Evaluation of SMT in localization to under-resourced inflected language

Raivis Skadiņš, Maris Puriņš, Inguna Skadiņa, Andrejs Vasiļjevs

EAMT 2011, Leuven

May 30-31, 2011

Let's MT! TILDE

Localization Work

- The localization process is generally related to the **cultural adaptation** and **translation** of software, video games, and websites, and less frequently to any written translation
- Translation Memory is commonly used
- MT is not yet widely used in practical localization, some use-cases and application areas for few larger languages

Previous Work – MS Research

- Evaluation with keyboard-monitoring program and Choice Network Analysis for measuring the effort involved in post-editing MT output (O´Brien, 2005)
- Productivity tests have been performed in translation and localization industry settings at Microsoft (Schmidtke, 2008)
  - SMT system of Microsoft Research trained on MS tech domain for 3 languages for Office Online 2007 localization task: Spanish, French and German
  - By applying MT to all new words on average 5-10% productivity was gained.

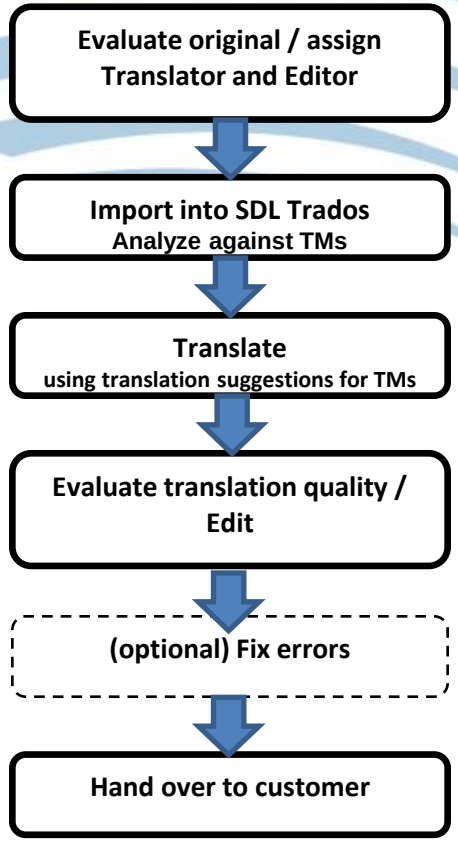
Previous Work – Adobe

- In Adobe two experiments were performed (Flournoy and Duran, 2009):
  - Small test set of 800-2000 words was machine translated and post-edited
  - Then, based on the positive results, about 200,000 words of new text were localized
  - The rule-based MT was used for translation into Russian (PROMT) and SMT for Spanish and French (Language Weaver)
  - Productivity increase between 22% and 51%

Previous Work – Autodesk

- Evaluation of Autodesk Moses SMT system (Plitt and Masselot, 2010):
  - Translation from English to French, Italian, German and Spanish with three translators for each language pair
  - To measure translation time special workbench was designed to capture keyboard and pause times for each sentence
  - MT allowed translators to improve their throughput on average by 74%
  - Varying increase in productivity: from 20% to 131%
- Optimum throughput has been reached for sentences of around 25 words in length

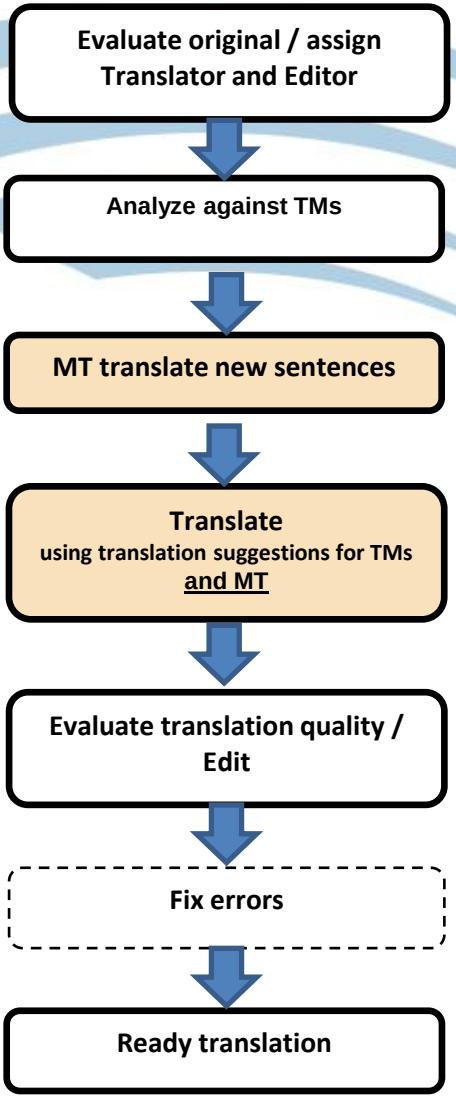
Localization workflow at Tilde



Let's MT!

TILDE

MT Integration into Localization Workflow



Evaluation of Productivity

- Productivity of translation process without degradation of quality is the most important measure that is interesting for localization industry
- Measured as translation output of average translator in words per hour
- Both experienced translators and new translators were involved

Evaluation of Quality

- Performed by human editors as part of their regular QA process
- Result of translation process was evaluated, editors did not know was or was not MT applied to assist translator
- Comparison to reference is not part of this evaluation
- Tilde standard QA assessment form will be used covering at least the following text quality areas:
  - Accuracy
  - Spelling and grammar
  - Style
  - Terminology
- Error score was calculated.
  Metric calculated by counting errors identified by the editor and applying a weighted multiplier based on the severity of the error type.

QA Evaluation Form

Error Category	Weight	Amount of errors	Negative points
1. Accuracy			
1.1. Understanding of the source text	3		0
1.2. Understanding the functionality of the product	3		0
1.3. Comprehensibility	3		0
1.4. Omissions/Unnecessary additions	2		0
1.5. Translated/Untranslated	1		0
1.6. Left-overs	1		0
Total			0
2. Language quality			
2.1. Grammar	2		0
2.2. Punctuation	1		0
2.3. Spelling	1		0
Total			0
3. Style			
3.1. Word order, word-for-word translation	1		0
3.2. Vocabulary and style choice	1		0
3.3. Style Guide adherence	2		0
3.4. Country standards	1		0
Total			0
4. Terminology			
4.1. Glossary adherence	2		0
4.2. Consistency	2		0
Total			0
Additional plus points for style (if applicable)			0
Grand Total			0
Negative points per 1000 words			0
Quality:			Resulting Evaluation

Let's MT!

TILDE

QA Scores

Error Score (sum of weighted errors)	Resulting Quality Evaluation
0…9	Superior
10…29	Good
30…49	Mediocre
50…69	Poor
>70	Very poor

Tilde Localization QA assessment to be applied in LetsMT! evaluations

The MT system used

- English-Latvian
- Factored phrase-based SMT system
- Tools: Moses, Giza++, SRILM,
  Latvian morphological analyzer
- Monolingual corpus: 391 M words
- Parallel corpora:

Corpora	Size (M, sentences)
DGT-TM	1.06
OPUS EMEA	0.97
Localization TM	1.27
Dictionary	0.51
Web comparable corpus	0.90
Books	0.66
Total:	**4.10**

Let's MT! TILDE

The MT system

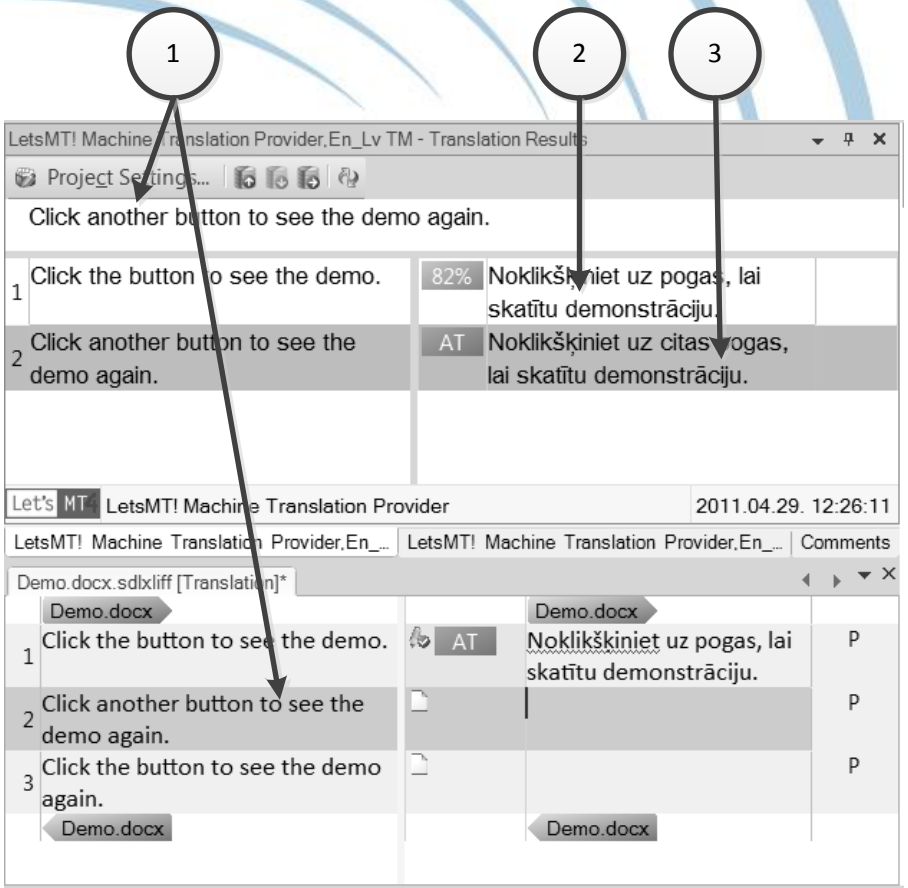
- Development and evaluation data
  - Development - 1000 sentences
  - Evaluation – 500 sentences
  - Balanced
  - BLEU score: 35.0
  - SMT system **is not** in domain

Topic	Percentage
General information about European Union	12%
Specifications, instructions and manuals	12%
Popular scientific and educational	12%
Official and legal documents	12%
News and magazine articles	24%
Information technology	18%
Letters	5%
Fiction	5%

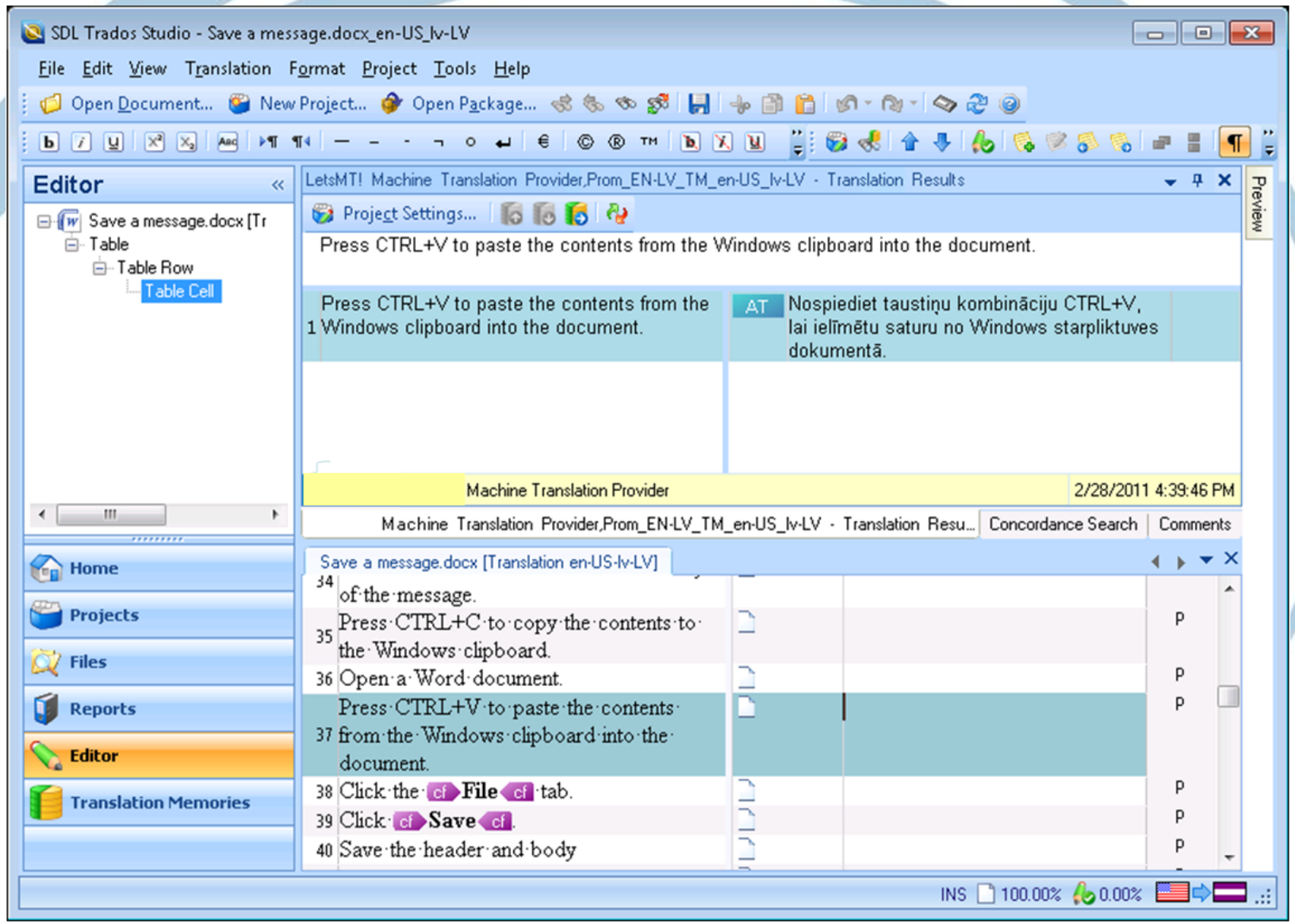
Let's MT! TILDE

Integration in SDL Trados

- SDL Trados 2009 plug-in
  using standard MT integration approach described in SDL Trados SDK



Integration of MT in SDL Trados



Results of the First Evaluation

- Moses based EN->LV SMT (BLEU score 35.0)
- 5 translators, 54 documents with 950-1050 adjusted words each
- Average increase of translators productivity: **32.9%**
- Increase of error score from **20.2** to **28.6** points but still at the level "**GOOD**" (<30 points)

Let's MT! TILDE

Let's MT!

Thank you!

The research within the LetsMT! project leading to these results has received funding from the ICT Policy Support Programme (ICT PSP), Theme 5 – Multilingual web, grant agreement no 250456