



Designing for International Users: Practical Tips

Richard Ishida
W3C Internationalization Activity Lead



The slide features a blue header bar at the top. On the left side of the header, there are two small squares, one light blue and one dark blue, followed by the text 'Text size'. On the right side of the header is the W3C logo. The main body of the slide is white and contains the text 'Text size' in a large, blue, sans-serif font. At the bottom of the slide, there is a thin grey bar containing the copyright text 'Copyright © 2005 W3C (MIT, ERCIM, Keio)' on the left and 'slide 2' on the right.

Text size

W3C®

Text size


Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 2

Text size


Text size in translation

W3C®

DochuLa Pass



Around 3,100 m high, in Bhutan.

 © This photo is public. [Change?](#)
 Uploaded on [Jan 28, 2007](#) | [Map](#) | [Delete](#)
 953 views / [comments](#)

Language	Translation	Ratio
Korean	조회	0.8
English	views	1
Chinese	次檢視	1.2
Portuguese	visualizações	2.6
French	consultations	2.6
German	-mal angesehen	2.8
Italian	visualizzazioni	3

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 3


This slide shows an example of how text can expand (and shrink) when translated (here from English).

Text size


Text size in graphics

W3C®

Seals and sealions at Pier 39



Seals and sealions at Pier 39



Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 4


If you are using graphic text, beware! Not only is it more difficult to translate graphic text, but graphic text is often squeezed into small spaces, making it difficult to replicate the source in another language.

Text size


Text size in graphics

W3C®

Seals and sealions at Pier 39



Seals and sealions at Pier 39



Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 5

Here Flickr has resorted to using just icons with mouse-overs in the translated UIs to get around the expansion issues they would have otherwise had.

Text size

Form field label positioning

W3C®

Global settings

Interface language	English
Search language	English
Number of results	10
Save preferences	

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 6

Here is another potential issue related to the length of text in translation.

Let's continue to assume a situation where text appears in a fixed width box. We will apply the same approach we discussed earlier to deal with the title of the box. The issue this time will be that we have used a table to apply form labels to the left side of the form entry field to which they apply. Our initial source text is in English.

W3C[®]
Text size
Form field label positioning

Global settings

Interface language English

Search language English

Number of results 10

Acuan Umum

Bahasa Pengantar di Antar Muka Ingriss

Bahasa Pengantar untuk Penelusuran Ingriss

Jumlah Hasil Penelusuran 10

Allgemeine Voreinstellungen

Sprache der Benutzeroberfläche Englisch

Suchsprache Englisch

Anzahl der Ergebnisse 10

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

The English looks nice enough. The Malay, on the other hand, looks pretty ugly. The large expansion factor produces unfortunate stacking of the text on the left, and large white spaces to the right. Although the box expands vertically to hold all the text, we are wasting a lot of space and decreasing the amount of information that will appear in the reader's initial screen (you can imagine that this would be compounded by other fixed width boxes on the page).

With the German translation we have a different problem. The long word Benutzeroberfläche doesn't wrap, and so pushes the select boxes beyond the width of the fixed box container. This has the potential to badly affect the layout of other parts of the screen.

W3C[®]

Text size

Form field label positioning

Global settings

Interface language
English ▼

Search language
English ▼

Number of results 10 ▼

Save preferences

Acuan Umum

Bahasa Pengantar di Antar Muka
Inggris ▼

Bahasa Pengantar untuk Penelusuran
Inggris ▼

Jumlah Hasil Penelusuran 10 ▼

Simpan Acuan

Allgemeine Voreinstellungen

Sprache der Benutzeroberfläche
Englisch ▼

Suchsprache
Englisch ▼

Anzahl der Ergebnisse 10 ▼

Einstellungen speichern

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

You may want to consider avoiding table cells in such constrained circumstances. This slide shows how the text would look if the input fields were just in a paragraph with the label text.

All the boxes now look fine, and although there is a very slight increase in vertical height overall, we have removed the problems seen with the Malay and German text on the previous slide.

Let's note, again, that this is down to the way the page is **designed/developed**, not the way it is localized. That's a fundamental message of this presentation. Internationalization during design and development removes significant barriers to deploying your content globally.

Text size

Graphics in fixed width boxes

W3C®

Idioma de la interfaz

☐ arabe

☐ inglés

☐ finlandés

☐ francés

☐ italiano

☐ japonés

☐ ruso

☐ español

☐ sueco

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 9

Let's assume that we want to implement a **fixed-width** box on our page.
The text can expand downwards, but not sideways.

Let's also assume that we want a background with a nice gradient behind
the title of the box, and that the background has a line across the bottom.

(This slide in Spanish has the title 'Interface Language', and a list of radio
buttons to select a language.)

Text size

Graphics in fixed width boxes

W3C®

Bahasa Pengantar
untuk Penelusuran

☐ Arab
☐ Inggris
☐ Finlandia
☐ Perancis
☐ Italia
☐ Jepang
☐ Rusia
☐ Spanyol
☐ Swedia

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 10

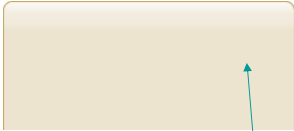
As our text expands during translation into Malay, the title occupies two lines. Unfortunately the graphic used for the gradient background is only one line deep, and things now begin to look a mess.

Text size
W3C®

Graphics in fixed width boxes

**Bahasa Pengantar
untuk Penelusuran**

- ☐ Arab
- ☐ Inggris
- ☐ Finlandia
- ☐ Perancis
- ☐ Italia
- ☐ Jepang
- ☐ Rusia
- ☐ Spanyol
- ☐ Swedia



```
.box h3 {
  margin: 0;
  padding: 6px 8px 4px 10px;
  font-size: 130%;
  color: #333;
  background: url(img/h3-bg.gif)
    no-repeat top left;
  border-bottom: 1px solid #E0CFAB;
}
```

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 11

A way to approach this issue is to use a graphic that is three or four lines deep behind the title. By attaching the graphic using the CSS background property, only the amount needed to view the title will actually be shown.

Text size
W3C

Graphics in fixed width boxes

Bahasa Pengantar untuk Penelusuran

- ☐ Arab
- ☐ Inggris
- ☐ Finlandia
- ☐ Perancis
- ☐ Italia
- ☐ Jepang
- ☐ Rusia
- ☐ Spanyol
- ☐ Swedia

```
.box h3 {
  margin: 0;
  padding: 6px 8px 4px 10px;
  font-size: 130%;
  color: #333;
  background: url(img/h3-bg.gif)
    no-repeat top left;
  border-bottom: 1px solid #E0CFAB;
}
```

Bulletproof Web Design, Dan Cederholm

slide 12

To get the line to appear in the right place, we simply create it as the bottom border of the heading.

This example uses a technique (and the exact same code and graphic) described in Dan Cederholm's book, *Bulletproof Web Design* (although the text is borrowed from Google's language preferences). This is significant! Dan is not writing about internationalization per se – he is more concerned with people pumping up the text size for accessibility reasons. It just so happens, however, that the same approach helps with localizability.

This is an example that you don't necessarily have to learn new information to deal with internationalization issues – just following existing best practices can be the key in many cases. Note again, however, that we are still talking about the **design and development** of content – not about work that the localizers will do!

Dan's book contains several other recommendations that will benefit internationalization.



When creating tabs, don't draw the background as a single graphic.



Instead use something like Doug Bowman's 'sliding doors' technique (<http://alistapart.com/articles/slidingdoors/>). This allows the graphic part of the tab to expand with the text.

Text size

Tabs

W3C[®]

首頁 你的足跡 管理 自己人 小組

Home Tu Organizza Contatti Gruppi

↓


首頁 你的足跡 管理 自己人 小組

Home Tu Organizza Contatti Gruppi

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

Note, again, that this also aids accessibility. The slide shows the result of increasing text size on the tabs.

One other thing to note, however, is that you should leave some space for expansion to the right of the tabs (or left in a right-to-left script). You may otherwise need to design the CSS so that the tabs can be wrapped and still look good.

Text size
Round cornered boxes


The value of internationalization

Internationalization significantly affects the ease of the product's localization. Retrofitting a linguistically- and culturally-centered deliverable for a global market is obviously much more difficult and time-consuming than designing a deliverable with the intent of presenting it globally. (Think back to the Y2K effort and trying to "undo" two-character year fields that were built on the assumption of "19xx").

So ideally, internationalization occurs as a fundamental step in the design and development process, rather than as an afterthought that can often involve awkward and expensive re-engineering.

further reading

- W3C i18n resources relating to internationalization

Author: Richard Ishida, W3C, Susan K. Miller, Boeing

Content first published 2005-12-05. Last substantive update 2005-12-05 GMT. This version 2006-10-19 19:25 GMT

For the history of document changes, search for [qac18n](#) in the i18n blog.

Copyright © 2005-2006 W3C® (MIT, ERCIM, Keio). All Rights Reserved. W3C liability, trademark, document use and software licensing rules apply. Your interactions with this site are in accordance with our [public](#) and [Member privacy statements](#).




Tell us what you think (English):

[Send us a comment](#)

Subscribe to an RSS feed:

[New resources](#)

[Home page news](#)

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 16


In the W3C's i18n articles we use a box with rounded corners.

W3C®


Text size


Round cornered boxes

Tell us what you think (English).


 **Send us a comment**

Subscribe to an RSS feed.


 **New resources**


 **Home page news.**

Powiedz nam co myślisz (po angielsku)


 **Wyślij komentarz**

Prenumeruj kanał RSS


 **Nowe źródła**


 **Wiadomość ze strony głównej.**

قل لنا رأيك (بالإنجليزية).


 **أبعث تعليق.**

الإنضمام إلى خدمة RSS.


 **موارد جديدة.**


 **الصفحة الرئيسية للأخبار.**

Выскажите своё мнение (по-английски).


 **Напишите нам**

Подпишитесь на ленту новостей RSS.


 **Новые публикации**


 **Новости главной страницы.**

告诉我们您的想法 (用英语) 。

 **将您的评论发送给我们**

订阅 RSS 摘要

 **新的资源**

 **主页新闻.**

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 17

This will be much easier to do when CSS3's rounded corners properties are available. In the meantime, people are using background graphics.

This box can use techniques similar to those described earlier to allow the box to contract or expand (up to a certain point) horizontally as well as vertically, to fit the text.

Text size

Round cornered boxes

W3C®

Tell us what you think (Feedback)

Send us a comment

Subscribe to an RSS feed

New resources

Home page new

Выскажите свое мнение (по-английски)

Напишите нам

Подпишитесь на ленту

Новые публикации

Новости главной страницы

قل لنا رأيك

الإنضمام إلى

مصادر

الصفحة

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 18

This shows the graphics used.

Text size
Round cornered boxes

English:
Input processing features

German:
Eingabeverarbeitungsfunktionen

Eingabeverarbeitungsfunktionen

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

One potential issue here is that some languages have very long words, which may not fit within the specified max-width.

Text size
Round cornered boxes

English:
Input processing features

German:
Eingabeverarbeitungsfunktionen

German:
Eingabe-verarbeitungsfunktionen


Eingabe&00AD;verarbeitungs&00AD;funktionen

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

An ideal solution to this would be to get the translator to insert soft-hyphens at appropriate places in the word. The word would then wrap to fit the box.

Unfortunately, although this works in browsers like IE and Opera, the soft hyphen is not yet recognized in Mozilla browsers (although it should do in version 3).

Until it is available widely as a solution, the translator will probably need to break the text manually, and the designer should try not to make such boxes too narrow.

What this means to me

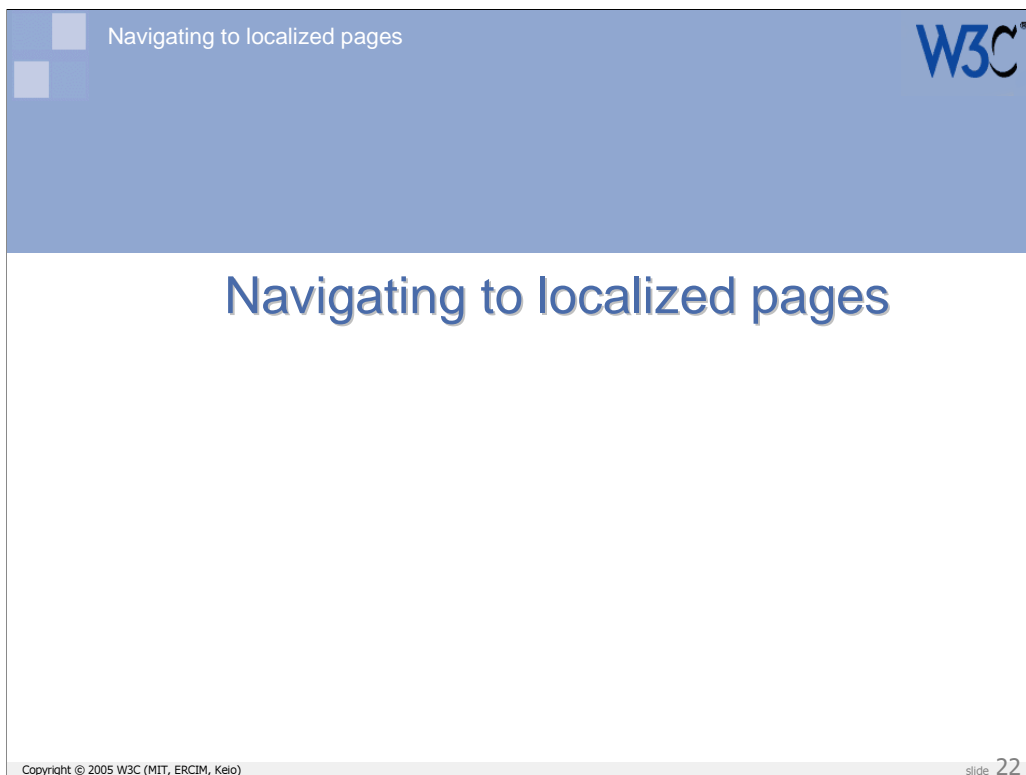
Some lessons from this section

- Think about how to cope with text that expands in translation
- Avoid placing labels to the side of form entry fields, icons, etc, if you can
- For text backgrounds use large images and 'sliding door' effects to allow coverage of text whatever the size
- Leave expansion space at the end of lines of tabs
- Consider how to deal with long compound nouns: don't make the column too narrow; warn translators
- Look out for and lobby the W3C CSS Working Group to publish rounded borders and multiple backgrounds parts of CSS3
- Lobby Mozilla to support the soft-hyphen character

<http://www.w3.org/International/questions/qa-resizing-backgrounds>

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

This slide summarizes some of the practical takeaways from this presentation.



The slide features a blue header bar with the text "Navigating to localized pages" on the left and the W3C logo on the right. The main content area is white with the title "Navigating to localized pages" in blue. The footer is a light gray bar containing copyright information on the left and "slide 22" on the right.

Navigating to localized pages

W3C®

Navigating to localized pages

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

In this first section we will look at a few ways in which languages differ, and then see those differences causing practical issues for localization where the developer/designer has not thought about internationalization.

Navigating to localized pages

Language links



Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 23

Note, in passing, an issue related to the Google text I used in the previous example. The dialogue allowed you to select a different language for the user interface from a pull-down list, presumably assuming that your reason for changing was that you couldn't read the current language.

The issue for me is that the names of all the languages are in the language of the current page. Let's assume, for example, that a curious person wanted to see what the interface looked like in Persian, so they selected that language from the list and clicked on the 'Save Preferences' button.

Navigating to localized pages

Language links





Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 24

Assuming that they would be able to find their way back to the appropriate dialogue box to get back to English (which would require them to remember which link to hit on the thankfully uncluttered Persian Google home page), that they can remember which is the required select list, and that they can do so in spite of the mirror-imaging of the page when using Arabic script, they would then be faced with what you see on the next slide.

Navigating to localized pages
Language links
W3C[®]



Copyright © 2005 W3C (MIT, ERCIM, Keio)
Google 2005D
slide 25

Note that the names of languages are all in Persian, and are sorted by Persian rules.

Which selection would get you back to English ??

(Hint: if you want to explore like this, use a different tab or window for your explorations, and leave the original dialogue available in another for when you want to reset to your current language.)

Of course, the point is really that a Persian person taken to the English site may have as much trouble finding their way to the appropriate user interface language as the curious explorer does in getting back. In my opinion it would help a great deal to write each language name in its own script and language. You can read more about that in the W3C Internationalization article at <http://www.w3.org/International/questions/qa-navigation-select>.

Navigating to localized pages
Pull down lists

The screenshot displays the Nortel website's localization interface. A pull-down menu titled 'CHANGE LOCATION' is open, showing a list of countries and regions. Each entry includes the country name in English followed by its native script in parentheses. The list includes: CHINA (中国), CIS, COLOMBIA, DENMARK (DANMARK), EASTERN EUROPE, ECUADOR, FINLAND (SUOMI), FRANCE, GERMANY (DEUTSCHLAND), GREECE, HONG KONG, INDIA, INDONESIA, IRELAND (REPUBLIC OF), ISRAEL (יִשְׂרָאֵל), ITALY (ITALIA), JAPAN (日本), KOREA (韓国), LUXEMBOURG, and MALAYSIA. The website header features the Nortel logo, navigation tabs for SOLUTIONS, PRODUCTS, SERVICES, and SUPPORT, and a search bar with a 'LOG IN' button. The main content area includes a 'UNIFIED COMMUNICATIONS' banner with the tagline 'it's as simple as' and several promotional tiles: 'HYPERCONNECTIVITY' (We make managing it simple), 'EXCLUSIVE OFFER' (The Value of Unified Communications), 'PARTICIPATE' (Phil Edholm, Enterprise, blogs on thoughts on the business), and 'CASE STUDIES' (The Economist Group reports: A cost-effective, high-performance, converged network provides the flexibility to add new advanced services while helping staff collaborate and meet critical deadlines).

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 26

One way of allowing users to access an alternative language or country site is to use pull-down lists (with the select element in HTML).

Here is an example. My personal preference would be to use the native translation of the country name first and put the English version in parentheses after. It can be useful, however, to have both as shown here.

Navigating to localized pages
Pull down lists

The screenshot displays the Nortel website's internationalization interface. A dropdown menu for 'CHANGE LOCATION' is open, showing a list of countries and regions. The website layout includes a top navigation bar with the Nortel logo and a search bar. Below this, there are sections for 'SOLUTIONS', 'PRODUCTS', 'SERVICES', and 'SUPPORT'. The main content area features a large banner for 'UNIFIED COMMUNICATIONS' with the tagline 'it's as simple as'. Below the banner, there are several promotional tiles: 'HYPERCONNECTIVITY' with the tagline 'We make managing it simple.', 'EXCLUSIVE OFFER' for a white paper on 'The Value of Unified Communications', 'PARTICIPATE' with a blog post by Phil Edholm, and 'CASE STUDIES' featuring a report from The Economist Group. The footer contains copyright information for 2005 W3C (MIT, ERCIM, Keio) and a slide number 27.

A potential issue with this approach is that the user may not have fonts to cover all the characters, resulting in something like the boxes shown above.

You may feel that this is acceptable, since the English text alongside gives sufficient cues to the reader about what those boxes are about.

On the other hand, some people may feel that this is not acceptable.

Navigating to localized pages
Selection pages

W3C®

Search | A-Z Index | Feedback | Help | Languages

Home | Languages

Choose a Language

▪ Bahasa Indonesia ▪ Hmoob ▪ Polski ▪ Vosa vakaviti ▪ 廣東話
 ▪ Česky ▪ Hrvatski ▪ Português ▪ Ellhniká ▪ 日本語
 ▪ Dansk ▪ Ilokano ▪ Reo Tahiti ▪ Български ▪ 한국말
 ▪ Deutsch ▪ Italiano ▪ Română ▪ Монгол
 ▪ Diné bizaad ▪ Kiribati ▪ Slovenčina ▪ Русский
 ▪ Español ▪ Kreyòl Ayisyen ▪ Slovenščina ▪ Українська
 ▪ Eesti keeles ▪ Lea Fakatonga ▪ Suomi ▪ ภาษาไทย
 ▪ Français ▪ Magyar ▪ Svenska ▪ བོད་སྐད་ཀྱི་སྐད་
 ▪ Gagana Samoa ▪ Nederlands ▪ Tagalog ▪ ภาษาไทย
 ▪ Hiligaynon ▪ Norsk ▪ Tiếng Việt ▪ 國語

© 2005 Intellectual Reserve, Inc. All rights reserved. [Rights and use information](#). [Privacy policy](#)

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

An alternative approach is to route the user to a selection page like this. On such a page, which is in HTML, you can resolve the missing font issue by presenting the text in graphic form.

I would not normally recommend using text in graphics, but there shouldn't be a big issue here, given that this text doesn't need translating and is perhaps less likely to be searched for.

One thing to bear in mind, however, is that you should route a user to an alternative language and (often) country site that doesn't remove them from their current context, ie. don't force them up to the home page if they entered the site from a link. If they go to a selection page like this, you should return them to the same page (if such exists) in the localized site.

W3C®

Navigating to localized pages

Using alternative selection lists

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

This example shows a hybrid approach, where what looks like a pull-down list has been implemented using javascript and HTML. Although they haven't done so in this case, this would allow you to use graphics in the pull-down.

W3C[®]

 Navigating to localized pages
Chatty selection sites

Country Sites

Either the webapp context your region JSP URI refers to doesn't exist or the calling web application does not have permission to access another web application context. If the web application exists in your application server, consult the Trouble Shooting Guide for details on how to configure your application server to support cross context forwarding.

American Samoa ▪ English	Estonia ▪ Estonian	Mexico ▪ Spanish ▪ English	Singapore ▪ English
Argentina ▪ Spanish	Fiji ▪ English	Netherlands ▪ Dutch	Slovak Republic ▪ Slovak
Australia ▪ English	Finland ▪ Finnish	New Zealand ▪ English	Slovenia ▪ Slovenian
Austria ▪ German	France ▪ French	Nicaragua ▪ Spanish	South Africa ▪ English
Belgium ▪ Dutch	French Polynesia ▪ English	Niue ▪ English	South Korea ▪ Korean


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

If you take users to a selection page, don't be too chatty on that page. If they are going to it because they don't speak the current language, it won't be a lot of help.

W3C®

Navigating to localized pages issues

Identifying the list



Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 31

In this example, the link to the selection page (though it could also be the location of a pull-down list) is not easy to spot from the text. The designer has, however, used a graphic – ie. a non-verbal cue – to help users locate the appropriate place.

W3C®

Navigating to localized pages

Using maps




The screenshot shows the Xerox website interface. At the top, there's a navigation bar with the Xerox logo and a 'Change Country' link. Below this, a search bar and a list of product categories (Office Equipment, Production Equipment, Document Outsourcing, Supplies, Support & Drivers) are visible. The main content area is titled 'Xerox Websites' and features a world map. Below the map, there are two sections of localized links: 'North America' with links for Canada (English and French) and the United States of America, and 'Latin America and Caribbean' with links for Argentina, Bermuda, Brazil, Cayman Islands, Chile, Colombia, Ecuador, Mexico, Nicaragua, Paraguay, Peru, Trinidad & Tobago, Uruguay, and Venezuela.

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 32

One issue that is difficult to resolve is how to order the list of languages or countries. You could use the Unicode Collation Algorithm, sort by population size or GNP, sort by BCP 47 language tags, etc. to be neutral, although many users would possibly not understand the rationale for such ordering.

For **country-specific** links it may help to use maps of regions to narrow the search to a list that may be easier to scan.

Don't use flags for language-specific selections.

What this means to me

Some lessons from this section

- For links to alternative language or country sites, translate link text into the language of the content pointed to
- Make it easy to find the current page when changing language or site
- Consider whether to use graphics or text, given that fonts may not be available on the user's platform for all options
- If you use text, you'll need the page encoding to be UTF-8
- Remember that pull-down lists don't support graphics, so consider alternatives
- Find a way to draw attention to a list that is recognizable regardless of language, and put it in a visible location
- Consider maps as an alternative selection mechanism

<http://www.w3.org/International/questions/qa-navigation-select>

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

This slide summarizes some of the practical takeaways from this presentation.



Composite messages

W3C®

Composite messages

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 34

This slide features a blue header bar with the text 'Composite messages' on the left and the W3C logo on the right. The main content area is white and contains the title 'Composite messages' in a large blue font. The footer is a light gray bar with copyright information on the left and 'slide 34' on the right.




Composite messages
Syntax differences




Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 35

This shows four different ways of writing one idea. In each case the order of 'words' and the number of 'words' is different.



Composite messages
Agreement & Concept mappings




On


Conectado -os -a -as

Encendido -os -a -as

Activado -os -a -as

This slide shows how the English word 'On' can map to three different words in Spanish. And then there are the masculine, feminine and plural forms of agreement that change the shape of the word according to its context.


Composite messages



Agreement of Russian plurals

Number of pages	Russian word for 'page'
1	страница
2-4	страницы
5-10	страниц
11-20 (irregular)	страниц
21	страница
22-24	страницы
25-30	страниц
> 30	Repeat pattern of endings for 1 to 10.

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 37

In Russian there is a complex plural system. Apart from the irregular teens, the word endings are applied in a rotating way.

W3C® Resource **Markup Validation Service** v1.7.2

Home About News Docs Help & FAQ Feedback

Result: Failed validation
There were 266 validation errors in the file myFirst.html

File: webdeveloper_localhost_1159293639132.html
Modified: (undefined)
Server: Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.8.0.7) Gecko/20060909 Firefox/1.5.0.7
Size: (undefined)
Content-Type: text/html
Encoding: utf-8
Doctype: XHTML 1.0 Transitional
Root Namespace: http://www.w3.org/1999/xhtml

Note: The Validator XML support has some limitations.

This page is not Valid XHTML 1.0 Transitional!

Below are the results of checking this document for XML well-formedness and validity.

1 **Error** Line 14 column 103: end tag for "br" omitted, but OMITTAG NO was specified

... 16:16:10 rishida: Exp: <p>p</p>
/body >

You may have neglected to close an element, or perhaps you meant to "self-close" an element, that is, ending it with ">" instead of ">".

2 **Error** Line 14 column 93: start tag was here


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

For this slide we imagine that the W3C Validator is altered slightly so that it tells you how many validation errors are in your file. It will do this using a 'composite message' whose parts are assembled using PHP code as the page is served.

Although we use PHP for these examples, the concepts can be applied to other scripting or coding environments.

Composite messages
Handling variables

W3C®



There were 268 validation errors in the file myFirst.html.

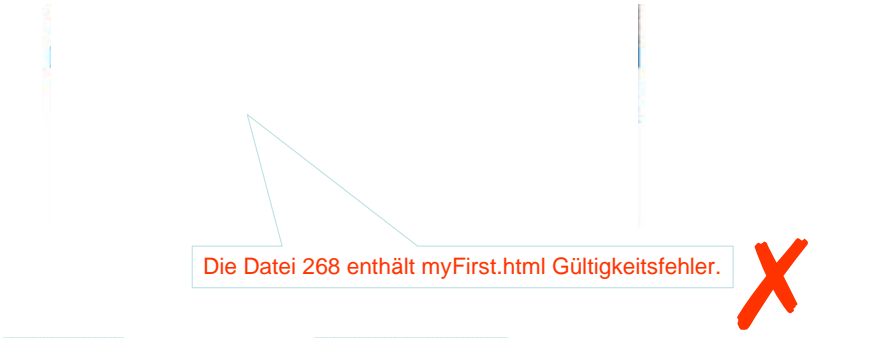
Die Datei myFirst.html enthält 268 Gültigkeitsfehler.

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 39

In the German translation, the order of the two variables may need to be changed.

Composite messages
Handling variables

W3C



Die Datei 268 enthält myFirst.html Gültigkeitsfehler. **X**

```
echo "There were ".$errors." validation errors in  
the file ".$filename."";
```

```
echo "Die Datei ".$errors." enthält ".$filename."  
" Gültigkeitsfehler.";
```

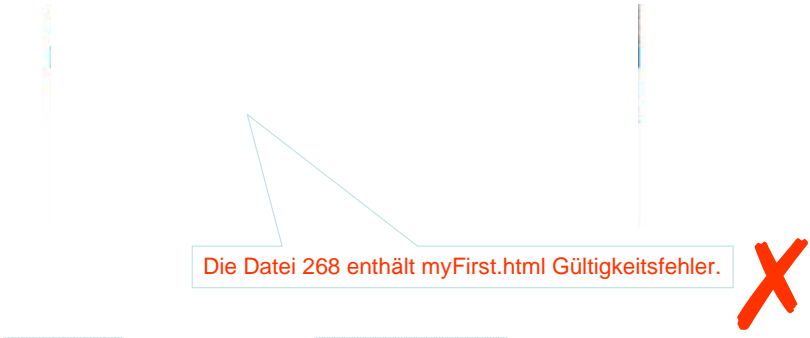
Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 40

Typically translators have no access to the actual code, to avoid them introducing bugs into the page. Either the text is extracted or a translation tool masks the code.

Although we are fortunate that we were able to add words after the second variable, due to the English string containing a period, this still didn't produce the right result. The German reads "File 268 contains myFirst.html validation errors."

Composite messages
Handling variables

W3C



Die Datei 268 enthält myFirst.html Gültigkeitsfehler. X

```
echo "There were ".$errors." validation errors in  
the file ".$filename."";
```


```
echo "Die Datei ".$errors." enthält ".$filename."  
" Gültigkeitsfehler.";
```

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 41

The reason is that the translation process didn't switch the order of the variables.

Composite messages
Handling variables

W3C



```
printf(
    "There were %d validation errors in the file %s.",
    $errors, $filename )

printf(
    "Die Datei %s enthält %d Gültigkeitsfehler.",
    $errors, $filename )
```


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

So next we try using a printf statement. This has the benefit that text and variable locators all sit within a single string, and the translator can access the items they want to reorder.

Unfortunately, this doesn't help, since PHP still replaces the variables in the string in the order of the variables cited in the following parameters to printf. This causes the 268 to be shown instead of the filename, by converting the integer value to a string. It is unable to find an integer value in the file name, and so presents us with the zero for the number of errors.

Composite messages
Handling variables

W3C®



```
printf(
    "There were $errors validation errors in the file $filename."
)

printf(
    "Die Datei $filename enthält $errors Gültigkeitsfehler."
)
```

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

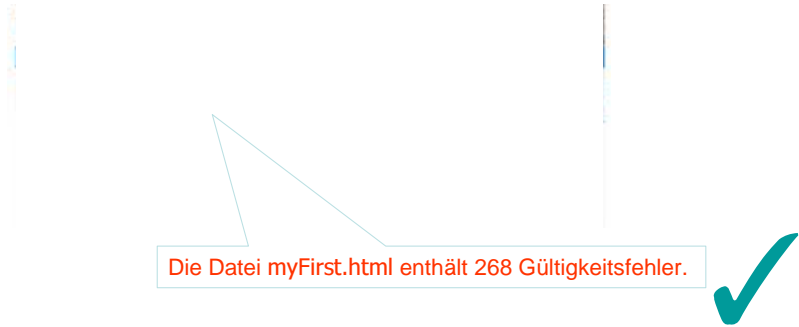
By embedding the variable names directly in the printf string, as shown in this slide, we finally achieve the desired result in German.

Nota bene: Successful, or at the very least, cost effective localization in this case is down to the **designer/developer** understanding the potential pitfalls of various approaches to coding. It is not the job of the localization vendor to get this right. It needs to be done as the initial content is created!

You should also be very careful of the assumption that 'This doesn't affect me, since we don't translate the content I develop.' I have seen many, many cases where the thing being developed was later so successful that people wanted to take it to other regions, only to find that they ran into major difficulties because of issues with the translatability of the code or content. It's best to just do it right from the start.

Composite messages
Handling variables

W3C



```
printf(
    "There were %1$d validation errors in the file %2$s.",
    $errors, $filename )


printf(
    "Die Datei %2$s enthält %1$d Gültigkeitsfehler.",
    $errors, $filename )
```

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 44

By the way, there is a way to produce the right effect while using the %d and %s variable markers in a PHP string, but it involves a slightly more complex syntax. This is shown in the above slide. The numeric markers refer to the relevant variable in the parameters that follow the string, even after reordering.

Composite messages
Handling variables

W3C



1 ошибка
2 ошибки
5 ошибок

При проверке файла myFirst.html была обнаружена 268 ошибка.

```
printf(
    "При проверке файла $filename была обнаружена $errors
    ошибка."
)

printf(
    "Файл: $filename. Количество ошибок при проверке: $errors."
)
```


Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 45

So now we know how to code this type of text in PHP... or do we?

Let's think back to our example of how plurality works in Russian, and we realize that we still have a problem for that language. We only have a single string and it can only be translated one way – yet the Russian requires three variants of the word ошибка, depending on the number that precedes it.

Composite messages
Handling variables

W3C®



1 ошибка
2 ошибки
5 ошибок

Файл: myFirst.html. Количество ошибок при проверке: 268.

```
printf(
    "При проверке файла $filename была обнаружена $errors
    ошибка."
)

printf(
    "Файл: $filename. Количество ошибок при проверке: $errors."
)
```

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 46

To deal with this, the Russian translator would probably resort to a completely different structure for the text, essentially equivalent to "File: X. Validation errors: Y". This approach requires only one form of ошибка in the invariable string. This is an example of what I call a 'topic-comment' composite message.

Composite messages

Sentence-like composite messages

W3C®

The < > is on.

printer stacker

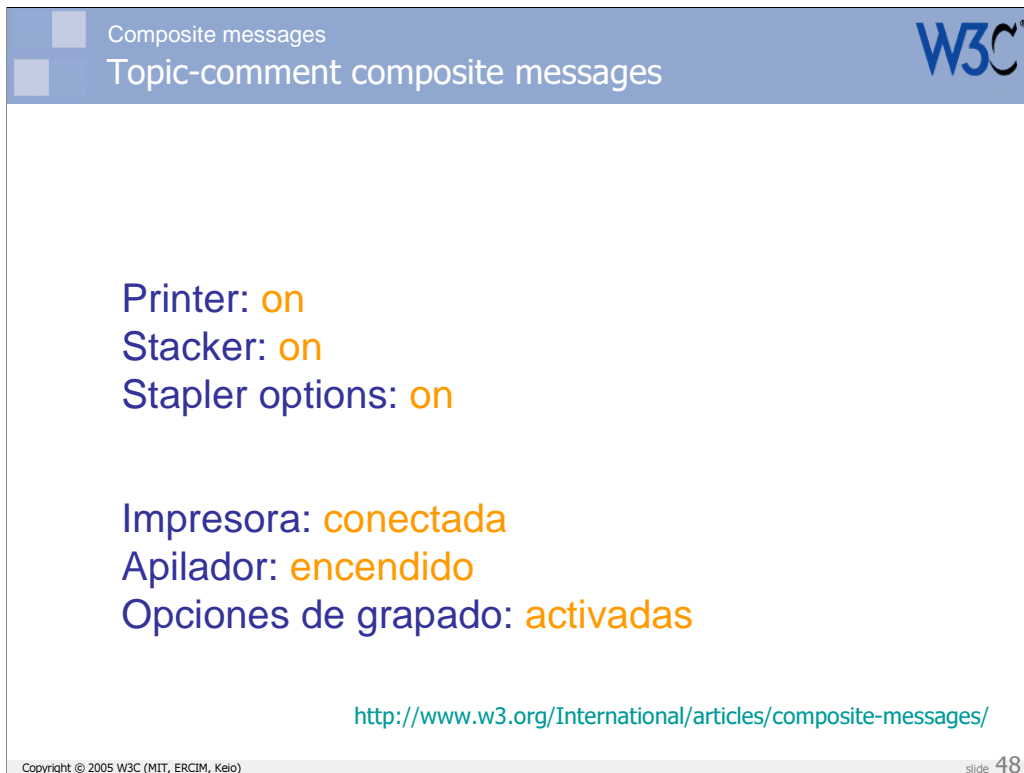
 stapler options

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 47

So we are beginning to see here that there are two distinct types of composite message. The first is based on a sentence-like approach, and the invariant string can be difficult to translate in some circumstances because of the need for agreement or different word mappings.

In the example above, 'The' should be translated 'el', 'la', or 'las' in Spanish, depending on what word follows it. Also the word 'on' should be translated using three different Spanish words (with different endings).



The slide features a blue header bar with the text 'Composite messages' and 'Topic-comment composite messages' on the left, and the W3C logo on the right. The main content area is white and contains two sets of text. The first set is in English: 'Printer: on', 'Stacker: on', and 'Stapler options: on'. The second set is in Spanish: 'Impresora: conectada', 'Apilador: encendido', and 'Opciones de grapado: activadas'. A URL is provided at the bottom: <http://www.w3.org/International/articles/composite-messages/>. The footer contains copyright information and the slide number 48.

Composite messages
Topic-comment composite messages

W3C

Printer: on
Stacker: on
Stapler options: on

Impresora: conectada
Apilador: encendido
Opciones de grapado: activadas

<http://www.w3.org/International/articles/composite-messages/>

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

The other approach to designing composite messages is what I like to call the 'topic-comment' approach: you state a topic, then you say something about it.

This approach works much better for the previous slide, since you each comment you associate with a topic can use a different word with the appropriate word endings.

There is a little more to this theory of composite messages than we have mentioned so far, but you can get more information from the W3C Internationalization site at the following URI:
<http://www.w3.org/International/articles/composite-messages/> .

W3C[®]
Composite messages
Text string re-use

Connectado	-os -a -as
Encendido	-os -a -as
Activado	-os -a -as

```

graph LR
    On[On] --> Printer[Printer: ]
    On --> Stacker[Stacker: ]
    On --> Stapler[Stapler options: ]
    
```

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

I should, however, mention just one other point. Many designers/developers looking at the English topic-comment arrangement on the previous slide might think to themselves that they could save a little bandwidth by reducing all those instances of the word 'On' to a single string that is used for all comments, ie. they want to re-use strings.

Composite messages
Text string re-use

W3C

```
<p>Printer: <?php print getStatus($printer);?></p>
<p>Stacker: <?php print getStatus($stacker);?></p>
<p>Stapler options: <?php print getStatus($sOptions);?></p>

<?php
function getStatus (device) {
    if ($device.status != 0) { return 'On'; }
    else { return 'Off'; }
}

?>
```


<http://www.w3.org/International/articles/text-reuse/>

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

Tempting as this idea may appear, it will unfortunately introduce insurmountable problems for translation, since the comment is likely to require different agreement forms at the least, and possibly different words altogether, depending on the context.

This slide shows an example of how such a problem may come about by returning the same text from a function for each comment.

Note that I do not want to rule out string re-use altogether – there are situations where it is a sensible approach. But re-use must not occur across different contexts. For more information about this, see the W3C Internationalization article at <http://www.w3.org/International/articles/text-reuse/>.

What this means to me
Some lessons from this section

- Use topic-comment arrangements for composite messages, if you can, rather than sentential arrangements
- Use coding approaches that allow syntactic flexibility when creating composite messages
- Be careful about reuse of text strings – ensure that the context of reuse is always the same


Copyright © 2005 W3C (MIT, ERCIM, Keio)slide 51

This slide summarizes some of the practical takeaways from this presentation.


The presentation is not designed to give you a thorough overview of potential internationalization and localization issues. It aims to provide you with a few practical takeaways, but more importantly it aims to get you thinking about what internationalization is all about – to take designers out of their comfort zone, and help them realize that if you want your content to wow people outside your own culture and language, you need to build in certain flexibilities and adopt certain approaches during the design and development – not as an afterthought. Otherwise you are likely to be creating substantial barriers for worldwide use.

The presentation also aims to show that, although using Unicode is an extremely good start to making your stuff world-ready, using a Unicode encoding such as UTF-8 throughout your content, scripts and databases is only a start. You need to worry about whether translators will be able to adapt your stuff linguistically, but you also need to also consider whether graphics and design are going to be culturally appropriate or can be adapted, and whether your approaches and methodologies fit with those of your target users.

Getting more information
W3C can help



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




The screenshot shows the W3C Internationalization Activity page. It features a sidebar with navigation links like 'Getting started', 'About the Activity', 'Working Groups', 'Topic Index', 'Technical Index', 'Resource Types', and 'RSS/Atom/Translators'. The main content area is titled 'W3C Internationalization (I18n) Activity' and includes sections for 'Recent highlights' and 'Setting language preferences in a browser'. The right sidebar contains 'Quick Categories' and 'Quick Archives'.

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 52

The W3C is trying to provide useful advice at
<http://www.w3.org/International/>

We could always do with help and support for this.



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

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 53