


A photograph of three computer keyboard keys: 'w', '3', and 'c'. The keys are white with black characters and are set against a blue background. The 'w' key is on the left, the '3' key is in the middle, and the 'c' key is on the right. They are arranged in a slightly descending line from left to right.

New Work on Japanese Layout Requirements

Richard Ishida, W3C
Tatsuo Kobayashi, JustSystem
Steve Zilles, Adobe Systems



Outline


- Kihon hanmen: laying out the page body
- Converting between horizontally and vertically set text
- Line layout
- Jukugo ruby

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 2

Introduction
W3C

The W3C's Japanese Layout Task Force



The screenshot shows the title page of a W3C Working Draft titled "Requirements of Japanese Text Layout". It includes the W3C logo, the document title, the date "11 April 2008", and the authors "Toshiaki Matsuda, Takanori Nishida". The page also contains an abstract, a status section, and a footer with copyright information for 2005 W3C (MIT, ERCIM, Keio).

- Requirements for general Japanese layout
- Latest version: <http://www.w3.org/TR/jlreq/>
- JIS X 4051
- Print and on-screen layout

slide 3


The goal of this task force is the creation of a document about requirements for general Japanese layout realized with technologies like CSS, SVG and XSL-FO. The document will be mainly based on a standard for Japanese layout, JIS X 4051. However, it will address also areas which are not covered by JIS X 4051. The document is currently in draft stage and is being developed further by the Japanese participants in the task force.

The task force will gather and integrate feedback from the participating working groups about the need and technical feasibility of various requirements.


The output of the task force will be an English document (a Working Group Note) describing requirements for Japanese Layout. The description will not be on a technology-specific level. It is assumed that the participating working groups (see below) will decide independently how (and if) to implement the requirements in their specifications.

Introduction

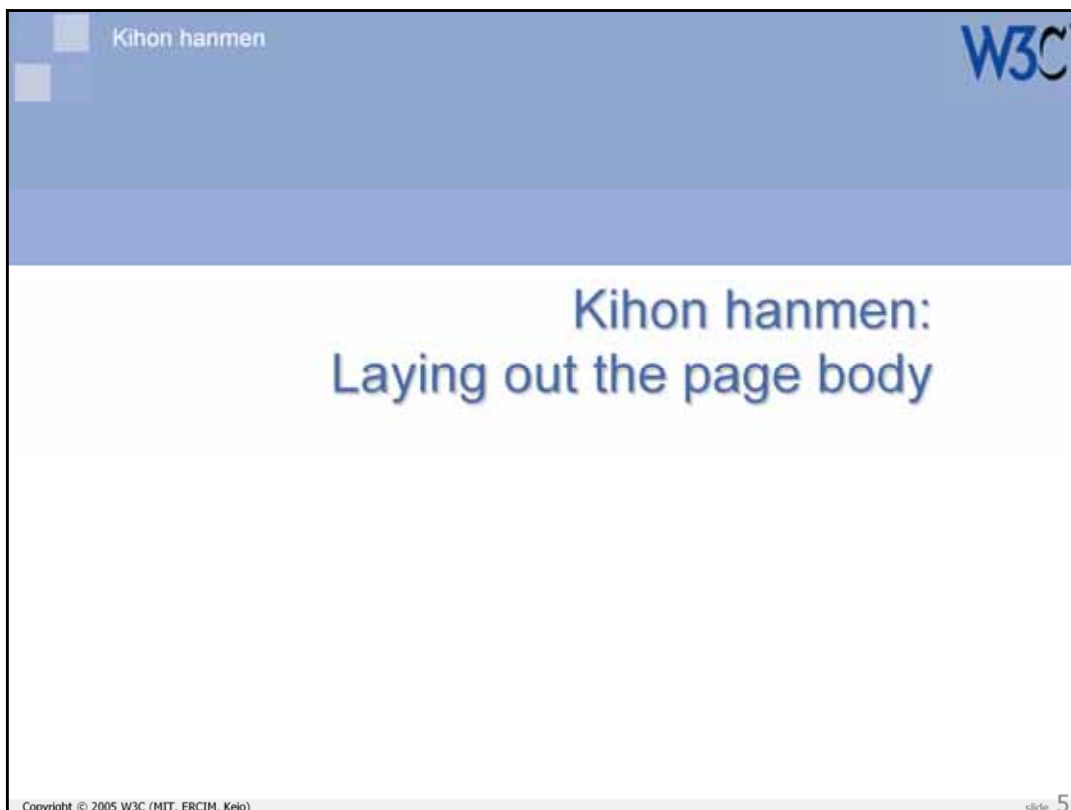
The W3C's Japanese Layout Task Force



- Editors: Toshi Kobayashi, JustSystems, Yasuhiro Anan, Microsoft
- Participants from I18n, CSS, SVG, XSL Working Groups
- Working in Japanese, with periodic meetings in English
- Output: W3C Note (in English and Japanese)



Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 4

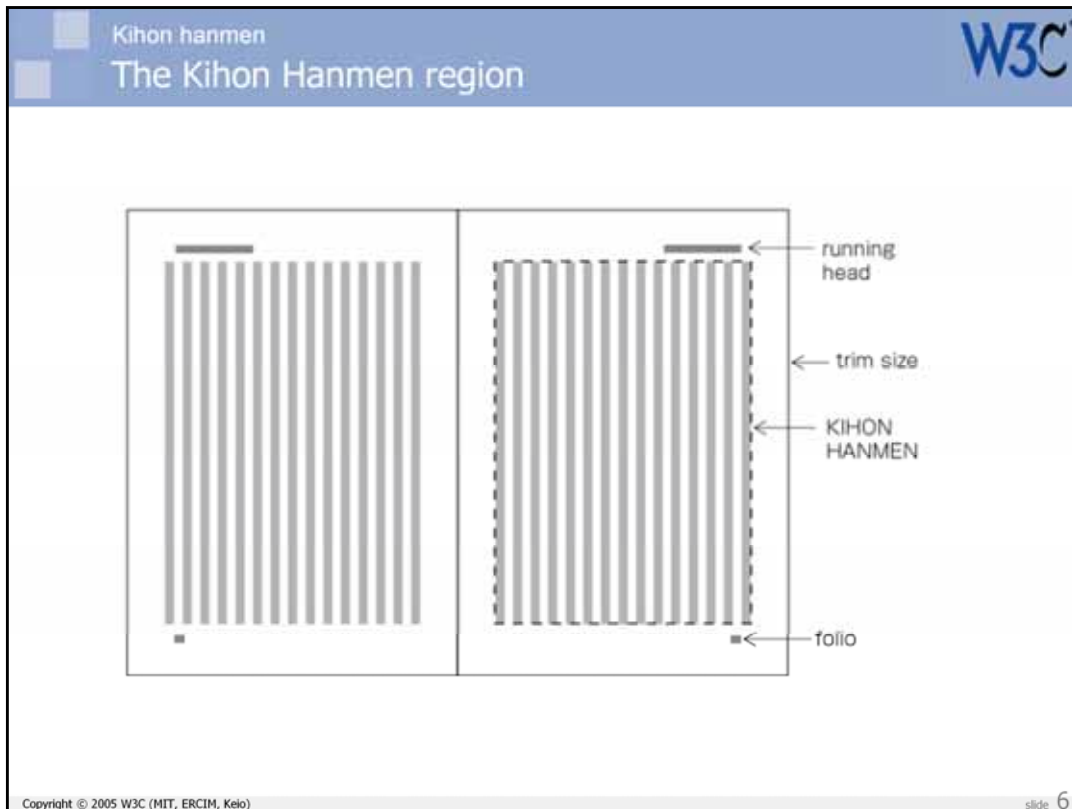


Kihon hanmen

W3C

Kihon hanmen: Laying out the page body

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 5



The Kihon hanmen is the area of the page that contains the main body of the Japanese text. It doesn't include page headers and footers.

The document in the slide is vertically set and typical of the layout of a Japanese novel.

In order to understand why the kihon hanmen is of interest, we must first understand some of the characteristics of Japanese characters and their use.

Kihon hanmen
Japanese script characteristics

W3C

コンピュータは、本
質的には数字しか扱
ることができません。
コンピュータは、文字
や記号などのそれぞれ
に番号を割り振ること
によって扱えるように
します。

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 7

Japanese kanji, hiragana and katakana characters normally sit in small square boxes that are all the same size. This leads to a very regular effect.

Even characters such as the Japanese period and comma shown on the slide sit within a square box of the same size.

There are no spaces between characters in the slide above.

Kihon hanmen
Japanese script characteristics

W3C

し	に	に	や	ン	こ	質	コ
ま	よ	番	記	ピ	と	的	ン
す	っ	号	号	ユ	が	に	ピ
°	て	を	な	ー	で	は	ユ
	扱	割	ど	タ	き	数	ー
	え	り	の	ー	ま	字	タ
	る	振	そ	は	せ	し	ー
	よ	る	れ	、	ん	か	は
	う	こ	ぞ	文	°	扱	、
	に	と	れ	字	コ	う	本

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 8

The regularity of the character sizes produces a grid-like effect. Lines could be drawn through the text both vertically and horizontally, as shown.

Kihon hanmen
Japanese script characteristics

W3C

ま	よ	番	記	co	こ	質	コ
す	っ	号	号	m	と	的	ン
。	て	を	な	p	が	に	ピ
	扱	割	ど	u	で	は	ユ
	え	り	の	t	き	数	ー
	る	振	そ	e	ま	字	タ
	よ	る	れ	は	ま	し	ー
	う	こ	ぞ	、	せ	か	は
	に	と	れ	文	ん	扱	本
	し	に	に	字	。	う	
				や			

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 9

In reality, life is not quite so simple. Embedded Latin characters, rules about what characters can begin or end a line, etc, tend to produce small variations in the placement of characters relative to the grid. Nevertheless, where characters come adrift from the grid cells, there is a strong tendency to return to the grid pattern as soon as possible. This is achieved using justification rules. If the grid layout is not recovered before, the last character box on a line will end flush with the bottom of the grid.

Kihon hanmen
Japanese script characteristics

W3C

コンピューターは、本質的には数字しか扱わることができません。記号などのそれぞれに番号を割り振ることによって扱えるようになります。

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 10

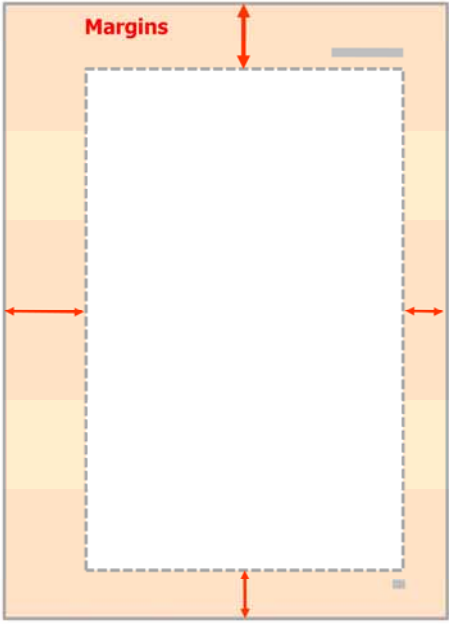
Actually, the grid cells are not usually square. Japanese text usually has some interline spacing, typically ranging from half a character box in width, for short lines, to a full character box for longer lines.

This tendency for characters on a page to fit to a grid makes the kihon hanmen a little special in a couple of ways, as we will now see.

Kihon hanmen

Specifying page layout

W3C



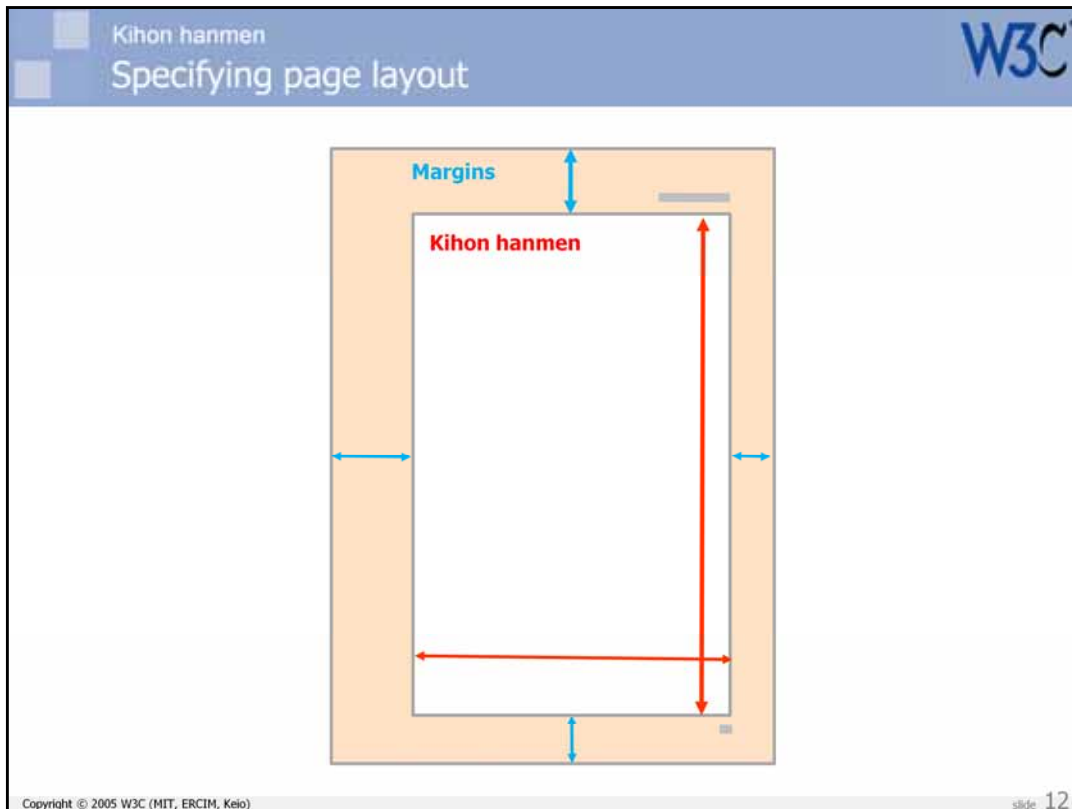
Margins

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 11

A typically Western way of creating the main text region on a page involves just specification of margins. For example, to create a simple document in XSL-FO, after specifying the printable area of the page using page margins, you specify the size, within the printable area, of the main body region by specifying its margins. (Those margins can then be used for page headers and footers.)

The key point is that you don't usually specify the size of the body region itself. The size of the body region is whatever is left over after you have specified its margins. In other words, you specify the size of the blank space, then just flow the text into the remaining space.

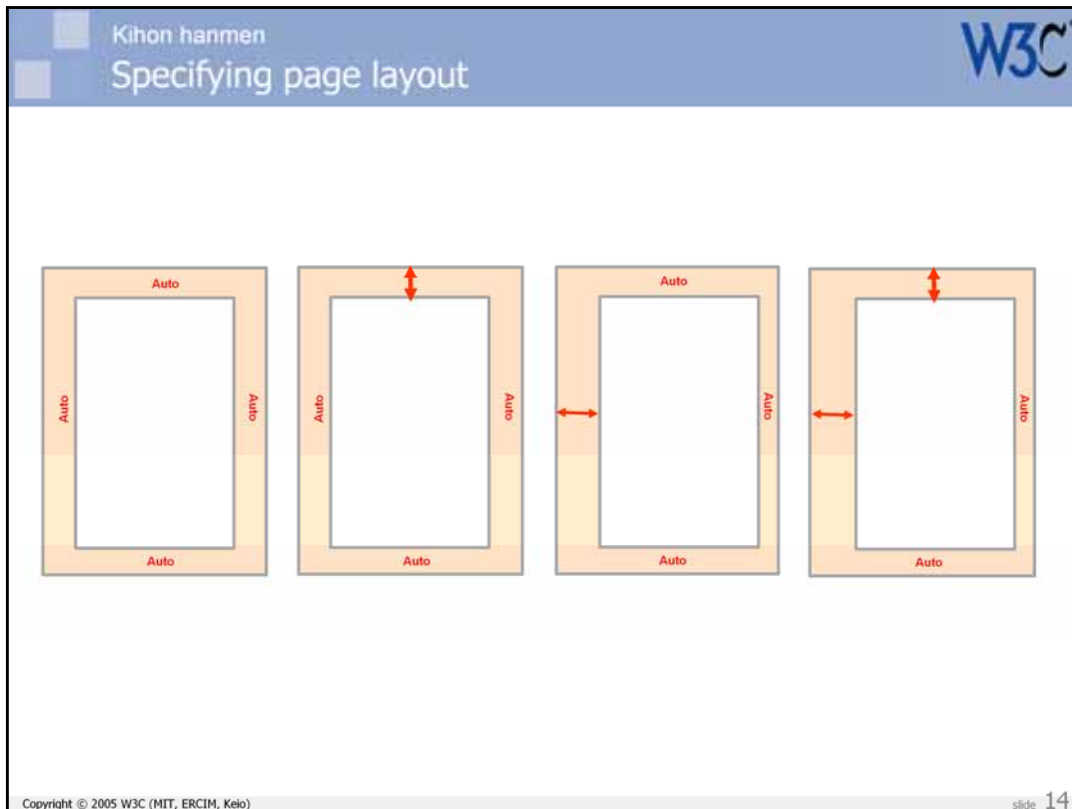


Because the kimon hanmen is a grid, Japanese people think of page layout slightly differently.

Their natural inclination is to first define the size of the main text area (ie. the kimon hanmen). Only after that are the margins considered.

For a page with a single column, defining the size of the kihon hanmen involves choosing a character size, specifying the length of the lines as a number of characters, then specifying the other dimension of the text area as the number of lines per page and the size of the interline gap. (There are preferred sizes and ratios for characters, line lengths and lines per page.)

For documents with multiple columns, the size of the gap between the columns is also taken into account.



Strategies for determining margins.

The common default is to center the kihon hanmen in the body region.

Other alternatives include:

1. Specifying the top space (for horizontally set pages) or the bottom (for vertically set pages), and centering the sides.
2. Specifying the size of the gutter to the side, and centering the top and bottom.
3. A mixture of 1 and 2.

In summary, to specify the position of the main text flow in a simple document using kihon hanmen you would specify the dimensions of the main text region itself, and zero to two margins.

There has been discussion at the W3C about making it easier to specify the body region in this way in XSL-FO 2.0.

Kihon hanmen
The grid

W3C

1 em line head indent at the beginning of paragraphs

← 原稿内容の吟味では、入手した原稿が出版に適しているかどうかを検討する。テーマや範囲、表現のスタイル、内容のレベル、構成、原稿の分量は問題ないだろうか、そして名誉毀損や著作権といった法規上の問題が含まれていないかを吟味し、刊行するかどうかを検討・決定する。

←この後、「原稿整理」または「原稿編集」とよばれる編集作業が行われる。内容や形式について、細かく点検し整理していく仕

見出しの処理

line head
line head indent is 4 ji
heading
The grid of the dotted line indicates the assigned positions of characters and lines used to determine KIHON-HANMEN
arranged in the center of three lines

見出しは標題ともいい、内容が一目でわかるようにしたものである。編・章・節のように見出しが三段階のレ

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 15

Another significant characteristic of the kihon hanmen as a type of body region goes back to its grid-like nature.

Elements positioned within the kihon hanmen tend to snap to the logical grid that underlies this region.

For example, the first line of a paragraphs tends to be indented, and the usual indentation is one character box.

Headings in the document also tend to be positioned relative to the grid, even though the size of the characters in the title is larger than that used for the logical grid of the kihon hanmen.

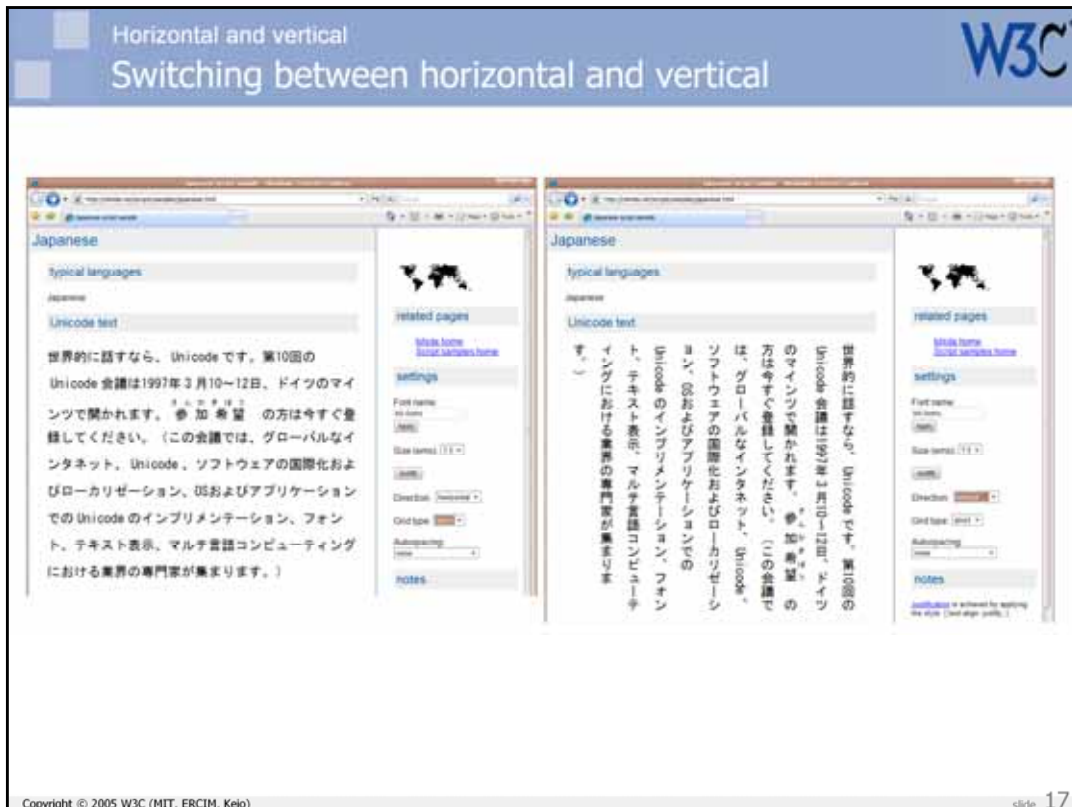
Figures and surrounding text tend to be positioned relative to the kihon hanmen grid.

Horizontal and vertical

W3C

Converting between horizontally
and vertically set text

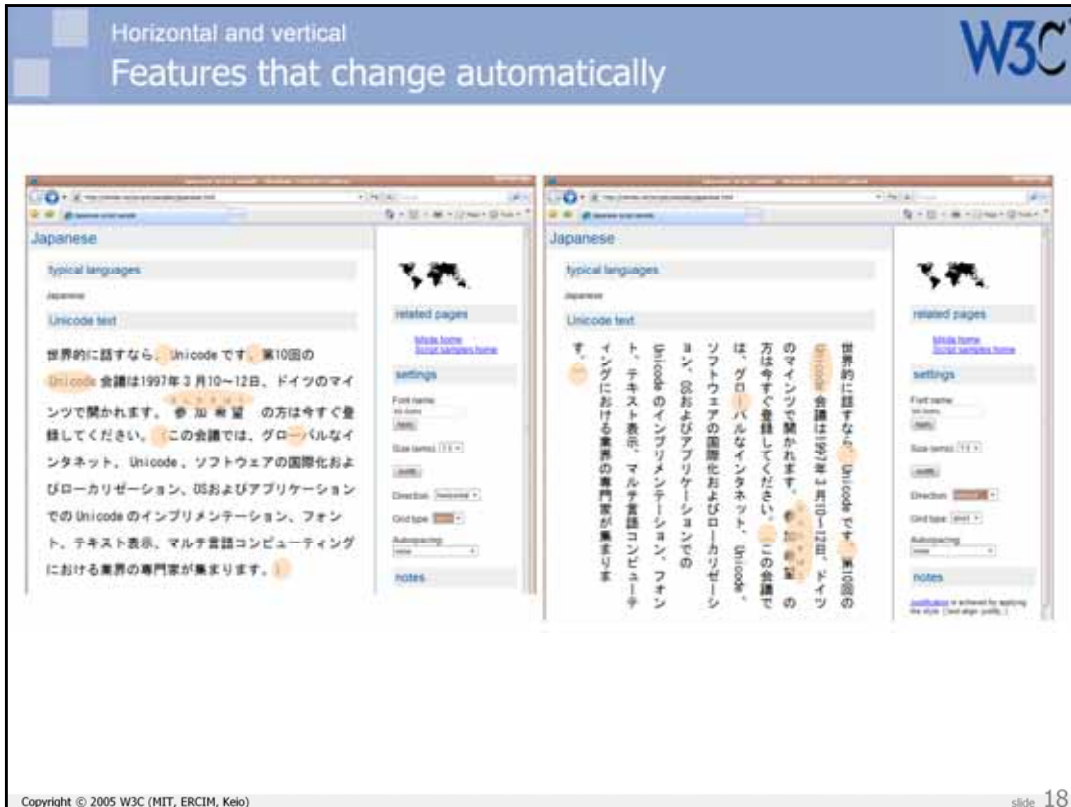
Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 16



You may think that you would be able to switch between horizontal and vertical text at the flick of a CSS switch. I even have a page that allows you to do that, using some JavaScript that changes the writing-mode property on a block of Japanese text (see <http://rishida.net/scripts/samples/japanese.html> - requires Internet Explorer 5.5+).

There are some differences between text in vertical and horizontal layouts that can be dealt with automatically, but there are others that are not so easy to handle.

The change shown on the slide is brought about by simply changing the value of the writing-mode property on the enclosing block from lr-tb (left-to-right block flow, top-to-bottom line progression) to tb-rl (top-to-bottom block flow, right-to-left line progression).



This slide shows some things that already work in Internet Explorer, using this simple switch.

The Japanese comma and full-stop move from the bottom left of the character square to the top right. This is achieved by choosing a vertical-text-specific glyph from the font. The character codes remain the same.

Sound elongation marks and punctuation such as parentheses are rotated.

Runs of Latin text flow down the page in vertical lines.

Ruby text moves from above the horizontal line to the right of the vertical line.

Horizontal and vertical
Other features that need to change

W3C

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 19

Unfortunately, there are other aspects of the text that should have changed, but didn't. For example, some numbers would have looked better as kanji characters, rather than Western digits, in vertical text; other numbers should run horizontally within the vertical line progression; and some Latin text that is used as an acronym would look better if the characters were not rotated.

But there are also other differences. Some can be worked around, others are less tractable. Let's take a look at some of the common differences between horizontally and vertically set Japanese text.

Horizontal and vertical
Other features that need to change

W3C

平成18年12月25日

の GNP は

日 平成
18
年
12
月
25

の
G
N
P
は

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 20

A Latin-script acronym or single letter is typically rendered without rotation in vertical text. It also uses a full-width shape, whereas horizontal text will typically use proportionally-spaced characters.

Short two-digit numbers are often horizontally set within the vertical line (this is called *tate-chuu-yoko*).

In both of these cases, you would need to have markup in the horizontal text to automatically achieve the desired result when the switch to vertical is flipped.

Horizontal and vertical
Other features that need to change

W3C

世界的に話すなら、Unicodeです。

世界的に話すなら、Unicodeです。

世界的に話すなら、Unicodeです。

世界的に話すなら、Unicodeです。

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 21

Punctuation characters tend to be used differently in horizontal and vertical text.

Due to Western influence, Japanese authors sometimes use Western commas and punctuation in text. Vertical text, on the other hand, should only use the ideographic comma and full stop.

Horizontal and vertical
W3C

Other features that need to change

<p>日本語の表記では“漢字” や“仮名”だけでなく、 “ローマ字”や“アラビア 数字”、さらに“句読点” や“括弧類”などの記述記 号を用いる。</p>	<p>日本語の表記では「漢字」 や「仮名」だけでなく、「 ローマ字」や「アラビア数 字」、さらに「句読点」や 「括弧類」などの記述記号 を用いる。</p>
---	---

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 22

Similarly, there are some differences in the usage of quotation marks in horizontal vs. vertical text. For example, horizontal text often uses double or single quotation marks as an alternative to left and right corner brackets. Vertical text, on the other hand, doesn't use the quotation marks (except sometimes for embedded Latin runs); it uses corner brackets or double prime marks. Double prime marks are not used for horizontal text.

Similar rules apply to parentheses. Left and right tortoise shell brackets (⌈
⌋) are a vertical equivalent of left and right square bracket ([]). Square brackets should be used in horizontal text except for special cases.

Horizontal and vertical
Other features that need to change

W3C

456cm

四
五
六
セ
ン
チ

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 23

The slide features a blue header with the text 'Horizontal and vertical' and 'Other features that need to change'. The W3C logo is in the top right. The main content area shows '456cm' in a large black font on the left and the Japanese ideographic numerals '四五六センチ' (456 centimeters) written vertically on the right. The footer contains copyright information and the slide number '23'.

Ideographic numerals were traditionally used in vertically set text, instead of Western-Arabic numerals. (Road numbers and car numbers are examples of exceptions). However, newspapers and other publications have been adopting Western-Arabic numerals more in vertical writing mode.

In vertically set text, symbols for units are usually described with katakana characters, eg. センチメートル (centimeter) or センチ (abbreviated form for centimeter). In horizontally set text, the International System of Units (SI) is usually used, eg. "cm".

Horizontal and vertical
Other features that need to change

W3C

The diagram illustrates four scenarios of text column balancing. The top row shows two vertical text examples: the left one is labeled 'example arrange without completing it with the upper column' and shows a full column of text with a shorter column below it; the right one is labeled 'example that it arrange only in the upper column' and shows a shorter column of text with a longer column below it. The bottom row shows two horizontal text examples: the left one shows two columns of text of different lengths; the right one is labeled 'same number of lines on average' and shows two columns of text where the total number of lines is balanced.

example arrange without completing it with the upper column

example that it arrange only in the upper column

same number of lines on average

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 24

Where a new page is started and the current page isn't full of text, there may be different strategies for balancing columns on the current page. Typically, text columns in vertically set text end without any balancing. Columns in horizontally set text, however, are typically balanced so that the two columns are the same length.

Horizontal and vertical
Other features that need to change

W3C

The image shows a screenshot of a Japanese document with vertical text. Annotations indicate a '9 pt' dimension for the 'KIHON HANMEN' (baseline) and another '9 pt' dimension for the running head. The document text includes a title '第3章 印刷製本...' and several paragraphs of text. The page number '123' is visible at the bottom of the document content.

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 25

There are also differences in page layout details between horizontally and vertically set pages that are not automatically achieved.

For example, when positioning running heads and page numbers horizontally with reference to kihon hanmen in vertical writing mode, the amount of vertical space between the edge of kihon hanmen and the running head is one full width character size based on the kihon hanmen character sizes. If kihon hanmen is horizontally set, you need to take more vertical space than the character size in kihon hanmen.

Horizontal and vertical
Other features that need to change

W3C

漢字廃止論には、前島密(一八三五―一九一九年。明治時代の政治家・実業家、郵便制度の創始者)などによるものがある。しかし、漢字の廃止は、困難であり、そこで主張されたのが漢字節減論である。その代表的なものが福沢諭吉(一八三四―一九〇一年。明治時代の啓蒙思想家、慶応義塾大学を創設)の「文字之教」などがある。

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 26

Typographic conventions such as warichu are much more common in vertically set text than horizontal.

Line layout

W3C

Line Layout

Copyright © 2005 W3C (MIT, ERCIM, Keio)

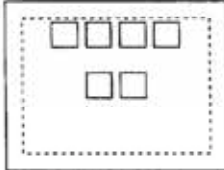
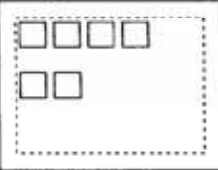
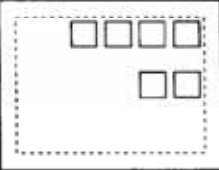
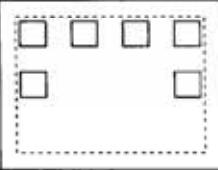
slide 27

Line layout

Rules for line layout in Japanese

W3C

- Three basic sets of rules
 - for spaces between characters
 - for places at which line breaks are allowed
 - for justifying a line by adding adjustable space

Center Alignment	Left/Top Alignment	Right/Bottom Alignment	Japanese Full Justification
			

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 28

Line layout

Rules are based on character class combinations

W3C

- 26 Character classes - examples

Character Class	Horizontal Characters
(1) Opening parentheses	‘ ’ “ ” () [] { } < > < > < >
(5) Bullets, colons, semi-colons	• : ;
(7) Non-separable characters	— … …
(9) Post fix unit symbols	° ’ ” % ‰ ø
(10) Japanese space character	
(11) Hiragana	あ い う …
(12) Other Japanese characters	
(15) Number run characters	
(18) Latin characters	a b c d e …

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 29

Line layout

Rules are expressed in tables


W3C

- Three tables
 - one for character width and spacing
 - one for allowed line breaks
 - one for adjustable space for justification
- Each row and column labeled with a character class
- Each cell describes the action for the combination of the row and column label classes for that cell
- Example: between two kanji characters
 - no normal space is inserted
 - line breaking can occur
 - adjustable space can be inserted

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 30

Line layout

Three kinds of inter-character space



- Fixed – always added independent of context (e.g. Japanese character space)
- Conditional – added depending on context (e.g. space around punctuation characters)
- Adjustable – extra space added to justify a line
Similar to extra word or inter-character space in Latin text

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 31

Line layout

Conditional Space Examples

W3C

② Closing brackets and parenthesis, etc. following comma / period

solid setting

1/2 em space

③ Comma / period following closing brackets and parenthesis, etc.

solid setting

1/2 em space

④ Opening brackets and parenthesis, etc. following comma / period

1/2 em space

⑤ Opening brackets and parenthesis following closing brackets and parenthesis

1/2 em space

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 32

Line layout

Line Breaking Rule Examples

W3C

- Not within number run or before post-fix unit symbol

break prohibited

は 70% で

break prohibited break prohibited

は 75.123 4 kg の

- Not within Em Dash or Ellipsis sequence

break prohibited

という……そういう

- Certain characters cannot end or start a line
(e.g. opening bracket at line end is not allowed)

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 33

Line layout

Adjustable Spacing to Justify a line

W3C

- The opening bracket cannot occur at the end of a line
- Therefore, by line breaking rules it must be moved to beginning of next line
- Then extra (adjustable) space is needed to justify the line
- Such space is added to existing (conditional and fixed) spaces and in between characters
- Amount of adjustment to existing spaces is proportional to amount of space already present

line end

© Example of line length adaptation via leaving space between characters

© Corner bracket at the end of the line makes adjustment necessary

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 34

The slide features a blue header bar at the top. On the left side of the header, there are two small light blue squares followed by the text "Ruby Annotations". On the right side of the header, the W3C logo is displayed. The main body of the slide is white and contains the text "Ruby Annotations" centered in a blue font. At the bottom left, there is a small copyright notice: "Copyright © 2005 W3C (MIT, ERCIM, Keio)". At the bottom right, the text "slide 35" is visible.

- Written in four scripts
 - Kanji (漢字)
based on (Han) Chinese ideographs
 - Hiragana (ひらがな)
phonetic for Japanese
 - Katakana (カタカナ)
phonetic for foreign words
 - Romanji (romanji) alphabetic

Japanese is a multisyllabic language that is very different than monosyllabic, tonal Chinese, but lacking a script of their own, the Japanese adopted the Chinese ideographic script to write Japanese. When this proved not adequate to the task they added two phonetic scripts (first hiragana and then katakana). Today, hiragana is used to express things like verb and adjective inflections (endings) (lacking in Chinese) and to write words that lacked Chinese character(s). And, katakana is used to express words imported in the 19th century or later and for proper names and/or emphasis. Finally, the Latin alphabet is used to write Japanese in a form that foreigners might understand.

- Native Japanese words (often multisyllabic)
 - e.g. “higashi” (東) meaning “east”
 - or “yama (山) meaning “mountain”
- Imported Chinese words (one syllable per ideograph)
 - e.g. “tō” (東) meaning “east”
 - or “san” (山) meaning “mountain”
 - or “futon”(布団) meaning “mattress”
- The latter is a compound word or “jukugo” (熟語)
- Note that many ideographs have two “readings”
 - one in the native language (Kun)
 - “higashi” and “yama”
 - one in the Chinese language (On)
 - “tō” and “san”
- So how do you “read” (東 山)? Use an annotation to say!

English is composed of words derived from Anglo-Saxon (pig, sheep) and Norman French (pork, mutton). Similarly, Japanese has words from the native spoken language, words imported from China, and words imported from Western (or other non-chinese) countries.

The native words are typically multi-syllabic and without tones. Typically a single kanji is assigned to each Japanese word, independently of how many syllables it has. So we have “higashi” (3 syllables) for “east” and “yama” (2 syllables) for “mountain”. Since at the time the Chinese script (Kanji) was adopted for writing Japanese, it typically had words, in Chinese, for all the things Japanese had words for, there are also Sino-Japanese pronunciations for the same kanji that represent Japanese words. Here there is “Tō” for “east” and “san” for “mountain”. The Chinese “readings” of the kanji characters are normally mono-syllabic (but they have lost the associated Chinese tone).

Since Chinese was more sophisticated than Japanese at the time of borrowing of Kanji, there are words that have multiple kanji in their “spelling”. These compound words are called “jukugo” in Japanese. The individual characters in “futon”: fu (character 737) means “spread” or “cloth” and “ton” (character 243) means “group or troupe”. So the literal combination would be “group of cloth” which is an OK description of a “mattress” which is what “futon” means as a compound word.

Another example in Japanese is the pair **Shinto** (神道 *Shintō*) meaning the “**path/way of the gods**” and

The word *Shinto*, from the original Chinese *Shendao* (神道), [1] combines two *kanji*: “*shen*” (神 “*shen*”) (Chinese *loanwords* usually retain their Chinese pronunciation, hence “*shin*” not “*kami*” the Japanese for “gods”), meaning gods or spirits; and “*dao*” (道 “*dao*”), meaning a philosophical way or path.

Ruby annotations

Ruby Annotations

W3C

- Small (1/2 size) characters that specify a pronunciation for or a meaning of the characters to which they are attached

Ruby → くんし わ どう
Base chars → 君子は和して同ぜず

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 38

The ruby says, "ku" "n" "shi" (ha) "wa" (shi) (te) "do" "u" (ze) (zu) where the ruby is quoted and the base kana is in parentheses

'kun' (character 479) is an honorific

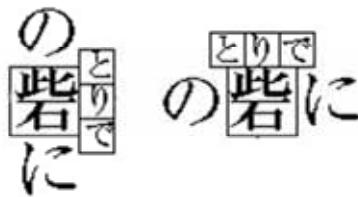
'shi' (character 61) means child or person

so together they mean "wise man", "man of virtue or high rank"

"wa" (character 277) means 'peace', "harmony", or 'Japan'

"do u" (character 185) seems to mean "same" (but normally without the "u" part)

- Annotations are attached to single characters
- Typically 1 to 5 (*3 below*) ruby characters per base character
- In vertical text, ruby characters are on the right
- In horizontal text, ruby characters are on top



The imaginary box (in which the character is centered) is drawn to more clearly show the positioning of the ruby characters relative to the base character.

The hiragana seems to say, "to n de"

- Ruby characters can be centered on the base character. This is called “nakatsuki”.
- Ruby characters can be aligned with start edge of of the base characters. This is called “katatsuki”



There are two approaches to positioning the ruby characters with respect to the base character. If there is an odd number of characters, as is the case here, then the center character can be aligned with the center of the base characters. This is called, “nakatsuki” alignment.

Alternatively, the initial character of the sequence of ruby characters can be aligned with the start edge of the base characters. This is called, “katatsuki” alignment. This approach was more often used with (hot) metal type setting and with digital text layout, the nakatsuki approach is more popular.

Note that if there are two ruby characters attached to the base character, then both approaches give the same results as long as the size of the ruby characters are $\frac{1}{2}$ the size of the base characters.

- Characters adjacent to base character affect the alignment of the ruby characters
- Ruby characters can overlap “kana” characters
- Ruby characters should not overlap “kanji” characters
- Extra space may be needed to prevent overlaps



The ruby annotation is: “shi” “yu” “n” (?)

Going from left to right:

The first sequence has kana before and after the annotated kanji so overlap is allowed on either side, but the in katatsuki the first ruby character is normally aligned with the start edge of the base characters

The second sequence as a kanji character (hand) following the annotated kanji so the only allowable overlap is with the hiragana character that is before the annotated base character

The third sequence has a kanji before the annotated base character and kana afterwards, so the normal katatsuki rule works (actually is forced) in this case.

In the fourth and final sequence, there are kanji characters before and after the annotated base character and, therefore, it is necessary to (a) use nakatsuki alignment and (b) introduce space between the annotated base character and the kanji characters on either side. Using nakatsuki alignment means that the extra space is distributed equally on both sides of the annotated base character.

Ruby annotations
W3C

Complex Ruby: Group Ruby

- Annotations attached to a sequence of characters
- Base character sequence has meaning as a phrase rather than as individual characters
- Most often used for imported words (e.g., motor, oasis)
- The ruby characters give the meaning or pronunciation
- Ruby characters are justified across the phrase

モ	デ	ル
模	型	

mo te ru

モ	デ	ル
模	型	

オ	ア	シ	ス
憩	い	の	場

o a shi su

オ	ア	シ	ス
憩	い	の	場

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 42

Ruby annotations

Romanji is used for Group Ruby is used for Romanji

W3C

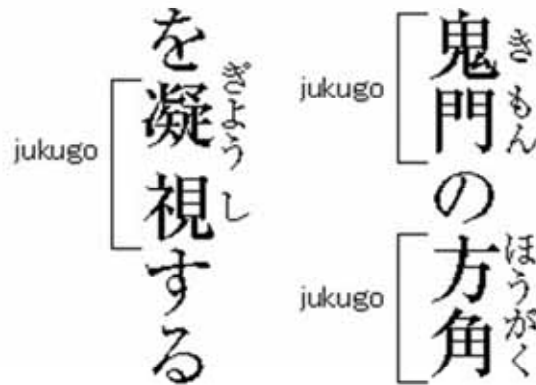
<p>編集者 editor</p>	<p>editor エディター</p>
<p>editor 編集者</p>	<p>エディター editor</p>
Meaning	Pronunciation

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 43

In the lefthand set, the kanji is given an English meaning, "editor". Note that the latin characters are formatted/justified as the normally would be rather than spreading them across the whole "width" of the kanji base character sequence.

In the righthand set, the pronunciation of the English word, "editor" is given in katakana, which is evenly distributed across the "width" of the base (romanji) characters.

- Annotations are attached to individual base characters
- Recall, “jukugo” means words consisting of more than 1 character



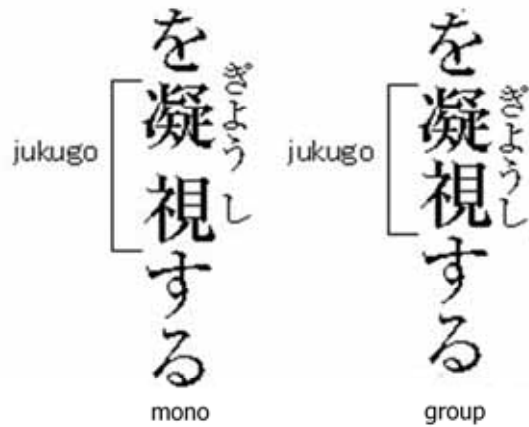
The example here is in two lines of Japanese and is read from right to left vertically

The first jukugo, the one on the top right, has only three ruby characters in its annotation: one on the first base character and two on the second base character

The second jukugo, the one on the bottom right, has two ruby characters on each of its two base characters

The third jukugo, the one on the left, has three ruby characters on its first base character, but only one ruby character on its second base character. Because the ruby cannot overlap a kanji character, there must be a space between the first and second kanji characters. But, this is somewhat ugly. so, next slide

- Jukugo Ruby can be set like mono ruby
- Or, it can be set like phrasal group ruby



In this slide, we have repeated the mono ruby setting of jukugo ruby, but show that a more beautiful setting results from treating the jukugo ruby as if it were group ruby.

Ruby annotations

Jukugo Ruby – additional examples

W3C

- Either the first base character has 3 ruby characters or, the second base character has 3 ruby characters
- Either mono ruby or jukugo (group-like) ruby can be used

mono

jukugo

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 46

Here are two examples of jukugo text with attached ruby annotations.

In the first example, the left one, the first base character has three attached ruby characters and the second base character only has one attached ruby character.

In the second example, the right one, the first base character has only one attached ruby character and the second base character only has three attached ruby character

In both cases, the first base character and its attached ruby characters are shown in blue and the second base characters of the jukugo and its attached ruby characters are shown in black.

The pair of examples on the left show a setting of both examples using the rules for mono ruby. The second pair of examples show that the rules for jukugo ruby allow a much more pleasing setting of the text. The jukugo setting avoids breaking the Kihonhanmen grid. Note that it does not matter which character has 3 ruby characters attached as long as the other character has only one ruby attached.

So, what happens if the jukugo base characters have more than a total of 4 ruby characters attached?

- Long annotations can overlap adjacent base characters



As for group or mono ruby, the overlaps are only allowed if the adjacent characters are not kanji; that is the case in these examples.

Ruby annotations
Complex Ruby – line breaking

W3C

- Breaking a group ruby is not allowed; it must all go on one line
- Breaking a jukugo ruby is allowed and it acts like mono ruby

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 48

One major difference between jukugo and group ruby is that a line break may not occur within a group ruby sequence. For jukugo ruby, on the other hand, line breaks can occur and the layout reverts to the mono ruby layout before and after the line break. The example shows the behavior of our previous 1-3 and 3-1 examples when a line break occurs within the jukugo base character sequence.

note that when there are 3 or more mono ruby characters at the start or end of a line, then the base characters may be moved down to give room, within the Kihonhanmen, for the ruby characters. There are other ways to set the ruby characters in that situation, but that is beyond the scope of this talk.


Concluding remarks

W3C

Concluding remarks


Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 49

This is a presentation slide with a blue header and footer. The header contains the text 'Concluding remarks' on the left and the W3C logo on the right. The main body of the slide is white and contains the text 'Concluding remarks' centered. The footer contains the copyright information 'Copyright © 2005 W3C (MIT, ERCIM, Keio)' on the left and 'slide 49' on the right.

In conclusion 

- Kihonhanmen provides an alternative way of specifying page layout
- Converting between horizontally and vertically oriented text is not so straightforward
- Tables of character classes provide details on spacial adjustment & line breaking
- There are three forms of ruby:
mono, group and jukugo
- The specification is located at:
<http://www.w3.org/TR/jlreq/>

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 50



Thank you
<http://www.w3.org/International/>

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 51