welocalize
doing things differently

MT SUMMIT 2013

Connectivity, Adaptability, Productivity, Quality, Price...

Getting the MT recipe right

Organic Approach

All aspects of the localization ecosystem are taken into consideration

Selecting the right MT provider

By using our MT Scorecard we make sure all important KPIs are taken into consideration at selection time.

Empowerment through education

Internal, by the use of customized Toolkits; external, through specialised Trainings.

The feedback loop

Constructive communication from post-editor to MT provider

MT KPIs:
- ✓ **Productivity: Throughputs**
- ✓ **Productivity: Delta**
- ✓ **Quality: LQA**
- ✓ **Quality: Automatic Scores**
- ✓ **Cost**
- ✓ **GlobalSight: Connectivity**
- ✓ **GlobalSight: Tagging**
- ✓ **Human Evaluation**
- ✓ **Customization: Internal/External**
- ✓ **Customization: Time**

welocalize
doing things differently

Selecting the right MT provider
The WL MT scorecard

Productivity - Throughputs

   Number of post-edited words per hour

Productivity - Delta

   Percentage difference between translation and post-
   editing time

Cost

   Extrapolation, cost per word

GlobalSight - Connectivity

   Is there a connector in place?

Quality (Final) - LQA

   Internal quality verification

Quality (MT) - Automatic Scores

   A set of automatic scoring systems is used

> *We have tested and used different engines so we've seen the good, the bad and the ugly; now we can better appreciate what we have*

The WL MT Scorecard, II

Quality (MT) - Human Evaluation

   Internal, performed by linguists, WL form

GlobalSight – Tagging

   Behaviour around tags

Customization - Internal/External

   Who performs the training?

Customization - Time

   How long will it take?

> *If you could ask your MT provider to fix 1 thing, what would it be?*

> *Tag placement*
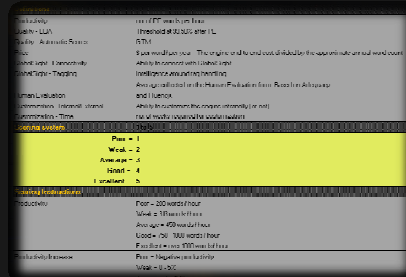
The WL MT Scorecard, III

Example
Overall data

German	#1	#2	#3	#4	French	#1	#2	#3	#4
KPIs					**KPIs**				
Productivity	4	4	4	4	Productivity	4	5	3	4
Productivity Increase	5	4	1	3	Productivity Increase	5	5	1	4
Quality - LQA	2	2	1	2	Quality - LQA	5	3	3	4
Quality - Automatic Scores	3	3	3	3	Quality - Automatic Scores	3	4	3	3
Cost	4	2	3	3	Cost	4	2	3	3
GlobalSight - Connectivity	4	3	2	4	GlobalSight - Connectivity	4	3	2	4
GlobalSight - Tagging	4	2	4	2	GlobalSight - Tagging	4	2	2	2
Human Evaluation	3	3	3	4	Human Evaluation	3	3	3	3
Customization - Internal/External	4	2	3	3	Customization - Internal/External	4	2	3	3
Customization - Time	3	1	2	1	Customization - Time	3	1	2	1
Total	**36**	**26**	**26**	**29**	**Total**	**39**	**30**	**25**	**31**

Automatic Scoring

	#3	#1	#2	#4	#3	#1	#2	#4
	FR	FR	FR	FR	DE	DE	DE	DE
BLEU	78.83	74.63	70.88	55.2	72.38	63.8	63.88	56.4
Nist	9.77	9.55	8.87	7.42	8.82	8.31	7.97	7.7
TER	13.66	20.52	19.91	34.11	16.39	23.53	22.8	29.2
METEOR	79.62	72.28	72.24	57.42	68.71	65.14	70.39	52.4
Prec.	82.33	76.63	78.5	66.19	73.03	73.18	76.6	64.6
Rec.	83.5	78.88	77.85	66.29	75.28	73.81	76.23	61.8
GLM	77.5	69.6	69.41	56.48	67.11	67.63	70.19	53.3

Definitions and instructions

Human Evaluation

Lena	Simon	Lena	Simon	Lena	Simon	Lena	Simon
4	4	3	3	4	3	3	3
4	3	4	4	2	3	3	3
3	3	1	3	2	1	4	3
2	1	2	4	4	4	4	4
3	2	3	3	4	3	4	4
1	2	1	1	2	3	3	3
2	3	4	3	2	2	2	3
2.917	2.9583	2.54167	2.7083	2.45833	2.70833	3.083333	3.0833
#3 DE 2.92		#1 DE 2.6		#4 DE 2.6		#2 DE 3.12	

Empowerment through education
Toolkits and Trainings

Our experience:

✓ Most language providers know and have experienced post-editing but they have limited knowledge of any other related aspect (automatic scoring, output differences between RBMT and SMT...)
✓ The majority of people who work in localization have heard about MT but most of them still find it a daunting subject.

Our answer:

✓ Continuous MT and PE related trainings and documentation for language providers
✓ Customized Toolkits for different internal departments (Production, Quality, Sales, Vendor Management)

Toolkits and Trainings II

Theory – knowledge foundations

Practice – customized PE sessions for different client accounts

- ✓ Master Glossary was *not* used during engine training! (QA checks)
- ✓ See Style Guide for section-specific usage of quotation marks/brackets; nominalization (headers); imperative/infinitive; taigen-dom style; formatting of UI options
- ✓ First segment of ITD has to remain untranslated!
- ✓ Check string length limits for audio files

Transparency – process, engine selection/customization, evaluations

Locales	Bing		Google		MSHub MT Training	
Spanish	4,13	3,6	4,13	3,73	4,13	4
Indonesian	4,56	3,76	4,52	3,84	4,72	4,08
Dutch	3,44	2,96	4,2	3,48	3,48	3,28
Danish	3,48	3,56	3,8	3,64	3,8	3,88

Training helps a lot - After I was told some of the background information and tips and tricks for certain engines/outputs, I was much more relaxed and happy to give MT a go.

Responsibility – valid evaluations, constructive feedback, quality ownership

---

Toolkits and Trainings III

Human Evaluations of MT Output (Adequacy + Fluency)

1. Train evaluators
   - ✓ adequacy versus fluency
   - ✓ error categories
   - ✓ comments + feedback
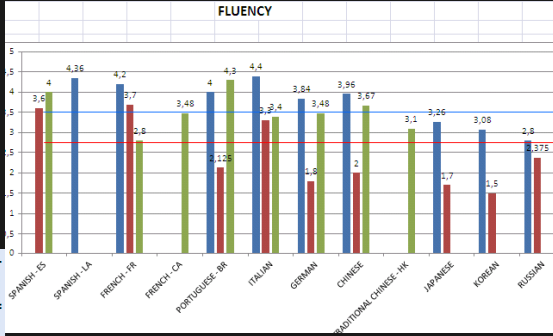
2. Controlled string selection

3. Verfication of scores

Click Ok when finished. // Click Add.	Very short and frequent structure in English "Click X"	Translations of this structure are usually more complicated, use prepositions etc.*
Alternatively, click … // Carefully perform … // However,… // once in a while… // Optionally,…	Adverbs / connectors	Can get omitted or translated literally
Enter a name for the NEW Configuration in the Configuration Name field.	Capitalized words / special formatting	Is it rendered correctly?
Attach each reel to the bracket with four screws (#2-56x .31 in).	Numbers / measurements / dates	Any numbers omitted? Units converted? Decimal dot converted?
When installing the ⅓ NPT pipe or cable gland on the AP, the use of a thread sealant (for example, Loctite® 565) is required.	Project-specific terminology	Is it translated appropriately for the domain?
Try adding donotreply@quickbooks.com to your email safe senders list.	URLs or email address	Are they inserted correctly, from a syntactical point of view?
Easily and conveniently run payroll in 3 easy steps - all within <A0/>.	Placeholders & tags	Are they inserted correctly, from a syntactical point of view?
Select at least one invoice to print or send.	Imperative structure	many Style Guides contain instructions of how to translate them (imperative / infinitive)

Insira a SKU do produto. | Insira a SKU do produto.

2   3 | 5   5 | please verify scores

Um menu suspenso exibe todos os produtos. | Um menu suspenso exibe todos os produtos.

5   5 | 5   5

Clique na seta suspensa ao lado de criado por e selecione a opção apropriada. | Clique na seta suspensa ao lado de criado por e selecione a opção apropriada.

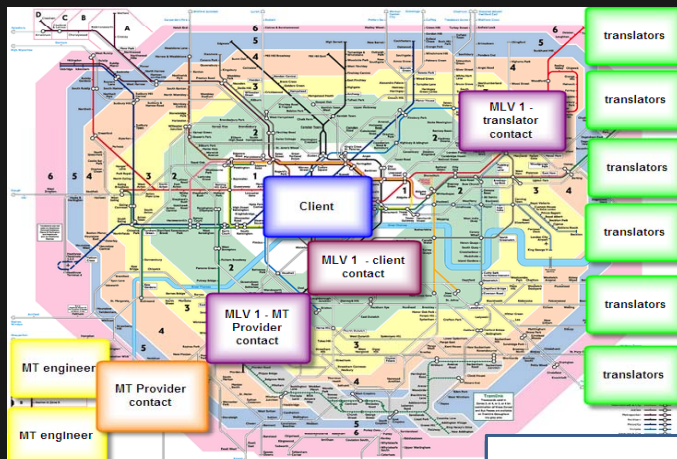3   3 | 2   2 | please verify scores

> Statistics from internal knowledge base

FLUENCY

Zeilenbeschriftungen	Count of Wrong terminolog	Count of Wrong Spelling	Count of Source not Translat	Count of Compliance with c	Count of literal translation	Count of Text/Information	Count of Capitalization	Count of Wrong Word Form	Count of Wrong Part of Spe	Count of Punctuation	Count of sentence Structur	Count of Tags and Markup	Count of Locale Adaptation	Count of Spacing	Count of Other (please spe
DE - Techdocs	3		5	1		4	2	1	1	6	4			6	1
FR - Techdocs	7	1	3		9		5	2	1		10				
IT Techdocs - Novolinguists			4		2			4	4		4			3	1
JA - Techdocs	3				2					2	6			5	6
KO - Techdocs	6		1		1	6		13	1	2	15			24	1
PTBR - Techdocs					7		11	1	1		7				
RU - Techdocs		1	8			1	6	7		6	6	1			
SP LA - Techdocs			1				8	3	1	3	1				1
Average/Total	19	2	22	1	21	11	32	31	9	19	53		1	38	10

Continued engine learning
The feedback loop

translators

MLV 1 - translator contact

translators

translators

Client

translators

MLV 1 - client contact

MLV 1 - MT Provider contact

translators

MT engineer

MT Provider contact

translators

MT engineer

For me the biggest advantage would be the possibility to implement a client terminology list [in SMT]

I wish we could easily fix the corpus for outdated terminology and characters

Teach the engine to properly cope with sentences containing more than one verb and/or verbs in progressive form

engine retraining improved significantly the handling of tags and spaces around tags, this is a productive achievement as it saves us a lot of manual corrections.

375

The feedback loop, II

Initial Productivity Test, 2012
Results

Locale ID	Productivity Delta
FRCA	33.40%
ZH-TW	-34.50%
ES-ES	27.50%
DE-DE	-16.20%
PT-BR	17.60%
RU-RU	8.10%
FR-FR	25%
ES-XL	4.30%

The feedback loop, III

Working towards productivity increase

✓ **Feedback loop** between language teams and MT provider is set up

✓ Symbiotic relationship between WL and MT provider
  Feedback helps the core engine to improve
  Better engines generate better quality output

✓The 27 engines undergo deep rebuild and full re-train
  Enhanced tag re-insertion algorithm
  Enhanced tag handling for better source format painting alignment
  Improved space handling around tags
  Improved punctuation and symbol formatting

✓The language teams attend training sessions

The feedback loop, IV

MT provider to complete rows F and G: Status - "Fixed", "Open", "More information required", "Cannot be fixed"
Instructions: highlight the error clearly in the MT output (in the case of omissions in the source); only highlight *1 issue per listed string* ; when logging errors around tags, ens
difficult to spot and/or automatically check during post-editing, overall high impact on productivity; "Medium" = fairly frequent, can be checked via QA checkers, but correcting th
is easy to spot and/or does not impact overall productivity too much

Date	Language	Issue	Source	MT Output	Corrected Translation
		Format [errors around spaces, tag positioning; content of tags has been translated]	<fmt id="1" tooltip="b" endtooltip="b">Interface Speed:</fmt> 6 Gb/s SAS	<fmt id="1" tooltip="b" endtooltip="b"> Interfacesnelheid: 6 </fmt> Gb/s SAS	<fmt id="1" tooltip="b" endtooltip="b">Interfacesnelheid:</fmt > 6 Gb/s SAS
		Locale Adaptation [quotation marks, date/time/measurement formats, parenthesis style not adapted to target locale]	Beginning June 2, 2012, ….	À compter de juin 2, 2012, …	À compter du 2 juin 2012, …
		Punctuation [wrong punctuation for target locale is used]	• Support for up to 12TB¹ maximum internal capacity in a compact 1U chassis enables you to choose the amount of storage that's right for your business.	*Prise en charge de jusqu'à 12TB¹ interne maximum de capacité dans un châssis 1U compact vous permet de choisir la quantité de stockage qui convient le mieux à votre entreprise.	• La prise en charge d'une capacité interne maximale de 12 To¹ dans un châssis 1U compact vous permet de choisir la quantité de stockage qui convient le mieux à votre entreprise.
		Locale Adaptation [quotation marks, date/time/measurement formats, parenthesis style not adapted to target locale]	5.2" x 2.53" (132.08mm x 64.26mm)	5.2 英寸 x 2,53 " （132,08mm x 64,26mm)	5.2英寸 x 2.53英寸 （132.08毫米 x 64.26毫米）
		Word form [e.g. gender, plural/singular mismatch in target, case mismatch]	Computer Start up issues	Tietokoneen Käynnistä ongelmat	Tietokoneen käynnistysongelmat

---

The feedback loop, V

Examples of engine improvements based on feedback from the language teams.

Improvements in tag handling related issues
- Reported by DE, ESES, FRFR, NL, PL, and TR teams

Improvements in syntax and formatting including
- Handling of hyphens and dashes - reported by DE, ESXL, FRCA, and PTBR teams
- Conversion of bullet points - reported by ESES, FRCA, FRFR, and TR teams
- Normalization of " " - reported by DE team
- Handling of "®" and "™" symbols - reported by ZHCN team
- Handling of "|" - reported by ESES, FI, and FRCA teams
- Handling of inches symbols - reported by CS, ESXL, and PTBR teams
- Forcing a capital letters segment start - reported by FRFR team
- Use of non-breaking spaces - reported by FRFR team
- Conversion of en-dashes and em-dashes - reported by DE and NL teams
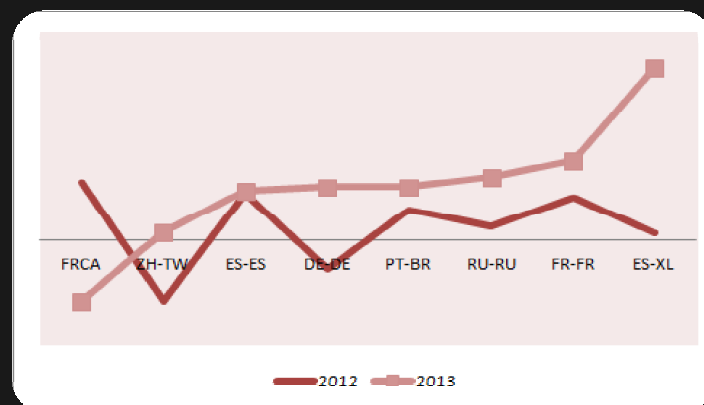
The feedback loop, VI

Second Productivity Test, 2013
Results

Locale ID	Productivity Delta
FRCA	-35.00%
ZH-TW	4.70%
ES-ES	28.50%
DE-DE	31.20%
PT-BR	31.30%
RU-RU	36.60%
FR-FR	46%
ES-XL	99.30%

The recipe works

Productivity increases after combining all factors

We are in a better position to offer and negotiate realistic discounts with clients and language providers.

Thank you!
Laura Casanellas, Lena Marg

welocalize
doing things differently