out can vary, depending on the scripts, the style, and sometimes the context. The preferred or desired orientation may also change over time.

While the choice of orientation for a character can vary across documents, it is important that the choice made by an author for a specific document be clearly established, so that a rendering system can display what the author intended. It is also important that this choice be established independently of the font resources, as the rendering systems may have to use other fonts than those intended or specified in the document. Finally, the expression of the author's choice should be relatively concise, to facilitate document authoring and minimize document size.

This report describes a Unicode character property which can serve as a stable default character orientation for the purpose of reliable document interchange.

For the purpose of reliable document interchange, this property defines an unambiguous default value, so that implementations could reliably render a character stream based solely on the property values, without depending on other information such as provided in the tables of the selected font.

The intent is that document formats should offer to the author the possibility of specifying the desired orientation of a given character (either all occurrences or a particular occurrence), and that in the absence of an explicit specification, the orientation is implicitly that defined by the property presented in this report.

In plain text, which by definition does not allow the recording of any data beyond the characters, the orientations are by necessity those specified by the property.

The actual choice for the property values should result in a reasonable or legible default, but it may not be publishing-material quality, and it may not be a good choice if used in a specific style or context.

The property values are chosen first to match existing practice in Japanese context in Japan and then in other East Asian contexts in their respective environments. For characters that are not generally used in such environments, similarity to existing characters has been taken into consideration. Commonly used characters of Latin and other scripts that appear in Japanese and other East Asian environments are also taken into account, but with the lower priority.

The scope of the property is limited by the scope of Unicode. For example, Unicode does not directly support the representation of texts and inscriptions using Egyptian Hieroglyphs. Instead, Unicode provides characters intended for use when writing *about* such texts or inscriptions, or for use in conjunction with a markup system such as the *Manuel de Codage*. While the property values are defined for Egyptian Hieroglyphs, they are meaningful only for occurrences of these characters in discursive texts; when the characters are used with markup, the markup ultimately controls the orientation. See [\[Unicode\]](#), Section 11.4 for a more complete discussion of the scope of Egyptian Hieroglyph characters.

2 Conformance

The property defined in this report is informative. The intent of this report is to provide, in the absence of other information, a reasonable way to determine the correct orientation of characters, but this behavior can be overridden by a higher-level protocol, such as through markup or by the preferences of a layout application. This default determination is defined

in the accompanying data file [Data50](#), but in no way implies that the character is used only in that orientation.

For more information on the conformance implications, see [\[Unicode\]](#), Section 3.5, Properties, in particular the definition (D35) of an informative property.

~~Note that the property is not part of the Unicode Character Database [\[UCD\]](#).~~

3 The Vertical_Orientation Property (vo)

3.1 Property Values

The possible property values are given in [Table 1](#).

Table 1. Property Values

U	Characters which are displayed upright, with the same orientation that appears in the code charts.
R	Characters which are displayed sideways, rotated 90 degrees clockwise compared to the code charts.
Tu	Characters which are not just upright or sideways, but generally require a different glyph than in the code charts when used in vertical texts. In addition, as a fallback, the character can be displayed with the code chart glyph upright.
Tr	Same as Tu except that, as a fallback, the character can be displayed with the code chart glyph rotated 90 degrees clockwise.

Note that the orientation is described with respect to the appearance in the code charts.

3.2 Grapheme Clusters

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

A possible choice for the notion of grapheme cluster is either that of legacy grapheme cluster or that of extended grapheme cluster, as defined in [\[UAX29\]](#).

The orientation for a grapheme cluster as a whole is then determined by taking the orientation of the first character in the cluster, with the following exception:

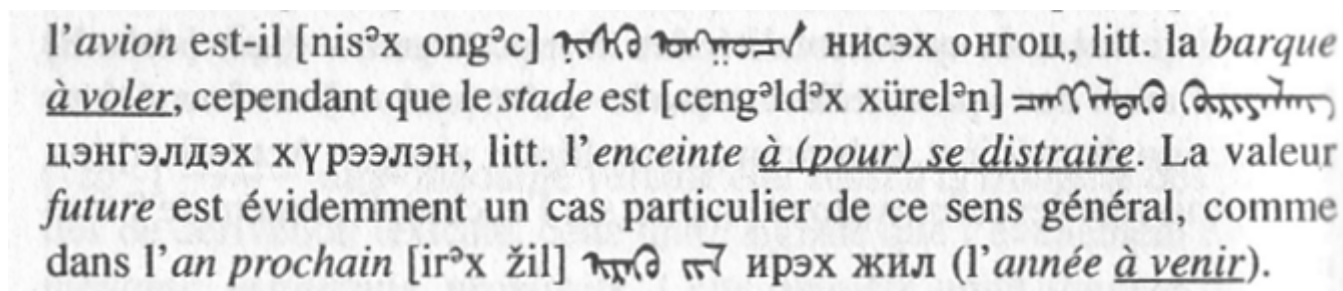
If the cluster contains an enclosing combining mark (general category Me), then the whole cluster has the Vertical_Orientation value U.

3.3 Vertical Glyphs in the Code Charts

The Unicode code charts generally show characters in the orientation they take when used in horizontal lines. However, prior to Unicode 7.0, there were a few exceptions, mostly for characters or scripts which are normally written in vertical lines; in those cases, the code charts used to show the characters in the same orientation as in vertical lines.

Furthermore, such characters are often rotated when displayed in horizontal lines; Figure 3 shows an example of Mongolian text in horizontal lines in which the Mongolian characters are rotated 90 degrees counterclockwise with respect to the code charts prior to Unicode 7.0.

Figure 3. Mongolian Text on Horizontal Lines



The Unicode 7.0 code charts changed the orientation of characters for Mongolian and Phags-pa by rotating counterclockwise so that they match the orientation in horizontal lines. This change makes the code charts more consistent with other scripts in terms of the orientation of characters. It also aligns the code charts with many recent rendering systems such as OpenType, and therefore it is expected to make implementations of the property easier. However, implementations should be aware that underlying rendering systems may not have exactly the same orientation of characters as the code charts.

While this property defines only default orientations compared to the code charts, high-level protocols or applications could combine information provided in a font's tables with the property values to more reliably calculate in which orientation they should render such glyphs, in order to achieve the desired visual result.

4 Tailorings

To facilitate tailorings, this report identifies sets of characters which behave similarly, and for which it can be useful to tailor the orientation as a group.

4.1 The Brackets

This set contains brackets, which while they appear rotated, are commonly implemented as if they were transformed.

Table 2. The Brackets Set

U+00AB
U+00BB
U+201C..U+201F
U+2039..U+203A
U+2045..U+2046

U+3008..U+3011
U+3014..U+301B
U+FE59..U+FE5E
U+FF08..U+FF09
U+FF3B
U+FF3D
U+FF5B
U+FF5D
U+FF5F..U+FF60
U+FF62..U+FF63

4.2 The Arrows

This set contains arrows.

Table 3. The Arrows Set

U+2190..U+21FF
U+261A..U+261F
U+2794
U+2798..U+27AF
U+27B1..U+27BE
U+27F0..U+27FF
U+2900..U+297F
U+2B00..U+2B11
U+2B30..U+2B4F
U+2B5A..U+2BB7
U+2BEC..U+2BEF
U+FFE9..U+FFEC
U+1F800..U+1F80B
U+1F810..U+1F847

U+1F850..U+1F859
U+1F860..U+1F887
U+1F890..U+1F8AD

5 Glyphs Changes for Vertical Orientation

Table 4 provides representative glyphs for the horizontal and vertical appearance of characters with the **Vertical Orientation** property values **Tr** and **Tr**.

The vertical glyphs that are shown in the table are exemplary, and their presence does not imply that font implementations should necessarily support them. Font developers should instead research the vertical glyph conventions for the intended regions to determine whether a vertical glyph is necessary for a particular character, and what the appropriate vertical glyph should be.

The Horizontal column may also specify more than one glyph when regional or other differences exist. Font developers should adhere to regional conventions when determining the appearance of horizontal glyphs.

Table 4. Glyph Changes for Vertical Orientation

Character	Horizontal	Vertical
U+2329 LEFT-POINTING ANGLE BRACKET		
U+232A RIGHT-POINTING ANGLE BRACKET		
U+3001 IDEOGRAPHIC COMMA		 
U+3002 IDEOGRAPHIC FULL STOP		 
U+3008 LEFT ANGLE BRACKET		
U+3009 RIGHT ANGLE BRACKET		
U+300A LEFT DOUBLE ANGLE BRACKET		
U+300B RIGHT DOUBLE ANGLE BRACKET		
U+300C LEFT CORNER BRACKET		

Character	Horizontal	Vertical
U+300D RIGHT CORNER BRACKET		
U+300E LEFT WHITE CORNER BRACKET		
U+300F RIGHT WHITE CORNER BRACKET		
U+3010 LEFT BLACK LENTICULAR BRACKET		
U+3011 RIGHT BLACK LENTICULAR BRACKET		
U+3014 LEFT TORTOISE SHELL BRACKET		
U+3015 RIGHT TORTOISE SHELL BRACKET		
U+3016 LEFT WHITE LENTICULAR BRACKET		
U+3017 RIGHT WHITE LENTICULAR BRACKET		
U+3018 LEFT WHITE TORTOISE SHELL BRACKET		
U+3019 RIGHT WHITE TORTOISE SHELL BRACKET		
U+301A LEFT WHITE SQUARE BRACKET		
U+301B RIGHT WHITE SQUARE BRACKET		
U+301C WAVE DASH		
U+301D REVERSED DOUBLE PRIME QUOTATION MARK		
U+301E DOUBLE PRIME QUOTATION MARK		
U+301F LOW DOUBLE PRIME QUOTATION MARK		

Character	Horizontal	Vertical
U+3030 WAVY DASH		
U+3041 HIRAGANA LETTER SMALL A		
U+3043 HIRAGANA LETTER SMALL I		
U+3045 HIRAGANA LETTER SMALL U		
U+3047 HIRAGANA LETTER SMALL E		
U+3049 HIRAGANA LETTER SMALL O		
U+3063 HIRAGANA LETTER SMALL TU		
U+3083 HIRAGANA LETTER SMALL YA		
U+3085 HIRAGANA LETTER SMALL YU		
U+3087 HIRAGANA LETTER SMALL YO		
U+308E HIRAGANA LETTER SMALL WA		
U+3095 HIRAGANA LETTER SMALL KA		
U+3096 HIRAGANA LETTER SMALL KE		
U+309B KATAKANA-HIRAGANA VOICED SOUND MARK		
U+309C KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK		
U+30A0 KATAKANA-HIRAGANA DOUBLE HYPHEN		

Character	Horizontal	Vertical
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+30FC KATAKANA-HIRAGANA PROLONGED SOUND MARK	ー	丨
U+3127 BOPOMOFO LETTER I	一 丨	一
U+31F0 KATAKANA LETTER SMALL KU	ク	ク
U+31F1 KATAKANA LETTER SMALL SI	シ	シ

Character	Horizontal	Vertical
U+31F2 KATAKANA LETTER SMALL SU	ス	ス
U+31F3 KATAKANA LETTER SMALL TO	ト	ト
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

Character	Horizontal	Vertical
U+3303 SQUARE AARU	アー ル	ルア ー
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
U+3307 SQUARE ESUKUUDO	エス クード	クエ ドス
U+3308 SQUARE EEKAA	エー カー	カエ ー
U+3309 SQUARE ONSU	オン ス	スオン
U+330A SQUARE OOMU	オー ム	ムオー
U+330B SQUARE KAIRI	カイ リ	リカイ
U+330C SQUARE KARATTO	カラ ット	ツカ トラ
U+330D SQUARE KARORII	カロ リー	リカロ
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	リキュ
U+3313 SQUARE GIRUDAA	ギル ダー	ダギ ール

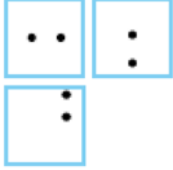
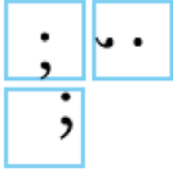
Character	Horizontal	Vertical
U+3314 SQUARE KIRO	キ ロ	キ ロ
U+3315 SQUARE KIROGURAMU	キロ グラム	グ キ ロ ム
U+3316 SQUARE KIROMEETORU	キロメ ートル	丨 キ ロ メ ー ル
U+3317 SQUARE KIROWATTO	キロ ワット	ワ ッ キ ト ロ
U+3318 SQUARE GURAMU	グラ ム	ム グ ラ
U+3319 SQUARE GURAMUTON	グラ ム トン	ム ト グ ラ ン
U+331A SQUARE KURUZEIRO	クル ゼイロ	ゼ ク ロ ル
U+331B SQUARE KUROONE	クロ ー ネ	丨 ク ロ ネ
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 チーム	チ ー サ ム ン
U+3321 SQUARE SIRINGU	シリ ング	ン シ グ 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

Character	Horizontal	Vertical
U+3325 SQUARE DESI	デ シ	シ デ
U+3326 SQUARE DORU	ド ル	ル ド
U+3327 SQUARE TON	ト ン	ン ト
U+3328 SQUARE NANO	ナ ノ	ノ ナ
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 ン ト パー
U+332C SQUARE PAATU	パー ツ	ツ パー
U+332D SQUARE BAARERU	バー レル	レ バ ル ー
U+332E SQUARE PIASUTORU	ピア スト ル	ス ピ ア ル
U+332F SQUARE PIKURU	ピク ル	ル ピ ク
U+3330 SQUARE PIKO	ピ コ	コ ピ
U+3331 SQUARE BIRU	ビ ル	ル ビ
U+3332 SQUARE HUARADDO	ファ ラッド	ラ ッ ド ファ
U+3333 SQUARE HUIITO	フィ ート	フ イ ト
U+3334 SQUARE BUSSYERU	ブッ シェル	シ ェ ッ ブ
U+3335 SQUARE HURAN	フラ ン	ン フラ

Character	Horizontal	Vertical
U+3336 SQUARE HEKUTAARU	ヘク タール	タヘ ールク
U+3337 SQUARE PESO	ペ ソ	ソペ
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 ー
U+333C SQUARE BEETA	ベー タ	タベ ー
U+333D SQUARE POINTO	ポイ ント	ンポ イ
U+333E SQUARE BORUTO	ボル ト	トボル
U+333F SQUARE HON	ホン	ホン
U+3340 SQUARE PONDO	ボン ド	ドボン
U+3341 SQUARE HOORU	ホー ル	ルホ ー
U+3342 SQUARE HOON	ホー ン	ンホ ー
U+3343 SQUARE MAIKURO	マイ クロ	クマ ロイ
U+3344 SQUARE MAIRU	マイ ル	ルマ イ
U+3345 SQUARE MAHHA	マッ ハ	ハマ ッ
U+3346 SQUARE MARUKU	マル ク	クマル

Character	Horizontal	Vertical
U+3347 SQUARE MANSYON	マン ション	シマ ンション
U+3348 SQUARE MIKURON	ミク ロン	ロミ ンク
U+3349 SQUARE MIRI	ミ リ	リ ミ
U+334A SQUARE MIRIBAARU	ミリ パール	バミ ルリ
U+334B SQUARE MEGA	メ ガ	ガ メ
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 ルツ
U+3352 SQUARE RIRA	リ ラ	リ ラ
U+3353 SQUARE RUPII	ルピ ー	ルピ ー
U+3354 SQUARE RUUBURU	ルー ブル	ブル ル
U+3355 SQUARE REMU	レ ム	レ ム
U+3356 SQUARE RENTOGEN	レン トゲン	トレ ン
U+3357 SQUARE WATTO	ワッ ト	トワ ッ

Character	Horizontal	Vertical
U+337B SQUARE ERA NAME HEISEI	平成	平成 𐐇
U+337C SQUARE ERA NAME SYOUWA	昭和	昭和 𐐈
U+337D SQUARE ERA NAME TAISYOU	大正	大正 𐐉
U+337E SQUARE ERA NAME MEIZI	明治	明治 𐐊
U+337F SQUARE CORPORATION	株式会社	株式会社 𐐋
U+FE50 SMALL COMMA	,	’ ,
U+FE51 SMALL IDEOGRAPHIC COMMA	,	、 、
U+FE52 SMALL FULL STOP	.	。 。
U+FE59 SMALL LEFT PARENTHESIS	(	⏟
U+FE5A SMALL RIGHT PARENTHESIS	)	⏞
U+FE5B SMALL LEFT CURLY BRACKET	{	⏟
U+FE5C SMALL RIGHT CURLY BRACKET	}	⏞
U+FE5D SMALL LEFT TORTOISE SHELL BRACKET	[	⏟
U+FE5E SMALL RIGHT TORTOISE SHELL BRACKET	]	⏞
U+FF01 FULLWIDTH EXCLAMATION MARK	!	! !
U+FF08 FULLWIDTH LEFT PARENTHESIS	(	⏟
U+FF09 FULLWIDTH RIGHT PARENTHESIS	)	⏞

Character	Horizontal	Vertical
U+FF0C FULLWIDTH COMMA		
U+FF0E FULLWIDTH FULL STOP		
U+FF1A FULLWIDTH COLON		
U+FF1B FULLWIDTH SEMICOLON		
U+FF1F FULLWIDTH QUESTION MARK		
U+FF3B FULLWIDTH LEFT SQUARE BRACKET		
U+FF3D FULLWIDTH RIGHT SQUARE BRACKET		
U+FF3F FULLWIDTH LOW LINE		
U+FF5B FULLWIDTH LEFT CURLY BRACKET		
U+FF5C FULLWIDTH VERTICAL LINE		
U+FF5D FULLWIDTH RIGHT CURLY BRACKET		
U+FF5E FULLWIDTH TILDE		
U+FF5F FULLWIDTH LEFT WHITE PARENTHESIS		
U+FF60 FULLWIDTH RIGHT WHITE PARENTHESIS		
U+FFE3 FULLWIDTH MACRON		

Character	Horizontal	Vertical
U+1F200 SQUARE HIRAGANA HOKA		
U+1F201 SQUARED KATAKANA KOKO		

6 Data File

The data file is in versioned directories under [\[Data50\]](#). The character repertoire of this revision is the repertoire of Unicode Version 9.0.

Starting with Version 10.0.0 of the Unicode Standard, the data file listing the Vertical_Orientation property value assignments [\[Data50\]](#) is formally included in the Unicode Character Database [\[UCD\]](#). In Revisions 17 and prior of this specification, the data file was provided in versioned directories under the following stable URL:

<http://www.unicode.org/Public/vertical/>

Acknowledgments

Thanks to the original editor Eric Muller, and reviewers: Julie Allen, Van Anderson, John Cowan, John Daggett, Mark Davis, Martin Dürst, Erika J. Etemad, Michael Everson, Asmus Freytag, [Laurentiu Iancu](#), Soji Ikeda, Norikazu Ishizu, Nozomu Katō, Yasuo Kida, Ken Lunde, Nat McCully, Shinyu Murakami, Addison Phillips, Roozbeh Pournader, Dwayne Robinson, Kyoko Sato, Hiroshi Takenaka, Bobby Tung, Philippe Verdy, Ken Whistler, Taro Yamamoto, the W3C CSS Working Group, the W3C I18N Interest Group, and the W3C Internationalization Working Group.

References

For references for this annex, see Unicode Standard Annex #41, “[Common References for Unicode Standard Annexes](#).”

~~[Data50] <http://www.unicode.org/Public/vertical/revision-17/>~~
~~[Errata] Updates and Errata~~
~~<http://www.unicode.org/errata/>~~
~~[Feedback] <http://www.unicode.org/reporting.html>~~
~~*For reporting errors and requesting information online.*~~
~~[LREQ] [Requirements for Japanese Text layout](#), W3C Working Group Note~~
~~[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.*~~
~~[UAX29] UAX #29: *Unicode Text Segmentation*~~
~~<http://www.unicode.org/reports/tr29/>~~

~~[UCD] *Unicode Character Database*
<http://www.unicode.org/ucd/>
For detailed documentation about the Unicode Character Database, see Unicode Standard Annex #44: Unicode Character Database
<http://www.unicode.org/reports/tr44/>~~

~~[Unicode] *The Unicode Standard, Version 9.0.0*, (Mountain View, CA: The Unicode Consortium, 2016. ISBN 978-1-936213-13-9
<http://www.unicode.org/versions/Unicode9.0.0/>~~

~~[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 18

- **Revised** for status change from UTR to Proposed Update UAX.
- Revised text in Sections 2 and 6 documenting the inclusion of the data file into the UCD.
- Updated for the character repertoire of Unicode 10.0.
- Changes made for new assigned characters and allocated blocks in Unicode 10.0:
 - Changes made in the [data file](#):
 - Characters assigned the property value vo=R: U+2BD2.
 - Characters assigned the property value vo=U: U+11A00..U+11AAF, U+1B100..U+1B12F, U+1B170..U+1B2FF.
- Other characters new to Unicode 10.0 have values unchanged from the previous revision of this document.
- Minor edits and cleanup.

Revision 17

- Updated for the character repertoire of Unicode 9.0.
- Changes made for new assigned characters and allocated blocks in Unicode 9.0:
 - Changes made in the [data file](#):
 - Characters assigned the property value vo=U: U+16FE0..U+16FFF, U+17000..U+187FF, U+18800..U+18AFF.
- Other characters new to Unicode 9.0 have values unchanged from the previous revision of this document.
- Updated table styles.

Revision 16 being a proposed update, only changes between revisions 15 and 17 are listed here.

Revision 15

- Updated for the character repertoire of Unicode 8.0.
- Changes made for new assigned characters and allocated blocks in Unicode 8.0:
 - Characters added to the arrows tailoring set: U+2BEC..U+2BEF.
 - Changes made in the [data file](#):
 - Characters assigned the property value vo=R: U+218A..U+218B, U+2BEC..U+2BEF.
 - Characters assigned the property value vo=U: U+14400..U+1467F, U+1D800..U+1DAAF, U+1F900..U+1F9FF.
 - Other characters new to Unicode 8.0 have values unchanged from the previous revision of this document.

Revision 14 being a proposed update, only changes between revisions 13 and 15 are listed here.

Revision 13

- Changed U+3127 from U to Tu, and added its horizontal variant to Table 4.
- Changed the representative vertical glyphs for U+301E and U+FF1A in Table 4.
- Updated for the character repertoire of Unicode 7.0.
- Changes made to designated code points as of Unicode 6.3:
 - Characters moved from U to R: U+2066..U+2069, U+FFFFE..U+FFFF, U+10FFFE..U+10FFFF.
- Changes made for assigned characters and allocated blocks in Unicode 7.0:
 - Characters moved from U to R: U+2B4D..U+2B4F, U+2B5A..U+2BB7.
 - Characters moved from R to U: U+11580..U+115FF, U+1F650..U+1F67F, U+1F780..U+1F7FF.
 - Characters added to the arrows tailoring set: U+2B4D..U+2B4F, U+2B5A..U+2BB7, assigned characters in U+1F800..U+1F8FF (U+1F800..U+1F80B, U+1F810..U+1F847, U+1F850..U+1F859, U+1F860..U+1F887, U+1F890..U+1F8AD).

Revision 12 being a proposed update, only changes between revisions 11 and 13 are listed here.

Revision 11

- First approved published version.
- The property value Tx was removed because it was no longer used.
- Characters moved from U to R: U+1800..U+18AF (Mongolian), U+A840..U+A87F (Phags-pa).
- Characters moved from U to Tu: U+FF1F.
- Characters moved from Tu to U: U+309D..U+309E.
- Changes to Table 4:
 - Added U+FF1F.
 - Removed U+309D..U+309E.
 - Removed erroneously included U+FF61..U+FF70.
 - Added a variant form to U+3001..U+3002, U+337B..U+337E, U+337F, U+FE50..U+FE52, U+FF01, U+FF0C, U+FF0E, U+FF1A, and U+FF1F.
 - Added two variant forms to U+FF1B.

Revision 10

- Table 4 is updated.
- Characters moved from U to R: U+2126, U+212A, U+212B.
- Characters moved from R to Tr: U+2329, U+232A.

Revision 9

- The property renamed back to vo.
- T renamed to Tx.
- Clarified priorities of contexts and environments.
- Overview and Scope edited to clarify what tailoring is better.
- Characters moved from U/T to R: U+00B6, U+02E5..U+02E9, U+2018..U+2019, U+203D, U+2044, U+2102, U+210A..U+210E, U+2110..U+2112, U+2115, U+2119..U+211D, U+2124, U+2128, U+212C..U+212D, U+212F..U+2131, U+2133..U+2134, U+2322..U+2323, U+2329..U+232A, U+23B4..U+23B6, U+23BA..U+23BD, U+23CE, U+2423.
- Characters moved from R to U: U+00B1, U+00D7, U+00F7, U+221E, U+2234..U+2235, U+10980..U+1099F.
- Characters that were undefined and are defined as R: U+E0000..U+E007F.
- Characters that were undefined and are defined as U: U+FFFC, U+F0000..U+10FFFF.

Revision 8

Editorial changes.

Revision 7

- The scope is clarified to serve as a stable default for reliable document interchange.
- The svo property is removed because supporting multiple styles and scripts is out of scope for document interchange.
- The ho property is removed because it is out of scope for document interchange.
- The 'L' value is removed from Table 1.

Revision 6

Data updated. No change to the text.

Revision 5

- TR renamed to include “Horizontal”.
- New property for horizontal text. The current assignment is L for Mongolian and Phags-pa, U for all the other characters.
- Proposal B has been accepted; removed proposal A.
- Characters moved from U or R to T: U+3008..U+3011, U+3014..U+301B, U+301D..U+301F, U+309B..U+309E, U+20A0, U+FF01, U+FF08..U+FF09, U+FF0C..U+FF0E, U+FFAA..U+FF1E, U+FF3B, U+FF3F, U+FF5B..U+FF60, U+FF62..U+FF63, U+FF70, U+FFE3, on the basis of small shift in the box, similar to small kana.

- T moved to Tr or Tu, following MS proposal. The only T characters remaining are U+2018 and U+2019, which are R/R in MS proposal.
- Arrow set introduced, as in MS proposal.
- Yi blocks changed from svo/mvo R to U.
- UCAS changed from mvo R to U, except for U+1400 □ CANADIAN SYLLABICS HYPHEN.

Revision 4

- Properties renamed to Stacked Vertical Orientation (previously Default Vertical Orientation) and Mixed Vertical Orientation (previously East Asian Vertical Orientation).
- Introduced sets of characters for tailoring.
- Property value S renamed to R.
- Property value Sb merged with R; set created for brackets.

Revision 3

- Mongolian and Egyptian Hieroglyphs changed to U.
- Implementation of the UTC decisions made during meeting #130, February 2012.
 - Removal of the East Asian Class property.
 - East Asian Orientation renamed East Asian Vertical Orientation.
 - New property, Default Vertical Orientation. The initial assignment is: T if EAVO=T, SB if EAVO=SB and the bracket is specific to CJK, S to align with CSS Sv value except for vertical presentation forms, Tibetan, Mongolian, sup/sub parens, sup punctuation, U+FD3E, U+FD3F, which remain U.

Revision 2

- Clarification of the status of the properties (end of Section 1).
- Clarification of the handling of grapheme clusters.
- Removed the “comments” column in Table 3.
- Hangul characters: new class cl-19.4, hangul, orientation U.
- Yijing Hexagram symbols are now cl-19-3, symbols, orientation U.
- Small forms variants are treated like their fullwidth counterparts.
- Superscripts and subscript characters are now cl-27, western, orientation S.
- Small kana: orientation U; class split in cl-11.1, smallHiragana and cl-11.2, smallKatakana.
- U+3030 ~ WAVY DASH has orientation T.
- The two alternatives for math etc. are described.

Revision 1

First working draft.

© 2017 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.

