Application of Machine Translation in Localization into Low-Resourced Languages

Raivis Skadiņš[1], Mārcis Pinnis[1], Andrejs Vasiļjevs[1], Inguna Skadiņa[1], Tomas Hudik[2]
Tilde[1], Moravia[2]
{raivis.skadins;marcis.pinnis;andrejs;inguna.skadina}@tilde.lv,
xhudik@gmail.com

Abstract

This paper evaluates the impact of machine translation on the software localization process and the daily work of professional translators when SMT is applied to low-resourced languages with rich morphology. Translation from English into six low-resourced languages (Czech, Estonian, Hungarian, Latvian, Lithuanian and Polish) from different language groups are examined. Quality, usability and applicability of SMT for professional translation were evaluated. The building of domain and project tailored SMT systems for localization purposes was evaluated in two setups. The results of the first evaluation were used to improve SMT systems and MT platform. The second evaluation analysed a more complex situation considering tag translation and its effects on the translator's productivity.

1 Introduction

In recent years, machine translation has received more and more interest from the localization industry. To stay competitive in the market, localization companies have to increase the volume of translation and decrease costs of services. For this reason, the localization industry is increasingly interested in combining translation memories (TM) with machine translation solutions adapted for the particular domain or customer requirements.

Building usable machine translation systems for less-resourced languages with complex morphology and syntax is difficult due to a lack of linguistic resources, on one hand, and the complexity of the language, on the other hand.

The benefits of the application of machine translation in localization are also recognized by developers of computer aided translation (CAT) tools. Such widely used CAT tools as SDL Trados Studio, Kilgray memoQ, ESTeam Translator, Swordfish, MemSource and Wordfast besides traditional translation memory support provides integration with machine translation systems. Several cloud-based platforms offer machine translation services for the localization industry: KantanMT[1], LetsMT[2] and tauyou[3], and others.

This paper describes the methodology used for MT evaluation in localization process and results of two experiments where MT was integrated into CAT tool and used in two professional localization companies – Tilde and Moravia.

In the first experiment we evaluated the impact of in-domain SMT on the productivity of translation of plain text, i.e., text without any formatting. Application of in-domain English-Latvian, English-Czech, English-Hungarian and English-Polish MT systems were evaluated by using MT plug-in to integrate them in the SDL Trados Studio translation environment.

In the second experiment, we set a more complex scenario where translatable documents are slightly out of the domain of the SMT system, contain formatting tags, and are written in a more technical language than in the previous experiment. The second experiment was carried out on English-Latvian, English-Lithuanian, and English-Estonian language pairs. In both experiments, in addition to the productivity evaluation we also performed assessment of the translation quality according to the standard internal quality assessment procedure.

[1] http://www.kantanmt.com
[2] https://www.letsmt.eu
[3] http://www.tauyou.com

2 Related Work

Although experiments on the application of MT for assisting humans in professional translation started more than four decades ago (e.g., Bisbey and Kay 1972; Kay, 1980), it got more attention from the research community only in the late 1990s, with various studies on post-editing and machine translatability (e.g., Berry, 1997; Bruckner and Plitt, 2001). A comprehensive overview of research on machine translatability and post-editing is provided by O´Brien (2005).

Several productivity tests have been performed in translation and localization industry settings at Microsoft, Adobe, Autodesk and others. The Microsoft Research trained SMT on MS tech domain and used it for Office Online 2007 localization into Spanish, French and German. By applying MT to all new words, on average a 5-10% productivity improvement was gained (Schmidtke, 2008).

In experiments performed by Adobe, about 200,000 words of new text were localized using rule-based MT for translation into Russian (PROMT) and SMT for Spanish and French (Language Weaver). Authors reported an increase of translator's daily output by 22% to 51% (Flournoy and Duran, 2009). They also found that quality of MT output varied significantly: while some sentences needed no editing and others required full retranslation.

At Autodesk, a Moses SMT system was evaluated for translation from English into French, Italian, German and Spanish (Plitt and Masselot, 2010). To measure translation time, a special workbench was designed to capture keyboard and pause times for each sentence. Authors reported that although by using MT all translators worked faster, it was in varying proportions from 20% to 131%.

For many years, the Directorate General for Translation (DGT) of the European Commission has probably been the largest user of MT. In 2010, DGT launched its MT@EC project to work on Moses-based SMT for all official EU languages. In July 2013, the first versions of MT@EC systems were released for use in everyday work of translators. The translator's survey (Fontes, 2013) showed that most of MT engines were rated as *'many words or partial phrases reusable with acceptable editing'*. Another conclusion was made regarding quality. According to the feedback for some translation directions, MT quality was excellent (e.g. English-Swedish) but useless for translation from English to Estonian and Hungarian (Verleysen, 2013).

We started our experiments in 2011 with a simplified scenario (Skadiņš et al., 2011). In the following years we extended this evaluation with new languages as described in Section 4 and made a numerous improvements followed by other evaluation experiment as described in Section 5.

3 Methodology

The aim for our experiments was to assess MT impact on translator's productivity and translation quality in a typical localization scenario. For MT application to be useful it has to bring significant improvement in the productivity of translation process - decrease the total time spent on translation while keeping the required level of quality. To assess this we measure:

- translator's productivity,
- quality of translation,
- time spent identifying and correcting errors in the translations.

Unlike in many other post-editing experiments (e.g. Plitt and Masselot, 2010; Teixeira, 2011) where automatic tools were used to measure time spent on individual activities, to log translator key strokes, etc., we evaluated productivity and quality in realistic working environment. In both localization companies, we applied the typical everyday translation workflow using the same tools for process management, time reporting and quality checking as in everyday work.

We ran experiments in two scenarios:

Scenario 1. Translation using TM only (the baseline scenario).

Scenario 2. Translation using TM and MT; MT suggestions are provided for every translation unit that does not have a 100% match in TM.

For training and running SMT systems we used the cloud-based platform LetsMT (Vasiļjevs et al., 2012).

3.1 Data for evaluation

Evaluation was made in the software localization domain for translations from English into target language(s). In this domain, the same sentences frequently appear in different texts (e.g., "Open file") and translators receive such translations (or translations of closely matching sentences) from translation memories of previously translated projects. To take this into account, the following criteria were applied in selecting the source text (documents) for evaluation:

- the documents have not been translated in the organization before;
- about 50% of the documents contain at least 95% new words (texts in less used sub-domain, TM does not contain many segments from this sub-domain);
- about 50% of documents contain sentences with different level of fuzzy matches (texts in typical sub-domains, TM contains segments from this sub-domain).
- The size of each document has to be about 1,000 weighted words on average.

The *weighted word count* is a metric widely used in localization; it means the word count adjusted to take into account the translation effort required. The translator spends less time checking or revising a sentence that has already been translated (exact or fuzzy matches to translation memory) than translating a new sentence (no match in the translation memory). The number of words in the document is therefore "weighted" by the matching rate to the translation memory.

All documents were split into 2 equally sized parts to perform two translation scenarios described above. Texts were selected from user assistance and user interface sub-domains. In the first experiment the following requirements were applied for the selection of the test set:

- Only plain text documents containing no formatting tags,
- Documents related to the topics of the data on which the SMT systems are trained (thus ensuring in-domain translation characteristics of SMT translation suggestions),
- Documents with a similar style and terminology as in the training data used for generating SMT.

For the second experiment a different test set was selected:

- Documents containing text with a mark-up (formatting or tags, placeholders, etc.),
- Documents have to be in the same domain as the data on which the SMT systems were trained, but sub-domains may differ,
- Documents that have different style and terminology to the training data.

The different approaches in the selection of the test sets make the two experiments not comparable. But that was to be expected, as the goals of the two experiments differ significantly.

3.2 Evaluation Process

The evaluation process was the same for all languages. At least 5 translators were involved with different levels of experience and average (or above average) productivity. All translators were trained to use MT systems and SDL Trados Studio 2009 or 2011 in their translation work before the evaluation process started.

In both scenarios, translators were allowed to use whatever external resources they needed (dictionaries, online reference tools, etc.), just as during regular operations.

Translators performed the test without interruption and without switching to other translation tasks during their working day – 8 hours – because splitting the time into short periods would not show reliable evaluation results. Each scenario was performed on a different working day. The time spent for translation was manually reported.

To avoid any "start-up" impact, in *Scenario 2* we removed from the result analysis the first translation task performed by each translator.

3.3 Productivity and Quality Assessment

The translator's productivity was calculated as a number of weighted words translated per hour.

The translation quality for each document was evaluated by at least 2 experienced editors. Editors were not aware of the scenario used (whether MT was applied or not). Editors reported the time spent on identifying and correcting errors and quality assessment. There was no inter-editor (inter-annotator) agreement measured, as this is not an everyday practice in localization.

The quality of translation is measured by filling in a Quality Assessment (QA) form in accordance with the QA methodology based on the Localization Industry Standards Association (LISA) QA model[4]. The evaluation process involves inspection of translations and classifying errors according to the error categories.

The productivity and quality of work was measured and compared for every individual translator. An error score was calculated for every translation task by counting errors identified by the editor and applying a weighted multiplier based on the severity of the error type. The error score is calculated per 1,000 weighted words and is calculated as:

$$ErrorScore = \frac{1000}{n} \sum_i w_i e_i$$

[4] LISA QA model: http://web.archive.org/web/20080124014404/http://www.lisa.org/products/qamodel/

where n is a weighted word count in a translated text, e_i is a number of errors of type i and w_i is a coefficient (weight) indicating severity of type i errors.

There are 15 different error types grouped in 4 error classes: accuracy, language quality, style and terminology. Different error types influence the error score differently because errors have a different weight depending on the severity of error type. For example, errors of type comprehensibility (an error that obstructs the user from understanding the information; very clumsy expressions) have weight 3, while errors of type omissions/unnecessary additions have weight 2.

Depending on the error score the translation is assigned a translation quality grade (Table 1).

Error Score	Quality Grade
0…9	Superior
10…29	Good
30…49	Mediocre
50…69	Poor
>70	Very poor

Table 1. Quality evaluation based on the score of weighted errors

3.4 Tools

The LetsMT (Vasiļjevs et al., 2012) plug-in for the SDL Trados 2009 (or 2011) CAT environment was used in all experiments. It was developed using standard MT integration approach described in SDL Trados SDK.

The plug-in was loaded when the user started SDL Trados Studio. During translation of a document, MT suggestions from the selected MT system are provided as shown in Figure 1.

The *Scenario 1* (baseline) establishes the productivity baseline of the current translation process using SDL Trados Studio when texts are translated unit-by-unit (sentence-by-sentence). The *Scenario 2* measures the impact of MT on the translation process when translators are provided with matches from the translation memory (as in baseline scenario) and with MT suggestions for every translation unit that does not have a 100% match in TM. Suggestions coming from the MT systems are clearly marked; according to Teixeira (2011), identification of suggestion origin helps increase translator performance.

We chose to mark MT suggestions clearly because it allows translators to pay more attention to these suggestions. Usually translators trust suggestions coming from the TM and they make only small changes if necessary. They usually do not double-check terminology, spelling and grammar, because the TM is supposed to contain good quality data. However, translators must pay more attention to suggestions coming from MT, because MT output may be inaccurate, ungrammatical, it may use wrong terminology, etc.

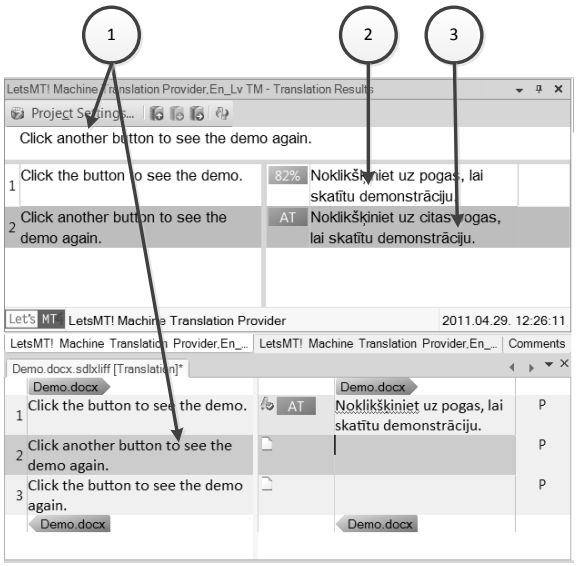

Figure 1. Translation suggestions in SDL Trados Studio; 1 – source text, 2 – a suggestion from the TM, 3 – a suggestion from the MT

4 Experiment 1

A goal of the first experiment was to test hypothesis that MT can be beneficial in a translator's everyday operations and can increase their productivity. The experiment was performed for four language pairs: English-Latvian, English-Polish, English-Czech and English-Hungarian with domain specific SMT systems.

4.1 MT Systems

The MT systems were slightly different for different language pairs depending on available training resources. We used domain specific training data available to the companies participating in the experiment. For English-Latvian MT we used the best available MT system (Skadiņš et al., 2010) that also includes knowledge about Latvian morphology and some out-of-domain publicly available training data, like DGT-TM (Steinberger et al., 2012) and OPUS EMEA (Tiedemann, 2009).

Two different SMT systems where trained for Polish and Czech. The first Polish MT engine (v1) was trained using all available parallel data from localization company production data (data of various clients); the second MT engine (v2)

was trained on smaller client specific data. The first Czech MT engine (v1) was trained using small client specific parallel data from localization company production data and the Czech National Corpus (topic: tech domain)[5]; the second MT engine (v2) was trained using only company production data (data of various clients).

We used the BLEU (Papineni et al., 2002) and METEOR (Denkowski and Lavie, 2011) metrics for the automatic MT system evaluation. The IT domain tuning (2,000 sentences) and testing (1,000 sentences) data were automatically filtered out from the training data before the training process. Table 2 shows details of the MT systems.

MT System	Size (sentences)	Eval. corpus	BLEU score	METEOR score
EN-LV	5.37 M*	IT	70.37	N/A
EN-PL v1	1.5 M	IT	70.47	0.48
EN-PL v2	0.5 M	IT	71.90	0.49
EN-CS v1	0.9 M	IT	67.97	0.46
EN-CS v2	1.5 M	IT	71.60	0.49
EN-HU	0.5 M	IT	59.50	0.41

Table 2. Details of the MT systems and results of automatic MT system quality evaluation. * 1.29 M in-domain data.

4.2 Evaluation Data Sets

The data sets for the productivity evaluation were created by selecting documents in the software localization domain from the tasks that had not been translated by the translators in the organizations before the SMT engines were built. This ensures that translation memories do not contain all the segments of texts used in evaluation.

Documents for translation were selected from the incoming work pipeline if they contained about 1,000 weighted words each. Each document was split in half; the first part was translated as described in the baseline scenario (*Scenario 1*), and the second half of the document was translated using the MT scenario (*Scenario 2*). Every document was entered in the translation project tracking system as a separate translation task. The size of evaluation data set varied from 33 to 54 documents, depending on language pair.

All MT systems used in the evaluation were trained using specific vendor translation memories as a significant source of parallel corpora. Therefore, the SMT systems may be considered slightly biased to a specific IT vendor, or a vendor specific narrow IT domain. The evaluation set contained texts from this vendor and another vendor whose translation memories were not included in the training of the SMT system. We will refer to these texts as narrow IT domain and broad IT domain for easier reference in the following sections. From 33% to 50% of texts (depending on language pair) translated in each scenario were in broad IT domain.

4.3 Results

The results are assessed by analysing average values of translator's productivity and an error score for translated texts.

Usage of MT suggestions in addition to the use of TMs increased productivity of the translators in all evaluation experiments (Table 3).

MT System	Scenario 1 (1)	Scenario 2 (2)	Increase (3)
EN-LV	550	731	32.9 %
EN-PL v1	305	392	28.5 %
EN-PL v2	294	357	21.5 %
EN-CS v1	315	394	25.1 %
EN-CS v2	291	351	20.8 %
EN-HU	287	339	18.0 %

Table 3. Productivity (weighted words translated per hour) evaluation results. (1) Average productivity, *Scenario 1*, (2) Average productivity, *Scenario 2*, (3) Average productivity increase.

There were significant productivity differences in the various translation tasks. The standard deviation of productivity for English-Latvian evaluation in the baseline and MT scenarios were 213.8 and 315.5, respectively. Significant differences in the results of different translators have been observed; the results for English-Latvian evaluation vary from a 64% increase in productivity to a 5% decrease in productivity for one of the translators. Further analysis is necessary, but most likely the differences are caused by the working patterns and skills of individual translators.

At the same time, the error score increased in all but one evaluation experiments (Table 4) still remaining at the quality grade "Good". We have not performed a detailed analysis of the reasons causing an increase in error score, but this can be explained by the fact that translators tend to trust suggestions coming from the CAT tool and do not sufficiently check them, even if they are marked as a MT suggestion.

[5] Institute of Formal and Applied Linguistics (ÚFAL) http://ufal.mff.cuni.cz

MT System	Error score, Scenario 1	Error score, Scenario 2
EN-LV	**20.2**	28.6
EN-PL v1	**16.8**	23.6
EN-PL v2	26.1	**24.2**
EN-CS v1	**19.0**	27.0
EN-CS v2	**19.0**	25.0
EN-HU	**16.9**	22.9

Table 4. Linguistic quality evaluation results

We also analysed how translator productivity and quality is affected by text domain for English-Latvian language pair. Grouping of the translation results by narrow/broad domain attribute reveals that MT-assisted translation provides a better increase in productivity for narrow domain (37%) than for broad domain texts (24%). Error scores for both text types are very similar – 29.1 and 27.6, respectively. The number of errors for each error class is shown in Table 5.

MT System	Accuracy		Language quality		Style		Terminology	
Scenario	S1	S2	S1	S2	S1	S2	S1	S2
EN-LV	6	9	6	10	3	4	5	7
EN-PL v1	2	4	1	2	3	4	2	3
EN-PL v2	4	4	3	3	4	3	2	3
EN-CS v1	4	6	1	3	3	3	1	2
EN-CS v2	3	5	1	3	2	3	2	3
EN-HU	3	5	2	3	3	4	3	2

Table 5. Comparison by error classes in both *Scenario 1* (S1) and *Scenario 2* (S2).

5 Experiment 2

Although our first experiment showed significant productivity increase, translators were reluctant to use MT in their everyday work. There reason was various mark-ups (tags, placeholders, etc.) which are very frequent in real-life translation segments but were not properly handled by the MT requiring a lot of additional post-editing efforts.

The goal of the second experiment was to evaluate a more complex translation scenario where source documents contain formatting tags, placeholders and differs in used terminology and language style, and thus are slightly out-of-domain for the SMT system than in the previous experiments. We performed this experiment to analyse the LetsMT platform and SMT systems trained on it in a difficult scenario, to find more detailed beneficial aspects of MT usage in localization workflows and to identify areas that require improvements. The experiment was performed for three language pairs: English-Estonian, English-Latvian and English-Lithuanian.

5.1 MT Systems

All three MT systems were trained on proprietary parallel corpora in the IT domain (consisting of user manuals, user interface strings, technical documents, etc.). See Table 6 for the size of the parallel corpora for translation model training.

All systems were trained as typical phrase-based SMT systems using the Moses SMT engine (Koehn et al., 2007) and tuned with the Minimum Error Rate Training (MERT) (Bertoldi et al., 2009). The sentence pairs used for tuning and also automatic evaluation of the SMT systems were randomly extracted from the parallel corpora and manually verified and cleaned by professional translators. The size of the tuning and automatic evaluation data sets were c.a. 2,000 and 1,000, respectively.

Two different English-Latvian MT systems were trained; the second MT system (v2) had much better support for different formatting tags, URLs, numbers and other non-translatable units. The results of the SMT system automatic evaluation are given in Table 6.

MT System	Size (sentences)	BLEU score	METEOR score
EN-LV (v1)	1.70 M	69.57	0.48
EN-LV (v2)	3.80 M	66.98	0.46
EN-LT	2.14 M	59.72	0.43
EN-ET	3.56 M	55.88	0.40

Table 6. Results of automatic MT system quality evaluation for the second experiment.

5.2 Evaluation Data Sets

For all three language pairs of the second experiment, we created the evaluation data sets by selecting documents in the IT domain that had not been translated by the translators before the evaluation. Similarly to the first experiment, this ensured that translation memories did not contain the translatable segments. We also selected documents aiming at different target audiences (system administrators, programmers, everyday users) as well as from vendors contrasting to the ones those translation memories were used in the training of SMT systems (usually having different translation guidelines and writing styles). This ensured that the selected texts were of different linguistic characteristics (including syntax, terminology usage, style, etc.), thus making the

translation task more difficult for the SMT systems.

Documents for translation were selected if they contained c.a. 1,000 weighted words each and had formatting tags (on average in ¼ to ⅓ of all translation segments). Similarly to the first experiment, each document was split in half and the first part was translated by the translators without SMT system support (*Scenario 1*) and the second part of the document – using SMT systems (*Scenario 2*). Altogether 100 documents were translated for each language pair by 5 professional translators. Every document was entered in the translation project tracking system as a separate translation task.

Documents for the experiment were selected from four different topics: (1) tablet computer manuals (aimed at general public); (2) programming language manuals (aimed at programmers); (3) navigations software manuals (aimed at general public); and (4) networking system set-up manuals (aimed at system administrators).

5.3 Results

Following the evaluation procedure of the first experiment, we analysed the average values for productivity and the error score for translated texts. We also asked translators to provide system-performance related feedback for more detailed analysis of the experiment.

Language pair	**Productivity changes**	**Standard deviation changes in %**
EN-LV (v1)	-3.10% ± 5.76%	20.80%
EN-ET	-4.70% ± 7.53%	27.17%
EN-LT	-3.76% ± 8.11%	29.28%

Table 7. Productivity changes from *Scenario 1* to *Scenario 2* with a 95% confidence interval

Bearing in mind the complexity of this experiment (formatting tags, more complex language and slight subdomain deviations from the data the SMT system is trained on), the results suggest that the average productivity slightly decreases for all language pairs; however, this cannot be statistically proved in a 95% confidence interval (as shown in Table 7). The large confidence interval is caused by the significant productivity differences (as shown by the changes of the standard deviation of productivity) in the various translation tasks. The average translator productivity with a 95% confidence interval in both translation scenarios is given in Table 8.

Language pair	**Scenario 1**		**Scenario 2**	
	Average productivity	Standard deviation	Average productivity	Standard deviation
EN-LV	576 ± 47	171	558 ± 49	178
EN-ET	470 ± 49	178	448 ± 40	143
EN-LT	728 ± 87	314	700 ± 67	240

Table 8. Average translator productivity and standard deviation of productivity results.

The quality review results for all three language pairs are given in Table 9. The results show a minor decrease of translation quality, from 18.7 to 23.0 points for English-Latvian and from 17.0 to 22.7 points for English-Lithuanian. For English-Estonian the quality of translated texts slightly increased (from 12.9 to 12.0), which is mainly because of "Superior" quality rating for two translators. Although for two language pairs we see a slight drop, the quality evaluation grade is still in the level "Good", which is acceptable for production.

Language pair	**Error score Scenario 1**	**Error score Scenario 2**
EN-LV (v1)	**18.7**	23.0
EN-LT	**17.0**	22.7
EN-ET	12.9	**12.0**

Table 9. Linguistic quality evaluation results of the second experiment

After evaluation, translators submitted informal feedback describing their SMT post-editing experience. Three main directions for further improvements were evident:

- In many cases segments with formatting tags were not translated correctly due to limitations and errors in our implementation of the tag translation functionality.
- As every segment was sent to MT system only at the time of its translation, translators had to wait up to 3 sec. while SMT translation suggestion was provided. Pre-translation or increase of MT speed would solve this problem.
- SMT made a lot of errors in handling and translating named entities, terminology, numbers, non-translatable phrases (e.g., URLs, file paths, etc.).

Since the second experiment, we have actively worked to address the issues raised by the translators. Bugs in the tag translation framework have been fixed, specific non-translatable named entity (e.g., directory paths, URLs, number sequences, etc.) as well as some structured named entity (e.g., dates, currencies) handling has been

implemented in the LetsMT platform, and most importantly SMT pre-translation was enabled for the translators. Our preliminary analysis on a small-scale evaluation scenario (following the guidelines of the second experiment) for English-Latvian with two involved translators and 16 translation tasks (8 translation tasks per scenario) shows that the average productivity using the improved LetsMT platform increases from 16.7% up to 35.0% (with a 95% confidence interval) when using SMT support over manual translation without SMT support. This suggests that even for very difficult scenarios SMT systems can be beneficial and lead to significant productivity increases.

Acknowledgements

The research leading to these results has received funding from the research project "2.6. Multilingual Machine Translation" of EU Structural funds, contract nr. L-KC-11-0003 signed between ICT Competence Centre and Investment and Development Agency of Latvia.

References

Berry M. 1997. Integrating Trados translator's workbench with Machine Translation. *Proceedings of Machine Translation Summit VI.*

Bertoldi N, Haddow B, Fouet J B. 2009. Improved Minimum Error Rate Training in Moses. *The Prague Bulletin of Mathematical Linguistics*, Vol. 91 (2009). Prague, Czech Republic, 7-16.

Bisbey R, and Kay M. 1972. The MIND translation system: a study in man-machine collaboration. *Tech. Rep. P-4786*, Rand Corp.

Bruckner C. and Plitt M. 2001. Evaluating the operational benefit of using machine translation output as translation memory input. *MT Summit VIII, MT evaluation: who did what to whom (Fourth ISLE workshop)*, 61–65.

Denkowski M, and Lavie A. 2011. Meteor 1.3: Automatic Metric for Reliable Optimization and Evaluation of Machine Translation Systems. *Proceedings of the EMNLP 2011 Workshop on Statistical Machine Translation.*

Flournoy R, and Duran C. 2009. Machine translation and document localization at Adobe: From pilot to production. *MT Summit XII: Proceedings of the Twelfth Machine Translation Summit*. Ottawa, Canada.

Fontes H L. 2013. Evaluating Machine Translation: preliminary findings from the first DGT-wide translators' survey. *Languages and translation*, 02/2013 #6, 6-9.

Kay M. 1980. The proper place of men and machines in language translation. *Tech. Rep. CSL-80-11*. Xerox Palo Alto Research Center (PARC).

Koehn P, Federico M, Cowan B, Zens R, Duer C, Bojar O, Constantin A, Herbst E. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. *Proceedings of the ACL 2007 Demo and Poster Sessions*. Prague, 177-180.

O´Brien S. 2005. Methodologies for measuring the correlations between post-editing effort and machine translatability. *Machine Translation*, 19(1):37–58.

Papineni K, Roukos S, Ward T, Zhu W. 2002. BLEU: a method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting of the Association of Computational Linguistics (ACL).*

Plitt M, and Masselot P. 2010. A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context. The Prague Bulletin of Mathematical Linguistics, 93 (January 2010), 7–16.

Schmidtke D. 2008. Microsoft office localization: use of language and translation technology.

Skadiņš, R., Goba, K., & Šics, V. 2010. Improving SMT for Baltic Languages with Factored Models. In *Proceedings of the Fourth International Conference Baltic HLT 2010, Frontiers in Artificial Intelligence and Applications, Vol. 2192* (pp. 125–132). Riga: IOS Press.

Skadiņš, R., Puriņš, M., Skadiņa, I., Vasiļjevs, A. 2011. Evaluation of SMT in localization to under-resourced inflected language. *Proceedings of the 15th International Conference of the European Association for Machine Translation EAMT 2011*, 35-40.

Steinberger R, Eisele A, Klocek S, Pilos S, Schlüter P. 2012. DGT-TM: A freely Available Translation Memory in 22 Languages. *Proceedings of the 8th international conference on Language Resources and Evaluation (LREC'2012)*, Istanbul.

Teixeira, C. 2011. Knowledge of provenance and its effects on translation performance in an integrated TM/MT environment. *Proceedings of the 8th International NLPCS Workshop - Special theme: Human-Machine Interaction in Translation*, 107-118.

Tiedemann J. 2009. News from OPUS - A Collection of Multilingual Parallel Corpora with Tools and Interfaces. *Recent Advances in Natural Language Processing* (vol V). John Benjamins, Amsterdam/Philadelphia, 237-248.

Vasiļjevs, A., Skadiņš, R., & Tiedemann, J. (2012). LetsMT!: a cloud-based platform for do-it-yourself machine translation. *Proceedings of the ACL 2012 System Demonstrations* (pp. 43–48). Jeju Island, Korea: Association for Computational Linguistics.

Verleysen P. 2013. MT@Work Conference: by practitioners for practitioners. *Languages and translation*, 02/2013 #6, 10-11.