

COACH

lingo24 
love your language

COACH: Consolidating the Localisation Industry through Technology

Sergio Penkale

Lingo24

`sergio.penkale@lingo24.com`

Outline of Presentation

- 1 Translators and Current CAT Tools
- 2 Gathering Translators' Feedback
- 3 Changes to COACH
- 4 Conclusions

Justified Translator Resistance

- CAT tools forced them to alter the way they work,
- Lack of consultation in tool design.



Introduction of CAT Tools Inevitable . . .

- Change of Paradigm: paper to electronic
- Ability to deal with file formats,
- Support for in-line formatting at appropriate points in target,
- Leveraging from previous translations.



COACH

General Acceptance

- Translators require professional tools, which develop and evolve over time.
- Makes sense to involve translators – actual *users* of tools – when developing new technology.
- This is the approach we took in the design of COACH.

Main Focal Points for Lingo24

- Bringing high volumes to the translator,
- Allowing clients to be more flexible with their quality expectations,
- Having more content translated at predictable quality,
- Creating a PEMT interface that allows more and differently educated users to approach this task.

Identification of Focus Group

- *Age & cultural background:*
 - ▶ COACH designed to be used by wide variety of translators.
- *Technical background*
 - ▶ COACH meant to be easy-to-use by any translator, irrespective of CAT tools experience.

Identification of Focus Group (cont.)

- *Previous working relationship with translators:*
 - ▶ Quality and speed two of the main areas COACH was intended to improve.
 - ▶ Use translators who had previously done a lot of work for Lingo24 with consistently good results.
- *PEMT Experience:*
 - ▶ COACH designed to be Lingo24's main environment for Post-Edited MT.
 - ▶ PEMT experience crucial to assess features that distinguish PEMT from HT projects.

Round 1: General Feedback on Usability and Utility

- *Speed:*
 - ▶ What features could increase efficiency during the translation process?
 - ▶ What would help speed up the decision-making process?
- *Quality:*
 - ▶ What features would improve the output quality, even with limited manual checks?
 - ▶ What types of automated quality checks would help the translation process and improve output quality?
- *Usability and Utility:*
 - ▶ How intuitive is the tool?
 - ▶ How can the user's experience be improved?
 - ▶ Does the tool cater to the user's requirements?

Translators' Feedback Loop: Round 2

- After Phase 1, COACH stable and offering proper support for European languages.
- Phase 2 focused on support for Asian and R-to-L languages.
 - ▶ Second focus group selected on the same criteria as in Phase 1.
 - ▶ Recruited translators to provide feedback on language-specific aspects:
 - ★ Segmentation rules,
 - ★ Reliability of QA checks.

Speed: Task Prioritisation

<input type="checkbox"/>	project_training	1897 training_file.htm	Translation	1-151	1052	English (UK)	English (US)	<div style="width: 0%; background-color: #ccc;">0%</div>	00:00:00	2013-02-26 17:14:30
<input type="checkbox"/>	LING130213_252	3274 FM_09_SL_10_First Draft Translation_Brief_DE.docx	Review1	1-72	449	German (Germany)	Spanish (Spain)	<div style="width: 0%; background-color: #ccc;">0%</div>	00:00:00	2013-02-28 10:25:00
<input type="checkbox"/>	LING130213_252	3277 FM_12_SL_11_Transcreation Brief_DE.docx	Review1	1-112	922	German (Germany)	Spanish (Spain)	<div style="width: 0%; background-color: #ccc;">0%</div>	00:00:00	2013-02-28 10:25:00
<input type="checkbox"/>	LING130221_295	3435 source_2 Spanish (Spain) - 3366.xlsx	Review1	1-334	971	German (Germany)	Spanish (Spain)	<div style="width: 0%; background-color: #ccc;">0%</div>	00:00:00	2013-02-22 11:07:00
<input type="checkbox"/>	LING130330_359	3829 12_12_10 Account Development Strategy.docx	Review1	1-172	1582	English (UK)	Romanian	<div style="width: 0%; background-color: #ccc;">0%</div>	00:00:00	2013-03-31 19:29:00
<input type="checkbox"/>	LING130331_361	3840 Management Review Report_April 2012.docx	Review1	1-160	1109	English (UK)	Romanian	<div style="width: 0%; background-color: #ccc;">0%</div>	00:00:00	2013-03-31 13:39:00
<input type="checkbox"/>	RESO130408_391	3943 Press Release.docx	Translation	1-73	1541	English (UK)	Romanian	<div style="width: 0%; background-color: #ccc;">0%</div>	00:00:00	2013-04-30 18:19:00

Priority 1 (passed deadline)
 Priority 2 (very close deadline)
 Priority 3 (close deadline)
 Priority 4


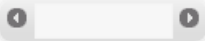
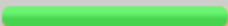
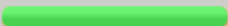
0 words translated, 1541 expected by now

Pending

You have no project items planned to start. If you're eager to work on something, please contact us right now or [take the training project item](#) to improve your skills

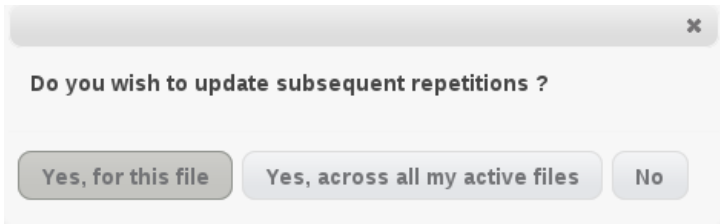
Project	File Name	Step	Segments	Words	Source	Target	Progress	Overall time	Deadline
Finished									

Speed: Measuring Real vs. Expected Progress

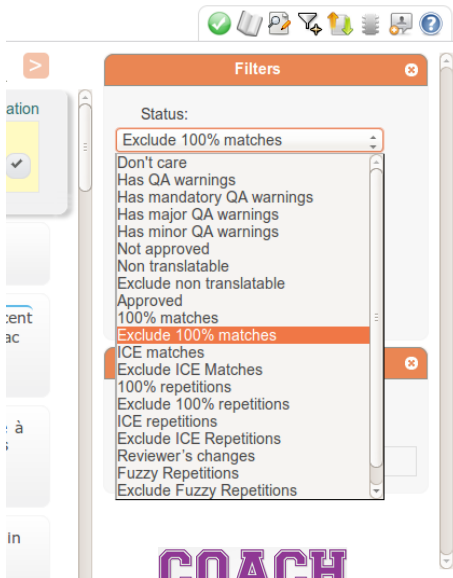
Progress	Task/Batch Deadline	Overall time	Editing distance
 100%			<input type="button" value="Reset"/> <input type="button" value="Filter"/>
 100%	2013-04-22 11:00:00	00:30:00	11.4%
 100%	2013-04-19 17:00:00	00:11:38	0.27%

230 words translated, 230 expected by now

Speed: Propagate Changes & Repetitions



Speed: Filter Translatable Content



Speed: Enhanced Internal Matching

- In-Context Repetitions
- Fuzzy Repetitions

A Martian is a native inhabitant of the planet Mars. 5 100% A Martian is a native inhabitant of the planet Mars.

Since there is no solid evidence of life there at present, all Martians known as of 2013 are fictional creatures 6 ICR No Translation

100% | not-approved
Source: Since there is no solid evidence of life there at present, all Martians known as of 2013 are fictional creatures

Since there is no solid evidence of life there at present, all Martians known as of 2013 are fictional creatures

100%
This is a translation 2

A Martian is a native inhabitant of the planet Mars. 7 100% A Martian is a native inhabitant of the planet Mars.

Since there is no solid evidence of life there at present, all Martians known as of 2013 are fictional creatures 8 ICR Since there is no solid evidence of life there at present, all Martians known as of 2013 are fictional creatures

A Martian is a native inhabitant of the planet Mars. 9 100% A Martian is a native inhabitant of the planet Mars.

Quality: Customisable Quality Checks

- Live highlighting of spelling mistakes and multiple spaces,
- Live terminology check,
- Inconsistent translations,
- Add words to dictionary,
- Language-specific QA checks:
Number formatting &
Punctuation.

Which QA Checks do you want to enable ?

<input type="checkbox"/>	Major QA Checks
<input checked="" type="checkbox"/>	Empty Translation
<input checked="" type="checkbox"/>	Terminology Check
<input checked="" type="checkbox"/>	Inconsistent TM Match
<input checked="" type="checkbox"/>	Spell Check
<input checked="" type="checkbox"/>	Inconsistent repetitions
<input type="checkbox"/>	Minor QA Checks
<input checked="" type="checkbox"/>	Consecutive Spaces
<input checked="" type="checkbox"/>	Consecutive Punctuation Marks
<input checked="" type="checkbox"/>	Different Brackets
<input checked="" type="checkbox"/>	End Punctuation
<input checked="" type="checkbox"/>	Entire Capitalisation
<input checked="" type="checkbox"/>	Inconsistent Spaces
<input checked="" type="checkbox"/>	Initial Capitalisation
<input checked="" type="checkbox"/>	Number Values
<input checked="" type="checkbox"/>	Non-matching Brackets
<input checked="" type="checkbox"/>	Non-matching Question Marks

Usability and Utility

- Customisable font types and sizes,
- Detailed tool tips for quality warnings.
- Customisable shortcuts:

Select previous segment shortcut:	<input type="text" value="ctrl+up"/>	default: ctrl+up
Select next segment shortcut:	<input type="text" value="ctrl+down"/>	default: ctrl+down
Approve segment shortcut:	<input type="text" value="ctrl+s"/>	default: ctrl+s
Copy source shortcut:	<input type="text" value="shift+ctrl+c"/>	default: shift+ctrl+c
Insert previous tag into translation:	<input type="text" value="alt+p"/>	default: alt+p
Insert next tag into translation:	<input type="text" value="alt+n"/>	default: alt+n
Insert TM match in target:	<input type="text" value="shift+ctrl"/>	default: shift+ctrl+(1-5)

Key Differentiators in COACH

- Important functions in a centralised tool, e.g. QA made available in the cloud.
- QA is a weakness in many specialised tools.
- Translator access to QA report. Information may be shared, depending on workflow.
- Connect content and translator: unlock content that would not be translated unless COACH was available.

Key Differentiators in COACH (cont.)

- Dedicated Language Technologies group
- State-of-the-art Machine Translation
- Automatic terminology extraction
- Automatic subject detection

Automatic Terminology Extraction

- Many customers have TMs, but no centralised terminologies.
- From TM, automatically determine most important source terms
 - ▶ Most frequent terms in TM
 - ▶ Most frequent terms in standard Source Language text
- Using by-product of MT, locate target equivalent of source terms
- Output: draft terminology bank

Automatic Terminology Extraction (cont.)

My work **sdas.txt, en-gb...** The coming of L... Zoltan Szabo TR | Preferences | Logout

Language **English (UK)** | Status | Language **Romanian** | Progress 0%

ID	Language: English (UK)	Status	Language Romanian	Progress 0%	Actions
1	Lorem ipsum		At vero eos		<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
2	dolor sit amet		et accusamus		<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
3	consectetur		voluptatum		<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
4	adipiscing		similique		<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
			est laborum		<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
			dolorum		<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
5	labore et dolore		nihil impedit		<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

Before user processing

Results of user processing

2	dolor sit amet	<input checked="" type="checkbox"/>	et accusamus		<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
3	consectetur	<input type="checkbox"/>	voluptatum		<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
4	adipiscing	<input checked="" type="checkbox"/>	similique		<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
		<input checked="" type="checkbox"/>	laborum est		<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
		<input type="checkbox"/>	dolorum		<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

View terms 1 / 35 from the page 1 of 4 Finish

Search

polarization 1 Polarizarea

Term Details for... "Polarizarea"

source polarization target Polarizarea

Definition: *Not known*
Reference: *Not known*
Abbreviation: *Not known*
Context: *Not known*

Less

Customer: Coach - user test
Source: User glossary
Subject: Arts and Culture
Added on: *Not known*
Added by: *Not known*
Status: Not Approved

Frontend: Sorin Pleava, Thu Jul 18 15:54:27 2013 +0300, Backend: Unknown build, 19.07.2013 @ 09:25:56 BST [Back to my account](#)

Automatic Subject Detection

- Given source document: which MT engine to use? Which supplier?
- Extract most relevant terms from document
- Use SVM classifier to find most likely subject
- Choose optimal MT engine according to subject

Summary

- COACH is a better tool thanks to translator cooperation during the build.
- Translators are better disposed towards COACH as a result.
- Makes translators *demonstrably* more efficient.
- Creative pricing models owing to complete control over both our CAT tool and our own MT technology.

Thanks for your attention!

`sergio.penkale@lingo24.com`