UNICODE PROPERTIES FOR HORIZONTAL AND VERTICAL TEXT LAYOUT

Editor	Koji Ishii 石井宏治 (koji.a.ishii@mail.rakuten.com)
Date	2013-04-25
This Version	http://www.unicode.org/reports/tr50/tr50-9.html
Previous Version	http://www.unicode.org/reports/tr50/tr50-8.html
Latest Version	http://www.unicode.org/reports/tr50/
Latest Proposed Update	http://www.unicode.org/reports/tr50/proposed.html
Revision	9

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*This is a **draft** document which may be updated, replaced, or superseded by other documents at any time.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

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 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ェンだが、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' に出演したワールド・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

However, 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. In many parts of the world, most characters are upright, as can be seen in figure 2.

Figure 2. Western vertical text

M~~However, m~~ost languages and scripts are written horizontally and vertical presentation is a special case, usually used for short runs of text (as in ~~the~~ figure 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. ~~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.~~

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 ~~in the absence of other information,~~ such as ~~information~~ 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 (or even better, of a given character 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.

~~Various factors affect how text is actually oriented in vertical text flows. This report assumes that high-level protocols can override the default orientations that are specified by this 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 ~~to~~ 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 default property values are useful for displaying plain text, but may not be sufficient for some styles or for other environments.~~

~~High-level protocols could implement desired result through the use of tailoring.~~

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 14.8 for a more complete discussion of the scope of Egyptian Hieroglyph characters.

2 Conformance

The property ~~and algorithms~~ defined~~presented~~ in this report ~~is~~are 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~~based on the most common use of a particular character~~, 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) ~~The mvo Property~~

3.1 Property values

The possible property values are given in table 1.

Table 1. Property Values

U	C~~c~~haracters which are displayed upright, with the same orientation as they appear~~s~~ in the code charts.
R	C~~c~~haracters which are displayed sideways, rotated 90 degrees clockwise compared to the code charts.
Tx~~, Tu, Tr~~	C~~c~~haracters which are not just upright or sideways, but require a different glyph than in the code charts when used in vertical texts. ~~In addition, Tu indicates that as a fallback, the character can be displayed with the code chart glyph upright; similarly, Tr indicates a possible fallback using the code chart glyph rotated 90 degrees clockwise.~~

Tu	In addition to Tx, as a fallback, the character can be displayed with the code chart glyph upright.
Tr	In addition to Tx, 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 exceptions:

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

3.3 Vertical Glyphs in the Code Charts

The Unicode code charts generally show characters in the orientation when used in horizontal lines. However, there are a few exceptions, mostly for characters or scripts which are normally written in vertical lines; in those cases, the code charts 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 counter-clockwise with respect to the code charts.

Figure 3. Mongolian text on horizontal lines
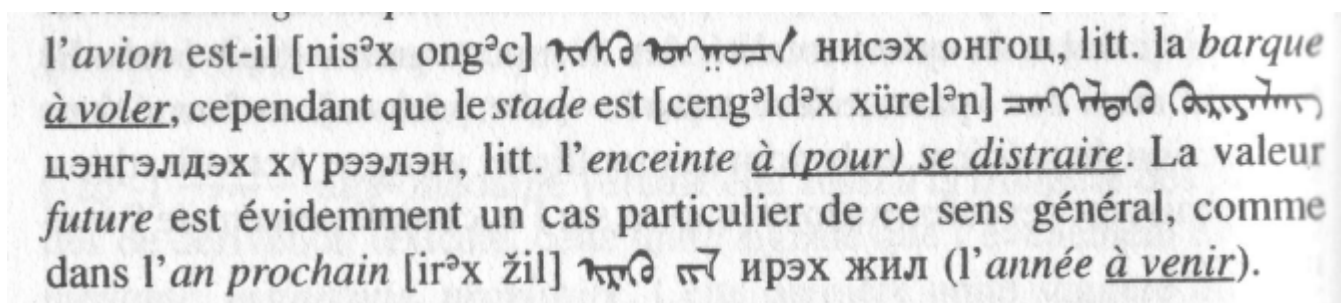
Some rendering~~font~~ systems include glyphs that are rotated counter-clockwise with respect to the characters as shown in the code charts, and expect applications to rotate them clockwise in vertical lines. For example, OpenType fonts that support Mongolian and Phags-pa typically follow this convention.

While this property defines only default orientations compared to the code charts, high-level protocols or applications that make use of such rendering font systems 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 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

00AB

00BB

201C..201F

2039..203A

2045..2046

3008..3011

3014..301B

FE59..FE5E

FF08..FF09

FF3B

FF3D

FF5B

FF5D

FF5F..FF60

FF62..FF63

4.2 The arrows

This set contains arrows.

Table 3. The arrows set

2190..21FF

261A..261F

2794

2798..27AF

27B1..27BE

27F0..27FF

2900..297F

2B00..2B11

2B30..2B4C

FFE9..FFEC

5 Glyphs Changes for Vertical Orientation

This section is not updated for this draft yet.

Table 4 provides representative glyphs for the horizontal and vertical appearance of characters with the property value T.

Need to find glyphs for a few entries.

Recently, the brackets have been made T, because they have a slightly different position in their box between horizontal and vertical. It is arguable whether characters for which the difference is only a slight position adjustment should be included in T.

Table 4. Glyph Changes for Vertical Orientation

character	H	V
U+2018 LEFT SINGLE QUOTATION MARK		
U+2019 RIGHT SINGLE QUOTATION MARK		
U+201A SINGLE LOW-9 QUOTATION MARK		
U+201B SINGLE HIGH-REVERSED-9 QUOTATION MARK		
U+201C LEFT DOUBLE QUOTATION MARK		
U+201D RIGHT DOUBLE QUOTATION MARK		
U+201E DOUBLE LOW-9 QUOTATION MARK		
U+201F DOUBLE HIGH-REVERSED-9 QUOTATION MARK		

character	H	V
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		
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		

character	H	V
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	〟	〟
U+3030 WAVY DASH	〰	〰
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	っ	っ
U+3083 HIRAGANA LETTER SMALL YA	ゃ	ゃ
U+3085 HIRAGANA LETTER SMALL YU	ゅ	ゅ
U+3087 HIRAGANA LETTER SMALL YO	ょ	ょ
U+308E HIRAGANA LETTER SMALL WA	ゎ	ゎ
U+3095 HIRAGANA LETTER SMALL KA	ゕ	ゕ
U+3096 HIRAGANA LETTER SMALL KE	ゖ	ゖ
U+309B KATAKANA-HIRAGANA VOICED SOUND MARK	゛	゛
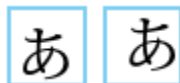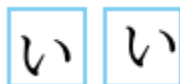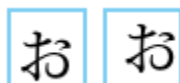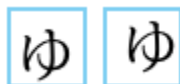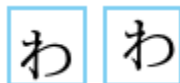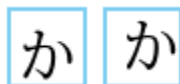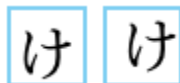
character	H	V
U+309C KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK	°	°
U+309D HIRAGANA ITERATION MARK	ヽ	⊠
U+309E HIRAGANA VOICED ITERATION MARK	ゞ	⊠
U+30A0 KATAKANA-HIRAGANA DOUBLE HYPHEN	＝	‖
U+30A1 KATAKANA LETTER SMALL A	ア	ア
U+30A3 KATAKANA LETTER SMALL I	イ	イ
U+30A5 KATAKANA LETTER SMALL U	ウ	ウ
U+30A7 KATAKANA LETTER SMALL E	エ	エ
U+30A9 KATAKANA LETTER SMALL O	オ	オ
U+30C3 KATAKANA LETTER SMALL TU	ッ	ッ
U+30E3 KATAKANA LETTER SMALL YA	ャ	ャ
U+30E5 KATAKANA LETTER SMALL YU	ュ	ュ
U+30E7 KATAKANA LETTER SMALL YO	ョ	ョ
U+30EE KATAKANA LETTER SMALL WA	ヮ	ヮ
U+30F5 KATAKANA LETTER SMALL KA	ヵ	ヵ
U+30F6 KATAKANA LETTER SMALL KE	ヶ	ヶ
U+30FC KATAKANA-HIRAGANA PROLONGED SOUND MARK	ー	│
U+31F0 KATAKANA LETTER SMALL KU	ク	ク
U+31F1 KATAKANA LETTER SMALL SI	シ	シ
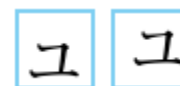
character	H	V
U+31F2 KATAKANA LETTER SMALL SU	ス	ス
U+31F3 KATAKANA LETTER SMALL TO	ト	ト
U+31F4 KATAKANA LETTER SMALL NU	ヌ	ヌ
U+31F5 KATAKANA LETTER SMALL HA	ハ	ハ
U+31F6 KATAKANA LETTER SMALL HI	ヒ	ヒ
U+31F7 KATAKANA LETTER SMALL HU	フ	フ
U+31F8 KATAKANA LETTER SMALL HE	ヘ	ヘ
U+31F9 KATAKANA LETTER SMALL HO	ホ	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	アパート
U+3301 SQUARE ARUHUA	アルファ	アルファ
U+3302 SQUARE ANPEA	アンペア	アンペア
U+3303 SQUARE AARU	アール	アール
U+3304 SQUAREININGU	イニング	イニング

character	H	V
U+3305 SQUARE INTI	インチ	チイン
U+3306 SQUARE UON	ウォン	ンウォ
U+3307 SQUARE ESUKUUDO	エスクード	クエードス
U+3308 SQUARE EEKAA	エーカー	カエーー
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	ッカトラ
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	リキュー
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	グキラロム
U+3316 SQUARE KIROMEETORU	キロメートル	ーキロトヌル
U+3317 SQUARE KIROWATTO	キロワット	ワキットロ

character	H	V
U+3318 SQUARE GURAMU	グラム	ムグラ
U+3319 SQUARE GURAMUTON	グラムトン	トングラム
U+331A SQUARE KURUZEIRO	クルゼイロ	ゼクロイル
U+331B SQUARE KUROONE	クローネ	ークロネ
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	チサンーム
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
U+3325 SQUARE DESI	デシ	デシ
U+3326 SQUARE DORU	ドル	ドル
U+3327 SQUARE TON	トン	トン
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

character	H	V
U+332B SQUARE PAASENTO	パーセント	パーセント
U+332C SQUARE PAATU	⊠	⊠
U+332D SQUARE BAARERU	バレル	バレル
U+332E SQUARE PIASUTORU	ピアストル	ピアストル
U+332F SQUARE PIKURU	ピクル	ピクル
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	ファラッド
U+3333 SQUARE HUIITO	フィート	フィート
U+3334 SQUARE BUSSYERU	ブッシェル	ブッシェル
U+3335 SQUARE HURAN	フラン	フラン
U+3336 SQUARE HEKUTAARU	ヘクタール	ヘクタール
U+3337 SQUARE PESO	ペソ	ペソ
U+3338 SQUARE PENIHI	ペニヒ	ペニヒ
U+3339 SQUARE HERUTU	ヘルツ	ヘルツ
U+333A SQUARE PENSU	ペンス	ペンス
U+333B SQUARE PEEZI	ページ	ページ
U+333C SQUARE BEETA	ベータ	ベータ
U+333D SQUARE POINTO	ポイント	ポイント

character	H	V
U+333E SQUARE BORUTO	ボルト	トボル
U+333F SQUARE HON	ホン	ホン
U+3340 SQUARE PONDO	ポンド	ドポン
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
U+3347 SQUARE MANSYON	マンション	シマョン
U+3348 SQUARE MIKURON	ミクロン	ロミンク
U+3349 SQUARE MIRI	ミリ	ミリ
U+334A SQUARE MIRIBAARU	ミリバール	バミリル
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

character	H	V
U+3351 SQUARE RITTORU	リットル	トリルッ
U+3352 SQUARE RIRA	リラ	リラ
U+3353 SQUARE RUPII	ルピー	ールピ
U+3354 SQUARE RUUBURU	ルーブル	ブルルー
U+3355 SQUARE REMU	レム	レム
U+3356 SQUARE RENTOGEN	レントゲン	トゲレンン
U+3357 SQUARE WATTO	ワット	トワッ
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	会株社式
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	⊠	⊠

character		H	V
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		）	⌣
U+FF0C FULLWIDTH COMMA		，	，
U+FF0E FULLWIDTH FULL STOP		．	˙
U+FF1A FULLWIDTH COLON		：	‥
U+FF1B FULLWIDTH SEMICOLON		；	⊠
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+FF61 HALFWIDTH IDEOGRAPHIC FULL STOP		｡	⊠
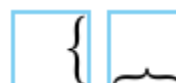
character	H	V
U+FF62 HALFWIDTH LEFT CORNER BRACKET	｢	⊠
U+FF63 HALFWIDTH RIGHT CORNER BRACKET	｣	⊠
U+FF64 HALFWIDTH IDEOGRAPHIC COMMA	､	⊠
U+FF67 HALFWIDTH KATAKANA LETTER SMALL A	ｱ	⊠
U+FF68 HALFWIDTH KATAKANA LETTER SMALL I	ｨ	⊠
U+FF69 HALFWIDTH KATAKANA LETTER SMALL U	ｩ	⊠
U+FF6A HALFWIDTH KATAKANA LETTER SMALL E	ｪ	⊠
U+FF6B HALFWIDTH KATAKANA LETTER SMALL O	ｫ	⊠
U+FF6C HALFWIDTH KATAKANA LETTER SMALL YA	ｬ	⊠
U+FF6D HALFWIDTH KATAKANA LETTER SMALL YU	ｭ	⊠
U+FF6E HALFWIDTH KATAKANA LETTER SMALL YO	ｮ	⊠
U+FF6F HALFWIDTH KATAKANA LETTER SMALL TU	ｯ	⊠
U+FF70 HALFWIDTH KATAKANA-HIRAGANA PROLONGED SOUND MARK	ｰ	⊠
U+FFE3 FULLWIDTH MACRON		
U+1F200 SQUARE HIRAGANA HOKA	⊠	⊠
U+1F201 SQUARED KATAKANA KOKO	⊠	⊠

6 Data File

This section is not updated for this draft yet.

To help during the review, a slightly more readable version is available.

Acknowledgments

References

[JLREQ]   Requirements for Japanese Text layout, W3C Working
            Group Note

[Errata]   Updates and Errata
            http://www.unicode.org/errata

[Feedback] http://www.unicode.org/reporting.html
            For reporting errors and requesting information online.

[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 6.2.0, defined by: The
Unicode Standard, Version 6.2.0 (Mountain View, CA: The
Unicode Consortium, 2012. ISBN 978-1-936213-07-8)
http://www.unicode.org/versions/Unicode6.2.0

[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 9

- The property renamed back to `vo`.

- T renamed to Tx.

- Clarified priorities of contexts and environments.

- Characters moved from U/T to R: 00B6, 02E5..02E9, 2018..2019, 203D, 2044, 2102, 210A..210E, 2110..2112, 2115, 2119..211D, 2124, 2128, 212C..212D, 212F..2131, 2133..2134, 2322..2323, 23B4..23B6, 23BA..23BD, 23CE, 2423.

- Characters moved from R to U: 00B1, 00D7, 00F7, 221E, 2234..2235, 10980..1099F.

- Characters that were undefined and are defined as R: E0000..E007F

- Characters that were undefined and are defined as U: FFFC, F0000..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. Refer to 3.3, Vertical Glyphs in the Code Charts for more details.

- 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: 3008..3011 3014..301B 301D..301F 309B..309E 20A0 FF01 FF08..FF09 FF0C..FF0E FFaA..FF1E FF3B FF3F FF5b..FF60 FF62..FF63 FF70 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 2018 and 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, FD3E, 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 r-cllkjljrparts.

- 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.