

XSL-FO meets the Tower of Babel

Tony Graham
Mentea
13 Kelly's Bay Beach
Skerries, Co Dublin, Ireland
info@mentea.net
<http://www.mentea.net>

Version 1.0 – MultilingualWeb Workshop – 15 March 2012
© 2012 Mentea



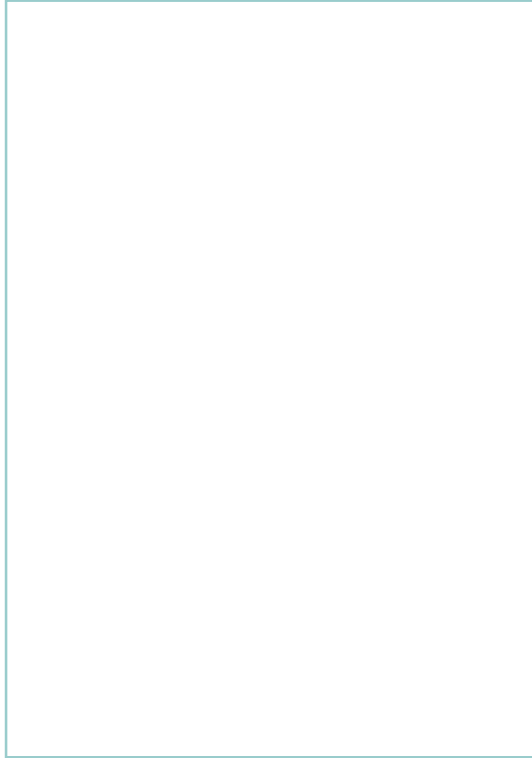
XSL-FO meets the Tower of Babel

Page	<i>5</i>
l18n in XSL 1.0 Requirements	<i>14</i>
Japanese layout	<i>17</i>
1948 and All That	<i>20</i>
What to do?	<i>23</i>
How hard can it be?	<i>25</i>
Further Information	<i>25</i>

I am here as a member of the W3C Working Group that is developing XSL-FO. XSL-FO is an XML vocabulary for specifying formatting semantics. Since some of you may not be familiar with XSL-FO, I'll start with a history of XSL-FO before getting to the Tower of Babel.

Page

1



In the beginning, there was the page. And the page was without objects and void, and not a whole lot was upon the face of the page

Text

2

XSL is Fun Lorem ipsum dolor sit amet, consectetur</to:inline> adipiscing elit. Donec ipsum justo, auctor ut euismod et, dignissim id arcu. Fusce sed odio urna, quis adipiscing justo. Aliquam sit amet magna augue. Morbi lectus neque, condimentum sit amet mattis vitae, euismod aliquet lectus. Donec semper nulla id ante venenatis laoreet. Nunc a quam ligula, id dapibus lectus. Curabitur turpis velit, consequat et dictum vitae, adipiscing Lorem ipsum dolor sit amet sed lorem. Nam fringilla, lacus consequat eleifend elementum, purus leo venenatis neque, a scelerisque justo lorem vitae dolor. Suspendisse in odio ut justo pulvinar congue vitae nec risus. Etiam ultricies eros non risus ultricies commodo. In vulputate, massa id pulvinar porta, mi metus ullamcorper libero, at lobortis felis libero et risus. Nullam id nisl turpis, quis rhoncus sapien. Donec magna magna, tristique at porttitor non, dapibus vitae urna. Mauris laoreet tincidunt lectus, nec tincidunt orci vestibulum ut. Quisque varius ante id arcu congue tempus. Donec at sodales diam. Etiam malesuada leo non felis congue sed porttitor augue aliquet. Proin cursus sollicitudin nisl, ut accumsan lectus vehicula a. Pellentesque a leo tortor. Integer sed est nisl. Duis laoreet, augue eget adipiscing gravida, metus elit varius justo, vitae tincidunt purus elit sit amet lorem. Aliquam suscipit, metus sit amet dignissim ultrices, arcu nibh vestibulum dui, a eleifend tortor tellus quis nunc. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Proin mollis tempus dui non mollis. Aenean lorem turpis, tincidunt ut luctus at, dictum id ipsum. Duis aliquam lacus sapien, quis scelerisque neque. Vivamus non nisl nec quam fermentum ultricies. Maecenas iaculis, tellus vitae rutrum bibendum, risus nisi blandit massa, ac molestie turpis risus ac felis. Nunc dignissim nunc non purus sodales vel dictum quam sodales. Aenean magna leo, mattis in lobortis vel, aliquet et nunc. Cras accumsan mi interdum orci vestibulum vulputate. Nam gravida cursus faucibus. Fusce libero libero, mattis at accumsan at, molestie in mauris. Integer id mi leo, sed pharetra nunc. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Sed luctus viverra turpis, ut egestas mi dictum vitae. Maecenas consequat condimentum iaculis. Integer sagittis magna et metus aliquam porttitor. Ut in tellus ac neque lacinia consectetur. Morbi fermentum cursus sodales. Donec iaculis tortor nec lorem scelerisque ac lobortis velit cursus. Phasellus dapibus urna sed mi dignissim fermentum. Nullam eget ligula vel ante interdum tincidunt nec sit amet urna. Phasellus sagittis, velit consectetur sodales blandit, tellus nulla molestie sapien, a aliquam dui nisl a purus. Fusce faucibus, nulla eu mattis consequat, ante metus pretium purus, sed scelerisque tellus dolor sed nisi. Integer varius magna

And XSL said, let there be text, and there was text.

Region-body

3

XSL is Fun Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec ipsum justo, auctor ut euismod et, dignissim id arcu. Fusce sed odio urna, quis adipiscing justo. Aliquam sit amet magna augue. Morbi lectus neque, condimentum sit amet mattis vitae, euismod aliquet lectus. Donec semper nulla id ante venenatis laoreet. Nunc a quam ligula, id dapibus lectus. Curabitur turpis velit, consequat et dictum vitae, adipiscing Lorem ipsum dolor sit amet sed lorem. Nam fringilla, lacus consequat eleifend elementum, purus leo venenatis neque, a scelerisque justo lorem vitae dolor. Suspendisse in odio ut justo pulvinar congue vitae nec risus. Etiam ultricies eros non risus ultricies commodo. In vulputate, massa id pulvinar porta, mi metus ullamcorper libero, at lobortis felis libero et risus. Nullam id nisl turpis, quis rhoncus sapien. Donec magna magna, tristique at porttitor non, dapibus vitae urna. Mauris laoreet tincidunt lectus, nec tincidunt orci vestibulum ut. Quisque varius ante id arcu congue tempus. Donec at sodales diam. Etiam malesuada leo non felis congue sed porttitor augue aliquet. Proin cursus sollicitudin nisl, ut accumsan lectus vehicula a. Pellentesque a leo tortor. Integer sed est nisl. Duis laoreet, augue eget adipiscing gravida, metus elit varius justo, vitae tincidunt purus elit sit amet lorem. Aliquam suscipit, metus sit amet dignissim ultrices, arcu nibh vestibulum dui, a eleifend tortor tellus quis nunc. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Proin mollis tempus dui non mollis. Aenean lorem turpis, tincidunt ut

And XSL saw the text, and they divided the text from the page. And XSL called the text the region-body, and the space they called the margins. And the region-body and the margins were the first formatting.

XSL is Fun

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec ipsum justo, auctor ut euismod et, dignissim id arcu. Fusce sed odio urna, quis adipiscing justo. Aliquam sit amet magna augue. Morbi lectus neque, condimentum sit amet mattis vitae, euismod aliquet lectus. Donec semper nulla id ante venenatis laoreet. Nunc a quam ligula, id dapibus lectus. Curabitur turpis velit, consequat et dictum vitae, adipiscing Lorem ipsum dolor sit amet sed lorem. Nam fringilla, lacus consequat eleifend elementum, purus leo venenatis neque, a scelerisque justo lorem vitae dolor. Suspendisse in odio ut justo pulvinar congue vitae nec risus. Etiam ultricies eros non risus ultricies commodo. In vulputate, massa id pulvinar porta, mi metus ullamcorper libero, at lobortis felis libero et risus. Nullam id nisl turpis, quis rhoncus sapien. Donec magna magna, tristique at porttitor non, dapibus vitae urna. Mauris laoreet tincidunt lectus, nec tincidunt orci vestibulum ut. Quisque varius ante id arcu congue tempus.

Donec at sodales diam. Etiam malesuada leo non felis congue sed porttitor augue aliquet. Proin cursus sollicitudin nisl, ut accumsan lectus vehicula a. Pellentesque a leo tortor. Integer sed est nisl. Duis laoreet, augue eget adipiscing gravida, metus elit varius justo, vitae tincidunt purus elit sit amet lorem. Aliquam suscipit, metus sit amet dignissim ultrices, arcu nibh vestibulum dui, a eleifend tortor tellus quis nunc. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur

And XSL said, let there be blocks in the midst of the text, and let them divide the text from the text. And it was so. And the text with blocks was the second formatting.

Outer Regions

5

	<p>XSL is Fun</p> <p>Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec ipsum justo, auctor ut euismod et, dignissim id arcu. Fusce sed odio urna, quis adipiscing justo. Aliquam sit amet magna augue. Morbi lectus neque, condimentum sit amet mattis vitae, euismod aliquet lectus. Donec semper nulla id ante venenatis laoreet. Nunc a quam ligula, id dapibus lectus. Curabitur turpis velit, consequat et dictum vitae, adipiscing Lorem ipsum dolor sit amet sed lorem. Nam fringilla, lacus consequat eleifend elementum, purus leo venenatis neque, a scelerisque justo lorem vitae dolor. Suspendisse in odio ut justo pulvinar congue vitae nec risus. Etiam ultricies eros non risus ultricies commodo. In vulputate, massa id pulvinar porta, mi metus ullamcorper libero, at lobortis felis libero et risus. Nullam id nisl turpis, quis rhoncus sapien. Donec magna magna, tristique at porttitor non, dapibus vitae urna. Mauris laoreet tincidunt lectus, nec tincidunt orci vestibulum ut. Quisque varius ante id arcu congue tempus.</p> <p>Donec at sodales diam. Etiam malesuada leo non felis congue sed porttitor augue aliquet. Proin cursus sollicitudin nisl, ut accumsan lectus vehicula a. Pellentesque a leo tortor. Integer sed est nisl. Duis laoreet, augue eget adipiscing gravida, metus elit varius justo, vitae tincidunt purus elit sit amet lorem. Aliquam suscipit, metus sit amet dignissim ultrices, arcu nibh vestibulum dui, a eleifend tortor tellus quis nunc. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur</p>
	2

And XSL said, let there be regions in the edges of the page to hold static content, and let them be for page numbers and running heads, and it was so.

Recto and Verso

6

<p>XSL is Fun</p> <p>Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec ipsum justo, auctor ut euismod et, dignissim id arcu. Fusce sed odio urna, quis adipiscing justo. Aliquam sit amet magna augue. Morbi lectus neque, condimentum sit amet mattis vitae, euismod aliquet lectus. Donec semper nulla id ante venenatis laoreet. Nunc a quam ligula, id dapibus lectus. Curabitur turpis velit, consequat et dictum vitae, adipiscing Lorem ipsum dolor sit amet sed lorem. Nam fringilla, lacus consequat eleifend elementum, purus leo venenatis neque, a scelerisque justo lorem vitae dolor. Suspendisse in odio ut justo pulvinar congue vitae nec risus. Etiam ultricies eros non risus ultricies commodo. In vulputate, massa id pulvinar porta, mi metus ullamcorper libero, at lobortis felis libero et risus. Nullam id nisl turpis, quis rhoncus sapien. Donec magna magna, tristique at porttitor non, dapibus vitae urna. Mauris laoreet tincidunt lectus, nec tincidunt orci vestibulum ut. Quisque varius ante id arcu congue tempus.</p> <p>Donec at sodales diam. Etiam malesuada leo non felis congue sed porttitor augue aliquet. Proin cursus sollicitudin nisl, ut accumsan lectus vehicula a. Pellentesque a leo tortor. Integer sed est nisl. Duis laoreet, augue eget adipiscing gravida, metus elit varius justo, vitae tincidunt purus elit sit amet lorem. Aliquam suscipit, metus sit amet dignissim ultrices, arcu nibh vestibulum dui, a eleifend tortor tellus quis nunc. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur</p> <p>2</p>	<p>ridiculus mus. Proin mollis tempus dui non mollis. Aenean lorem turpis, tincidunt ut luctus at, dictum id ipsum. Duis aliquam lacus sapien, quis scelerisque neque. Vivamus non nisl nec quam fermentum ultricies.</p> <p>Maecenas iaculis, tellus vitae rutrum bibendum, risus nisi blandit massa, ac molestie turpis risus ac felis. Nunc dignissim nunc non purus sodales vel dictum quam sodales. Aenean magna leo, mattis in lobortis vel, aliquet et nunc. Cras accumsan mi interdum orci vestibulum vulputate. Nam gravida cursus faucibus. Fusce libero libero, mattis at accumsan at, molestie in mauris. Integer id mi leo, sed pharetra nunc. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Sed luctus viverra turpis, ut egestas mi dictum vitae. Maecenas consequat condimentum iaculis. Integer sagittis magna et metus aliquam porttitor. Ut in tellus ac neque lacinia consectetur. Morbi fermentum cursus sodales. Donec iaculis tortor nec lorem scelerisque ac lobortis velit cursus. Phasellus dapibus urna sed mi dignissim fermentum. Nullam eget ligula vel ante interdum tincidunt nec sit amet urna. Phasellus sagittis, velit consectetur sodales blandit, tellus nulla molestie sapien, a aliquam dui nisl a purus. Fusce faucibus, nulla eu mattis consequat, ante metus pretium purus, sed scelerisque tellus dolor sed nisi. Integer varius magna tortor, ac vehicula tellus. Cras tempor nulla ut nisi tempor nec vestibulum orci suscipit. In mollis lacinia augue sit amet viverra. Nullam non ipsum nibh, id hendrerit ante. Nunc pretium vulputate odio</p> <p>3</p>
--	--

And XSL made the two page sides, the recto for the right hand side and the verso for the left.

And XSL set them in the page masters to apply to all pages. And the recto and the verso were the third formatting.

Inline FOs and Properties

7

<p style="text-align: center;">XSL is Fun</p> <p> Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec ipsum justo, auctor ut euismod et, dignissim id arcu. Fusce sed odio urna, quis adipiscing justo. Aliquam sit amet magna augue. Morbi lectus neque, condimentum sit amet mattis vitae, euismod aliquet lectus. Donec semper nulla id ante venenatis laoreet. Nunc a quam ligula, id dapibus lectus. Curabitur turpis velit, consequat et dictum vitae, adipiscing Lorem ipsum dolor sit amet sed lorem. Nam fringilla, lacus consequat eleifend elementum, purus leo venenatis neque, a scelerisque justo lorem vitae dolor. Suspendisse in odio ut justo pulvinar congue vitae nec risus. Etiam ultricies eros non risus ultricies commodo. In vulputate, massa id pulvinar porta, mi metus ullamcorper libero, at lobortis felis libero et risus. Nullam id nisl turpis, quis rhoncus sapien. Donec magna magna, tristique at porttitor non, dapibus vitae urna. Mauris laoreet tincidunt lectus, nec tincidunt orci vestibulum ut. Quisque varius ante id arcu congue tempus. </p> <p> Donec at sodales diam. Etiam malesuada leo non felis congue sed porttitor augue aliquet. Proin cursus sollicitudin </p> <p style="text-align: center;">2</p>	<p> nisi, ut accumsan lectus vehicula a. Pellentesque a leo tortor. Integer sed est nisl. Duis laoreet, augue eget adipiscing gravida, metus elit varius justo, vitae tincidunt purus elit sit amet lorem. Aliquam suscipit, metus sit amet dignissim ultrices, arcu nibh vestibulum dui, a eleifend tortor tellus quis nunc. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Proin mollis tempus dui non mollis. Aenean lorem turpis, tincidunt ut luctus at, dictum id ipsum. Duis aliquam lacus sapien, quis scelerisque neque. Vivamus non nisl nec quam fermentum ultricies. </p> <p> Maecenas iaculis, tellus vitae rutrum bibendum, risus nisi blandit massa, ac molestie turpis risus ac felis. Nunc dignissim nunc non purus sodales vel dictum quam sodales. Aenean magna leo, mattis in lobortis vel, aliquet et nunc. Cras accumsan mi interdum orci vestibulum vulputate. Nam gravida cursus faucibus. Fusce libero libero, mattis at accumsan at, molestie in mauris. Integer id mi leo, sed pharetra nunc. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Sed luctus viverra turpis, ut egestas mi dictum vitae. Maecenas consequat condimentum iaculis. <i>Integer sagittis</i> magna et metus aliquam porttitor. Ut in tellus ac neque lacinia consectetur. Morbi fermentum cursus sodales. Donec iaculis tortor nec lorem scelerisque ac lobortis velit cursus. Phasellus dapibus urna sed mi dignissim fermentum. Nullam eget ligula vel ante interdum tincidunt nec sit amet urna. Phasellus sagittis, velit consectetur sodales blandit, tellus nulla </p> <p style="text-align: center;">3</p>
--	--

And XSL said, let the blocks bring forth abundantly inline objects and properties that apply to them.

And XSL created `fo:character` and every inline FO after their kind, and every property after its kind, and it was good.

And the inline FOs and properties were the fourth formatting.

XSL is Fun

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec ipsum justo, auctor ut euismod et, dignissim id arcu. Fusce sed odio urna, quis adipiscing justo. Aliquam sit amet magna augue. Morbi lectus neque, condimentum sit amet mattis vitae, euismod aliquet lectus. Donec semper nulla id ante venenatis laoreet. Nunc a quam ligula, id dapibus lectus. Curabitur turpis velit, consequat et dictum vitae, adipiscing Lorem ipsum dolor sit amet sed lorem. Nam fringilla, lacus consequat eleifend elementum, purus leo venenatis neque, a scelerisque justo lorem vitae dolor. Suspendisse in odio ut justo pulvinar congue vitae nec risus. Etiam ultricies eros non risus ultricies commodo. In vulputate, massa id pulvinar porta, mi metus ullamcorper libero, at lobortis felis libero et risus. Nullam id nisl turpis, quis rhoncus sapien. Donec magna magna, tristique at porttitor non, dapibus vitae urna. Mauris laoreet tincidunt lectus, nec tincidunt orci vestibulum ut. Quisque varius ante id arcu congue tempus.

Lorem ipsum dolor sit amet



Donec at Aliquam suscipit, mus non nisl nec sodales diam. metus sit amet quam fermentum ultricies. Maecenas arcu nibh vestibulum dui, a eleifend iaculis, tellus vitae rutrum bibendum, risus nisi blandit massa, ac molestie turpis risus ac felis. Nunc dignissim nunc non purus sodales vel dictum quam sodales. Aenean magna leo, mattis in lobortis vel, aliquet et nunc. Cras accumsan mi interdum orci vestibulum vulputate. Nam gravida Donec 12 12 14 Morbi 12 24 83 augue eget adipiscing gravida, metus elit varius justo, vitae tincidunt purus elit sit amet lorem.

And XSL said, let users make graphics, and tables, and columns, and side floats and other areas. And the graphics and tables and columns and side floats were the fifth formatting.

Change Bars, Table Markers

9

<h2 style="text-align: center;">XSL is Fun</h2> <p> Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec ipsum justo, auctor ut euismod et, dignissim id arcu. Fusce sed odio urna, quis adipiscing justo. Aliquam sit amet magna augue. Morbi lectus neque, condimentum sit amet mattis vitae, euismod aliquet lectus. Donec semper nulla id ante venenatis laoreet. Nunc a quam ligula, id dapibus lectus. Curabitur turpis velit, consequat et dictum vitae, adipiscing Lorem ipsum dolor sit amet sed lorem. Nam </p> <p> fringilla, lacus consequat eleifend elementum, purus leo venenatis neque, a scelerisque justo lorem vitae dolor. Suspendisse in odio ut justo pulvinar congue vitae nec risus. Etiam ultricies eros non risus ultricies commodo. In vulputate, massa id pulvinar porta, mi metus ullamcorper libero, at lobortis felis libero et risus. Nullam id nisl turpis, quis rhoncus sapien. Donec magna magna, tristique at porttitor non, dapibus vitae urna. Mauris laoreet tincidunt lectus, nec tincidunt orci vestibulum ut. Quisque varius ante id arcu congue tempus. </p> <p style="text-align: center;">2</p>	<div style="text-align: center;">  </div> <p> Donec at Aliquam suscipit, mus non nisl nec sodales diam. metus sit amet quam fermentum ultricies. Etiam malesuada dignissim ultrices, Maecenas leo non felis arcu nibh vestibulum dui, a eleifend iaculis, tellus congue sed porttitor augue aliquet. lorum dui, a eleifend vitae rutrum Proin cursus nunc. Cum sociis bibendum, risus cursus natoque penatibus nisi blandit massa, et magnis dis ac molestie turpis sollicitudin nisl, ut parturient montes, risus ac felis. Nunc vehicula a. Pellentesque a leo tortor, nascetur ridiculus dignissim nunc Integer sed est mus. Proin mollis non purus sodales nisl. Duis laoreet, tempus dui non vel dictum quam Donec 12 12 14 mollis. Aenean sodales. Aenean Morbi 12 24 83 lorem turpis, magna leo, mattis tincidunt ut luctus in lobortis vel, augue eget adipiscing at, dictum id aliquet et nunc. gravidam, metus elit varius ipsum. Duis Cras accumsan mi justo, vitae tincidunt aliquam lacus interdum orci amet purus elit sit sapien, quis scelerisque neque. Vivat vestibulum vulputate. Nam gravida </p> <p style="text-align: center;">3</p>
--	--

Then XSL did errata until XSL 1.1, which added change bars, table markers, indexing FOs and much that was good.

I promised you a history, not the history, of XSL-FO. I'm now going to just slide past the whole temptation and fratricide thing and go straight to the Tower part of the story.

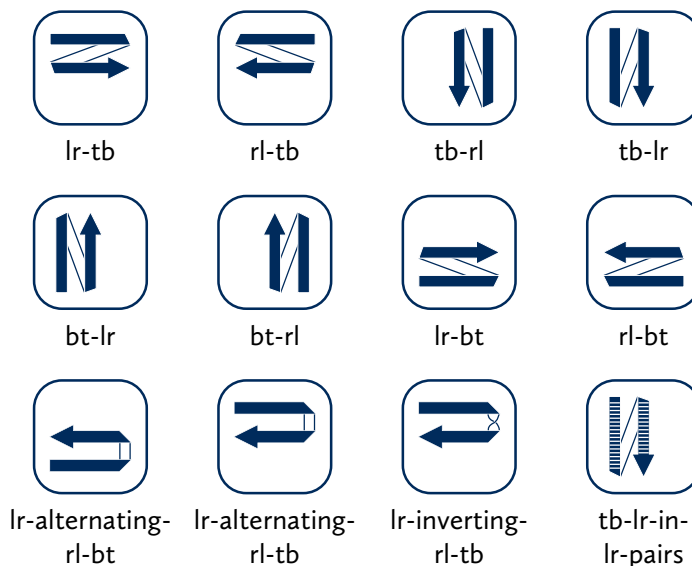
i18n in XSL 1.0 Requirements



Unlike in the original Tower of Babel, the designers of XSL-FO knew it was coming, and XSL-FO has from the beginning been designed to support multiple writing modes, etc.

This is the “bird’s eye” view of the requirements for XSL 1.0 from 1998 with the i18n section highlighted. As you can see, it’s 8 out of the 21 pages.

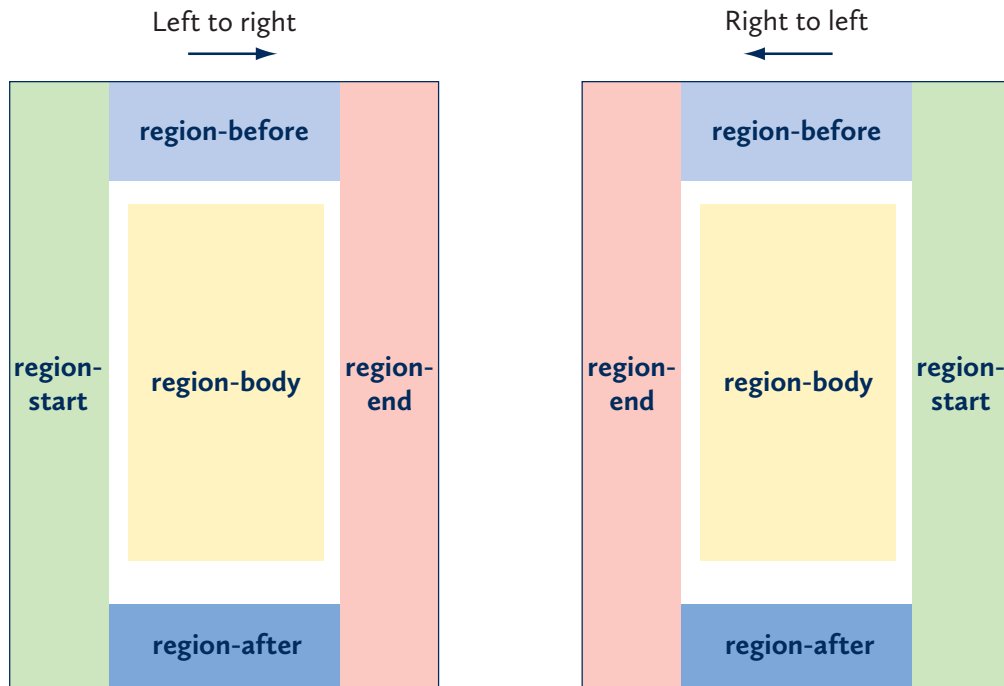
Writing modes



XSL-FO has always been good at the “big picture” of multilingualism. For example, XSL-FO defines multiple writing modes to cover the requirements of just about every known script.

Writing-mode-neutral names

12

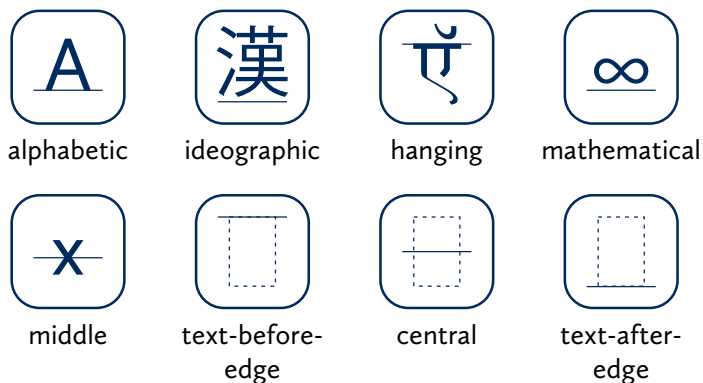


XSL-FO also has `margin-*`, `border-*`, `padding-*`, and other properties much like CSS, but they are defined in terms of `-before`, `-after`, `-start`, and `-end` rather than fixed `-top`, `-bottom`, `-left`, and `-right`. XSL-FO also has the `-top`, `-bottom`, `-left`, and `-right` properties for compatibility with CSS, but they are defined as mapping to `-before`, `-after`, `-start`, or `-end` depending on the current writing mode, reference orientation, etc.

Font baselines

13

Common with CSS



XSL-FO and CSS share the same concepts of different baselines for different styles of text.

I18n in XSL-FO 2.0 Requirements

14

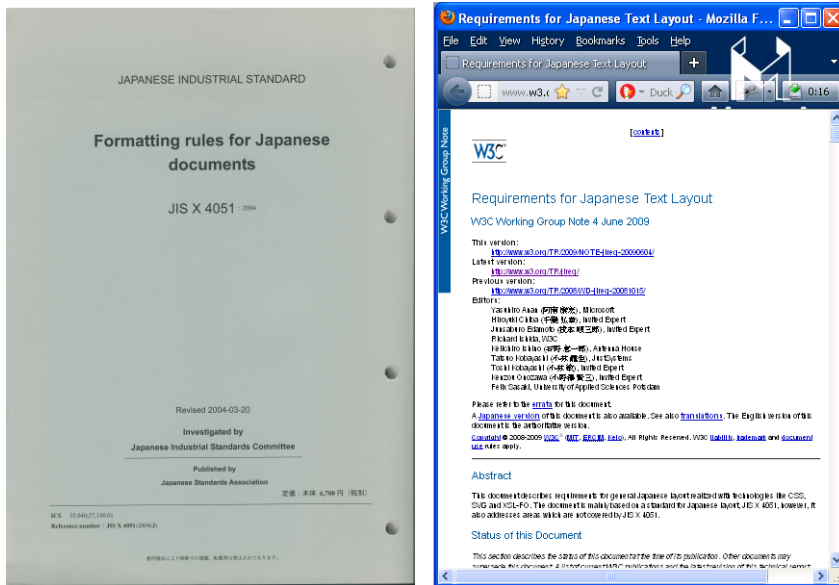


In contrast to what you saw from the XSL 1.0 requirements, this is the section on non-Western processing for the 2008 XSL 2.0 requirements.

It's not that we gave up on i18n, it's that we recognised that there wasn't much point repeating what remained from the XSL 1.0 requirements if we still weren't able to implement them. The other thing that changed since 1998 was that the W3C was working on its first layout task force.

Japanese layout

15

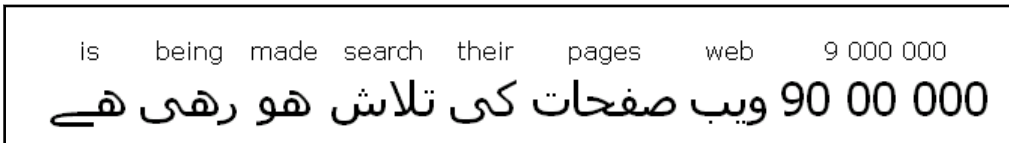


The Requirements for Japanese Layout was based on JIS-X-4051 and on the collected experience of the Task Force members. It's useful to: authors and designers producing Japanese content; to spec writers who try to capture its recommendations in terms of objects and properties; and to implementers, who try to take care of the rest of the details so you don't have to. It's also more accessible than the JIS standard since it's in English as well as Japanese.

Ruby

16

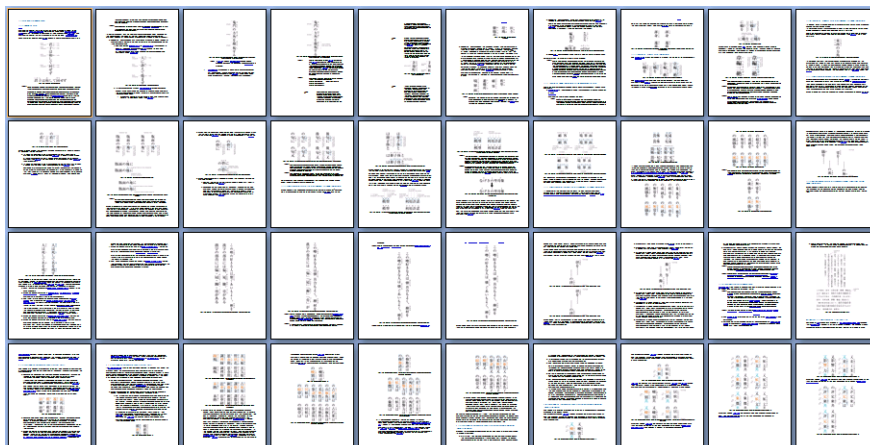
- “small-sized, supplementary text attached to a character or a group of characters in the main text”
- Used with:
 - Japanese
 - Chinese + Bopomofo
 - Chinese + Pinyin
 - Other languages



To illustrate the benefit of the Japanese Layout Requirements, I'm going to talk a bit about Ruby, hopefully without stealing too much of Richard and Jirka's thunder. Ruby is small annotations attached to a base text, and I'm talking about it because it's used not just with Japanese but with Chinese and bopomofo tone marks, with Chinese with pinyin pronunciations, and, this being the World Wide Web, with other languages as well.

What I know about Japanese Ruby

17

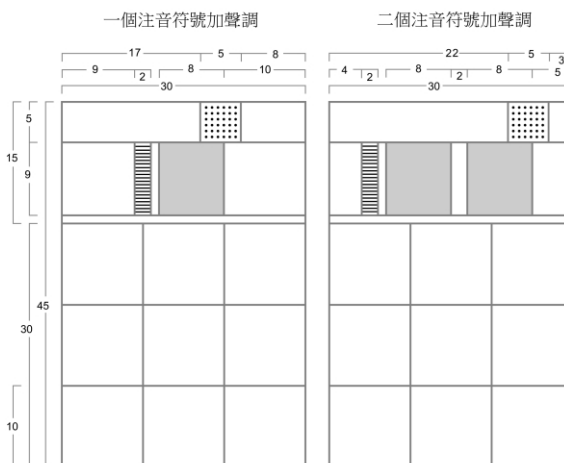


This is a bird's eye view of the 40 pages about Ruby in the Japanese Layout Requirements document. If Felix or Richard squint a little, they may even be able to recognise a few of the graphics.

What I know about Bopomofo Ruby

18

Twelve diagrams like this:



In contrast, this is an excerpt of the only details that I know about formatting bopomofo as Ruby. Neither XSL-FO nor CSS has this level of detail. Bopomofo Ruby is mostly used in children's textbooks, so it's a problem if someone wants to produce material for children, a web page for children, or even an EPUB for children.

What I know about Pinyin Ruby

19

Chūn Cán Dào Sǐ Sī Fāng Jìn
 春 蚕 到 死 丝 方 尽 ，
 Là Jù Chéng Huī Lèi Shǐ Gān
 蜡 炬 成 灰 泪 始 干 。

If you thought I didn't know much about bopomofo as Ruby, this is all I know about Pinyin as Ruby. I can tell you this is a poem about silkworms and candles, but I can't tell you anything about the finer points of formatting it.

1948 and All That

20

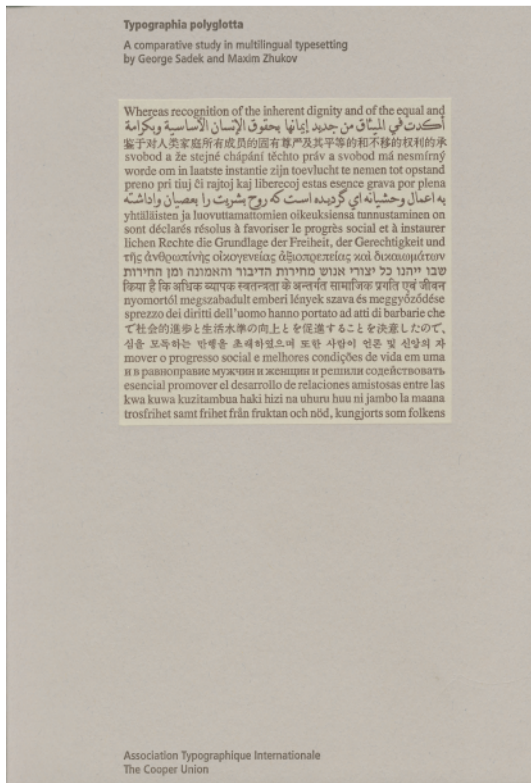
- UDHR adopted 10 December 1948
- Designed to give meaning to “fundamental freedoms” and “human rights” phrases in UN Charter
- Not universally liked
- Guinness World Record as “Most Translated Document”

Universal Declaration of Human Rights		English
Preamble		
Whereas recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world,		
Whereas disregard and contempt for human rights have resulted in barbarous acts which have outraged the conscience of mankind, and the advent of a world in which human beings shall enjoy freedom of speech and belief and freedom from fear and want has been proclaimed as the highest aspiration of the common people,		
Whereas it is essential, if man is not to be compelled to have recourse, as a last resort, to rebellion against tyranny and oppression, that human rights should be protected by the rule of law,		
Whereas it is essential to promote the development of friendly relations between nations,		
Whereas the peoples of the United Nations have in the Charter reaffirmed their faith in fundamental human rights, in the dignity and worth of the human person and in the equal rights of men and women and have determined to promote social progress and better standards of life in larger freedom,		
Whereas Member States have pledged themselves to achieve, in co-operation with the United Nations, the promotion of universal respect for and observance of human rights and fundamental freedoms,		
Whereas a common understanding of these rights and freedoms is of the greatest importance for the full realization of this pledge,		
Now, therefore,		
The General Assembly		
Proclaims this Universal Declaration of Human Rights as a common standard of achievement for all peoples and all nations, to the end that every individual and every organ of society, keeping this Declaration constantly in mind, shall strive by teaching and education to promote respect for these rights and freedoms and by progressive measures, national and international, to secure their universal and effective recognition and observance, both among the peoples of Member States themselves and among the peoples of territories under their jurisdiction.		
eng code	eng language	UK country

If you fast-forward an indeterminate number of years to 1948, you come across a Tower of Babel of our own making. The Universal Declaration of Human Rights may have started out as a way to fill in the holes in the UN Charter, but according to the Guinness Book of World Records, it's been translated into more languages than any other document.

Typographia Polyglotta (1991, 1997)

21



Every written language has its own typographic conventions, and these must be respected if printed material is to appear natural to its intended audience. Such rules include differences in punctuation, spacing, and diacriticals; standards for hyphenation and justification; existence or nonexistence of upper- and lowercase letterforms; traditions regarding format; and many other major and minor conventions of typographic styling.

This, of course, makes the UDHR ideal for people doing comparisons between languages or scripts. The oldest of these that I know of is the *Typographia Polyglotta*, published in 1991 with a second edition in 1997. The 1997 edition has the preamble of the UDHR in 22 languages, including Arabic, Hebrew, Chinese, Japanese, and Korean. To do that in 1991, when Unicode was little more than a good idea, or even in 1997 was an achievement, even if we fully expect our web browsers and even word processors to handle it today.

So it's interesting that the book includes some comments by its typesetter about the need to adhere to the conventions of the script that you're formatting.

“UDHR in Unicode”

- Demonstrates the use of Unicode for a wide variety of languages
- See <http://www.unicode.org/udhr/index.html>
- XML source for nearly 400 languages/scripts
- Over 2,000 pages when formatted

Another one of the uses with which you may be more familiar is the “UDHR in Unicode” project that has the dual purpose of promoting both the UDHR and Unicode. It suits my purposes that its files are available as XML, and when you format them in one go, as some of us like to do, they come out to over 2,000 pages, although different formatters will do better or worse with some of the languages.

UDHR in Khmer

“Nice try, shame about the superscripts and leading”

Khmer, Central	<p>សចេក្កដីប្រកាសជាសកលសុដីពីសិទ្ធិមនុស្ស</p> <p>អនុម័តនិងប្រកាសដោយសចេក្កដីសម្រេចចិត្តជនប្រហុសនុនិហាតលខេ ២១៧ A (III) នៅថ្ងៃទី ១០ ខែ ធ្នូ ឆ្នាំ ១៩៤៨</p> <p>ប្រព្រឹត្តិការណ៍</p> <p>ដោយយល់ឃើញថា ការទទួលស្គាល់សចេក្កដីថ្លៃថ្នូរជាប់ពីកំណើត និងសិទ្ធិមនុស្ស ភាពតុល្យភាព និងសិទ្ធិមនុស្ស អាចលក់ ដូរ ផ្លូវ ឬដកហូតបានរបស់សមាជិកទាំងអស់នៃគ្រួសារមនុស្ស គឺជាគ្រឹះនៃសេរីភាព យុត្តិធម៌ និងសន្តិភាពក្នុងពិភពលោក ។</p> <p>ដោយយល់ឃើញថា ការមិនទទួលស្គាល់ និងការប្រមាថមើលងាយសិទ្ធិមនុស្ស នាំឱ្យមានអំពើពុករលួយសោហ៊ុយយុទ្ធសាស្ត្រ ធ្វើឱ្យខូចខាតប្រជាជនដល់សតិសម្បជញ្ញៈមនុស្សជាតិ និងថា ការឈានដល់ពិភព លោកមួយ ដ៏ល្អមនុស្សទាំងឡាយមានសេរីភាពក្នុងការនិយាយសុដី និងជំនឿ និងការរួចផុតពីការកើតឡើង និងទុក្ខវេទនា ត្រូវបានប្រកាសថាជាសចេក្កដីប្រកាសដ៏ខ្ពស់បំផុតរបស់មនុស្សទូទៅ ។</p> <p>ដោយយល់ឃើញថា ជាការចាំបាច់ ដល់សិទ្ធិមនុស្សត្រូវតែមានការការពារដោយនីតិវិធី ដើម្បីជំរុញ រាងក្រឡាមនុស្សបង្កឱ្យបានជាច្រើនក្នុងការយល់ឃើញរបស់ ប្រជាជនប្រកាសទាំងអស់និងអំណាចផ្តាច់ការ និងការការពារសង្គម ។</p> <p>ដោយយល់ឃើញថា ជាការចាំបាច់ ដល់ត្រូវលើកស្ទួយការពង្រីកទំនាក់ទំនងជាមិត្តភាព រវាង ប្រជាជាតិទូទាំងពិភពលោក ។</p> <p>ដោយយល់ឃើញថា ប្រជាជាតិទាំងឡាយនៃប្រជាជាតិ មានប្រកាសបញ្ជាក់សារជាថ្មីក្នុង ធម្មនុញ្ញសហប្រជាជាតិ នូវជំនឿរបស់ខ្លួនទៅលើសិទ្ធិមនុស្សដ៏មានសេរីភាព និងលើសចេក្កដីថ្លៃថ្នូរ លើកម្មវិធីរបស់ មនុស្ស និងលើសម្រាប់និយោជន៍សុដី ហើយប្រជុំជាញាតិសាស្ត្រយុទ្ធសាស្ត្រសម្រាប់សង្គម និងកម្មវិធី ជីវភាពរស់នៅឱ្យបានល្អប្រសើរឡើង ក្នុងសេរីភាពកាន់តែទ្រុឌទ្រោម ។</p>
khm code	
khm language	

I was contacted by a guy last year about formatting Khmer, but I was away from the office that week so the best I could do was forward the UDHR in Khmer, though I hadn't looked at it previously. He emailed back that the lines were too close together and the subscripts and superscripts (as he put it) were wrong. I didn't hear from him again, so I couldn't try improving the Khmer formatting and getting another review.



What to do?

24

- More Task Forces?
- Multilingual Layout Community Group?
- Both?

So now I've demonstrated some of the gaps in my knowledge and, by extension, some of the gaps in the ability of current stylesheet languages and current formatters to properly represent non-Western scripts. The next question is what to do about it. The two possibilities that have occurred to me are for the W3C to create more task forces along the lines of the Japanese Layout Task Force or to create a Multilingual Layout Community Group to capture knowledge as it becomes available.

More Task Forces?

25

- A lot of work by experts
- JLReq influencing XSL-FO and CSS
 - Also useful to authors, designers, and implementers
- TF visible and co-ordinated
- Easier to hold/justify F2F reviews

The Japanese Layout Requirements document represents a mammoth amount of work (that is still going on) by a group of experts. As you would expect, the document is packed with useful information. If the W3C was able to put together more task forces, it would be a major effort to find the right groups of people and major efforts by those people to produce sets of layout requirements. But the stature of the group and the quality of the deliverables can make it easier for others to justify reviewing the work and easier to include it in other standards.

Multilingual Layout CG?

26

- New facet of W3C



- Per-CG wiki, mailing lists, blog, & IRC channel
- Anyone able to join
- Easy progression to standards track
- Dilute existing i18n mailing lists?
- Could it last beyond initial enthusiasm?

Community and business groups are a new venture for the W3C where there's a very low overhead for creating a community of interest around any particular topic. A Multilingual Layout Community Group could be a crowd-sourcing of the task of collecting information about best practices for laying out different scripts, but if it happened it would come with very real concerns about its effect on other, existing W3C i18n efforts and about whether it would start with a lot of enthusiasm but not achieve much.

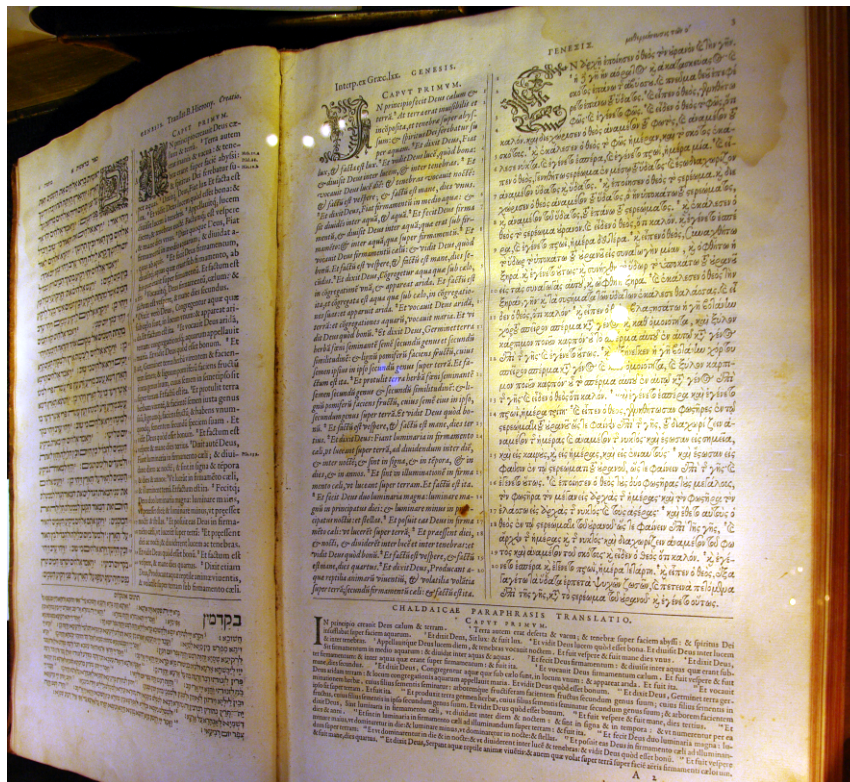
Task Forces and Community Group?

27

- The cathedral *and* the bizarre?
 - Task Forces for major efforts
 - CG wiki accretes information about lesser-known traditions
- CG as incubator for Task Forceable efforts
- Contributor Agreement makes it easy to use CG output

Then there's the possible middle ground of task forces and a Multilingual Layout Community Group co-existing. In the best of all possible worlds, there could be task forces to do the heavy lifting on defined areas and a Community Group as the incubator for potential task forces and the place to collect information about lesser-known scripts, but that best of all possible worlds assumes that the W3C can form new task forces and that it's possible to have a viable Community Group.

How hard can it be?



The Plantin Polyglot Bible was published just up the road in Antwerp between 1568 and 1573. It contains five languages, so they handled the Tower of Babel five times. However, we just need to solve it once. How hard can it be?

Further Information

- Extensible Stylesheet Language (XSL) 2.0
<http://www.w3.org/TR/xslfo20>
- Extensible Stylesheet Language (XSL) Requirements Version 2.0
<http://www.w3.org/TR/xslfo20-req/>
- XSL Requirements Summary
<http://www.w3.org/TR/WD-XSLReq>
- Requirements for Japanese Text Layout
<http://www.w3.org/TR/jlreq/>
- Summary of Japanese Layout Note
<http://www.tcworld.info/index.php?id=64>

UDHR References

30

- UDHR in Unicode
<http://www.unicode.org/udhr/>
- Wikipedia on UDHR
https://secure.wikimedia.org/wikipedia/en/wiki/Universal_Declaration_of_Human_Rights
- UDHR record
<http://www.guinnessworldrecords.com/records-1000/most-translated-document/>

Credits

31

- slide 16
<http://www.user.uni-hannover.de/nhtcapri/ruby-annotation.html>
- slide 26
<http://www.w3.org/community/>
- slide 28
http://en.wikipedia.org/wiki/File:Heritage_day_2011_gnangarra-68.jpg
Photograph by Gnangarra, Creative Commons Attribution 2.5 Australia licence



MENTEA