

Translation features of SharePoint 2013. Will they make your site useful or hilarious?

❖ Martin Laplante, CTO IceFire Studios

Machine Translation

❖ Risks

Machine Translation Risk to your brand



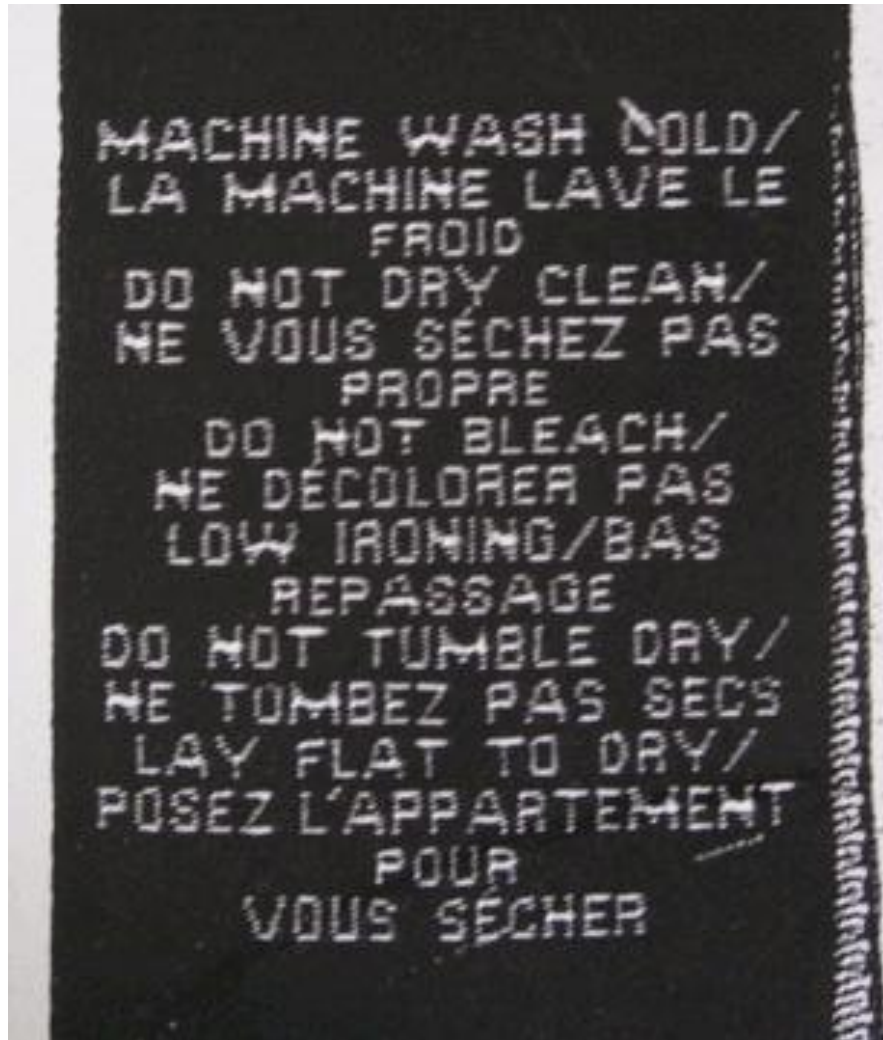
Machine Translation: Useful or hilarious?



French says “Buff the sausage.”

Machine translation

Useful or hilarious?



The machine cleans
the cold

Do not dry yourself
when clean

Do not fall down
when dry

Put down the
apartment to dry
yourself

Machine translation or incomprehensible?

Auto Puissance de temps à autre

Afin de réserver les pile. L'ordinateur seront auto puissance
tourné quand y a pas de signé pour cinq procès-verbal
L'ordinateur seront auto puissance one sans urgent tout
boutonner quand utilisateur commencer cyclisme

- ❖ save batteries: *reserve batteries*
- ❖ 5 minutes: *five meeting minutes*
- ❖ without pressing any button:
without urgent all buttoning

Machine translation technological problems

❖ Sears catalog

“Pull on pant regular in black” →

“Shoot the career soldier of panties in the dark”



**Tradition®/MD Pull on Pant Regular
in Black** **Sears®**

**Tirez sur le militaire de
carrière de culotte dans
le noir Tradition®/MD**

Soyez le premier à écrire un commentaire

- tirez sur la culotte
- ajustement formé qui étire vos courbes normales
- polyester de 92%, spandex de 8%
- 31" ; inseam

You can see
this translation
at SYSTRANet.com

Machine Translation

❖ Technology

Machine translation (MT) Technology

- ❖ Rule-based translation engines
 - ❖ Systran (BabelFish)
 - ❖ TAUM-METEO
 - ❖ Decodes the grammar
- ❖ Statistical translation engines
 - ❖ Google Translate
 - ❖ Microsoft Translator (Bing)
 - ❖ Little concept of grammar, only the probability that these words end-to-end are translated by those words end-to-end

Rule Based Translation engines

Pull on pant regular in black

v pp n/v a/n pp a/n

Eliminate parts of speech using grammar

-v- pp |—np—| pp n

"pant regular" is a noun phrase where the first word modifies the last

A verb without a subject is imperative

Statistical translation engines

- ❖ Based on millions of examples of translations, per language pair
 - ❖ Probability $\underline{\text{word1}} \rightarrow \underline{\text{wordA}}$
 - ❖ $\underline{\text{word2}} \rightarrow \underline{\text{wordB}}$
 - ❖ Probability $\underline{\text{word1 word2}} \rightarrow \text{wordA wordB}$
 - ❖ Probability $\underline{\text{word1 word2 word3}} \rightarrow \text{etc..}$
- ❖ Single-language statistics
 - ❖ Probability $\underline{\text{wordA wordB}}$ followed by $\underline{\text{wordC}}$
- ❖ Find the solution having maximum likelihood

Statistical translation engines

Quality varies by language pair

❖ Good news!

❖ English-French translation is excellent!

❖ Why?



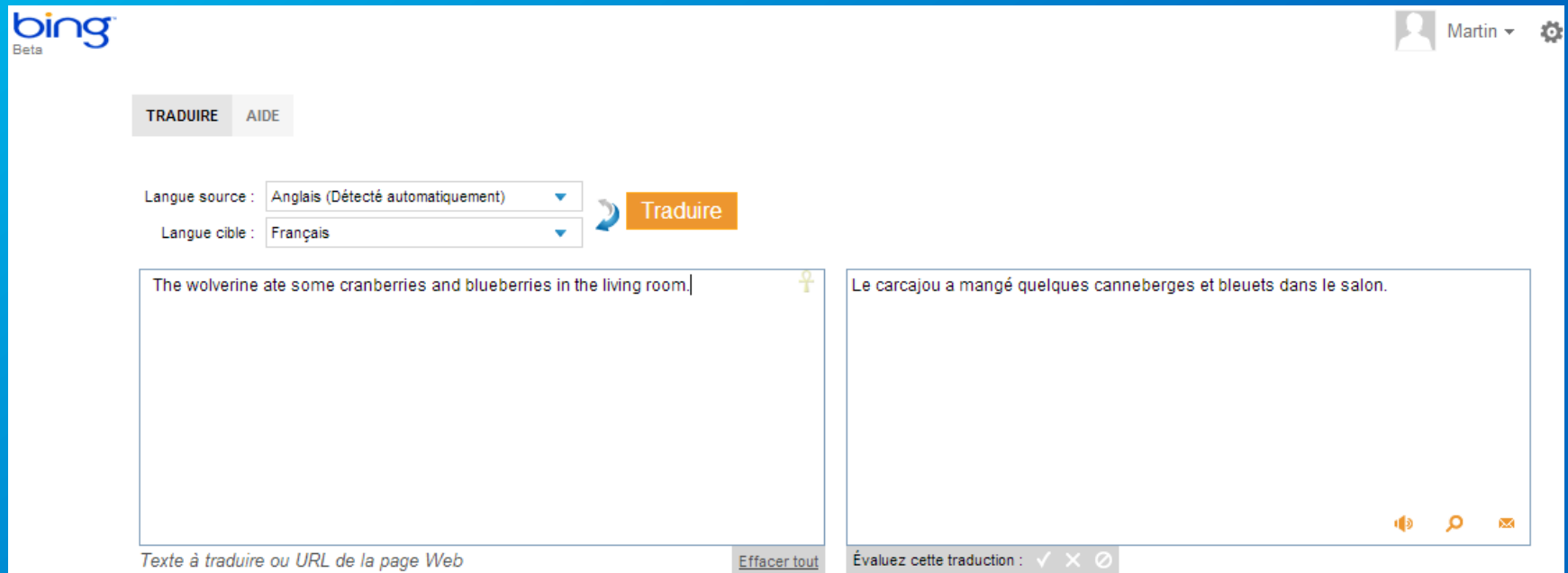
❖ Hansard, journal of parliamentary debates

Statistical translation engine no concept of grammar

The screenshot shows the Bing Translator interface. At the top, there are navigation links: WEB, IMAGES, VIDEOS, MAPS, NEWS, SHOPPING, TRANSLATOR, and MORE. The Bing logo is on the left. Below it are buttons for TRANSLATE and HELP. The 'Translate from' dropdown is set to 'English (Auto-Detected)' and the 'Translate to' dropdown is set to 'French'. An orange 'Translate' button is to the right. The input text box contains 'Please ring house boat door bell to check in'. The output text box contains 'S'il vous plaît anneau sonnette de porte bateau maison de vérifier'. At the bottom left of the input box is the text 'Enter text or webpage URL' and a 'Clear All' button. At the bottom right of the output box is a 'Rate this translation' section with check, X, and refresh icons, and a speaker icon.

Systran says "veuillez sonner la cloche de porte de bateau de maison afin de vérifier"

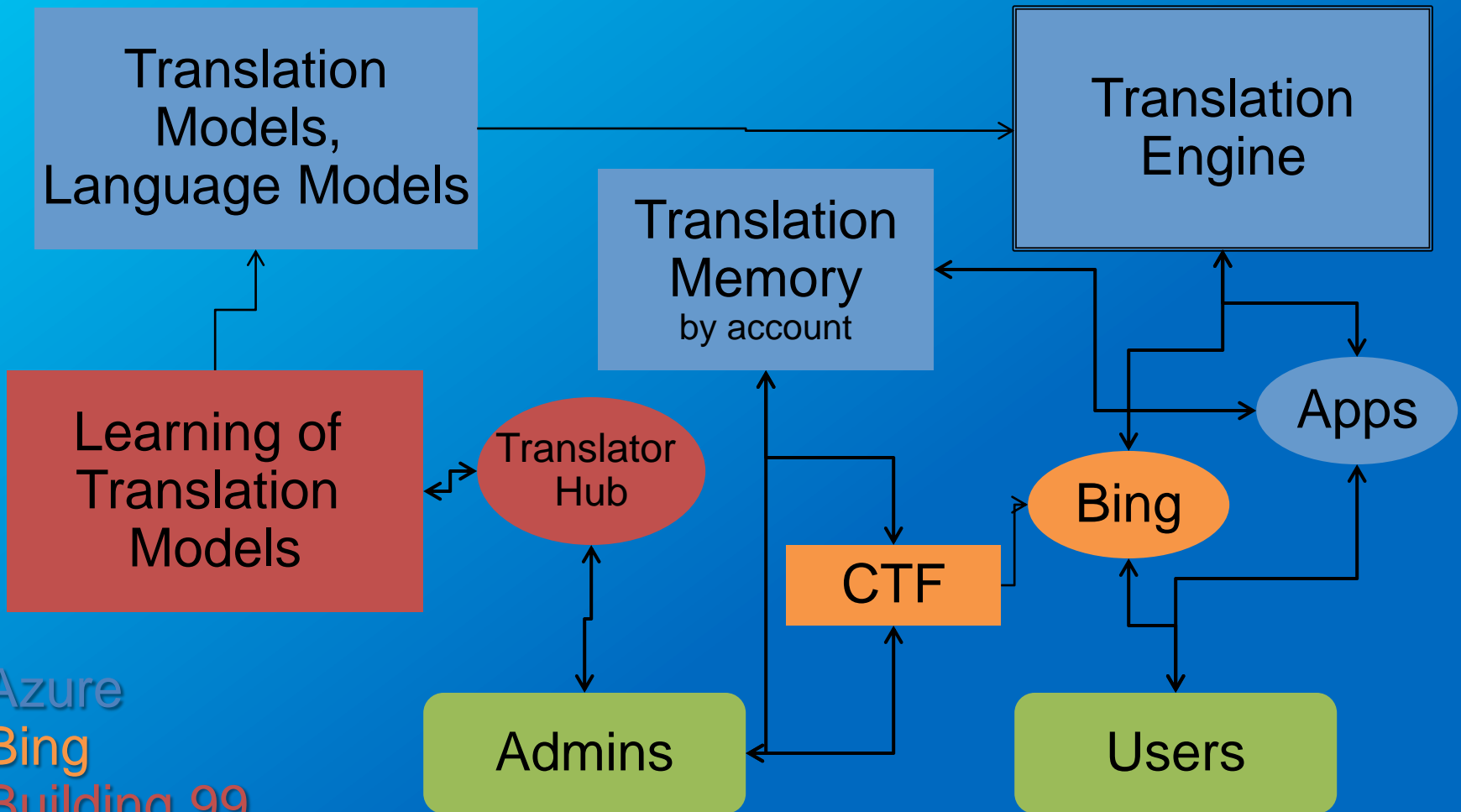
Statistical translation engine French, Canadian version



The screenshot shows the Bing Beta translation interface. At the top left is the "bing Beta" logo. At the top right is a user profile icon labeled "Martin" and a settings gear icon. Below the logo are two buttons: "TRADUIRE" (highlighted) and "AIDE". The interface features two dropdown menus: "Langue source : Anglais (Détecté automatiquement)" and "Langue cible : Français". A blue double-headed arrow icon is positioned between the dropdowns, and an orange "Traduire" button is to its right. Below these elements are two large text boxes. The left box contains the English text: "The wolverine ate some cranberries and blueberries in the living room." with a small yellow icon to its right. Below this box is the label "Texte à traduire ou URL de la page Web" and an "Effacer tout" button. The right box contains the French translation: "Le carcajou a mangé quelques canneberges et bleuets dans le salon." Below this box is the label "Évaluez cette traduction : ✓ ✕ ⌂" and three small icons: a speaker, a magnifying glass, and an envelope.

Not "le glouton a consommé des airelles et des myrtilles dans la salle de séjour"

The Microsoft Translator engine



Azure
Bing
Building 99

Machine & Human Translation

❖ New in SharePoint 2013

Variations: improvements

- ❖ Only for language

 - Up to 209 LocaleID, 39 languages

- ❖ More flexible

 - All separated into small bites

 - Small timer jobs

 - Flexibility: what, when, where propagated – pages, documents, lists

- ❖ Integration of Translation

 - Automatic (MT)

 - Human (XLIFF)

 - "Translatable columns" now has a function

Machine Translation:

In general, translate or not?

❖ What not to translate with MT

- ❖ Public web pages
- ❖ Internal published web pages
- ❖ Documents on site
 - ❖ Why? Waiting for reliability & quality
- ❖ Confidential Information

❖ What to translate with MT

- ❖ on-demand ephemeral content
- ❖ External content for crude translation (gisting)
- ❖ First drafts of internal translation
 - ❖ If the translator does not have better tools
- ❖ Short & repetitive texts
 - ❖ Especially if there is a translation memory

Machine Translation: What **SP 2013** translates

❖ Variations

- ❖ Public web pages
- ❖ Internal published web pages
- ❖ Documents on site

❖ Managed Metadata

- ❖ Controlled vocabulary
 - ❖ Must use the official term
 - ❖ No access to the translation memory

Demo

Managed Metadata

Machine translation

Standard transmission format



The Welsh says

"I am not in the office at the moment. Send any work to be translated."

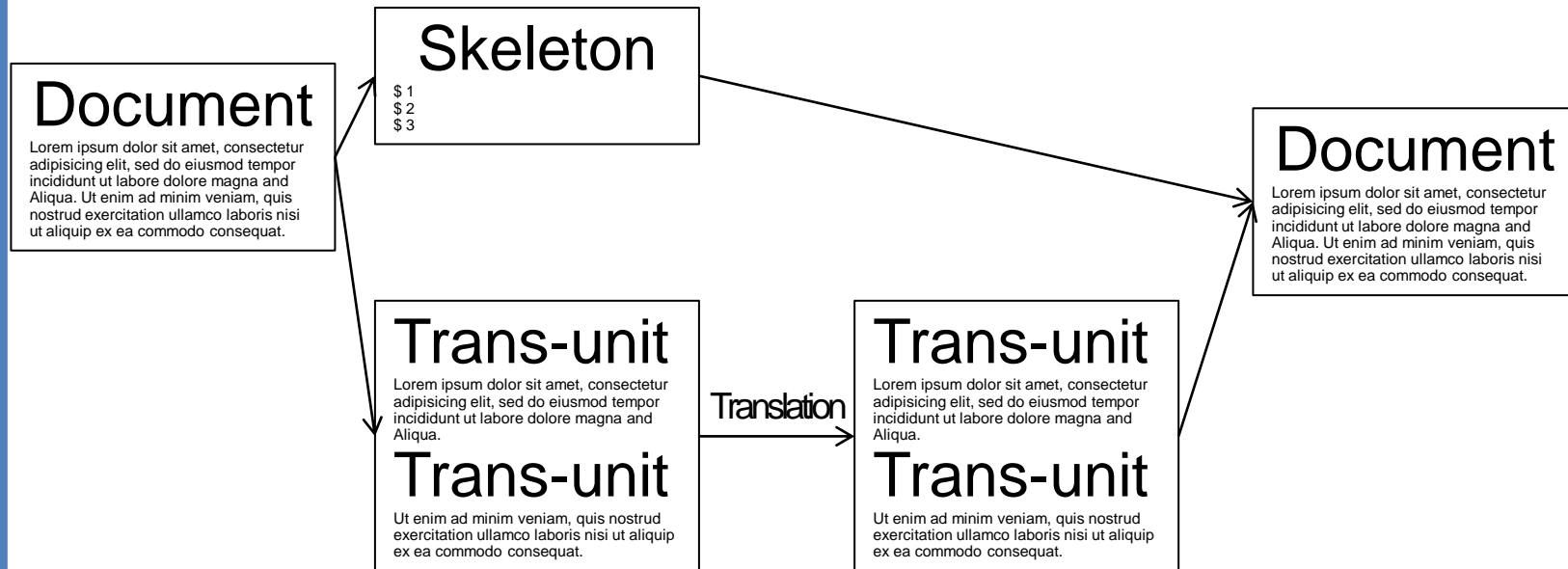
The XLIFF 1.2

What it is

- ❖ XLIFF standard
 - ❖ XML Localization Interchange File Format
 - ❖ Transfer information from one tool to another
 - ❖ Can transmit:
 - ❖ glossaries, translation memory
 - ❖ user interface software resources
 - ❖ elements to translate, groups of elements
 - ❖ documents, since version 1.2
- ❖ Not the same as "Export translations"
 - ❖ Which is still a .resx file

The XLIFF 1.2

The intent



- <skl> A skeleton not translate
- <trans-unit> text to translate & translations
- <bin-unit> binary object to be translated or not

The tools do not support all the possible elements of XLIFF



XLIFF SUPPORT IN CAT TOOLS

Results of the survey, January 2012

For html SharePoint 2013 uses bpt/ept, ph, sub



3.3.5 Inline elements

Virtaal leads this category with full score, closely followed by Okapi and MultiTrans Prism that miss only "sub". XTM has always been vocal on supporting "g" and "x" only.

Tool Name	Elements								
	bpt	bx/	Ept	ex/	g	It	Ph	sub	x/
Okapi Framework	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
MultiTrans Prism (5.0)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
SDL Trados Studio 2011	No	No	No	No	Yes	No	No	Yes	Yes
Swordfish III (3.0-6)	No	No	No	Yes	No	No	No	Yes	Yes
Transifex	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Translator Workspace XLIFF Editor	Yes	No	Yes	No	No	Yes	Yes	Yes	No
Virtaal 0.7.0	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
XTM 6.0	No	No	No	No	Yes	No	No	No	Yes

Table 9. Inline elements support

“The XLIFF specification allows for marking "beginning tags" and "ending tags" (<bpt, <ept)... Since there are efficient alternate ways to preserve the HTML code, it is not recommended to use the <bpt and <ept tags.”

For documents (.docx,.pptx,.pdf,etc.)SP2013 uses only binunit/binsource

3.3.4 Structural elements support

Overall good support of elements except for the notoriously unsupported binary ones. Notably, Multitrans and Trados do not support the important “alt-trans” element.

Tool Name	Elements									
	alt-trans	bin-source	bin-target	bin-unit	body	group	seg-source	source	target	trans-unit
Okapi Framework	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
MultiTrans Prism (5.0)	No	No	No	No	Yes	Yes	No	Yes	Yes	Yes
SDL Trados Studio 2011	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Swordfish III (3.0-6)	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Transifex	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Translator Workspace XLIFF Editor	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Virtaal 0.7.0	Yes	No	No	No	Yes ⁴	No	No	Yes	Yes	Yes
XTM 6.0	Yes	No	No	No	Yes	Yes	No	Yes	Yes	Yes

Table 8. Structural elements support

Translator tools are unable to open SP2013's
xliff-embedded documents

Demo

Variations

How to use the functionality Managed Metadata

- ❖ In general, do not pre-translate with MT
 - ❖ Can make the task more difficult for translators
- ❖ Extract an XLIFF file
- ❖ Send to translator
- ❖ Once back, load the translated XLIFF file into managed metadata service
- ❖ Edit if necessary

How to use the functionality

Variations: MT then HT?

- ❖ For some pages/lists send to MT first
- ❖ When the MT is back
 - ❖ Accept or modify or
 1. extract an XLIFF file
 2. Wait ...
 3. Send to a translator
 4. Once translation is back, return to the same site same list
 5. Load the translated XLIFF file
 6. Check, edit, either by hand or return to 1.
 7. Publish!

How to use the functionality

Variations

- ❖ Documents : Send document directly to the translator
 - ❖ In general, don't translate documents using MT
 - ❖ Accept or return or change outside SharePoint
 - ❖ Title and document metadata
- ❖ Writing for translation:
 - ❖ Microsoft Global English experiment
 - ❖ short sentences, correct grammar.
- ❖ Accumulates XLIFF files on the hard disk