Same-language MT
for local flavours/flavors

Janice Campbell - Adobe
Gema Ramírez - Prompsit

Intro: brief history of AltLang 1/3

A call from the industry:

not Prompsit's idea but Autodesk's

Intro: brief history of AltLang 2/3

Based on FOSS Apertium technology and data: easy to reuse, extend and adapt
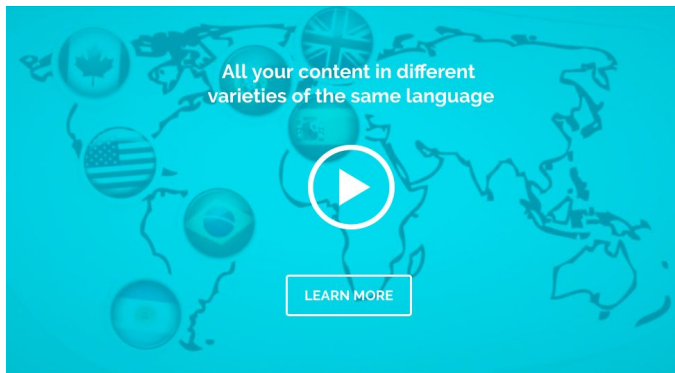
Automatic, reliable conversion for language varieties

Intro: brief history of AltLang 3/3

www.altlang.net

All your content in different varieties of the same language

Languages and formats

EN	FR	ES	PT
American English ↓↑ British English	Canadian French ↓↑ European French	LATAM Spanish ↓↑ European Spanish	Brazilian Portuguese ↓↑ European Portuguese

| Plain text .txt, .latex | (Libre)Office .odt, .docx .odp, .pptx .ods, .xslx | Website .html, .xml | Localisation .xliff, .po |

Adobe Business Case

Personalize the international customer experience by delivering local language variants to Adobe.com country sites.

Avoid time-consuming, costly manual editing of high-volume, constantly-changing content.

Requested by UK field sales & marketing.
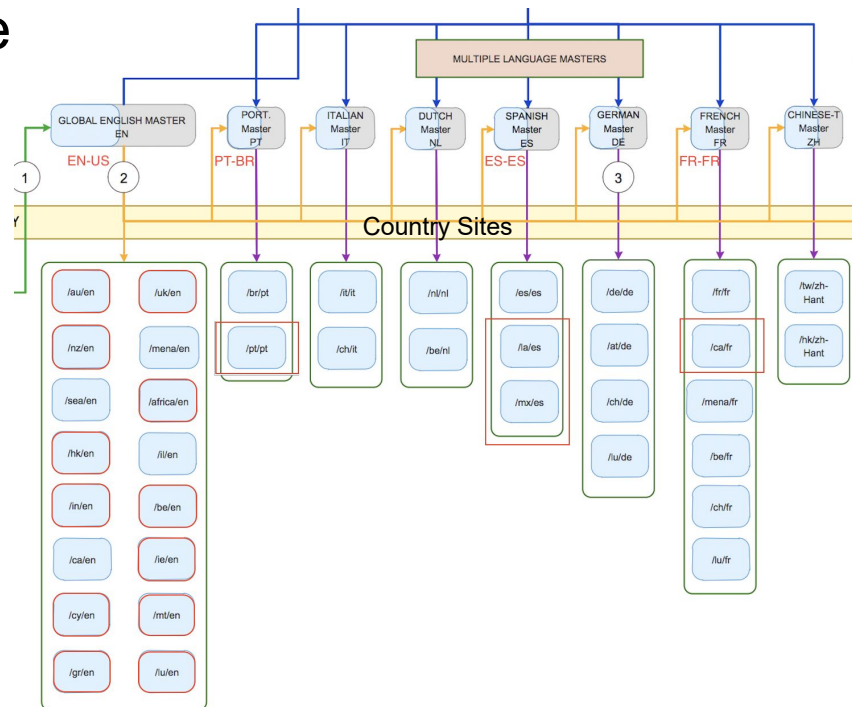
How does AltLang work? The basics… 1/3

- **automatically** and **quickly replaces differences** among two variants of the same language → nice for dynamic content
- performs **only controlled changes** → no (or low) risks
- highly **customisable** → can adapt to DNT, lexical choices, etc.
- easily **accessible** → full integration (JSON API), out-of-the-box testing (web-based demo), professional use (CAT tools and CMS)
- AltLang is **SaaS, a GPLv2- based service**

Variants is all about letters, words, some local changes...
It looks like we can automate it!

How does AltLang work? It deals with… 2/3

spelling differences

en_GB (-i**s**e)	en_US (-i**z**e)
customise	customize
realise	realize
wise	~~wize~~
advertise	~~advertize~~

word choices

a lift!
Please, give me ~~an elevator!~~

How does AltLang work? It deals with… 3/3

grammar structures

style/numeric conventions

CAN YOU PLEASE...

help + pronoun + infinitive =
help + pronoun + **to** + infinitive

Deliver by ~~28~~ **February 28**

2.5" = 6.4 cm (2.5")

AltLang all together... in its web-based demo

1) Go to www.altlang.net and log in

26/01/2015 - 10:34
The latest tool to help you to optimise your localisation process is here!

AltLang is the new programme to adapt your digitised content to a specific language variant.

Its colouring differences feature will render your content review easier.

01/26/2015 - 10:34
The latest tool to help you optimize your localization process is here!

AltLang is the new program to adapt your digitized content to a specific language variant.

Its coloring differences feature will render your content review easier.

British – American English

Add your own translation memory (.tmx)   *Optional*

Add your own terminology/glossaries   *Optional*

Translation memory file sample

Terminology file sample

AltLang Analytics   ⦿ Yes   ○ No

Highlight differences   ⦿ Yes   ○ No

2) Test short texts or documents
in all supported languages

OK

Implementation

Goal: Automatic, dynamic, low-risk conversion without any human intervention.

US English authoring → Copy to Global EN language master DIR → Push to EN country sites → Publish content to country sites

AltLang
Prompsit

Prompsit is a translation technology provider that cares about the impact of your content wherever it goes.
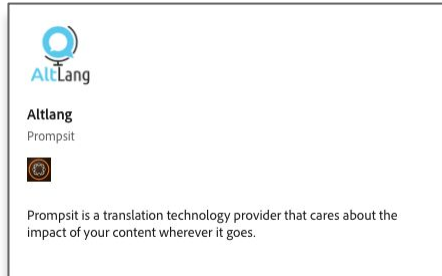
The conversion is **dynamically** triggered each time the EN content is **pushed** from the langmaster directory to the (12) country sites for publishing **live**, and is then **cached** locally (jcr node).

AltLang cloud service is delivered through integration with the CMS (AEM) as a connector available on the Adobe Exchange.

https://www.adobe.io/apis/marketingcloud/aem/aemconnectors.html

Challenges - Customization

- ## DNT - Adobe product & brand names
  - Lumetri Color
  - Media Optimizer
    (Media Manager in Japan)

English (United Kingdom)	English (United States)
Adobe Stock	Adobe Stock
Xpress	Xpress
Elements Organizer	Elements Organizer
ActionScript for Flash Lite	ActionScript for Flash Lite
Acrobat 8	Acrobat 8
Adobe Photoshop	Adobe Photoshop
Adobe Media Server Extended	Adobe Media Server Extended
Adobe Media Gateway	Adobe Media Gateway
Adobe Media Optimizer	Adobe Media Optimizer
Adobe Eazel	Adobe Eazel
Adobe Eazel for Photoshop	Adobe Eazel for Photoshop
Adobe Color	Adobe Color

- ## No Adobe-specific aligned bilingual corpus
  - zero target language content for language variations

- ## Adobe Style Guides lack language variations

Test Strategy

- ● Three test cycles to customize for Adobe
  - ○ Review staged content based on test cases
    - ■ Conversion is working correctly and consistently
    - ■ Check for over/under conversions via search and diff
    - ■ Check for regionalisms (state vs county)
    - ■ Check for anomalies
      - ● "leveraging" > "leverage";
        "resizable" > "resisable", but "sizable" did not convert, as expected
  - ○ Create dummy content pages (with sample test cases) for troubleshooting & debugging

Challenges - Post-Testing

- Language conversion for multiple countries not applicable to all; too many "exceptions" can create other issues.
  - [EN-US] "state" (administrative division) <> [EN-GB] "county" applies to just a few of the 12 country sites that prefer UK English.
  - [EN-US] "advertising dollars". How to deal with currencies? Best to change the source to "advertising budget", for example.
- How to scale fixing one-off feedback from web producers, field
- Website Architecture/Infrastructure Issues
  - Pulling and aligning source/converted content directly from the CMS so quality scoring can be performed.
  - Only a page with a change made to it, will trigger the conversion. (Bug fixes to AltLang)

Metrics

Language Quality

- Prompsit scoring (BLEU, WER, Edit Distance)

Customer Impact

- 12 of 16 Global English country sites prefer UK English
  (Africa, BE_EN, CY_EN, GR_EN, IE, LU_EN, MT, UK, AU, NZ, HK_EN, IN)
- ~15K pages converted in the initial launch
- Traffic: Avg. 77% VISITS/MONTH UK English/Global English-language websites

Customer Engagement - SEO (Future)

  on converted pages since the change from EN-US to EN-GB

- Bounce rate
- CTR
- Time/pages per visit

Language quality evaluation: general numbers

- American to British English
- 190 web pages (HTML)
- 5k translation units (after deduplication)

General numbers			
	# of words	# of lines that change from source to...	# of words that change from source to…*
source (US)	65,857	-	-
Altlang UK, before customisation	65,956	1,257 (≈25%)	1,618
AltLang UK, after customisation	66,025	1,535 (≈31%)	1,957

*sum of substitutions + insertions + deletions

Language quality evaluation: automatic metrics

- BLEU is almost 1 (between 0 -- worst and 1 -- best)
- WER is lower than 1% (between 0% -- best and 100% -- worst)
- Edit distance (character level, between before and after customisation):

Automatic metrics	
BLEU	0,98
WER	0,82%
EDIT DISTANCE	541

Are these differences good or bad?
Need for an qualitative evaluation

Language quality evaluation: a human insight

- sampling of 500 sentences with 84 changes
- human inspection of changes classified as correct, incorrect and missing

Qualitative results		
	UK before	**UK after**
Correct changes	62	72
Missing changes	20	10
Incorrect changes	1	2

- DNT (Adobe Color, Acrobat Catalog)
- -ising → - izing, offline → off-line, specialty → especiality, right away → straightaway

- Add "segment", "harness" and "retarget" as verbs to apply "help + pron. + to + infinitive" rule
- Adobe Experience Cloud → Adobe Marketing Cloud (7 times)

- a useful → ~~an~~ useful
- Facebook Custom~~ised~~ Audience

Conclusions about AltLang @ Adobe

- Easy integration through the AltLang connector for AEM
- Hands-off automation from authoring to publishing
- Cost-effective and "sim-ship" geo customization
- Customi[z|s]ation is key but not an obstacle
  - With no training corpora, customization is done post-implementation
- Metrics creation is a journey

Thanks!
Questions & Comments
Welcome

Same-language MT
for local flavours/flavors

Janice Campbell - Adobe
Gema Ramírez - Prompsit