

# Internationalization 101

How to Develop Localization-Ready  
Websites & Applications

# Preparation Prevents Unnecessary Work

## Bad Public Example

» Japanese VOIP software

ppphone (“Pee Pee Phone”)



# Preparation Prevents Unnecessary Work

## Good Public Examples

### » Proctor & Gamble

- Dreck > Dreft



### » Kodak



# Benefits of Internationalization

- » These public examples show how internationalization can save time, money, and embarrassment
- » “Internationalization 101” is going to show you how to save time and money like this “behind the curtain”

# Internationalization 101 Topics

- » About Us
- » Key Definitions
- » 7 Elements of Localizability (Internationalization)
- » 4 Elements of Cultural Internationalization
- » How to Quickly Test Localizability
- » Important Elements to Have in Your Localization-Friendly CMS
- » Q&A



# Adam Wooten



Director of Translation Services at Lingotek



Columnist writing on international business, language and culture at [DeseretNews.com](http://DeseretNews.com)



Adjunct professor teaching a course on translation technology at Brigham Young University



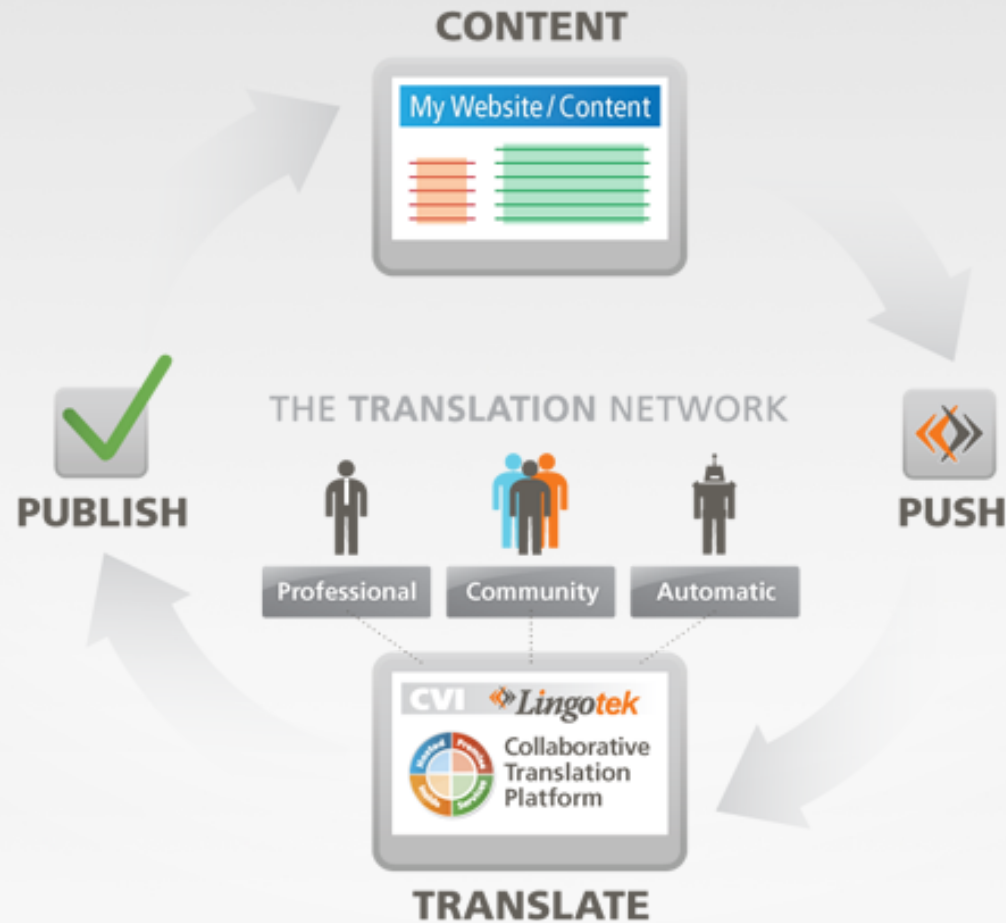
THE TRANSLATION NETWORK



 **Lingotek**

THE TRANSLATION NETWORK

# Lingotek's Collaborative Translation Platform™



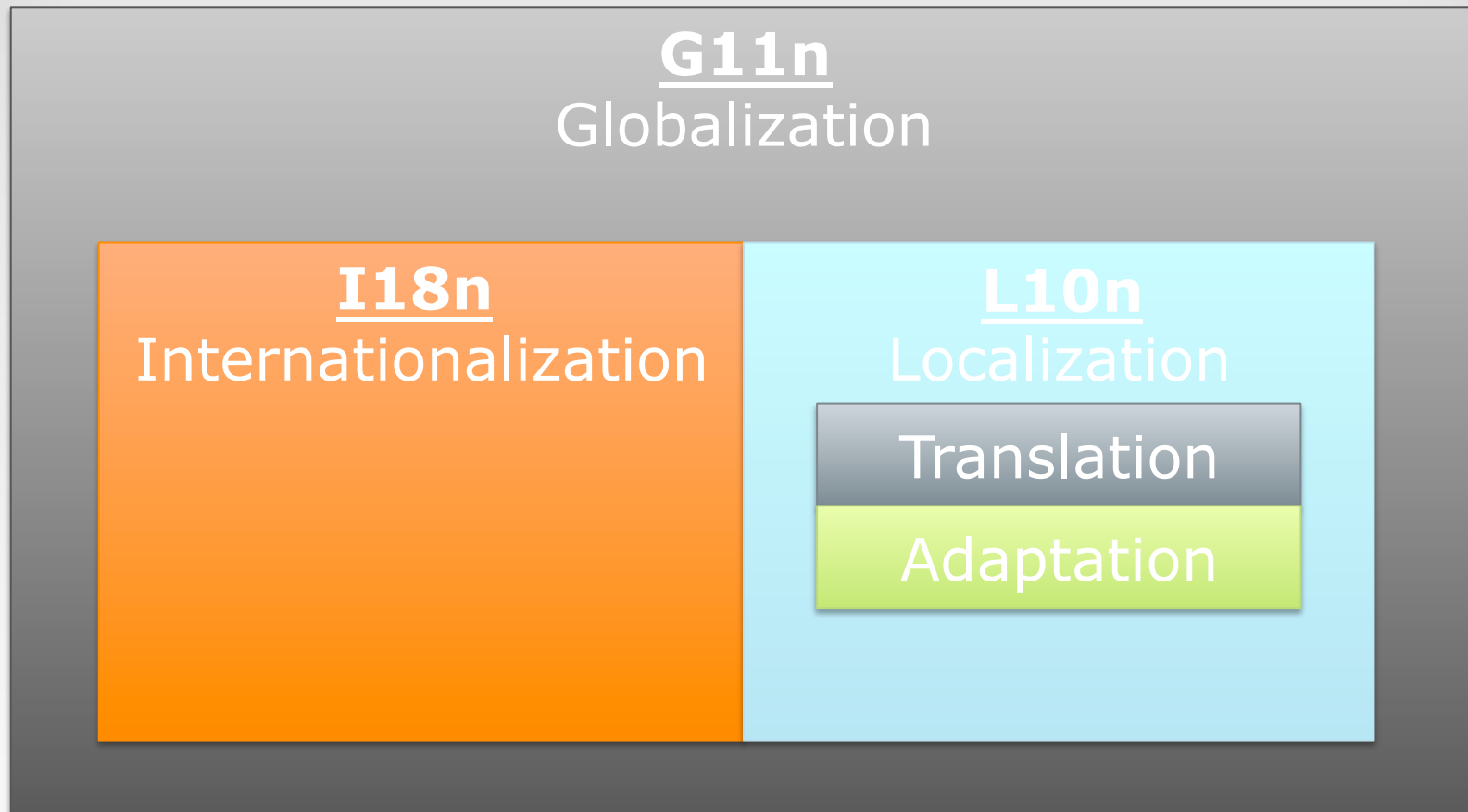
## Lingotek's Collaborative Translation Platform

We give Enterprise business customers new ways to engage their global communities by providing trusted, rapid translations to expand and grow business in a cost effective way.



THE TRANSLATION NETWORK

# Definitions



# Is there evidence that internationalization saves money?

Years of anecdotal evidence show that proper i18n cuts l10n time & cost by 50%

One way to look at it...

**\$1,000** fix in 1 language in internationalization **OR** **\$20,000** fix in 20 languages in localization **+** **\$20,000** fix for each update in 20 languages



Source: Localization Industry Standards Association



THE **TRANSLATION** NETWORK

# 7 Elements of Localizability (Internationalization)

1. Regional settings
2. Character sets
3. Translatable text
4. Over-externalization
5. Text expansion
6. Image text
7. Concatenated strings



# 1. Regional Settings

Can settings for the following be changed according to locale?



\$, €, £, ¥, etc.

**Currency**



e.g. metric

**Measurements**



e.g. decimal/comma

**Numbers**



e.g. day before month

**Times/Dates**



e.g. ZIP vs postal code

**Address/Phone**



A-Z vs character sets

**Sorting Rules**



## 2. Character sets

» Do various character sets display correctly?

Example Set	Example Characters
European	Español, Français, Português, Русский
Double byte & Asian	中文简体 中文繁體 日本語
Right-to-left	עברית العربية

# Incorrect Character display can be embarrassing

BBC TV Series recently displayed the epitaph on a gravestone... backwards & mistranslated



Photo credit: BBC

# Example of Garbage Characters

<b>Spanish</b>	Como esta usted?  
<b>German</b>	Die Informationen können jederzeit ohne vorherige Ankündigung geändert werden  
<b>Czech</b>	Při pokusu o zálohování pomocí některé ze služeb průběžného zálohování svazku (VSS) doslo k chybě.  



# The Solution:

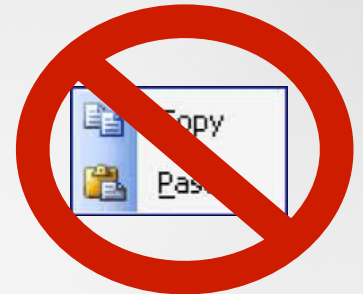
- » Be *AWARE!* Track your encoding!
- » Unicode supports most (90) of the world's scripts (writing systems)
- » Ensure that your CMS is Unicode compliant
- » Check out [Unicode.org](http://Unicode.org) for more info



# 3. Translatable Text

**Is all translatable text externalized into resource files and/or tagged so that it can be extracted with filters?**

- » Reduce/eliminate the manual labor needed for text extraction and replacement
- » Reduce human touch points to reduce human error (changes/omissions)



# For example...

- » Is your web content or software strings stored in external resource files? In XML?
- » Can you export all your website Flash to XML or text files?
- » Or is it nothing but buttons with text that can't be exported?
- » Do you have the original source files for embedded video and audio?
- » Here is how we preserve code and extract text for translation...



## Extraction Example

### Step 1: English XML Source

```
- <scriptedContentFragmentLanguageResources>
- <theme version="6.0.121.20682" name="424eb7d9138d417b994b64bff44bf274">
- <scriptedContentFragment name="Event Calendar - Calendar Title" instanceIdentifier="29aa41eefc8e406b9196f989452c5f22">
- <languageResources>
- <language key="en-us">
  <resource name="CF_Title">Widget Title</resource>
  <resource name="Options">Options</resource>
  <resource name="Calendar_Title_Name">Event Calendar - Calendar Title</resource>
  <resource name="Calendar_Title_Title">Title</resource>
  <resource name="Calendar_Title_Description">Sets and Displays Title for calendar pages</resource>
  <resource name="Home">Home</resource>
  <resource name="ViewMore">...</resource>
  <resource name="BreadCrumb_Seperator">&raquo;</resource>
  <resource name="TagBreadCrumb_Root">All Tags</resource>
  <resource name="calendar_Url_AddEvent">Add Event</resource>
  <resource name="calendar_Url_EditEvent">Edit Event</resource>
  <resource name="calendar_Url_AddCalendar">Add Calendar</resource>
  <resource name="calendar_Url_EditCalendar">Edit Calendar</resource>
  <resource name="calendar_Url_Registration">Registrations</resource>
  <resource name="calendar_Url_AllCalendars">Calendars</resource>
  <resource name="TagBrowser_Title">Browse by Tags</resource>
</language>
</languageResources>
</scriptedContentFragment>
- <scriptedContentFragment name="Event Calendar - Event Comment Add" instanceIdentifier="d9da6dde56d43c7a06a5a0cde0e36e7">
- <languageResources>
- <language key="en-us">
  <resource name="Calendar_EventCommentAdd_Name">Event Calendar - Event Comment Add</resource>
```



## Step 2: XML Code Separated from Translatable Text, (Hidden & Protected)

Targets Segments TM Statistics Tags Notes (0) Settings

Segments Load Template Save Template Clear All

Label	Segment	
resource	Widget Title	Ignore
resource	Options	Ignore
resource	Event Calendar - Calendar Title	Ignore
resource	Title	Ignore
resource	Sets and Displays Title for calendar pages	Ignore
resource	Home	Ignore
resource	&raquo;	Ignore
resource	All Tags	Ignore
resource	Add Event	Ignore
resource	Edit Event	Ignore

Page Size 10 << < 1 2 3 ... 6 > >>



## Step 3: Translation Added

The screenshot displays two panels of a translation tool. The top panel is titled "Source - English" and contains the text: "Widget Title", "Options", "Event Calendar - Calendar Title", "Title", "Sets and Displays Title for calendar pages", and "Home". The bottom panel is titled "Target - Japanese" and shows the same text translated into Japanese: "Widget Title" (ウイジェットタイトル), "Options", "Event Calendar - Calendar Title", "Title", "Sets and Displays Title for calendar pages", and "Home". Both panels have a toolbar with icons for navigation, saving, and editing.

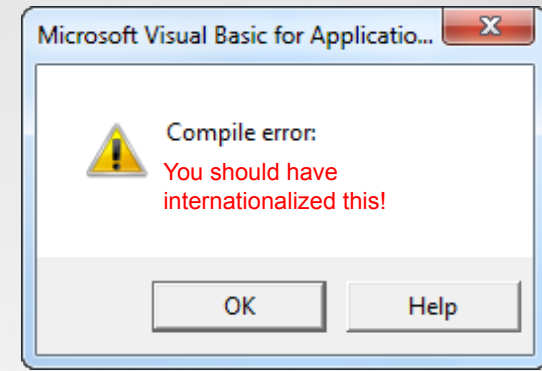
## Step 4: Translation Replaces Original Source Text in Recreated XML File

```
- <scriptedContentFragmentLanguageResources>
- <theme version="6.0.121.20682" name="424eb7d9138d417b994b64bff44bf274">
- <scriptedContentFragment name="Event Calendar - Calendar Title" instanceIdentifier="29aa41eefc8e406b9196f989452c5f22">
- <languageResources>
- <language key="en-us">
  <resource name="CF_Title">みんを代安 明次海要規道む</resource>
  <resource name="Options">ねのわみんを</resource>
  <resource name="Calendar_Title_Name">次海要規道む保 島教有現番販あがきね ネのわみんを代安 明次海要規道む</resource>
  <resource name="Calendar_Title_Title">んを代安明</resource>
  <resource name="Calendar_Title_Description">代安明次海 要規道む 保島教有現番販あがき ねのわみんを 代安明次海要規道む保島 教有現番販あが</resource>
  <resource name="Home">安明次海要</resource>
  <resource name="ViewMore">規道む保</resource>
  <resource name="BreadCrumb_Separator">のわみんを代</resource>
  <resource name="TagBreadCrumb_Root">島教有現 番販あがき</resource>
  <resource name="calendar_Url_AddEvent">ねのわ みんを代</resource>
  <resource name="calendar_Url_EditEvent">販あがきね ネのわみんを</resource>
  <resource name="calendar_Url_AddCalendar">教有現番 販あがきねネのわみん</resource>
  <resource name="calendar_Url_EditCalendar">次海要規道 む保島教有現番販あが</resource>
  <resource name="calendar_Url_Registration">んを代安明次海要規道む保島教有</resource>
  <resource name="calendar_Url_AllCalendars">ネのわみんを代安明</resource>
  <resource name="TagBrowser_Title">ぎねネのわみんを 代安明次海</resource>
</language>
</languageResources>
</scriptedContentFragment>
- <scriptedContentFragment name="Event Calendar - Event Comment Add" instanceIdentifier="d9da6dde56d43c7a06a5a0cde0e36e7">
- <languageResources>
- <language key="en-us">
  <resource name="Calendar_EventCommentAdd_Name">教有現番販あが きねネのわみんを 代安明次海要規 道む保島教有現番販 あがきね</resource>
```



# 4. Over-Externalization

» Avoid over-externalization. Ensure non-translatable strings (i.e. code) are not included with the text extracted for translation



- Translators are best at translation, not programming, so don't let them touch code
- When over-externalized code is translated, it can break the application and prevent the localized version from compiling correctly

# 5. Text Expansion

» Leave room for text expansion, either with existing sizing or dynamic resizing

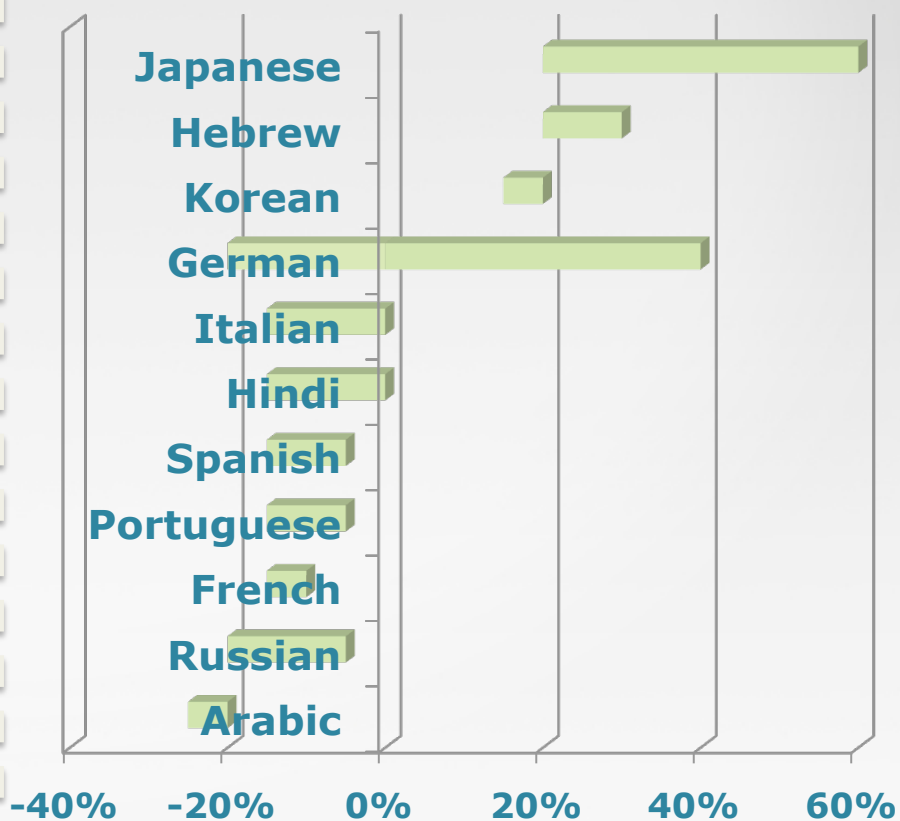
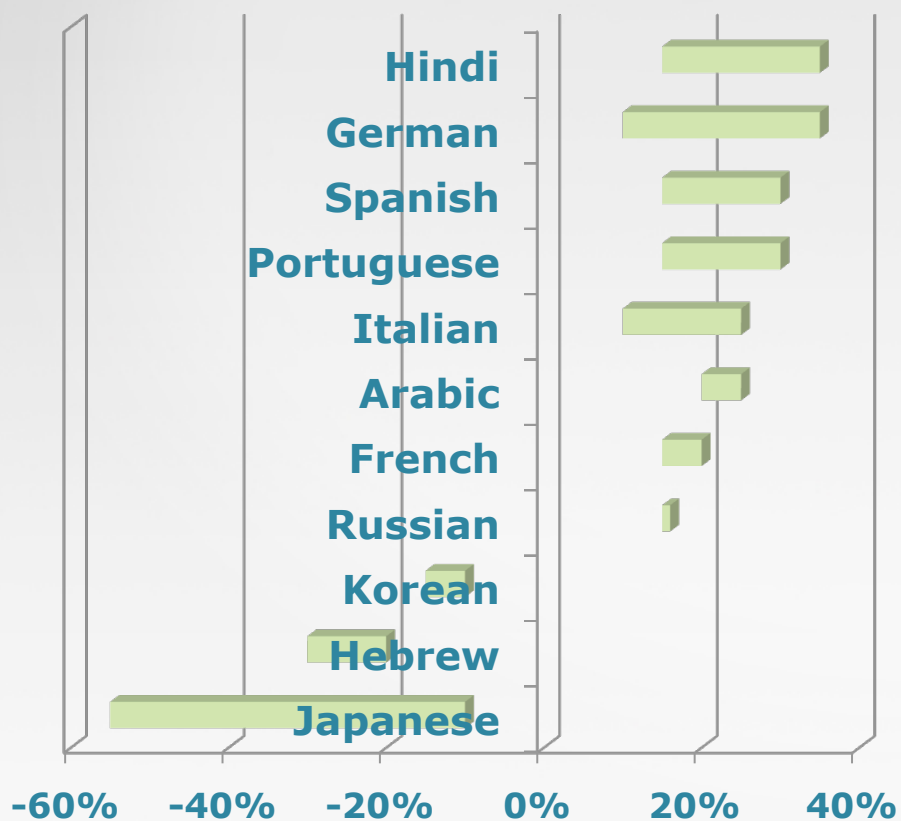
- i.e. Spanish usually expands 20% to 30% longer than English. English text usually expands 20% to 60% longer than Japanese.
- Expansion percentages are even greater for very short strings of text. Strings <7 characters can expand 100% to 300%)
- This applies to audio too – leave time for expansion



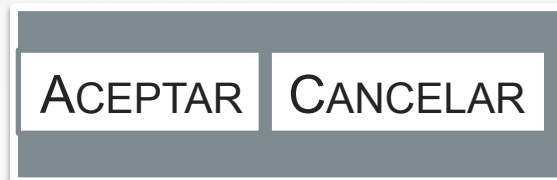
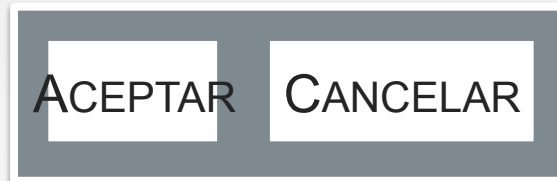
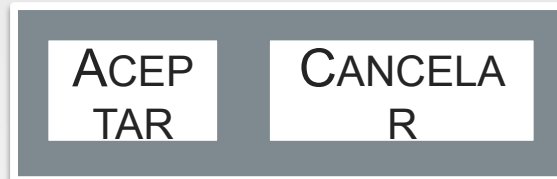
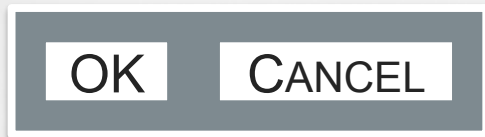
# Common Expansion and Contraction % in Translated Text Length

From English into \_\_\_\_\_

From \_\_\_\_\_ into English

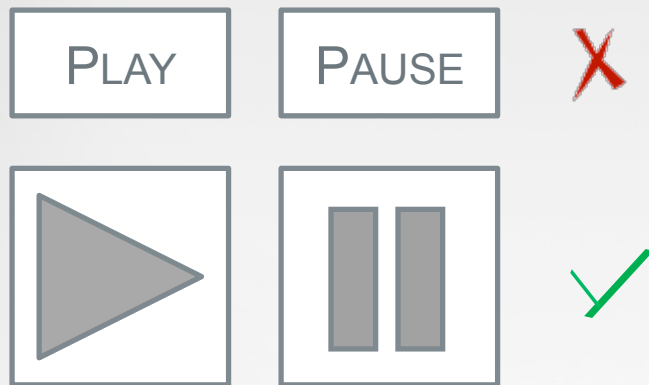


# Bad Examples

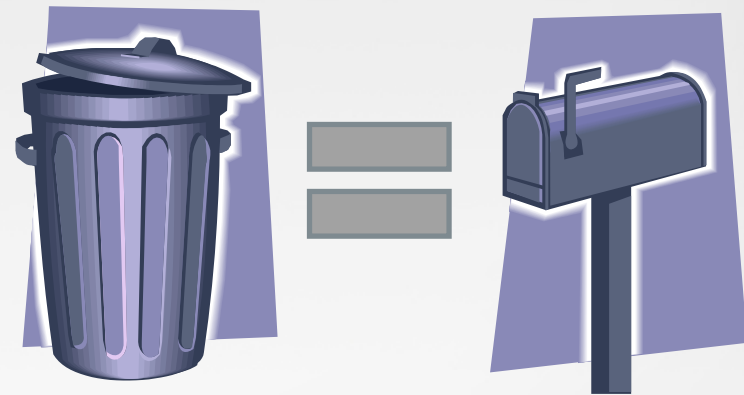


# Universally Understood Images Help Avoid Expansion Problems

## Some Images Are Universal



## Unchecked Images Are Dangerous



## 6. Image Text

- Embedded or hard-coded text requires more time & money for localization
- Layering can still create an embedded appearance

» At a minimum, **SAVE** the editable image files (PSD, AI, Visio, etc.)



# 7. Concatenated Strings

- » String concatenation is the operation of linking two or more character strings end to end
- » Over-the-phone example: dynamic feedback from your bank



# Concatenation Examples

OK	Better!
<p>1: The files will be deleted</p> <p>2: The profiles will be deleted</p> <p>3: one after the other</p>	<p>1: The files will be deleted one after the other</p> <p>2: The profiles will be deleted one after the other</p>

# Concatenation Example including Logic

```
{user_1} approved
```

```
<if about you>
```

```
you
```

```
<else>
```

```
{user_2}
```

```
</if>
```

```
as
```

```
<if about you>
```

```
a
```

```
<else>
```

```
    {user_1}'s
```

```
</if>
```

```
{relationship}.
```

What potential problems might this pose in translation?



# Manual Formatting Can Be as Problematic as Concatenations

Use standard/automatic formatting  
.....and styles instead of manual  
.....formatting. In particular,  
.....manually inserted line breaks  
.....and whitespace can require  
.....expensive reformatting after  
.....translation. ¶



## 7. Concatenated Strings (continued)

- » Many concatenations can cause problems for translators. Translators often need to restructure sentences or indicate word gender/plurality, but that can be awkward or impossible with many variables.
- » Even when accommodating concatenations is “possible,” it is still less cost-effective.
- » Provide complete strings and externalize variables where possible



# Important Elements to Have in Your CMS

- » Proper display of character sets
- » Automatically adjust regional settings
- » Content import/export  
(NO COPY/PASTE)



# Watch Out For Cultural Differences



Cultural  
References

Nontranslatable  
Names



Symbols &  
Images

Colors



# Examples of Cultural References

- » Sports analogies
- » Marketing references to cultural values including:
  - diversity common in the U.S.
  - how Chinese leaders conservatively avoid extremes
- » Advertising references to a “white Christmas”



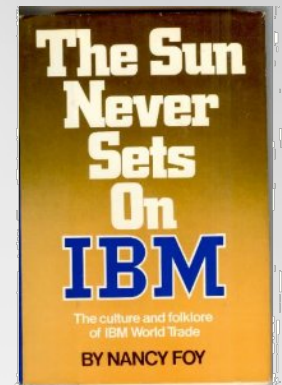
# Cultural References are More Avoidable in Some Content

More Avoidable	In Between / Either	Less Avoidable
Tech documentation	eLearning	Marketing Campaigns
UI strings	Press releases	Advertisements
Product descriptions	Legal content	

Original content may be required when many cultural references are unavoidable

# Cultural references in IBM training...

“How do you like the training stuff we are sending out from New York these days? ...We are putting a lot of work and money into that stuff... How about our September selection of aids for the Fall Kickoff Meeting?”



Page 68

## 3 Cultural Problems in Just the Title:

1. “First of all, the word in English-speaking countries out here would be 'autumn', not 'fall'.
2. “And below the equator, it is coming up to spring, not autumn.
3. “Finally, the word 'kickoff' relates to a uniquely American sport.”

# Example: Sometimes Content Adaptation is Unavoidable

Toronto-based AlertDriving, (Sonic e-Learning Inc.) expanded into 20+ countries in 2005-2007 before realizing the product had serious localization flaws.

- » Lesson dialogue was poorly-translated
- » Content failed to address geographic nuances.

Example: AlertDriving used to teach that the center lane is the safest on a multi-lane highway, but that is untrue in Dubai, where the center lane is used exclusively for passing.



Spent \$1 million over 18 months revamping its existing product line:

- » Language dialects
- » Local driving habits

Source: Maltby, Emily. "Expanding Abroad? Avoid Cultural Gaffes"  
*Wall Street Journal*, Jan 19, 2010

# Nontranslatable Names: International Brand Check

- » Like the examples given at the beginning...
  - PPPhone?
- » Solution = international brand check  
*(survey of linguists & country-specific brand checkers)*



# Microsoft's Bing Brand in China



- » Solution: add 1 extra Chinese character
- » New pronunciation: "**bee-ying**"
- » Origin: last 2 characters of Chinese proverb meaning "ask and you shall find."
- » New direct translation: "certain to respond or ready to answer."

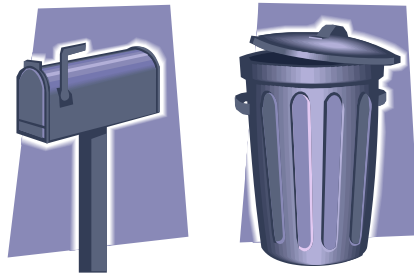
Source: Mehdi, Yusuf. "Microsoft on Bing in China" *The New York Times*, 12 Jun 2009

<[http://www.nytimes.com/2009/06/08/opinion/lweb08soft.html?\\_r=2&scp=3&sq=saul%20gitlin&st=cse](http://www.nytimes.com/2009/06/08/opinion/lweb08soft.html?_r=2&scp=3&sq=saul%20gitlin&st=cse)>

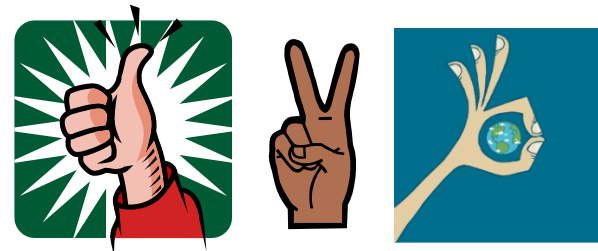


# International Symbol & Icon Check

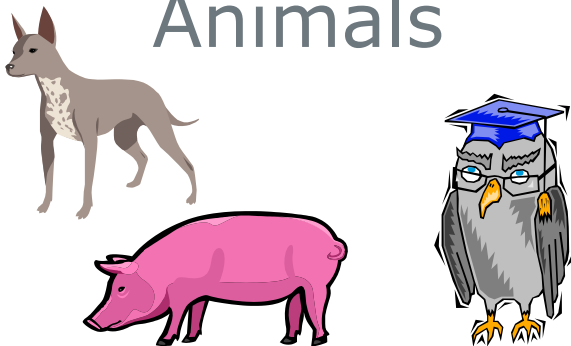
## Region-specific



## Hand gestures



## Animals



Religion, politics,  
race, ethnicity,  
sex, gender,  
slang, humor, etc.

# If You Are Big Enough, Sometimes Cultures Adapt a Little to You...

Facebook still uses an otherwise offensive thumbs-up “like” button in Thailand...

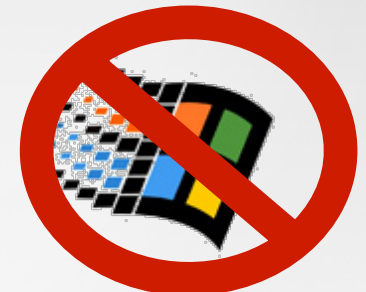


 เป็นคนแรกของคุณที่ถูกใจสิ่งนี้

# Political References: Microsoft Offends India

On 800,000-pixel map of India, Microsoft colored 8 pixels a different shade of green to represent the disputed Kashmiri territory

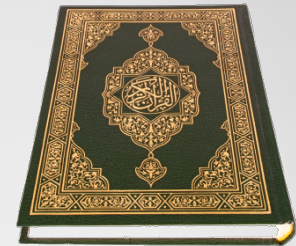
- » Kashmir was shown as non-Indian
- » India banned the Windows 95 software
- » 200,000 copies recalled to heal diplomatic wounds



Source: Edwards, Tom. "Kashmir Blunder: Microsoft Loses Millions" SiliconIndia.com, 23 Aug 2004  
< [http://www.siliconindia.com/shownews/Kashmir\\_blunder\\_Microsoft\\_loses\\_millions-nid-25271.html](http://www.siliconindia.com/shownews/Kashmir_blunder_Microsoft_loses_millions-nid-25271.html)>

# Religious References (Audio): Microsoft Offends Saudi Arabia

- » Chanting of the Quran included on soundtrack of a Microsoft computer game
- » Linguists say, “No! Wait!”
- » Senior management ships it anyway, saying, “Who’s going to notice?”
- » Saudi government bans game and demands official apology
- » 75,000 copies of the game are recalled and destroyed



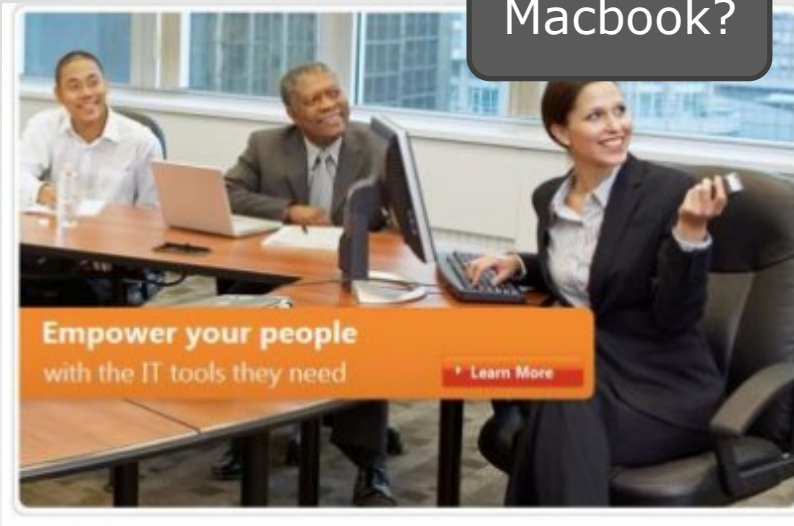
Source: Edwards, Tom. “Kashmir Blunder: Microsoft Loses Millions” SiliconIndia.com, 23 Aug 2004  
< [http://www.siliconindia.com/shownews/Kashmir\\_blunder\\_Microsoft\\_loses\\_millions-nid-25271.html](http://www.siliconindia.com/shownews/Kashmir_blunder_Microsoft_loses_millions-nid-25271.html) >



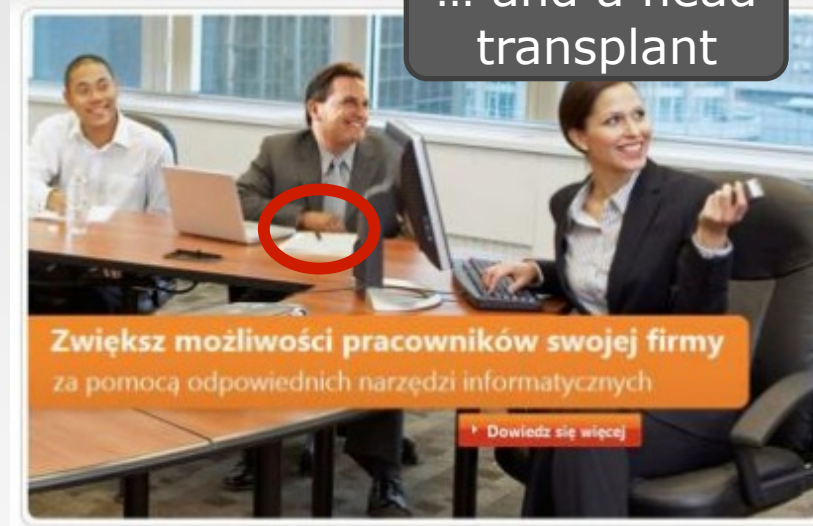
THE **TRANSLATION** NETWORK

# Adapt, but Don't Be Sloppy...

## United States



## Poland



As seen on the Microsoft website,  
poor Photoshop is no substitute for localization



THE TRANSLATION NETWORK

# Religious References

## Microsoft Prevents Image Gaffe

- » Avoided release of Office XP software that contained a moon and stars astrology icon resembling the Islamic Hila symbol.

Source: Edwards, Tom. "Kashmir Blunder: Microsoft Loses Millions" SiliconIndia.com, 23 Aug 2004  
< [http://www.siliconindia.com/shownews/Kashmir\\_blunder\\_Microsoft\\_loses\\_millions-nid-25271.html](http://www.siliconindia.com/shownews/Kashmir_blunder_Microsoft_loses_millions-nid-25271.html)>



THE **TRANSLATION** NETWORK

# International Color Check

## Color Connotations by Culture

» Joy in parts of Asia

VS

Mourning in parts of Africa

» Environmental awareness in U.S.

VS

Danger or forbidden color in parts of Pacific

» Feminine in Western Culture

VS

Appropriate for both genders in Japan

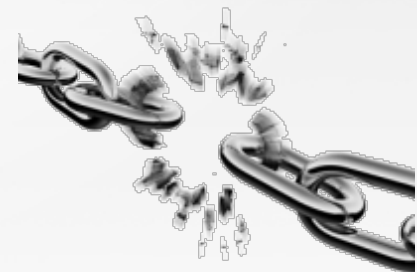
» Considered the "safest" global color



# Content Authors Can Save Time & Cost

## Write with the Translator in Mind

- » Avoid ambiguity
- » Avoid references and allusions that would not be understood by other cultures – where possible
- » Limit concatenations
- » Take advantage of repetition – repurpose existing elements using an authoring tool



# Test Localizability with a Pseudo Localization

- » Pseudo Localization
  - = Pseudo Translation
  - = Dummy Translation
  - = Test Translation
- » Replaces text with test characters to simulate translation
- » Enables quick testing for at least 5/7 elements of localizability



# Elements of Replaceables


- » Automatic replacement of all translatable text
- » Special character categories
  - European {óóóóóó}
  - Double-byte Asian {剣剣剣剣剣}
  - Right-to-Left {ي ي ي ي ي}
  - Or even language-specific characters
- » Simulated expansion %
- » Prefix and suffix markers: {} or []





# 7 Elements of Localizability

1. Regional settings
2. Character sets
3. Translatable text
4. Over-externalization
5. Text expansion
6. Image text
7. Concatenated strings



Pseudo localization helps test at least 5/7 elements



# Internationalization 101 in Summary...

- » Simply do it once in one language, not 20X in 20 languages
- » 7 Technical Elements of Localizability (I18n)
  1. Regional settings
  2. Character sets
  3. Translatable text
  4. Over-externalization
  5. Text expansion
  6. Image text
  7. Concatenated strings
- » I18n Can Be Tested with Pseudo Localization
- » Cultural I18n
  - Cultural references
  - Nontranslatable names
  - Symbols, icons & images
  - Colors
- » I18n Benefits
  - Save Time
  - Reduce Costs
  - Enable Greater Focus on Quality
  - Reduce Opportunity Costs...

(discussed more in “Localization 101”)

## Other topics you might like to learn about?

- » Website localization 101: Best practices
- » Global navigation: How to lose or keep international website visitors
- » Excellent business uses and disastrous misuses for machine translation
- » How to perform your own 10-step international brand check
- » Lessons learned from hilarious international blunders

# Free Trial Offer

- Free 30 day trial community for up to 10 users

<http://www.lingotek.com/>

- Code: webinar



**Any questions?**

[www.lingotek.com](http://www.lingotek.com)

# Contact Us

Lingotek contact information:

Adam Wooten  
**Director of Translation Services**  
awooten@lingotek.com  
801.331.7777  
@AdamWooten  
@Lingotek