LetsMT!

Platform for Online Sharing of Training Data and Building User Tailored MT

www.letsmt.eu/

Project no. 250456

Deliverable D6.4
Evaluation report on usability of SMT in CAT tools

Version No. 2.0

31/08/2012

Document Information

Deliverable number:	D6.4
Deliverable title:	Evaluation report on usability of SMT in CAT tools
Due date of deliverable according to DoW:	31/08/2012
Actual submission date of deliverable:	31/08/2012
Main Author(s):	Nansija Lībiete (Tilde), Raivis Skadiņš (Tilde), Mārcis Pinnis (Tilde), Josef Šarman (MOR), Tomáš Hudik (MOR)
Participants:	Andrejs Vasiļjevs, Māris Puriņš, Inguna Skadiņa, Indra Sāmīte
Reviewer	Tilde
Workpackage:	WP6
Workpackage title:	MT usage in localisation: facilities and evaluation
Workpackage leader:	MOR
Dissemination Level:	PU
Version:	V2.0
Keywords:	Evaluation, localization, CAT tools

History of Versions

Version	Date	Status	Name of the Author (Partner)	Contributions	Description/Approval Level
0.1	24.01.2012.	draft	Tilde	-	Document structure and description of evaluation methodology prepared.
0.2	22.02.2012.		Tilde	-	Evaluation results from Tilde. Conclusions from Tilde evaluation.
0.3	28.02.2012.		MOR	-	Evaluation results from Tilde. Conclusions from Tilde evaluation.
1.0	29.02.2012.		Tilde	-	Summary of Evaluation results.
2.0	31.08.2012		MOR, Tilde	-	Final

EXECUTIVE SUMMARY

LetsMT! system integration with CAT tools implemented in Task 6.1 has been evaluated. Industry partners Moravia and Tilde evaluated this application scenario and demonstrated its impact on the software localisation process and professional translators' daily work. Quality, usability, increase of productivity of translation process has been evaluated. Building of domain and project tailored SMT systems for localisation purposes also has been evaluated.

Above mentioned evaluations have been performed the two times during project. Results of the first evaluation have been used to improve translation quality in general as well as MT usability in CAT tools. The second evaluation looks into a more complex evaluation considering tag translation and its effects on the translation performance.

Table of Contents

1 Introduction 6
2 Methodology for Evaluation of Machine Translation in Localization 8
2.1 Evaluation approach 8
2.2 Scenarios 8
2.3 Test set selection 9
2.4 Test 10
2.4.1 Translation performance and quality assessment 11
2.5 Tools 12
3 Evaluation results 14
3.1 Experiment 1 results 14
3.1.1 English-Latvian 14
3.1.2 English-Polish 17
3.1.3 English-Czech 18
3.2 Experiment 2 (part 1) results 19
3.2.1 English-Polish 20
3.2.2 English-Czech 21
3.2.3 English-Hungarian 23
3.3 Experiment 2 (part 2) results 24
3.3.1 Test set 24
3.3.2 SMT systems 25
3.3.3 Results 28
4 Conclusions and Future work 33
References 34
Appendix 1 35
Appendix 2 36
Appendix 3 40
Appendix 4 41
Appendix 5 43
Appendix 6 45
Appendix 7 47
Appendix 8 49
Appendix 9 51
Appendix 10 53
Appendix 11 56
Appendix 12 59

1 Introduction

Growing pressure to reduce translation costs and to increase translation volumes motivates the localization industry to embrace machine translation in addition to other widely used computer assisted translation tools (CAT).

For several decades the most widely used CAT tools in the localization industry have been Translation Memory systems (TM). Since Translation Memories contain fragments of previously translated texts, they can significantly improve the efficiency of localization in cases when the new text is similar to the previously translated material. However, if the text is in a different domain than the TM or in the same domain from a different customer using different terminology, support from the TM is minimal.

The localization industry has experienced increased pressure to provide more efficient and better performing products, particularly due to the fact that volumes of texts that need to be translated are growing at a greater rate than the availability of human translation, and translation results are expected in real-time. For this reason the localization industry is increasingly interested in combining translation memories with machine translation solutions adapted for the particular domain or customer requirements.

Benefits of the application of machine translation in the localization industry are recognized by developers of TM systems. Some developers have already integrated MT in their products or provide such solutions for MT developers. For instance, SDL Trados Studio 2009 supports 3 machine translation engines: SDL Enterprise Translation Server, Language Weaver, and Google Translate. ESTeam TRANSLATOR and Kilgrey's memoQ are other systems providing integration of MT.

For the development of MT in the localization and translation industry, huge pools of parallel texts in a variety of industry formats have been accumulated. The most successful data collection effort is the online repository of TM data by the TAUS Data Association. However, the use of this data alone does not fully utilize the benefits of modern MT technology.

Although the idea to use MT in the localization process is not new, it has not been explored widely in the research community. Different aspects of post-editing and machine translatability have been researched since the nineties (e.g., Berry 1997, Bruckner and Plitt 2001). A comprehensive overview of research on machine translatability and post-editing has been provided by O´Brien (2005). However this work mainly concentrates on the cognitive aspects, not so much on productivity in the localization industry.

Increasing the efficiency of the translation process without a degradation of quality is the most important goal for a localization service provider.

In recent years several productivity tests have been performed in translation and localization industry settings at Microsoft (Schmidtke, 2008), Adobe (Flournoy and Duran, 2009) and Autodesk (Plitt and Masselot, 2010).

The Microsoft Research trained SMT on MS tech domain was used for 3 languages for Office Online 2007 localization: Spanish, French and German. By applying MT to all new words, on average a 5-10% productivity improvement was gained.

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%.

At Autodesk, a Moses SMT system was evaluated for translation from English to French, Italian, German and Spanish by three translators for each language pair. 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%. They concluded that MT allowed translators to improve their throughput on average by 74%.

This document describes the methodology used for MT evaluation in localization in the LetsMT! project and results of two experiments for using translation SMT integrated into TM in a professional localization company.

The first experiment evaluates the application of English-Latvian, English-Czech and English-Polish SMT in localization using the LetsMT! plug-in into the SDL Trados 2009 translation environment measuring the performance of a translators translating with and without MT. The goal of the first experiment was to evaluate the change of translator productivity by adding in-domain SMT support for plaintext translation. The first experiment was carried out early in 2012.

The second experiment has two goals: (1) to confirm that in-domain translations of plaintext documents with added SMT support really do increase productivity when using MT systems trained explicitly on the LetsMT! platform, and (2) to evaluate a more complex translation scenario where translatable documents are slightly out of domain, contain formatting tags and contain more complex language than in the previous experiment. The first part of the second experiment was carried out on English-Czech, English-Hungarian, and English-Polish language pairs and the second part of the second experiment was carried out on English-Latvian, English-Lithuanian, and English-Estonian language pairs.

A quality assessment for texts was also performed in all experiments according to the standard internal quality assessment procedure.

2 Methodology for Evaluation of Machine Translation in Localization

This procedure describes the process and requirements of evaluation of LetsMT! machine translation (MT) in localization scenario.

2.1 Evaluation approach

Evaluation of MT is based on:

1. the measurement of translation performance or productivity,
2. the measurement of translation quality,
3. the time spent for identifying and correcting errors in the translations.

MT systems will be tested against productivity and quality of day-to-day translation using translation memories (TM).

2.2 Scenarios

Translations are performed in SDL Trados Studio 2009 CAT tool environment. There are 2 scenarios:

1. Translation using TM only (baseline; shown in Figure 1).
2. Translation using TM and MT (shown in Figure 2).

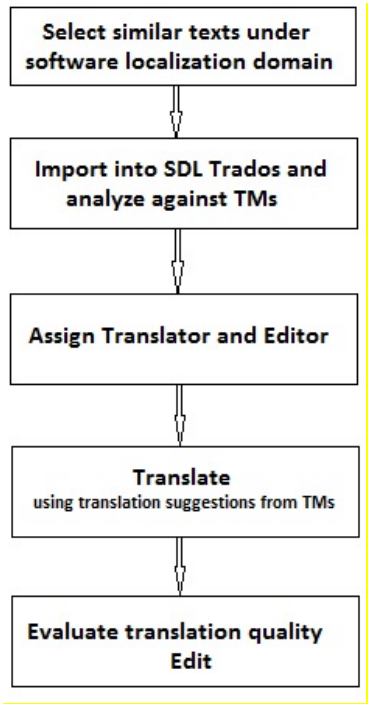

Figure 1. Scenario 1

MT suggestions are provided for every translation unit that does not have a 100% match in TM. Suggestions coming from the MT systems are clearly marked.

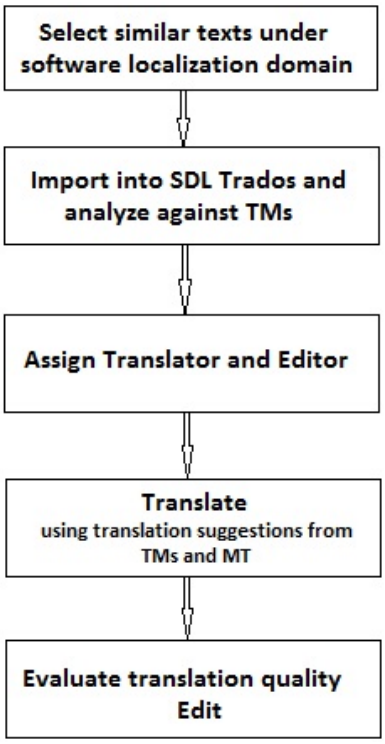

Figure 2. Scenario 2

2.3 Test set selection

Evaluation is made in the software domain for translations from English into target language(s). Translations for evaluation are selected from texts that have not been translated in the organization before.

The texts (documents) should be selected so that ca. 50 % documents contain at least 95% of new words (texts in less used sub-domain, TM does not contain many segments from this sub-domain) and ca. 50% documents contain different fuzzy matches (texts in typical sub-domains, TM contains many segments from this sub-domain).

Texts for tests should be used from the following sub-domains of software localization:

- User assistance
- User interface (should not be included in corpora)

For the first experiment and the first part of the second experiment the following requirements are set to the test set selection:

- Documents without mark-up (plaintext documents) should be used.
- All documents have to be split into 2 equally sized parts to perform the two translation scenarios described below. The first part of a document has to be translated as per scenario 1 and the second part of a document - as per scenario 2.
- The volume of each part of a document has to be 500 (±5%) weighted words on average.

- The selected documents have to be related to the topics of the data on which the SMT systems are trained on (thus insuring in-domain translation characteristics of SMT translation suggestions).
- The documents also have to feature similar language constructions (for instance, sentence syntax, style, terminology usage, etc.) to the training data.

For the first part of the second experiment the same requirements to the test set selection as to the first experiment are set.

For the second part of the second experiment the following requirements are set to test set selection:

- Documents containing text with mark-up (formatting or tags) should be used; markup should be included in ¼ to ⅓ of translation segments.
- All documents have to be split into 2 equally sized parts to perform the two translation scenarios described below. The first part of a document has to be translated as per scenario 1 and the second part of a document as per scenario 2.
- The volume of each part of a document has to be 500 (±5%) weighted words on average.
- The selected documents have to be in the same domain as the data on which the SMT systems are trained on, but the sub-domains (or topics) may be different (for instance, productivity software user manuals vs. navigation software user manuals).
- The documents may also feature different language constructions (for instance, sentence syntax, style, terminology usage, etc.) to the training data.

The different test set selection approaches make the two parts of the second experiment not comparable, but that was to be expected as goals of the two parts differ significantly.

2.4 Test

The evaluation process involves at least 5 translators with different levels of experience and average productivity performance. All translators are well trained to use the MT systems and SDL Trados Studio 2009 in their translation work before measuring their performance in evaluation.

Translators are allowed to use external resources (dictionaries, online reference tools, etc.), just as during regular operations.

Translators perform the test without interruption and switching to other translation tasks on their working day – 8 hours, because splitting the time into short periods would not show trustable performance results. The time spent for translation is reported to the nearest minute.

The first (document) translation made by each translator in scenario 2 should be removed from the result analysis to avoid any “start-up" impact.

Each scenario (scenario 1 and scenario 2) is performed on different working days.

Translators fill in a questionnaire (Appendix 1) when scenario 2 has been completed for each document.

2.4.1 Translation performance and quality assessment

After a document is translated it is evaluated for translation performance and translation quality by editors. The evaluation process involves at least 2 experienced editors. Editors are not aware of the scenario used by the translators. Editors also report their time spent for identifying and correcting errors of the translations and quality assessment to the nearest minute.

There is no inter-editor (inter-annotator) agreement as it is not an everyday practice in localization.

The measurement of translation performance is calculated as a number of weighted words translated per hour. Weighted wordcount is a CAT tool count of words that applies percentage according to various types of matches (new words, fuzzy matches, repetitions, 100% matches). The percentage used in the evaluation is shown in Table 1.

Table 1. The percentage used to calculate weighted word count

CAT categories	Count
new words	100%
50% - 74% matches	100%
75% - 84% matches	50%
85% - 94% matches	50%
95% - 99% matches	30%
repetitions	10%
100% matches	10%

Quality of translation is measured by filling in a Quality Assessment (QA) form in accordance with the Tilde QA methodology (*Appendix 2*) based on the industry standard – the Localization Industry Standards Association (LISA) QA model[1]. QA methodology provides a method of measuring the quality of translation. The evaluation process involves inspection of translations and classifying errors according to the following error categories:

- Accuracy
- Language quality
- Style
- Terminology

Preferential changes are not considered as errors.

Performance and quality of work in every of the two translation scenarios is measured and compared for every individual translator. Individual productivity of each translator in the

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

test is measured and compared against his or her own standard productivity. An error score is calculated for every translation task. The error score is a metric calculated 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 it is calculated as:

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

where

n is a number of weighted words in a translated text,

e_i is a number of errors of type *i*,

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 (*Appendix 2*). 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: Superior, Good, Mediocre, Poor, or Very poor (*Table 2*).

Table 2. Quality evaluation based on the score of weighted errors

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

Editors perform quality assessment by marking error categories electronically in the text and filling in a QA form for each translation. Editors inform the project manager when QA is completed.

2.5 Tools

For application in the localization scenario, LetsMT! provides a plug-in for the SDL Trados 2009 (or 2011) CAT environment to use created MT systems. The MT systems are running on the LetsMT! platform and are accessible using a web service interface based on the SOAP protocol. Connectivity with additional localisation environments will be ensured by providing web services for further integration efforts either by partners or the user community of the LetsMT! service.

The plug-in has been developed using standard MT integration approach described in SDL Trados SDK. It has been written in .NET (C#), using .NET framework 3.5. The setup is compiled using Nullsoft Install System (NSIS).

To use the plug-in, the user needs to download a setup file from the LetsMT! website (https://www.letsmt.eu/Integration.aspx) and run it. When the user starts SDL Trados Studio the plug-in is loaded. Machine translation suggestions from the selected LetsMT! system appears on screen during the translation of the document or can be used to pre-translate documents in the batch process. A SMT system must be specified manually for each language direction.

The baseline scenario establishes the productivity baseline of the current translation process using SDL Trados Studio 2009 when texts are translated unit-by-unit (sentence-by-sentence). The MT scenario measured the impact of using MT in the translation process when translators are provided with not only matches from the translation memory (as in baseline scenario), but also with MT suggestions for every translation unit that does not have a 100% match in translation memory. Suggestions coming from the MT were clearly marked (Figure 3).

We chose to mark MT suggestions clearly because it allows translators to pay more attention to these suggestions. Typically translators trust to suggestions coming from the TM and they make only small changes if it is not a 100% match. Translators are not double-checking terminology, spelling and the grammar of TM suggestions, because the TM contains good quality data. However, translators must pay more attention to suggestions coming from MT, because MT output may be inaccurate, ungrammatical, it may use the wrong terminology, etc.

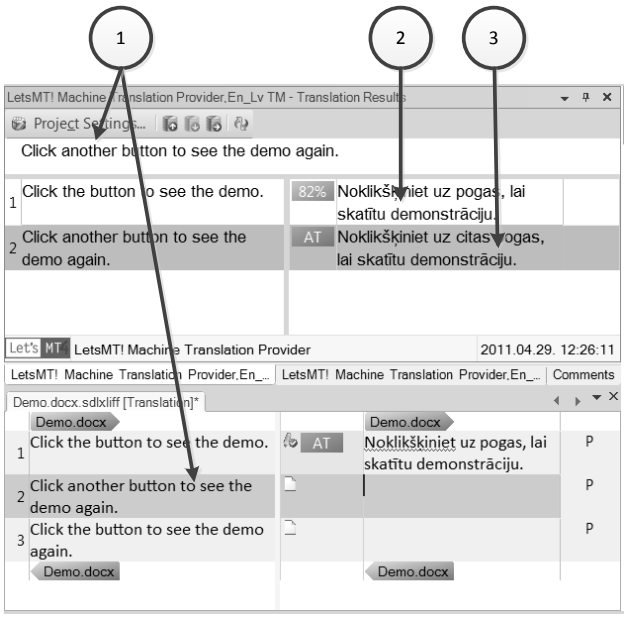

Figure 3. Translation suggestions in SDL Trados Studio 2009; 1 – source text, 2 – a suggestion from the TM, 3 – a suggestion from the MT.

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

3 Evaluation results

3.1 Experiment 1 results

The first experiment (finished early in 2012) is an initial evaluation of MT applicability in software localisation. The goal of the experiment was to prove that MT can be beneficial in translator everyday operations and that it can lead to increased translation productivity. The experiment has been performed for three language pairs: English-Latvian, English-Polish and English-Czech. Results of the initial evaluation are given in the following subsections.

3.1.1 English-Latvian

3.1.1.1 SMT system

The total size of the English-Latvian parallel data used to train the translation model is 5.37 M sentence pairs (Table 3). The parallel corpus includes publicly available DGT-TM[2] (1.06 M sentences) and OPUS EMEA (0.97 M sentences) corpora (Tiedemann, 2009), as well as a proprietary localization corpus (1.29 M sentences) obtained from translation memories that were created during the localization of interface and user assistance materials for software and user manuals for IT&T appliances. To increase word coverage, word and phrase translations were included from bilingual dictionaries (0.51 M units) from reliable sources with high quality. A larger selection of parallel data was used which was automatically extracted from comparable web corpus (0.9 M sentences) and from 104 works of fiction (0.66 M sentences).

Table 3. Bilingual corpora for English-Latvian system

Bilingual corpus	Parallel units
Localization TM	~1.29 M
DGT-TM	~1.06 M
OPUS EMEA	~0.97 M
Fiction	~0.66 M
Dictionary data	~0.51 M
Web corpus	~0.9 M
Total	**5.37 M**

The monolingual corpus was prepared from news articles from the Web and the monolingual part of the parallel corpora. The total size of the Latvian monolingual corpus was 391 M words (Table 4).

[2] http://langtech.jrc.it/DGT-TM.html

Table 4. Latvian monolingual corpora

Monolingual corpus	Words
Latvian side of parallel corpus	60 M
News (web)	250 M
Fiction	9 M
Total, Latvian	**319 M**

Since Latvian belongs to the class of highly inflected languages with a complex morphology, the SMT system was extended within the Moses (Koehn et al., 2007) framework by integrating morphologic knowledge (Skadiņš et al., 2010). The high inflectional variation of the target language increases data sparseness at the boundaries of translated phrases and a language model over surface forms might be inadequate to reliably estimate the probability of target sentences.

We used the BLEU (Papineni et al., 2002) metric for automatic evaluation. The BLEU score of the SMT system is 35.0 evaluating on a general domain balanced evaluation set and 70.37 evaluating on an IT domain evaluation set. The detailed description of test and development sets and system comparison to other English-Latvian systems are given by Skadiņš et al. (2010).

3.1.1.2 Test set

The test set for the evaluation was created by selecting documents in the IT domain from the tasks that have not been translated by the translators in the organization before the SMT engine was built. This ensures that translation memories do not contain all the segments of texts used for testing

Documents for translation were selected from the incoming work pipeline if they contained 950-1,050 adjusted words each. Each document was split in half and the first part of it was translated as described in the baseline scenario and the second half of the document – using the MT scenario.

Altogether 54 documents were translated. Every document was entered in the translation project tracking system as a separate translation task. An adjusted word is a metric used for quantifying work to be done by translators. Larger documents were split into several fragments.

Although a general purpose SMT system was used, it was trained using specific vendor translation memories as a significant source of parallel corpora. Therefore, the SMT system may be considered slightly biased to a specific IT vendor, or a vendor specific narrow IT domain. The test set contained texts from this vendor and another vendor whose translation memories were not included in the training of the SMT system. We will call these texts as *in narrow IT domain* and *in broad IT domain* for easier reference in the following sections. Approximately 33% of texts translated in each scenario were *in broad IT domain*.

3.1.1.3 Results

The results were analyzed for 46 translation tasks (23 tasks in each scenario) by analyzing average values for translation performance (translated words per hour) and an error score for translated texts.

Usage of MT suggestions in addition to the use of the translation memories increased productivity of the translators in average from 550 to 731 words per hour (32.9% improvement).

There were significant performance differences in the various translation tasks; the standard deviation of productivity in the baseline and MT scenarios were 213.8 and 315.5 respectively.

At the same time the error score increased for all translators. Although the total increase in the error score was from 20.2 to 28.6 points, it still remained at the quality evaluation grade "Good". We have not performed detailed analysis of reasons causing error score increase yet, but it can be explained by the fact, that translators are tended to trust suggestions coming from the CAT tool and they are not double checking them even if they are marked as MT suggestion.

Grouping of the translation results by narrow/broad domain attribute reveals that MT-assisted translation provides better increase in translation performance 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.

Grouping of errors identified by error classes reveal the increase of number of errors shown in Table 5.

Table 5. Comparison by error classes, English-Latvian

Error Class	Baseline scenario	MT scenario
Accuracy	6	9
Language quality	6	10
Style	3	4
Terminology	5	7

There were significant differences in the results of different translators from performance increase by 64% to decreased performance by 5% for one of the translators.

Analysis of these differences requires further studies but most likely they are caused by working patterns and the skills of individual translators.

Detailed results of evaluation for English-Latvian are given in Appendix 4.

3.1.2 English-Polish

3.1.2.1 SMT System

English-Polish translation engine was trained on 1.5M parallel sentences from Moravia's production data (data of various clients). All the clients were IT companies. The same data was used as a source for monolingual corpus.

The engine was trained without any additional adjustments and parameters, it is a baseline. This means tuning, as well as testing set were filtered out before the training started. Tuning set contained 2000 sentences, while testing set contained 1000 randomly selected sentences (segments). The trained engine achieved: 70.47 BLEU and 0.4812 METEOR score.

3.1.2.2 Test set

The test set for evaluation of the English – Polish engine was created from Moravia's production data. All the documents belong to IT domain and have not been translated in the organization before.

Segments for translation were taken from real-project data. All documents were divided into fragments with similar size of weighted word count - around 500 words. For every single document half of its fragments were translated as described in the baseline scenario. The remaining fragments were translated using the MT engine. In total 46 fragments were translated.

Even though the MT engine was trained on Moravia production data, most of the testing documents come from broad IT domain (approx. 60%). Client specific translation memories were incorporated in the translation package. So the translators could use inputs from TMs together with MT suggestions.

3.1.2.3 Results

The results were analyzed for 42 translation tasks (21 tasks in each scenario) by analyzing average values for translation performance (translated words per hour) and an error score for translated texts.

Even though most of the translators reported "poor" quality of MT suggestion the results shows an increased productivity across all documents. The average performance rose from 305 to 392 adjusted words per hour (28.5% improvement).

A significant performance variety has been observed while using MT scenario with 181 words difference compare to 86 under baseline scenario.

Slight decrease of translation quality was recorded. The overall error score increased from 16.8 to 23.6 points. Nevertheless the quality evaluation grade remains "Good". Grouping of errors identified by error classes reveal the increase of number of errors shown in Table 6. Comparison by error classesTable 5. Comparison by error classes, English-Latvian.

Table 6. Comparison by error classes, English-Polish

Error Class	Baseline scenario	MT scenario
Accuracy	2	4
Language quality	1	2
Style	3	4
Terminology	2	3

Results of MT are very sensitive to the training set. The accuracy can be improved by training the engine with more specific data or to have a client dedicated engine.

Language style is the major weakness of automated translations. Even though human translators were supposed to edit the target strings to ensure an appropriate language style is used, Table 6 shows that MT suggestion affected the style in general. Study of this phenomenon and improvement of MT in this area would help to use MT in commercial translations more often.

Detailed results for English-Polish evaluation are attached as Appendix 5.

3.1.3 English-Czech

3.1.3.1 SMT System

English-Czech engine was trained on 0.9M sentences. A larger part (1.6M sentences) was taken from Czech National Corpus (topic: **tech domain**) – Institute of Formal and Applied Linguistics (ÚFAL) - http://ufal.mff.cuni.cz/. And the rest (0.5M sentences) were Moravia's production data – different users, all of them were IT companies. LetsMT! filters out duplicate or somehow damaged segments, therefore engine's size (0.9M sentences) is lower than the sum of its constituents.
This is also base-line system, which means that no additional parameters were used. Tuning (2000 sentences) and testing (1000 sentences) were filtered out before the training process has started. After training, tuning and testing took a place. The trained engine achieved: 67.97 BLEU and 0.4668 METEOR score.

3.1.3.2 Test set

The test set for the evaluation was extracted from Moravia's production data. All the source documents belong to the IT domain and have not been translated in the organization prior the SMT system was trained.

Segments for translation were taken from real-project data. All documents were split into fragments with similar size of weighted word count - around 500 words. For every single document half of its fragments were translated according to the baseline scenario. The remaining part was translated using the MT engine. In total 39 files were translated.

Approximately 70% of the testing content comes from broad IT domain not directly linked with the training data. Therefore client specific translation memories were incorporated in the translation package. Hence the vendors can work with both MT and TM inputs.

3.1.3.3 Results

The results are based on 34 translation tasks (17 tasks in each scenario) by analyzing average values for translation performance (translated words per hour) and an error score for translated texts.

An increase of productivity by 25.1% was captured while using MT scenario. The average volume of adjusted words per hour rose from 315 to 394.

A quality review discovered minor decrease of translation quality from 19 to 27 error points per 1000 words. Nevertheless the quality evaluation grade is still “Good”. Grouping of errors identified by error classes reveal the increase of number of errors shown in Table 7.

Table 7. Comparison by error classes, English-Czech

Error Class	Baseline scenario	MT scenario
Accuracy	4	6
Language quality	1	3
Style	3	3
Terminology	1	2

Detailed analysis of quality degradation might be subject of another study.

Despite the fact that 59% of Czech translators evaluated generally the MT input as inefficient, the evaluation results shows that use of MT system significantly contributes to increase the translation productivity.

Results of MT are very sensitive to the training set. The accuracy can be improved by training the engine with more specific data or to have a client dedicated engine.

Language style is the major weakness of automated translations. Even though human translators were supposed to edit the target strings to ensure an appropriate language style is used, Table 7 shows that MT suggestion affected the style in general. Study of this phenomenon and improvement of MT in this area would help to use MT in commercial translations more often.

Detailed results of English-Czech evaluation are provided below as Appendix 6.

3.2 Experiment 2 (part 1) results

The second experiment (carried out in August, 2012) had two parts. The goal of the first part of the experiment was to confirm results of the first experiment using systems trained explicitly on the LetsMT! platform. The experiment has been performed for three language

pairs: English-Czech, English-Hungarian, and English-Polish. Results of the experiment are given in the following subsections.

3.2.1 English-Polish

3.2.1.1 SMT System

English-Polish translation engine was trained on 0.5M parallel sentences from Moravia's production data (IT domain). All the clients were IT companies. The same data was used as a source for monolingual corpus.

The engine was trained without any additional adjustments and parameters, it is a baseline. This means tuning, as well as testing set were filtered out before the training started. Tuning set contained 2000 sentences, while testing set contained 1000 randomly selected sentences (segments). The trained engine achieved: 71.9% BLEU and 0.49 METEOR score.

3.2.1.2 Test set

The test data used for evaluation of the English – Polish engine were based on Moravia's production data. All of them belong to the IT domain and have not been used or translated in the company before.

All the translation segments are real-project data. For test purposes all of them were split into fragments with similar size of 500 weighted words. For every single document half of its fragments were translated as described in the baseline scenario. The remaining fragments were translated using the MT engine. In total 35 fragments were translated.

Even though the MT engine was trained on Moravia production data, most of the testing documents came from broad IT domain (approx. 50%). Client specific translation memories were included in the translation package. Hence the translators could use inputs from TMs along with MT suggestions.

3.2.1.3 Results

The results were analysed for 35 translation tasks (18 tasks for baseline and 17 for MT scenario). Average values of translation performance (translated words per hour) and error score were analysed.

All the translators evaluated the MT inputs as "average" or even "poor" so most of the text had to be manually edited or removed. On the other hand the results show that productivity rose for all the tested fragments. The average performance rose from 294 to 357 adjusted words per hour (21.5% improvement).

A performance variety has been observed while using MT scenario with 144 words difference compare to 83 for the baseline scenario.

The average translation quality using the MT scenario was slightly better - overall error score rose from 26.1 to 24.2 points. Both results correspond to "Good" language quality. Grouping of errors identified by error classes reveal the increase of number of errors shown in Table 8.

Table 8. Comparison by error classes, English-Polish (experiment 2)

Error Class	Baseline scenario	MT scenario
Accuracy	4	4
Language quality	3	3
Style	4	3
Terminology	2	3

Generally the testing results proved increase of translation productivity while using MT scenario. Issues with decreased language quality that were reported in the first round did not appear during the second test. The difference in productivity comes from the fact that various segments used for evaluation are Marketing based texts which are not easy to translate. Therefore our translators had to remove the MT inputs quite often. On the other hand fragments with easy text strings were translated quickly. To sum up translation with MT scenario could be highly recommended mostly for localization of short strings.

Detailed results for English-Polish evaluation are attached as Appendix 9.

3.2.2 English-Czech

3.2.2.1 SMT System

English-Czech engine was also trained on Moravia production data – approximately 1.5M sentences coming from IT domain (one client).
This is also base-line system, hence no additional parameters were used. Tuning (2000 sentences) and testing (1000 sentences) were filtered out before the training process had started. After that the evaluation took place. The trained engine achieved: 71.6 BLEU and 0.49 METEOR score.

3.2.2.2 Test set

The test set for the evaluation came from production team and was based on real Moravia's production data. All the source documents belong to the IT domain and have not been translated in the organization prior the SMT system was trained.

All the testing documents were split into fragments with similar size of weighted word count - around 500 words. For every single document half of its fragments were translated according to the baseline scenario. The remaining part was translated using the MT engine. In total 33 files were translated.

Most of the testing content comes from broad IT domain not directly linked with the training data. Translation memories were incorporated in each localization package so the vendors could use either MT or TM input.

3.2.2.3 Results

The results are based on 33 translation tasks (17 tasks for Baseline and 16 tasks for MT scenario). Average values for translation performance (translated words per hour) and linguistic quality (error score) were compared.

According to the test results the translation productivity increased by 20.8% while using MT scenario. The average volume of adjusted words per hour rose from 291 to 351.

A quality review discovered minor decrease of translation quality from 19 to 25 error points per 1000 words. So both scenarios keep the output quality on average evaluation grade "Good". The error score difference is 2 points smaller than during the first test run which is a positive trend. Grouping of errors identified by error classes reveal the increase of number of errors shown in Table 9.

Table 9. Comparison by error classes, English-Czech (experiment 2)

Error Class	Baseline scenario	MT scenario
Accuracy	3	5
Language quality	1	3
Style	2	3
Terminology	2	3

Detailed analysis of quality degradation might be subject of another study.

Even though 44% of Czech translators evaluated the MT input as inefficient, the evaluation results shows that use of MT system contributes to increase the translation productivity.

Results of MT are very sensitive to the training set. Various sentence structure and inflection rules make the MT less usable for translation of marketing texts. More monolingual data and continuous training of customized engines should lead to improvements in this area. Accuracy can be improved by training the engine with more specific data or to have a client dedicated engine.

Language style is the major weakness of automated translations. Even though human translators were supposed to edit the target strings to ensure an appropriate language style is used, Table 7 shows that MT suggestion affected the style in general. Study of this phenomenon and improvement of MT in this area would help to use MT in commercial translations more often.

Detailed results for English-Polish evaluation are attached as Appendix 7.

3.2.3 English-Hungarian

3.2.3.1 SMT System

English-Hungarian translation engine was trained on 0.5M parallel sentences coming from Moravia's production data. All the data are from IT domain and various clients. The monolingual corpus has the same origin.

The engine is baseline. No additional adjustments or parameter changes were made. This means tuning, as well as testing set were filtered out before the training started. Tuning set contained 2000 sentences, testing set 1000 randomly selected segments. The trained engine achieved the lowest score compare to other Moravia's engines: 59.5% BLEU and 40.72% METEOR score.

3.2.3.2 Test set

The test data used for evaluation of the English – Hungarian engine were based on Moravia's production data. All of them belong to the IT domain and have not been translated in the company before.

All the segments used for translation were real-project data. Following the same principle as for other languages all documents were split into fragments with similar size of approximately 500 localization units (weighted words). Half of fragments were translated as following the baseline scenario and the remaining part was translated using the MT input. Total number of fragments used for evaluation of Hungarian engine is 36.

As the MT engine was trained on Moravia production data related to one client, the testing documents were collected from different project related to other customers. All of them belong to broad IT domain. Client specific translation memories were included in the translation package. So the linguists could use both TM and MT inputs.

3.2.3.3 Results

The results are based on 36 translation tasks (18 tasks for each scenario). Average values of translation performance (translated words per hour) and error score were analysed.

All the translators reported issues with language quality and accuracy so most of the text had to be manually edited or removed. However the MT input obviously had a positive effect on translation productivity. The average performance for Hungarian rose from 287 to 339 adjusted words per hour (18.0% improvement).

Analysis of linguistic quality for all the segments discovered minor decrease (6 error points difference) while using MT scenario. Nevertheless both scenarios resulted into the "Good" final quality grade. Grouping of errors identified by error classes reveal the increase of number of errors shown in Table 10.

Table 10. Comparison by error classes, English-Hungarian (experiment 2)

Error Class	Baseline scenario	MT scenario
Accuracy	3	5
Language quality	2	3
Style	3	4
Terminology	2	2

Evaluation of English-Hungarian SMT system confirmed the positive impact of MT scenario to translation productivity. Even though the MT outputs could not be considered as final and had to manually edited or deleted by a linguist, the time savings could be significant.

Further improvements of automated translation systems in language quality area would help to propagate this localization principle in the future for all European languages.

Detailed results for English-Polish evaluation are attached as Appendix 8.

3.3 Experiment 2 (part 2) results

The goal of the second part of the experiment is to evaluate a more complex translation scenario where translatable documents contain formatting tags, more complex language (in terms of terminology, sentence structure, writing style, etc.) and thus are slightly out-of-domain for the SMT system than in the previous experiments. We have performed this experiment to analyse the performance of the translation platform as a whole in a difficult scenario, to find more detailed beneficial aspects of MT usage in localisation workflow and to find possible areas of future improvements in various stages of the whole LetsMT! platform (including the CAT tool's plugin). The experiment has been performed for three language pairs: English-Estonian, English-Latvian and English-Lithuanian. Results of the experiment are given in the following subsections.

3.3.1 Test set

For all three language pairs of the second part of the second experiment a fixed test set was used. The test set for the evaluation was created by selecting documents in the IT domain that have not been translated by the translators in the organization before the SMT engine was built. This ensures that translation memories do not contain all the segments of texts used for testing. Also documents aiming at different target audiences (system administrators, programmers, everyday users) as well as from vendors contrasting to the ones on which TMs the SMT systems are trained (usually having different translation guidelines and writing styles) were selected. This ensures that the selected texts are 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 950-1,050 adjusted words each and had formatting tags (in average in ¼ to ⅓ of all translation segments where one

segment usually is one sentence long). Each document was split in half and the first part of it was translated as described in the baseline scenario and the second half of the document – using the MT scenario. Altogether 100 documents were translated. Every document was entered in the translation project tracking system as a separate translation task.

Documents 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). The English language test corpora statistics for the two translation scenarios are given in the following table.

Table 11 English test corpora statistics for the second experiment (part 2)

Document source	Scenario 1			Scenario 2		
	Adjusted words	Translation segments	Tags	Adjusted words	Translation segments	Tags
Tablet manual	6,041	656	661	5,953	616	476
Programming language manual	3,105	244	583	3,037	250	547
Navigation software manual	1,535	166	192	1,517	192	256
Networking set-up manual	4,477	433	396	4,131	413	438
All	**15,158**	**1,499**	**1,832**	**14,638**	**1,471**	**1,717**

3.3.2 SMT systems

In the initial experiment (experiment 1) the SMT systems were trained on out-of-domain (general language or publicly available) and in-domain (software related) corpora. In this experiment we used explicitly in-domain corpora to train the SMT systems. For training of the three systems proprietary parallel and monolingual in-house corpora acquired from software localisation related TMs was used.

3.3.2.1 English-Estonian SMT system

The system's name in the LetsMT! platform is "*LetsMT IT EN-ET v2.2*". The total size of the English-Estonian parallel data used to train the translation model is 3.56 M unique sentence pairs (Table 12). The parallel corpus includes four different versions of in-house translation memory data.

Table 12. Bilingual corpora for English-Estonian system

Bilingual corpus	Parallel units
Tilde Localization TMs EN-ET	2,749,077
Tilde Localization TMs EN-ET v1	3,920,807
Tilde Localization TMs EN-ET v1.extra	134,048
Tilde Localization TMs EN-ET v2	499,811
Total (unique, cleaned and after removal of the tuning and evaluation sets)	**3,562,063**

The monolingual corpus consists of the monolingual Estonian sentences of the parallel corpora that were used for the translation model training. The monolingual corpora (after duplicate removal, cleaning and filtering) consisted of 3,545,670 unique Estonian sentences.

Although Estonian belongs to the class of highly inflected languages with a complex morphology, the SMT system was not extended to incorporate morphological information (as in the initial experiment for English-Latvian). This is to test the performance of SMT systems in respect to current LetsMT! platform's functionality.

For evaluation and tuning respectively 999 and 1990 unique sentence pairs randomly extracted from the parallel corpora (and manually verified and cleaned) were used. The English-Estonian SMT system's automatic evaluation results are included in the table below.

Table 13 English-Estonian SMT system's automatic evaluation results

	BLEU score	NIST score	METEOR score
Case insensitive	55.88	9.2724	0.3963
Case sensitive	54.39	9.0421	0.3416

3.3.2.2 English-Latvian SMT system

The system's name in the LetsMT! platform is "*LetsMT IT EN-LV v2.1*". The total size of the English-Latvian parallel data used to train the translation model is 1.7 M unique sentence pairs (Table 14). The parallel corpus includes three different versions of in-house translation memory data.

Table 14. Bilingual corpora for English-Latvian system

Bilingual corpus	Parallel units
Tilde Localization TMs EN-LV	1,292,850
Tilde Localization TMs EN-LV v1.extra	2,804,994
Tilde Localization TMs EN-LV v2	338,401
Total (unique, cleaned and after removal of the tuning and evaluation sets)	**1,702,827**

The monolingual corpus consists of the monolingual Latvian sentences of the parallel corpora that were used for the translation model training. The monolingual corpora (after duplicate removal, cleaning and filtering) consisted of 1,669,856 unique Latvian sentences. Also for the English-Latvian SMT system morphological information was not incorporated in the system's training process.

For evaluation and tuning respectively 926 and 1837 unique sentence pairs randomly extracted from the parallel corpora (and manually verified and cleaned) were used. The English-Latvian SMT system's automatic evaluation results are included in the table below.

Table 15 English-Latvian SMT system's automatic evaluation results

	BLEU score	NIST score	METEOR score
Case insensitive	69.57	10.5757	0.4766
Case sensitive	68.22	10.3696	0.4272

3.3.2.3 English-Lithuanian SMT system

The system's name in the LetsMT! platform is "*LetsMT IT EN-LT v2.1*". The total size of the English-Lithuanian parallel data used to train the translation model is 2.14 M unique sentence pairs (Table 16). The parallel corpus includes three different versions of in-house translation memory data.

Table 16. Bilingual corpora for English-Lithuanian system

Bilingual corpus	Parallel units
Tilde Localization TMs EN-LT	2,219,759
Tilde Localization TMs EN-LT v1.extra	314,915
Tilde Localization TMs EN-LT v2	491,957
Total (unique, cleaned and after removal of the tuning and evaluation sets)	**2,138,909**

The monolingual corpus consists of the monolingual Lithuanian sentences of the parallel corpora that were used for the translation model training. The monolingual corpora (after duplicate removal, cleaning and filtering) consisted of 2,105,074 unique Lithuanian sentences. Also for the English-Lithuanian SMT system morphological information was not incorporated in the system's training process.

For evaluation and tuning respectively 966 and 1947 unique sentence pairs randomly extracted from the parallel corpora (and manually verified and cleaned) were used. The English-Lithuanian SMT system's automatic evaluation results are included in the table below.

Table 17 English-Lithuanian SMT system's automatic evaluation results

	BLEU score	NIST score	METEOR score
Case insensitive	59.72	9.5515	0.4291
Case sensitive	58.06	9.2854	0.3766

3.3.3 Results

The results are based on 100 translation tasks (50 tasks in each scenario) by analysing average values for translation performance (translated words per hour) and an error score for translated texts. Translators were also asked to provide system performance related feedback for more detailed analysis of the experiment.

Bearing in mind the complexity of this experiment (formatting tags, more complex language and slight subdomain deviations than in the data the SMT system is trained on), the results with a 95% confidence interval show that it is not possible to statistically prove that the overall translation performance of translators increased or decreased (see Table 18). The relatively large confidence interval is caused by the significant performance differences (as shown by the standard deviation of productivity changes) in the various translation tasks. The average translator performance with a 95% confidence interval in both translation scenarios is given in Table 19.

Table 18 Translation performance changes in per cents from scenario 1 to scenario 2 with a 95% confidence interval

Language pair	Translation performance changes	Standard deviation of productivity changes in %
English-Latvian	-3.10% ± 5.76%	20.80%
English-Estonian	-4.70% ± 7.53%	27.17%
English-Lithuanian	-3.76% ± 8.11%	29.28%

Table 19 Average translator performance and standard deviation of performance results

Language pair	Scenario 1		Scenario 2	
	Average productivity (words/hour)	Standard deviation	Average productivity	Standard deviation
English-Latvian	576 ± 47	171	558 ± 49	178
English-Estonian	470 ± 49	178	448 ± 40	143
English-Lithuanian	728 ± 87	314	700 ± 67	240

Detailed results of the English-Estonian, English-Latvian, and English-Lithuanian localisation experiment are included in Table 20. The table shows results grouped by various document sources and individual translators. Summarised total results are also given. These results show the different impact of different translation topics for different language pairs. For instance, the performance for English-Latvian translators in average increased when translating networking software related documents while for English-Lithuanian translators the performance increased when translating navigation software related documents. This shows that MT helps improving translation if the translatable data is in the same subdomain (and preferably from a similar topic) as the data on which the SMT system is trained on. For instance, the English-Latvian TMs did not feature navigation related texts, thus MT could not improve upon translation performance (as the MT system is not familiar with the text's characteristics including terminology and writing styles). On the other side, the TMs

contained data related to networking, which allowed the MT system to be beneficial in the second translation scenario. Results of MT in general are very sensitive to the training set and the accuracy could be improved by training the system with more specific data or to have a client dedicated engine.

The results also show that for all language pairs the average performance of the slowest translators increased and at the same time the performance did not degrade for all translators with the best performances. This shows the potential of MT when applied to translators with below-average performances.

Table 20 Results of the second part of the second English-Estonian, English-Latvian, and English-Lithuanian localisation experiment

Document Source	Trans-lator	Scena-rio	Total adjusted words	Latvian		Estonian		Lithuanian		Performance Change		
				Total hours	Perfor-mance	Total hours	Perfor-mance	Total hours	Perfor-mance	Latvian	Estonian	Lithuani-an
Programming	T1	S1	669.20	1.50	446.13	1.25	535.36	0.95	704.42			
		S2	550.00	1.50	366.67	0.91	604.40	1.20	458.33	-17.8%	12.9%	-34.9%
	T1	S1	619.00	1.35	458.52	1.10	562.73	0.50	1,238.00			
		S2	629.00	1.40	449.29	1.25	503.20	0.60	1,048.33	-2.0%	-10.6%	-15.3%
	T3	S1	614.00	1.02	601.96	1.00	614.00	1.32	465.15			
		S2	611.00	1.01	604.95	1.00	611.00	1.28	477.34	0.5%	-0.5%	2.6%
	T4	S1	628.00	0.85	738.82	2.40	261.67	1.38	455.07			
		S2	615.10	0.80	768.88	2.10	292.90	1.04	591.44	4.1%	11.9%	30.0%
	T5	S1	575.00	1.04	552.88	1.17	491.45	0.55	1,045.45			
		S2	631.80	1.17	540.00	1.67	378.32	0.70	902.57	-2.3%	-23.0%	-13.7%
	All	**S1**	**3,105.20**	**5.76**	**539.10**	**6.92**	**448.73**	**4.70**	**660.68**			
		S2	**3,036.90**	**5.88**	**516.48**	**6.93**	**438.23**	**4.82**	**630.06**	**-4.2%**	**-2.3%**	**-4.6%**
Navigation	T1	S1	306.00	0.50	612.00	0.50	612.00	0.35	874.29			
		S2	313.00	0.50	626.00	0.75	417.33	0.33	948.48	2.3%	-31.8%	8.5%
	T2	S1	306.00	0.65	470.77	0.58	527.59	0.30	1,020.00			
		S2	293.00	0.65	450.77	0.50	586.00	0.30	976.67	-4.2%	11.1%	-4.2%
	T3	S1	318.00	0.45	706.67	0.60	530.00	0.74	429.73			
		S2	293.00	0.53	552.83	0.70	418.57	0.51	574.51	-21.8%	-21.0%	33.7%
	T4	S1	306.00	0.45	680.00	1.20	255.00	0.55	556.36			
		S2	318.00	0.55	578.18	1.00	318.00	0.50	636.00	-15.0%	24.7%	14.3%
	T5	S1	299.10	0.41	729.51	0.67	446.42	0.26	1,150.38			
		S2	299.60	0.55	544.73	0.75	399.47	0.28	1,070.00	-25.3%	-10.5%	-7.0%
	All	**S1**	**1,535.10**	**2.46**	**624.02**	**3.55**	**432.42**	**2.20**	**697.77**			
		S2	**1,516.60**	**2.78**	**545.54**	**3.70**	**409.89**	**1.92**	**789.90**	**-12.6%**	**-5.2%**	**13.2%**
Tablet	T1	S1	1,143.80	1.55	737.94	1.50	762.53	1.90	602.00			
		S2	1,199.40	2.50	479.76	2.08	576.63	2.00	599.70	-35.0%	-24.4%	-0.4%
	T2	S1	1,264.90	2.90	436.17	2.10	602.33	1.00	1,264.90			
		S2	1,201.40	2.55	471.14	2.15	558.79	1.20	1,001.17	8.0%	-7.2%	-20.9%
	T3	S1	1,276.40	1.77	721.13	2.20	580.18	1.83	697.49			
		S2	1,233.50	1.47	839.12	2.20	560.68	2.10	587.38	16.4%	-3.4%	-15.8%
	T4	S1	1,198.00	1.63	734.97	3.90	307.18	2.01	596.02			
		S2	1,149.90	1.48	776.96	3.70	310.78	1.96	586.68	5.7%	1.2%	-1.6%
	T5	S1	1,157.80	2.01	576.02	2.42	478.43	1.12	1,033.75			
		S2	1,169.00	2.36	495.34	2.50	467.60	1.14	1,025.44	-14.0%	-2.3%	-0.8%
	All	**S1**	**6,040.90**	**9.86**	**612.67**	**12.12**	**498.42**	**7.86**	**768.56**			
		S2	**5,953.20**	**10.36**	**574.63**	**12.63**	**471.35**	**8.40**	**708.71**	**-6.2%**	**-5.4%**	**-7.8%**
Networking	T1	S1	901.20	2.00	450.60	1.41	639.15	0.93	969.03			
		S2	834.90	1.75	477.09	1.74	479.83	1.50	556.60	5.9%	-24.9%	-42.6%
	T2	S1	947.70	2.20	430.77	1.50	631.80	0.75	1,263.60			
		S2	903.00	1.85	488.11	1.80	501.67	0.90	1,003.33	13.3%	-20.6%	-20.6%
	T3	S1	889.60	1.10	808.73	2.00	444.80	1.74	511.26			
		S2	791.70	1.04	761.25	1.60	494.81	1.58	501.08	-5.9%	11.2%	-2.0%
	T4	S1	867.00	1.17	741.03	3.10	279.68	1.58	548.73			
		S2	885.40	1.15	769.91	2.30	384.96	1.07	827.48	3.9%	37.6%	50.8%
	T5	S1	871.50	1.77	492.37	1.67	521.86	1.07	814.49			

Document Source	Trans-lator	Scena-rio	Total adjusted words	Latvian		Estonian		Lithuanian		Performance Change		
				Total hours	Perfor-mance	Total hours	Perfor-mance	Total hours	Perfor-mance	Latvian	Estonian	Lithuani-an
		S2	716.30	1.42	504.44	2.00	358.15	0.71	1,008.87	2.5%	-31.4%	23.9%
	All	**S1**	**4,477.00**	**8.24**	**543.33**	**9.68**	**462.50**	**6.07**	**737.56**			
		S2	**4,131.30**	**7.21**	**573.00**	**9.44**	**437.64**	**5.76**	**717.24**	**5.5%**	**-5.4%**	**-2.8%**
All	**All**	**S1**	**15,158.20**	**26.32**	**575.92**	**32.27**	**469.73**	**20.83**	**727.71**			
		S2	**14,638.00**	**26.23**	**558.06**	**32.70**	**447.65**	**20.90**	**700.38**	**-3.1%**	**-4.7%**	**-3.8%**

For English-Latvian the localisation experiment has been performed multiple times and we have found that for all translators that have participated in the localisation experiment repeatedly (T2, T3 and T4) the performance has increased. This suggests that the performance and the ability to benefit from MT in translation workflow depend also on the experience and consequently the time spent analysing MT results and knowing how to best use the MT suggestions.

The quality review results for all three language pairs are given in the Table 21. The column title abbreviations denote the following categories: (1) "A" – accuracy, (2) "LQ" – language quality, (3) "S" – style, and (4) – "T" – terminology.

Table 21 Quality review results for English-Latvian, English-Estonian and English-Lithuanian

Trans-lator	Sce-nario	English-Latvian					English-Estonian					English-Lithuanian				
		A	LQ	S	T	Total	A	LQ	S	T	Total	A	LQ	S	T	Total
T1	S1	6.3	2.3	0.3	3.3	**12.3**	0.0	2.0	0.7	0.0	**2.6**	0.0	8.3	3.0	2.0	**13.2**
	S2	6.6	3.5	0.0	10.4	**20.4**	0.0	2.1	1.0	0.0	**3.1**	3.1	6.9	0.7	0.7	**11.4**
T2	S1	3.8	9.2	0.0	3.8	**16.9**	1.0	3.8	2.9	1.6	**9.2**	0.0	3.5	2.9	3.2	**9.6**
	S2	5.9	5.6	0.3	7.9	**19.8**	1.0	3.0	2.0	2.3	**8.3**	0.0	4.3	1.3	4.0	**9.6**
T3	S1	5.8	9.7	0.3	5.8	**21.6**	3.9	5.5	3.2	4.8	**17.4**	0.0	7.4	3.2	2.6	**13.2**
	S2	5.8	12.6	1.4	4.8	**24.6**	1.0	1.4	2.4	0.0	**4.8**	3.8	12.6	6.8	8.9	**32.1**
T4	S1	0.7	5.3	0.7	0.0	**6.7**	0.0	2.7	1.0	0.0	**3.7**	4.0	18.7	9.7	6.3	**38.7**
	S2	1.0	9.1	0.0	5.4	**15.5**	0.0	2.4	1.7	0.3	**4.4**	7.4	21.9	12.5	10.8	**52.6**
T5	S1	6.5	12.7	0.0	17.2	**36.5**	3.8	14.8	9.0	4.8	**32.4**	0.7	5.9	0.7	3.4	**10.7**
	S2	7.8	11.7	0.7	14.9	**35.1**	3.6	22.0	8.2	7.1	**40.8**	1.4	1.8	1.1	2.8	**7.1**
All	**S1**	**4.6**	**7.9**	**0.3**	**5.9**	**18.7**	**1.7**	**5.7**	**3.3**	**2.2**	**12.9**	**0.9**	**8.7**	**3.9**	**3.5**	**17.0**
	S2	**5.4**	**8.5**	**0.5**	**8.6**	**23.0**	**1.1**	**6.0**	**3.0**	**1.9**	**12.0**	**3.1**	**9.6**	**4.5**	**5.5**	**22.7**

The results show a minor decrease of translation quality from 18.7 to 23.0 error points per 1000 words for English-Latvian and from 17.0 to 22.7 error 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 for translators T2 and T3. The quality evaluation grade is still in the level "Good", which is acceptable for production.

After analysing translator feedback several areas of further improvement were evident:

- The overall platform was in some cases not performing correctly for translation segments with formatting tags because of bugs or limitations in the current implementation of tag translation. The following issues were found by the translators:
    - De-tokenisation in some cases malfunctioned for segments with tags (punctuation tokens were detached from word tokens and spaces in multiple cases between two phrases with tags were removed).
    - In cases, where a phrase with multiple (more than two) tokens within a tag would get reordered by the SMT system such that the first or last token would change its position (of being the first or the last one in that phrase), some tokens would be left out of the tags in the respective translation.
- The current implementation of SMT translation could be increased by pre-translation of translation segments or caching of further translations while the translator edits a previous translation segment.
- Improvements in SMT systems and in the capabilities of SMT systems to identify special text fragments (and treat those accordingly) are necessary to improve MT suggestion accuracy. The translators reported issues in the following text categories:
    - Named entities are often translated wrong, including translatable (for instance “United Nations” as an organization) and non-translatable phrases (for instance, “Microsoft” in “Microsoft wireless mouse”) in named entities. This, however, is a more complex issue as the translation or non-translation may depend on client specified guidelines (including product naming guidelines).
    - Terminology usage in some cases may be inconsistent, that is, a single term in two contexts can be translated by the SMT system with two different translations.
    - SMT reordering may in specific contexts cause equal sentence constituents to be reordered in a different sequence than in the source text (for instance, “*previous, current, and next*” may be translated to something like “*current, next, and previous*”).
    - Numbers in some contexts may be translated to different numbers (because of statistic probabilities caused by noise in the SMT training data).
    - Non-translatable fragments as URLs, directory paths, etc. get de-tokenized and certain parts may be even translated.
    - Statistical noise in the training data caused quotations to be wrongly reordered adjacent to one another (for instance, “"cat"” could be translated to “"" cat”) in the English-Latvian SMT system.

Many of the issues may be improved by further research and development of the overall LetsMT! platform. The SMT related issues, the translators reported are not specific to this project, but a challenge for SMT in general and work on these issues may significantly improve SMT system performance.

Detailed results of English-Estonian, English-Latvian, and English-Lithuanian evaluation are provided below as Appendix 6, Appendix 11, and Appendix 12 respectively.

4 Conclusions and Future work

Current development of SMT tools and techniques in the LetsMT! project has reached the level where they can be implemented in practical applications addressing the needs of large user groups in a variety of application scenarios.

Results promise important advances in the application of SMT in localization by integrating available tools and technologies into an easy-to-use cloud-based platform for data sharing and generation of customized MT. Building of domain and project tailored SMT systems for localisation purposes has been evaluated and results show that the current LetsMT! platform allows to train SMT systems which are practically usable in localization and help to increase translator productivity.

The results of our experiment clearly demonstrate that it is feasible to integrate the current state of the art SMT systems for 6 different highly inflected languages into the localization process.

The use of the SMT suggestions in addition to the translation memories in the SDL Trados CAT tool lead to the increase of translation performance by 18% to 32.9% while maintaining an acceptable quality of translation of documents without significant formatting. Even better performance results are achieved when using a customized SMT system that is trained on a specific domain and/or same customer parallel data.

The experiment shows that the current implementation of MT integration in CAT tools does not yet allow getting statistically significant translator performance improvement when translating texts with a high density of tags, URLs, numbers, IDs, control symbols and other non-translatable elements. Although the LetsMT! translation service supports tags in input, this support is not enough to ensure stable translator performance increase in more complex localization scenarios and more powerful treatment of non-translatable elements could be needed.

Error rate analysis shows that overall usage of MT suggestions decreases the quality of translation in all error categories, particularly in language quality. At the same time this degradation is not critical and the result is acceptable for production purposes.

References

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

Bruckner, Christine and Mirko Plitt. 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.

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

Koehn, Philipp, 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, Sharon. 2005. Methodologies for measuring the correlations between post-editing effort and machine translatability. *Machine Translation*, 19(1):37–58, March 2005.

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 Mirko, François Masselot. 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, Dag. 2008. Microsoft office localization: use of language and translation technology.

Skadiņš, Raivis, Kārlis Goba and Valters Šics. 2010. Improving SMT for Baltic Languages with Factored Models. *Proceedings of the Fourth International Conference Baltic HLT 2010*, Riga.

Tiedemann, Jörg. 2009. News from OPUS - A Collection of Multilingual Parallel Corpora with Tools and Interfaces. *N. Nicolov, K. Bontcheva, G. Angelova, R. Mitkov (eds.) Recent Advances in Natural Language Processing* (vol V), John Benjamins, Amsterdam/Philadelphia, 237-248.

Appendix 1

DOCUMENT TITLE: [enter the file name here]

QUESTION 1: What was the average quality of MT suggestions in cases where text was not found in TM (100% match)?

3. Good - minimal changes needed

2. Average - MT suggested translation was useful but had to be edited

1. Poor - MT suggestion was misleading or translation from TM (non-100% match) was used, or translation "from scratch" was more efficient)

QUESTION 2: What was the average quality of MT suggestions in cases where text contained formatting or other tags/mark-up?

3. Good - tags were almost correctly translated (some minor changes needed)

2. Average - Tags in MT suggested translation were useful but had to be edited

1. Poor - Tags in MT suggestion were mainly wrongly placed. That caused significant slowdown or translation "from scratch" was more efficient.

The format of this questionnaire is just informative, it can be replaced with 3 column spreadsheet or other technical implementation.

Appendix 2

Tilde Localization QA form - Translation Quality Assessment

This form is filled out by an Editor or a Language Specialist.

Please see procedural notes and description of error categories in **Error categories** sheet.

Fill in the **Basic information** section, **Amount of errors** column and **General comment** field.

Basic information	
Project name:	
File name:	
Source language:	
Target language:	
Translator:	
Validated by:	
Validation date:	
Stylistic type (please, select):	
Number of words checked:	1000

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**
Grand Total			**0**
Error Score (negative points) per 1000 words			**0**
Quality:			**Superior**

General comment:

Final assessment is done as follows:	**Score scale**	
Negative points for errors of each category are calculated according to the formula:	**Error score**	**Quality grade**
"Number of errors of given type" x "Error weight"	0…9	Superior
Weighted score is calculated according to the following formula:	10…29	Good
(Total negative points / Wordcount) x 1000	30…49	Mediocre
Final quality assessment is done according to the **Score Scale.**	50…69	Poor
	70…	Very poor

Notes:

In case of recurring errors (double space, the same spelling or terminology error) they should only be counted once.

Each error is counted once, by the most appropriate category. If in doubt, use the first appropriate category (top-down).

Preferential changes should not be counted as negative points, but they may be listed in a separate Comments spreadsheet.

Category	Description
Accuracy	
Understanding of the source text	A lack of comprehension of the source text resulting in incorrect meaning of the translation.
Understanding the functionality of the product	Translation does not comply with the actual function of the product. The translation of the word is OK as such but incorrect in the context.
Comprehensibility	Any error that obstructs the user from understanding the information. Very clumsy expressions.
Omissions/unnecessary additions	Words, part of sentences, sentences, paragraphs are missing. No relevant information in the source language should be omitted in the translation, unless specifically requested. The translation should not contain any unnecessary text.
Translated/Untranslated	Parts that were supposed to be translated were not translated or parts that should not be translated were translated.
Left-overs	Redundant words resulting from sentence change, wrong declinations resulting from correcting one word only but not the rest. Unnecessary question marks or asterisks left in translated text.
Language quality	
Grammar	Grammar, syntax or morphology rules are broken.
Punctuation	Incorrect usage of punctuation marks - full stops missing, opening or closing punctuation marks (quote, parenthesis), double spaces, etc.
Spelling	The translation should contain no spelling errors.
Style	
Word order, word-for-word translation	Functional sentence perspective (theme, rheme), word order. Word for word translation, resulting in stylistically inappropriate expression.
Vocabulary and style choice	Archaisms, jargon, colloquial words, verbosity, inappropriate style.
Style Guide adherence	Product Style Guide rules are ignored. In case of absence of Product Style Guide definite company style rules must be observed. Standard phrases must be used - in case of technical documentation.
Country standards	Adaptation of country standards (date and time formats, units of measurement, currency, number formats, sorting order, capitalization etc.). Examples (of names, streets, etc.) are not localized.
Terminology	
Glossary adherence	Translation does not adhere to the terms in the glossary of project/product, or does not use generally available industry terminology. Technical documentation does not use the correct translation of interface elements.
Consistency	Inconsistent usage of translation for one term or title (for cross-references).

Quality Assessment form, Values for form fields

Yes/No	Yes No
Languages	English Estonian Latvian Lithuanian
Text Type	User interface User assistance, tech. documentation Medicine Legal Marketing or Web material
Quality	Superior Good Mediocre Poor Very poor
Error category	Accuracy Language quality Style Terminology Preferential

Appendix 3

Form: Summary of the evaluation results

Task ID	Translator name	Translator qualification	Estimated time (h)	Planned performance	Actual time	Actual performance	Quality assessment (Appendix 2)						Quality grade	MT quality (References)	Editor name	Editing estimated time	Editing planned performance	Editing actual time	Editing actual performance
		(translator, senior translator)	h	(weighted words/h)	h	(weighted words/h)	Accuracy	Language quality	Style	Terminology	Count of weighted errors	Error score (total (per 1000 weighted words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3 (where 3 – the best)		h	(total words/h)	h	(total words/h)

Appendix 4

Detailed results of evaluation for English-Latvian in the first experiment

Task ID	Scenario,	Text size,	Text origin,	Translator name	Translator qualification	Estimated time	Planned performance,	Actual time	Actual performance,	Quality assesment, negative points						Quality total valuation	MT quality feedback
(in LPS)	(S1, S2)	(adjusted words)	(MS, Nokia, Oracle, other)			h	(adjusted words/h)	h	(adjusted words/h)	Accuracy	Language quality	Style	Terminology	Total	Total (per 1000 words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3(best)
Sc1_Tr4_D13-1 (Dana)	S1	486.6	MS	T1	Translator	1.39	350	1.00	487	3	4	0	2	9	17	Good	n/a
Sc1_Tr4_D14-1 (Dana)	S1	484	MS	T1	Translator	1.38	350	0.80	605	5	2	0	2	9	19	Good	n/a
Sc1_Tr3_D6-1 (Artūrs)	S1	512	MS	T2	Translator	1.46	350	1.50	341	0	5	0	2	7	14	Good	n/a
Sc1_Tr3_D7-1 (Artūrs)	S1	507	MS	T2	Translator	1.45	350	1.30	390	0	3	2	2	7	14	Good	n/a
Sc1_Tr5_D16-1 (Mārtiņš) - MS UA	S1	466.5	MS	T3	Translator	1.33	350	1.25	373	8	4	2	2	16	33	Mediocre	n/a
Sc1_Tr5_D17-1 (Mārtiņš) - MS UA	S1	497.1	MS	T3	Translator	1.42	350	1.40	355	7	2	2	4	15	28	Good	n/a
Sc1_Tr2_D3-1 (Jānis)	S1	482.1	MS	T4	Senior Translator	1.38	350	1.00	482	5	1	0	8	14	28	Good	n/a
Sc1_Tr2_D4-1 (Jānis)	S1	490.4	MS	T4	Senior Translator	1.40	350	1.10	446	9	6	3	4	22	42	Mediocre	n/a
Sc1_Tr1_D1-1 (Juris)	S1	522	MS	T5	Senior Translator	1.49	350	0.57	916	1	2	1	4	8	15	Good	n/a
Sc1_Tr1_D2-1 (Juris)	S1	496	MS	T5	Senior Translator	1.42	350	0.62	800	4	2	0	0	6	12	Good	n/a
Sc1_Tr1_D10-1 (Juris)	S1	511.4	Promethean	T5	Senior Translator	1.46	350	1.05	487	3	2	0	0	5	10	Superior	n/a
Sc1_Tr1_D10-2 (Juris)	S1	490.8	Promethean	T5	Senior Translator	1.40	350	0.80	614	0	3	2	2	7	14	Good	n/a
Sc1_Tr2_D8-1 (Jānis)	S1	496.3	Promethean	T4	Senior Translator	1.42	350	1.40	355	2	2	0	6	10	19	Good	n/a
Sc1_Tr2_D9-1 (Jānis)	S1	496.3	Promethean	T4	Senior Translator	1.42	350	1.00	496	0	4	0	0	4	8	Superior	n/a
Sc1_Tr3_D18-1 (Artūrs) - Promethean UA	S1	464.4	Promethean	T2	Translator	1.33	350	0.90	516	0	4	2	2	8	16	Good	n/a
Sc1_Tr3_D19-1 (Artūrs) - Promethean UA	S1	454.5	Promethean	T2	Translator	1.30	350	1.00	455	2	8	3	4	17	36	Mediocre	n/a
Sc1_Tr1_D15-1 (Juris) - Oracle UA	S1	509	Oracle	T5	Senior Translator	1.45	350	0.58	878	0	3	1	2	6	12	Good	n/a
Sc1_Tr1_D15-3 (Juris) - Oracle UA	S1	492	Oracle	T5	Senior Translator	1.41	350	0.52	946	0	3	2	2	7	14	Good	n/a
Sc1_Tr1_D15-5 (Juris) - Oracle UA	S1	489	Oracle	T5	Senior Translator	1.40	350	0.47	1040	4	2	4	0	10	20	Good	n/a
Sc1_Tr2_D12-5 (Jānis) - Oracle UA	S1	518.3	Oracle	T4	Senior Translator	1.48	350	1.25	415	3	0	1	0	4	8	Superior	n/a
Sc1_Tr2_D12-7 (Jānis) - Oracle UA	S1	505	Oracle	T4	Senior Translator	1.44	350	1.40	361	4	0	5	4	13	26	Good	n/a
Sc1_Tr2_D12-9 (Jānis) - Oracle UA	S1	479	Oracle	T4	Senior Translator	1.37	350	1.25	383	0	4	2	6	12	25	Good	n/a
Sc1_Tr5_D12-3 (Mārtiņš)	S1	501	Oracle	T3	Translator	1.43	350	1.00	501	11	2	2	2	17	34	Mediocre	n/a
		11351				**1.41**	**350.0**	**1.01**	**549.6**	**3.1**	**3.0**	**1.5**	**2.6**	**10.1**	**20.2**		
		total				avg	avg	avg	avg	avg	avg	avg	avg	avg	avg		
								23.16							***Good***		

Task ID	Scenario,	Text size,	Text origin,	Translator name	Translator qualification	Estimated time	Planned performance,	Actual time	Actual performance,	Quality assesment, negative points						Quality total valuation	MT quality feedback
(in LPS)	(S1, S2)	(adjusted words)	(MS, Nokia, Oracle, other)			h	(adjusted words/h)	h	(adjusted words/h)	Accuracy	Language quality	Style	Terminology	Total	Total (per 1000 words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3(best)
Sc2_Tr4_D13-2 (Dana)	S2	487	MS	T1	Translator	1.39	350	1.10	443	2	8	3	6	19	37	Mediocre	2
Sc2_Tr4_D14-2 (Dana)	S2	474.4	MS	T1	Translator	1.36	350	1.10	431	0	4	3	4	11	20	Good	2
Sc2_Tr4_D16-1 (Dana)	S2	466.5	MS	T1	Translator	1.33	350	0.60	778	6	7	1	0	14	29	Good	3
Sc2_Tr3_D20-1 (Artūrs) MS UA	S2	484	MS	T2	Translator	1.38	350	0.80	605	8	3	1	4	16	31	Mediocre	2
Sc2_Tr3_D20-2 (Artūrs) MS UA	S2	474.4	MS	T2	Translator	1.36	350	0.55	863	6	12	0	6	24	44	Mediocre	2
Sc2_Tr1_D6-2 (Juris)	S2	505	MS	T5	Senior Translator	1.44	350	0.55	918	6	5	6	0	17	34	Mediocre	2
Sc2_Tr1_D7-2 (Juris)	S2	513.2	MS	T5	Senior Translator	1.47	350	0.43	1193	1	4	0	2	7	14	Good	2
Sc2_Tr2_D2-2 (Jānis)	S2	529.5	MS	T4	Senior Translator	1.51	350	1.20	441	3	0	3	2	8	14	Good	2
Sc2_Tr5_D17-2 (Mārtiņš) - MS UA	S2	543.5	MS	T3	Translator	1.55	350	1.15	473	7	10	1	0	18	32	Mediocre	2
Sc2_Tr1_D8-2 (Juris) - Promethean UA	S2	461	Promethean	T5	Senior Translator	1.32	350	0.52	887	0	4	0	10	14	28	Good	2
Sc2_Tr1_D9-2 (Juris) - Promethean UA	S2	496.3	Promethean	T5	Senior Translator	1.42	350	0.43	1154	2	6	3	2	13	26	Good	2
Sc2_Tr2_D8-2 (Jānis) - Promethean UA	S2	461	Promethean	T4	Senior Translator	1.32	350	1.00	461	4	2	0	0	6	12	Good	2
Sc2_Tr2_D9-2 (Jānis) - Promethean UA	S2	513	Promethean	T4	Senior Translator	1.47	350	1.15	446	0	0	1	6	7	14	Good	2
Sc2_Tr3_D18-2 (Artūrs) - Promethean UA	S2	455.5	Promethean	T2	Translator	1.30	350	0.65	701	12	5	6	10	33	65	Poor	2
Sc2_Tr3_D19-2 (Artūrs) - Promethean UA	S2	476.1	Promethean	T2	Translator	1.36	350	0.80	595	3	6	4	6	19	37	Mediocre	1
Sc2_Tr1_D15-2 (Juris) - Oracle UA	S2	491	Oracle	T5	Senior Translator	1.40	350	0.60	818	2	2	2	4	10	20	Good	1
Sc2_Tr1_D15-4 (Juris) - Oracle UA	S2	486	Oracle	T5	Senior Translator	1.39	350	0.37	1314	0	7	2	2	11	23	Good	2
Sc2_Tr1_D15-6 (Juris) - Oracle UA	S2	486	Oracle	T5	Senior Translator	1.39	350	0.33	1473	1	4	1	2	8	16	Good	2
Sc2_Tr2_D12-10 (Jānis) - Oracle UA	S2	501.1	Oracle	T4	Senior Translator	1.43	350	1.15	436	6	11	1	0	18	35	Mediocre	1
Sc2_Tr2_D12-6 (Jānis) - Oracle UA	S2	514	Oracle	T4	Senior Translator	1.47	350	1.45	354	0	3	4	2	9	18	Good	1
Sc2_Tr2_D12-8 (Jānis) - Oracle UA	S2	520	Oracle	T4	Senior Translator	1.49	350	1.10	473	0	2	2	2	6	12	Good	1
Sc2_Tr5_D12-2 (Mārtiņš)	S2	505	Oracle	T3	Translator	1.44	350	0.60	842	16	3	0	4	23	46	Mediocre	2
Sc2_Tr5_D12-4 (Mārtiņš)	S2	495	Oracle	T3	Translator	1.41	350	0.70	707	13	6	2	4	25	51	Poor	2
		11339 total				**1.41** avg	**350.00** avg	**0.80** avg	**730.6** avg	**4.3** avg	**5.0** avg	**2.0** avg	**3.4** avg	**14.6** avg	**28.6** avg		**1.8**
								18.33 total							***Good***		

Appendix 5

Detailed results of evaluation for English-Polish in the first experiment

Task ID	Scenario	Text size	Text origin	Translator name	Translator qualification	Estimated time (h)	Planned performance	Actual time	Actual performance	Quality assessment (Appendix 2)						Quality grade	MT quality (References)
	S1,S2	(adjusted words)			(translator, senior translator)	h	(weighted words/h)	h	(weighted words/h)	Accuracy	Language quality	Style	Terminology	Count of weighted errors	Error score (total (per 1000 weighted words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3 (where 3 – the best)
A_PL_01-1	S1	456,2	cust-A	Andrzej Sawicki	Translator	1,63	280	1,45	315	1	0	7	2	10	22	Good	-
A_PL_01-2	S1	491,4	cust-A	Andrzej Sawicki	Translator	1,76	280	1,57	313	1	1	1	2	5	10	Good	-
A_PL_01-5	S1	472,1	cust-A	Andrzej Sawicki	Translator	1,69	280	1,50	315	5	0	3	4	12	25	Good	-
A_PL_01-7	S1	542,5	cust-A	Maksymilian Nawrocki	Senior Translator	1,94	280	1,75	310	2	2	4	0	8	15	Good	-
A_PL_02-1	S1	481,1	cust-A	Maksymilian Nawrocki	Senior Translator	1,72	280	1,40	344	0	1	2	0	3	6	Superior	-
A_PL_03-1	S1	505	cust-A	Maksymilian Nawrocki	Senior Translator	1,80	280	1,53	330	0	2	2	2	6	12	Good	-
A_PL_04-2	S1	522,4	cust-B	Agata Reszke	Senior Translator	1,87	280	1,73	302	0	1	3	2	6	11	Good	-
A_PL_05-2	S1	514	cust-A	Maksymilian Nawrocki	Senior Translator	1,84	280	1,72	299	2	1	0	2	5	10	Good	-
A_PL_06-1	S1	466,1	cust-B	Maksymilian Nawrocki	Senior Translator	1,66	280	1,55	301	3	0	2	0	5	11	Good	-
A_PL_06-3	S1	472,5	cust-B	Maksymilian Nawrocki	Senior Translator	1,69	280	1,57	301	1	2	0	0	3	6	Superior	-
A_PL_06-5	S1	562,3	cust-B	Agata Reszke	Senior Translator	2,01	280	1,85	304	4	1	2	4	11	20	Good	-
A_PL_07-2	S1	487	cust-C	Andrzej Sawicki	Translator	1,74	280	1,53	318	3	2	2	2	9	18	Good	-
A_PL_07-3	S1	541,2	cust-C	Andrzej Sawicki	Translator	1,93	280	1,75	309	2	3	3	0	8	15	Good	-
A_PL_10-1	S1	529,1	cust-C	Andrzej Sawicki	Translator	1,89	280	1,68	315	4	2	1	0	7	13	Good	-
A_PL_10-2	S1	534	cust-C	Maksymilian Nawrocki	Senior Translator	1,91	280	1,72	310	3	0	3	2	8	15	Good	-
A_PL_10-4	S1	492,5	cust-C	Maksymilian Nawrocki	Senior Translator	1,76	280	1,60	308	0	0	0	0	0	0	Superior	-
A_PL_11-1	S1	467,2	cust-C	Andrzej Sawicki	Translator	1,67	280	1,48	316	5	1	3	2	11	24	Good	-
A_PL_09-1	S1	503,1	cust-C	Witold Grzebinski	Translator	1,80	280	1,93	261	4	3	6	4	17	34	Mediocre	-
A_PL_09-3	S1	507,5	cust-C	Witold Grzebinski	Translator	1,81	280	1,95	260	6	1	8	6	21	41	Mediocre	-
A_PL_12-1	S1	536	cust-D	Agata Reszke	Senior Translator	1,91	280	1,68	319	2	2	3	2	9	17	Good	-
A_PL_13-1	S1	482,5	cust-D	Witold Grzebinski	Translator	1,72	280	1,87	258	3	2	4	4	13	27	Good	-
Total		**10565,7**				**1,80**	**280,00**	**1,66**	**305,14**	**2,43**	**1,29**	**2,81**	**1,90**	**8,43**	**16,76**	**Good**	

Task ID	Scenario	Text size	Text origin	Translator name	Translator qualification	Estimated time (h)	Planned performance	Actual time	Actual performance	Quality assessment (Appendix 2)						Quality grade	MT quality (References)
	S1,S2	(adjusted words)			(translator, senior translator)	h	(weighted words/h)	h	(weighted words/h)	Accuracy	Language quality	Style	Terminology	Count of weighted errors	Error score (total (per 1000 weighted words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3 (where 3 – the best)
B_PL_01-3	S2	488,2	cust-A	Agata Reszke	Senior Translator	1,74	280	0,97	503	1	2	5	0	8	16	Good	2
B_PL_01-4	S2	518	cust-A	Andrzej Sawicki	Translator	1,85	280	1,23	421	1	0	5	2	8	15	Good	1
B_PL_01-6	S2	505,2	cust-A	Maksymilian Nawrocki	Senior Translator	1,80	280	1,23	411	0	0	3	2	5	10	Good	1
B_PL_02-2	S2	512,4	cust-A	Maksymilian Nawrocki	Senior Translator	1,83	280	1,25	410	1	0	2	1	4	8	Superior	1
B_PL_04-1	S2	480,5	cust-B	Agata Reszke	Senior Translator	1,72	280	1,18	407	6	0	2	4	12	25	Good	1
B_PL_05-1	S2	489,2	cust-A	Maksymilian Nawrocki	Senior Translator	1,75	280	1,06	462	0	0	2	4	6	12	Good	2
B_PL_05-3	S2	506,2	cust-A	Maksymilian Nawrocki	Senior Translator	1,81	280	1,15	440	2	4	2	4	12	24	Good	2
B_PL_06-2	S2	473,5	cust-B	Maksymilian Nawrocki	Senior Translator	1,69	280	1,12	423	2	3	3	2	10	21	Good	1
B_PL_06-4	S2	502,1	cust-B	Agata Reszke	Senior Translator	1,79	280	1,20	418	1	2	5	2	10	20	Good	1
B_PL_07-1	S2	479	cust-C	Andrzej Sawicki	Translator	1,71	280	1,30	368	2	0	0	4	6	13	Good	1
B_PL_07-4	S2	509,8	cust-C	Andrzej Sawicki	Translator	1,82	280	1,35	378	3	4	3	4	14	27	Good	1
B_PL_08-1	S2	499,4	cust-C	Andrzej Sawicki	Translator	1,78	280	1,40	357	6	3	8	4	21	42	Mediocre	1
B_PL_10-6	S2	533,1	cust-C	Agata Reszke	Senior Translator	1,90	280	1,18	452	6	0	5	4	15	28	Good	2
B_PL_10-3	S2	486	cust-C	Andrzej Sawicki	Translator	1,74	280	1,37	355	4	1	6	2	13	27	Good	1
B_PL_10-5	S2	541,4	cust-C	Andrzej Sawicki	Translator	1,93	280	1,45	373	6	3	2	0	11	20	Good	1
B_PL_08-2	S2	453,4	cust-C	Andrzej Sawicki	Translator	1,62	280	1,30	349	7	4	4	0	15	33	Mediocre	1
B_PL_09-2	S2	487,2	cust-C	Andrzej Sawicki	Translator	1,74	280	1,38	353	8	1	12	4	25	51	Poor	1
B_PL_09-4	S2	496,1	cust-C	Andrzej Sawicki	Translator	1,77	280	1,42	349	5	2	3	2	12	24	Good	1
B_PL_12-2	S2	513	cust-D	Witold Grzebinski	Translator	1,83	280	1,53	335	2	2	2	6	12	23	Good	1
B_PL_13-2	S2	507,6	cust-D	Witold Grzebinski	Translator	1,81	280	1,50	338	7	2	4	4	17	33	Mediocre	1
B_PL_14-1	S2	579	cust-D	Witold Grzebinski	Translator	2,07	280	1,80	322	5	1	4	4	14	24	Good	2
Total		**10560**				**1,80**	**280,00**	**1,30**	**391,62**	**3,57**	**1,62**	**3,90**	**2,81**	**11,90**	**23,62**	**Good**	**1,24**

Appendix 6

Detailed results of evaluation for English-Czech in the first experiment

Task ID	Scenario	Text size	Text origin	Translator name	Translator qualification	Estimated time (h)	Planned performance	Actual time	Actual performance	Quality assessment (Appendix 2)						Quality grade	MT quality (References)
	S1,S2	(adjusted words)			(translator, senior translator)	h	(weighted words/h)	h	(weighted words/h)	Accuracy	Language quality	Style	Terminology	Count of weighted errors	Error score (total (per 1000 weighted words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3 (where 3 – the best)
A_CZ_01-1	S1	527	cust-A	Barbora Zlama	Translator	1,88	280	1,75	301	4	0	3	2	9	17	Good	-
A_CZ_01-3	S1	511,4	cust-A	Milan Vesely	Translator	1,83	280	1,58	324	2	1	3	2	8	16	Good	-
A_CZ_03-1	S1	569	cust-A	Jan Trhlik	Translator	2,03	280	1,95	292	5	4	6	4	19	33	Mediocre	-
A_CZ_04-2	S1	501,7	cust-A	Milan Vesely	Translator	1,79	280	1,48	339	1	2	3	0	6	12	Good	-
A_CZ_04-4	S1	504,4	cust-A	Barbora Zlama	Translator	1,80	280	1,62	311	5	2	3	0	10	20	Good	-
A_CZ_05-1	S1	547,2	cust-A	Ales Horak	Senior Translat	1,95	280	1,43	383	3	0	3	1	7	13	Good	-
A_CZ_05-3	S1	552	cust-A	Ales Horak	Senior Translat	1,97	280	1,33	415	3	0	1	0	4	7	Superior	-
A_CZ_06-2	S1	574,5	cust-B	Daniela Skotni	Senior Translat	2,05	280	1,72	334	2	0	1	2	5	9	Superior	-
A_CZ_06-4	S1	549,8	cust-B	Daniela Skotni	Senior Translat	1,96	280	1,67	329	3	1	2	0	6	11	Good	-
A_CZ_06-6	S1	477,1	cust-B	Jan Trhlik	Translator	1,70	280	1,93	247	9	6	8	2	25	52	Poor	-
A_CZ_07-2	S1	452	cust-B	Jan Trhlik	Translator	1,61	280	1,78	254	4	4	8	2	18	40	Mediocre	-
A_CZ_09-1	S1	554,6	cust-B	Milan Vesely	Translator	1,98	280	1,80	308	2	3	6	0	11	20	Good	-
A_CZ_09-3	S1	561,5	cust-B	Milan Vesely	Translator	2,01	280	1,73	325	1	1	2	2	6	11	Good	-
A_CZ_09-5	S1	471,3	cust-B	Milan Vesely	Translator	1,68	280	1,48	318	2	1	3	1	7	15	Good	-
A_CZ_10-2	S1	560,2	cust-C	Ales Horak	Senior Translat	2,00	280	1,82	308	4	0	2	0	6	11	Good	-
A_CZ_10-4	S1	509,1	cust-C	Barbora Zlama	Translator	1,82	280	1,98	257	9	0	4	2	15	29	Good	-
A_CZ_11-2	S1	518	cust-C	Daniela Skotni	Senior Translat	1,85	280	1,67	310	2	0	1	0	3	6	Superior	-
Total		**8940,8**				**1,88**	**280,00**	**1,69**	**315,00**	**3,59**	**1,47**	**3,47**	**1,18**	**9,71**	**18,94**	**Good**	

Task ID	Scenario	Text size	Text origin	Translator name	Translator qualification	Estimated time (h)	Planned performance	Actual time	Actual performance	Quality assessment (Appendix 2)						Quality grade	MT quality (References)
	S1,S2	(adjusted words)			(translator, senior translator)	h	(weighted words/h)	h	(weighted words/h)	Accuracy	Language quality	Style	Terminology	Count of weighted errors	Error score (total (per 1000 weighted words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3 (where 3 – the best)
B_CZ_01-2	S2	505	cust-A	Barbora Zlama	Translator	1,80	280	1,38	366	8	2	2	0	12	24	Good	1
B_CZ_02-1	S2	463,5	cust-A	Daniela Skotni	Senior Translat	1,66	280	0,90	515	2	2	3	4	11	24	Good	2
B_CZ_04-1	S2	495,2	cust-A	Milan Vesely	Translator	1,77	280	1,22	406	10	6	1	2	19	38	Mediocre	1
B_CZ_04-3	S2	517	cust-A	Milan Vesely	Translator	1,85	280	1,13	458	7	0	4	0	11	21	Good	2
B_CZ_04-5	S2	484,1	cust-A	Barbora Zlama	Translator	1,73	280	1,25	387	11	4	7	2	24	50	Poor	2
B_CZ_05-2	S2	493	cust-A	Ales Horak	Senior Translat	1,76	280	0,98	503	3	1	2	2	8	16	Good	2
B_CZ_06-1	S2	481,2	cust-B	Daniela Skotni	Senior Translat	1,72	280	1,12	430	4	0	4	4	12	25	Good	1
B_CZ_06-3	S2	469	cust-B	Daniela Skotni	Senior Translat	1,68	280	1,08	434	3	0	1	2	6	13	Good	1
B_CZ_06-5	S2	480,3	cust-B	Daniela Skotni	Senior Translat	1,72	280	1,17	411	8	0	2	2	12	25	Good	2
B_CZ_07-1	S2	521	cust-B	Jan Trhlik	Translator	1,86	280	1,78	293	5	9	4	6	24	46	Mediocre	2
B_CZ_08-1	S2	495	cust-B	Daniela Skotni	Senior Translat	1,77	280	1,28	387	8	2	1	0	11	22	Good	1
B_CZ_09-2	S2	563,2	cust-B	Milan Vesely	Translator	2,01	280	1,57	359	2	7	0	2	11	20	Good	1
B_CZ_09-4	S2	525,7	cust-B	Milan Vesely	Translator	1,88	280	1,53	344	10	6	4	0	20	38	Mediocre	1
B_CZ_10-1	S2	474	cust-C	Ales Horak	Senior Translat	1,69	280	1,13	419	7	0	2	0	9	19	Good	2
B_CZ_10-3	S2	500,5	cust-C	Barbora Zlama	Translator	1,79	280	1,68	298	14	2	4	4	24	48	Mediocre	1
B_CZ_11-1	S2	572,5	cust-C	Daniela Skotni	Senior Translat	2,04	280	1,62	353	3	1	1	0	5	9	Superior	1
B_CZ_11-3	S2	489	cust-C	Daniela Skotni	Senior Translat	1,75	280	1,43	342	4	2	2	2	10	20	Good	1
Total		**8529,2**				**1,79**	**280,00**	**1,31**	**394,41**	**6,41**	**2,59**	**2,59**	**1,88**	**13,47**	**26,94**	**Good**	**1,41**

Appendix 7

Detailed results of evaluation for English-Czech in the second experiment

Task ID	Sce-nario	Text size	Text origin	Trans-lator	Translator qualification	Estimated time (h)	Planned perfor-mance	Actual time	Actual perfor-mance	Quality assessment (Appendix 2)						Quality grade	MT quality (References)	Editor name	Editing estimated time	Editing planned performance	Editing actual time	Editing actual performance
	S1, S2	(adjust-ed words)			(translator, senior translator)	h	(weighted words/h)	h	(weighted words/h)	Accu-racy	Lan-guage quality	Style	Termi-nology	Count of weighted errors	Error score (total (per 1000 weighted words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3 (where 3 – the best)		h	(total words/h)	h	(total words/h)
A_CZ_01-1	S1	301.5	cust-A	T1	Senior translator	1.08	280	0.98	308	2	0	0	0	2	7	Superior	-	E1	0.22	1400	0.17	1774
A_CZ_01-3	S1	492.25	cust-A	T1	Senior translator	1.76	280	1.62	304	3	1	1	2	7	14	Good	-	E1	0.35	1400	0.3	1641
A_CZ_02-1	S1	317	cust-A	T1	Senior translator	1.13	280	1.05	302	5	1	1	2	9	28	Good	-	E1	0.23	1400	0.25	1268
A_CZ_03-1	S1	442	cust-A	T2	Translator	1.58	280	1.53	289	3	4	5	4	16	36	Mediocre	-	E1	0.32	1400	0.38	1163
A_CZ_04-1	S1	506.25	cust-A	T3	Translator	1.81	280	1.80	281	3	3	3	0	9	18	Good	-	E1	0.36	1400	0.35	1446
A_CZ_05-1	S1	512	cust-A	T2	Translator	1.83	280	1.70	301	1	1	4	6	12	23	Good	-	E1	0.37	1400	0.35	1463
A_CZ_05-3	S1	520.25	cust-A	T2	Translator	1.86	280	1.75	297	7	5	1	4	17	33	Mediocre	-	E1	0.37	1400	0.45	1156
A_CZ_06-1	S1	493.75	cust-A	T3	Translator	1.76	280	1.67	296	5	2	2	2	11	22	Good	-	E1	0.35	1400	0.33	1496
A_CZ_06-3	S1	424	cust-A	T3	Translator	1.51	280	1.50	283	4	0	0	2	6	14	Good	-	E1	0.3	1400	0.25	1696
A_CZ_06-5	S1	405	cust-A	T3	Translator	1.45	280	1.52	266	5	2	3	0	10	25	Good	-	E1	0.29	1400	0.28	1446
A_CZ_08-1	S1	660.5	cust-B	T4	Senior translator	2.36	280	2.20	300	0	0	2	2	4	6	Superior	-	E2	0.47	1400	0.35	1887
A_CZ_09-1	S1	723	cust-B	T5	Translator	2.58	280	2.65	273	4	0	2	2	8	11	Good	-	E2	0.52	1400	0.45	1607
A_CZ_09-3	S1	477	cust-B	T5	Translator	1.70	280	1.80	265	5	1	6	2	14	29	Good	-	E2	0.34	1400	0.30	1590
A_CZ_09-5	S1	711.5	cust-B	T4	Senior translator	2.54	280	2.42	294	3	2	5	0	10	14	Good	-	E2	0.51	1400	0.43	1655
A_CZ_10-1	S1	408	cust-B	T1	Senior translator	1.46	280	1.38	296	3	0	0	0	3	7	Superior	-	E2	0.29	1400	0.22	1855
A_CZ_10-3	S1	683.5	cust-B	T1	Senior translator	2.44	280	2.32	295	4	0	1	2	7	10	Good	-	E2	0.49	1400	0.38	1799
A_CZ_12-1	S1	402.75	cust-B	T4	Senior translator	1.44	280	1.37	294	2	3	3	0	8	20	Good	-	E2	0.29	1400	0.28	1438
Total		**8480.25**				**1.78**	**280.00**	**1.72**	**290.82**	**3.47**	**1.47**	**2.29**	**1.76**	**9.00**	**18.65**	**Good**			**0.36**	**1400.00**	**0.32**	**1551.76**

Task ID	Sce-nario	Text size	Text origin	Trans-lator	Translator qualification	Estimated time (h)	Planned perfor-mance	Actual time	Actual perfor-mance	Quality assessment (Appendix 2)						Quality grade	MT quality (References)	Editor name	Editing estimated time	Editing planned performance	Editing actual time	Editing actual performance
	S1, S2	(adjust-ed words)			(translator, senior translator)	h	(weighted words/h)	h	(weighted words/h)	Accu-racy	Lan-guage quality	Style	Termi-nology	Count of weighted errors	Error score (total (per 1000 weighted words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3 (where 3 – the best)		S1, S2	(adjust-ed words)		
B_CZ_01-2	S2	742	cust-A	T1	Senior translator	2.65	280	1.75	424	3	1	2	2	8	11	Good	2	E1	0.53	1400	0.47	1579
B_CZ_03-2	S2	445	cust-A	T2	Translator	1.59	280	1.08	412	9	2	7	2	20	45	Mediocre	2	E1	0.32	1400	0.48	927
B_CZ_04-2	S2	575	cust-A	T4	Senior translator	2.05	280	1.45	397	2	1	3	6	12	21	Good	2	E1	0.41	1400	0.42	1369
B_CZ_05-2	S2	549	cust-A	T2	Translator	1.96	280	1.62	339	7	1	0	6	14	26	Good	1	E1	0.39	1400	0.40	1373
B_CZ_05-4	S2	390.25	cust-A	T2	Translator	1.39	280	1.03	379	0	6	2	4	12	31	Mediocre	2	E1	0.28	1400	0.33	1183
B_CZ_06-2	S2	595	cust-A	T3	Translator	2.13	280	1.82	327	8	0	1	4	13	22	Good	1	E1	0.43	1400	0.4	1488
B_CZ_06-4	S2	417	cust-A	T3	Translator	1.49	280	1.25	334	9	1	5	4	19	46	Mediocre	1	E1	0.3	1400	0.38	1097
B_CZ_06-6	S2	517	cust-A	T3	Translator	1.85	280	1.53	338	4	6	2	2	14	27	Good	1	E1	0.37	1400	0.42	1231
B_CZ_07-1	S2	612	cust-B	T5	Translator	2.19	280	1.80	340	7	3	3	0	13	21	Good	2	E2	0.44	1400	0.45	1360
B_CZ_08-2	S2	668.5	cust-B	T4	Senior translator	2.39	280	1.95	343	1	3	2	0	6	9	Superior	1	E2	0.48	1400	0.35	1910
B_CZ_09-2	S2	646	cust-B	T5	Translator	2.31	280	2.02	320	9	4	5	4	22	34	Mediocre	2	E2	0.46	1400	0.53	1219
B_CZ_09-4	S2	569	cust-B	T5	Translator	2.03	280	1.75	325	5	5	6	4	20	35	Mediocre	2	E2	0.41	1400	0.45	1264
B_CZ_09-6	S2	678	cust-B	T5	Translator	2.42	280	2.10	323	2	4	4	2	12	18	Good	2	E2	0.48	1400	0.45	1507
B_CZ_10-2	S2	533	cust-B	T1	Senior translator	1.90	280	1.72	310	3	3	1	2	9	17	Good	1	E2	0.38	1400	0.35	1523
B_CZ_11-1	S2	381	cust-B	T4	Senior translator	1.36	280	1.02	374	4	1	0	2	7	18	Good	2	E2	0.27	1400	0.25	1524
B_CZ_12-2	S2	487	cust-B	T4	Senior translator	1.74	280	1.45	336	2	4	8	0	14	29	Good	1	E2	0.35	1400	0.33	1476
Total		**8804.75**				**1.97**	**280.00**	**1.58**	**351.31**	**4.69**	**2.81**	**3.19**	**2.75**	**13.44**	**25.63**	**Good**	**1.56**		**0.39**	**1400.00**	**0.40**	**1376.88**

Appendix 8

Detailed results of evaluation for English-Hungarian in the second experiment

Task ID	Sce-nario	Text size	Text origin	Trans-lator	Translator qualification	Estimated time (h)	Planned perfor-mance	Actual time	Actual perfor-mance	Quality assessment (Appendix 2)						Quality grade	MT quality (References)	Editor name	Editing estimated time	Editing planned performance	Editing actual time	Editing actual performance
	S1, S2	(adjust-ed words)			(translator, senior translator)	h	(weighted words/h)	h	(weighted words/h)	Accu-racy	Lan-guage quality	Style	Termi-nology	Count of weighted errors	Error score (total (per 1000 weighted words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3 (where 3 – the best)		h	(total words/h)	h	(total words/h)
A_HU_01-1	S1	301.5	cust-A	T1	Senior translator	1.10	275	0.95	317	7	0	1	2	10	33	Mediocre	-	E1	0.22	1400	0.25	1206
A_HU_01-3	S1	492.25	cust-A	T1	Senior translator	1.79	275	1.62	304	2	1	5	0	8	16	Good	-	E1	0.35	1400	0.3	1641
A_HU_02-1	S1	317	cust-A	T1	Senior translator	1.15	275	1.02	311	2	0	2	2	6	19	Good	-	E1	0.23	1400	0.22	1441
A_HU_03-1	S1	442	cust-A	T2	Translator	1.61	275	1.45	305	4	0	7	2	13	29	Good	-	E1	0.32	1400	0.35	1263
A_HU_04-1	S1	506.25	cust-A	T2	Translator	1.84	275	1.88	269	1	3	2	0	6	12	Good	-	E1	0.36	1400	0.30	1688
A_HU_05-1	S1	512	cust-A	T1	Senior translator	1.86	275	1.70	301	3	2	1	2	8	16	Good	-	E1	0.37	1400	0.32	1600
A_HU_05-3	S1	520.25	cust-A	T3	Translator	1.89	275	1.83	284	4	3	2	0	9	17	Good	-	E1	0.37	1400	0.35	1486
A_HU_06-1	S1	493.75	cust-A	T4	Senior translator	1.80	275	1.65	299	3	1	3	2	9	18	Good	-	E1	0.35	1400	0.38	1299
A_HU_06-3	S1	424	cust-A	T4	Senior translator	1.54	275	1.45	292	2	0	2	0	4	9	Superior	-	E1	0.3	1400	0.22	1927
A_HU_06-5	S1	405	cust-A	T4	Senior translator	1.47	275	1.30	312	1	0	5	2	8	20	Good	-	E1	0.29	1400	0.28	1446
A_HU_08-1	S1	660.5	cust-B	T4	Senior translator	2.40	275	2.18	303	4	1	1	4	10	15	Good	-	E2	0.47	1400	0.40	1651
A_HU_09-1	S1	723	cust-B	T5	Senior translator	2.63	275	2.55	284	4	1	1	0	6	8	Superior	-	E2	0.52	1400	0.42	1721
A_HU_09-3	S1	477	cust-B	T5	Senior translator	1.73	275	1.72	277	3	3	2	6	14	29	Good	-	E2	0.34	1400	0.42	1136
A_HU_10-1	S1	408	cust-B	T2	Translator	1.48	275	1.70	240	3	1	3	0	7	17	Good	-	E2	0.29	1400	0.33	1236
A_HU_10-3	S1	683.5	cust-B	T2	Translator	2.49	275	2.78	246	2	4	5	4	15	22	Good	-	E2	0.49	1400	0.52	1314
A_HU_11-1	S1	491.5	cust-C	T3	Translator	1.79	275	1.75	281	4	3	1	0	8	16	Good	-	E2	0.35	1400	0.30	1638
A_HU_11-3	S1	522.25	cust-C	T3	Translator	1.90	275	1.98	264	8	3	9	4	24	46	Mediocre	-	E2	0.37	1400	0.45	1161
A_HU_11-5	S1	580	cust-C	T3	Translator	2.11	275	2.05	283	3	2	0	4	9	16	Good	-	E2	0.41	1400	0.42	1381
Total		**8959.75**				**1.81**	**275.00**	**1.75**	**287.33**	**3.33**	**1.56**	**2.89**	**1.89**	**9.67**	**19.89**	**Good**			**0.36**	**1400.00**	**0.35**	**1457.50**

Task ID	Scenario	Text size	Text origin	Translator	Translator qualification	Estimated time (h)	Planned performance	Actual time	Actual performance	Quality assessment (Appendix 2)						Quality grade	MT quality (References)	Editor name	Editing estimated time	Editing planned performance	Editing actual time	Editing actual performance
	S1, S2	(adjusted words)			(translator, senior translator)	h	(weighted words/h)	h	(weighted words/h)	Accuracy	Language quality	Style	Terminology	Count of weighted errors	Error score (total (per 1000 weighted words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3 (where 3 – the best)		h	(total words/h)	h	(total words/h)
B_HU_01-2	S2	742	cust-A	T1	Senior translator	2.70	275	1.80	412	5	0	0	4	9	12	Good	2	E1	0.53	1400	0.43	1726
B_HU_03-2	S2	445	cust-A	T2	Translator	1.62	275	1.20	371	6	2	2	6	16	36	Mediocre	2	E1	0.32	1400	0.37	1203
B_HU_04-2	S2	575	cust-A	T2	Translator	2.09	275	1.88	306	2	1	5	2	10	17	Good	1	E1	0.41	1400	0.40	1438
B_HU_05-2	S2	549	cust-A	T1	Senior translator	2.00	275	1.45	379	4	3	6	0	13	24	Good	2	E1	0.39	1400	0.37	1484
B_HU_05-4	S2	390.25	cust-A	T3	Translator	1.42	275	1.13	345	9	3	6	2	20	51	Poor	2	E1	0.28	1400	0.42	929
B_HU_06-2	S2	595	cust-A	T4	Senior translator	2.16	275	1.75	340	2	3	2	4	11	18	Good	1	E1	0.43	1400	0.45	1322
B_HU_06-4	S2	417	cust-A	T4	Senior translator	1.52	275	1.18	353	1	0	3	2	6	14	Good	1	E1	0.3	1400	0.25	1668
B_HU_06-6	S2	517	cust-A	T4	Senior translator	1.88	275	1.40	369	6	1	3	0	10	19	Good	2	E1	0.37	1400	0.35	1477
B_HU_07-1	S2	612	cust-B	T5	Senior translator	2.23	275	1.82	336	1	2	6	2	11	18	Good	2	E2	0.44	1400	0.46	1330
B_HU_08-2	S2	668.5	cust-B	T4	Senior translator	2.43	275	2.18	307	1	6	3	0	10	15	Good	1	E2	0.48	1400	0.38	1759
B_HU_09-2	S2	646	cust-B	T5	Senior translator	2.35	275	2.10	308	4	1	3	2	10	15	Good	1	E2	0.46	1400	0.4	1615
B_HU_09-4	S2	569	cust-B	T5	Senior translator	2.07	275	1.80	316	3	0	2	2	7	12	Good	1	E2	0.41	1400	0.3	1897
B_HU_10-2	S2	533	cust-B	T2	Translator	1.94	275	1.95	273	11	7	4	6	28	53	Poor	1	E2	0.38	1400	0.52	1025
B_HU_10-4	S2	491.25	cust-B	T2	Translator	1.79	275	1.72	286	7	7	9	2	25	51	Poor	1	E2	0.35	1400	0.50	983
B_HU_11-2	S2	447.75	cust-C	T3	Translator	1.63	275	1.43	313	5	6	0	0	11	25	Good	1	E2	0.32	1400	0.37	1210
B_HU_11-4	S2	472	cust-C	T3	Translator	1.72	275	1.42	332	4	3	2	2	11	23	Good	1	E2	0.34	1400	0.40	1180
B_HU_12-1	S2	436.5	cust-B	T2	Translator	1.59	275	1.12	390	7	3	4	2	16	37	Mediocre	2	E2	0.31	1400	0.42	1039
B_HU_12-2	S2	423	cust-B	T2	Translator	1.54	275	1.15	368	6	5	3	0	14	33	Mediocre	2	E2	0.3	1400	0.37	1143
Total		**9529.25**				**1.93**	**275.00**	**1.58**	**339.11**	**4.67**	**2.94**	**3.50**	**2.11**	**13.22**	**26.28**	**Good**	**1.44**		**0.38**	**1400.00**	**0.40**	**1357.11**

Appendix 9

Detailed results of evaluation for English-Polish in the second experiment

Task ID	Sce-nario	Text size	Text origin	Trans-lator	Translator qualification	Estimated time (h)	Planned perfor-mance	Actual time	Actual perfor-mance	Quality assessment (Appendix 2)						Quality grade	MT quality (References)	Editor name	Editing estimated time	Editing planned performance	Editing actual time	Editing actual performance
	S1, S2	(adjust-ed words)			(translator, senior translator)	h	(weighted words/h)	h	(weighted words/h)	Accu-racy	Lan-guage quality	Style	Termi-nology	Count of weighted errors	Error score (total (per 1000 weighted words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3 (where 3 – the best)		h	(total words/h)	h	(total words/h)
A_PL_01-1	S1	301.5	cust-A	T1	Translator	1.08	280	1.02	296	4	2	5	2	13	43	Mediocre	-	E1	0.22	1400	0.25	1206
A_PL_01-3	S1	492.25	cust-A	T1	Translator	1.76	280	1.68	293	3	1	2	2	8	16	Good	-	E1	0.35	1400	0.28	1758
A_PL_02-1	S1	317	cust-A	T2	Senior translator	1.13	280	1.02	311	1	0	3	0	4	13	Good	-	E1	0.23	1400	0.18	1761
A_PL_03-1	S1	442	cust-A	T2	Senior translator	1.58	280	1.40	316	3	2	1	2	8	18	Good	-	E1	0.32	1400	0.28	1579
A_PL_04-1	S1	506.25	cust-A	T2	Senior translator	1.81	280	1.58	320	7	8	9	0	24	47	Mediocre	-	E1	0.36	1400	0.45	1125
A_PL_05-1	S1	512	cust-A	T3	Senior translator	1.83	280	1.63	314	4	0	3	0	7	14	Good	-	E1	0.37	1400	0.30	1707
A_PL_05-3	S1	520.25	cust-A	T3	Senior translator	1.86	280	1.70	306	5	2	4	2	13	25	Good	-	E1	0.37	1400	0.38	1369
A_PL_06-1	S1	493.75	cust-A	T1	Translator	1.76	280	1.65	299	4	1	3	2	10	20	Good	-	E1	0.35	1400	0.37	1334
A_PL_06-3	S1	424	cust-A	T1	Translator	1.51	280	1.47	288	1	3	7	4	15	35	Mediocre	-	E1	0.3	1400	0.33	1285
A_PL_06-5	S1	405	cust-A	T1	Translator	1.45	280	1.32	307	1	3	0	0	4	10	Good	-	E1	0.29	1400	0.23	1761
A_PL_08-1	S1	660.5	cust-B	T4	Translator	2.36	280	2.25	294	7	8	1	2	18	27	Good	-	E2	0.47	1400	0.50	1321
A_PL_09-1	S1	723	cust-B	T2	Senior translator	2.58	280	2.38	304	8	5	5	2	20	28	Good	-	E2	0.52	1400	0.60	1205
A_PL_09-3	S1	477	cust-B	T2	Senior translator	1.70	280	1.62	294	7	4	7	0	18	38	Mediocre	-	E2	0.34	1400	0.43	1109
A_PL_10-1	S1	408	cust-B	T5	Translator	1.46	280	1.72	237	10	8	3	4	25	61	Poor	-	E2	0.29	1400	0.42	971
A_PL_10-3	S1	683.5	cust-B	T5	Translator	2.44	280	2.80	244	5	3	7	2	17	25	Good	-	E2	0.49	1400	0.52	1314
A_PL_11-1	S1	572.5	cust-C	T4	Translator	2.04	280	1.87	306	3	3	2	2	10	17	Good	-	E2	0.41	1400	0.38	1507
A_PL_11-3	S1	412.25	cust-C	T4	Translator	1.47	280	1.32	312	1	0	1	0	2	5	Superior	-	E2	0.29	1400	0.22	1874
A_PL_12-1	S1	617	cust-C	T5	Translator	2.20	280	2.43	254	5	1	5	6	17	28	Good	-	E2	0.44	1400	0.48	1285
Total		**8967.75**				**1.78**	**280.00**	**1.71**	**294.17**	**4.39**	**3.00**	**3.78**	**1.78**	**12.94**	**26.11**	**Good**			**0.36**	**1400.00**	**0.37**	**1415.06**

Task ID	Sce-nario	Text size	Text origin	Trans-lator	Translator qualification	Estimated time (h)	Planned perfor-mance	Actual time	Actual perfor-mance	Quality assessment (Appendix 2)						Quality grade	MT quality (References)	Editor name	Editing estimated time	Editing planned performance	Editing actual time	Editing actual performance
	S1, S2	(adjust-ed words)			(translator, senior translator)	h	(weighted words/h)	h	(weighted words/h)	Accu-racy	Lan-guage quality	Style	Termi-nology	Count of weighted errors	Error score (total (per 1000 weighted words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3 (where 3 – the best)		h	(total words/h)	h	(total words/h)
B_PL_01-2	S2	742	cust-A	T1	Translator	2.65	280	1.68	442	3	2	0	2	7	9	Superior	2	E1	0.53	1400	0.38	1953
B_PL_03-2	S2	445	cust-A	T2	Senior translator	1.59	280	1.08	412	3	1	2	2	8	18	Good	2	E1	0.32	1400	0.30	1483
B_PL_04-2	S2	575	cust-A	T2	Senior translator	2.05	280	1.35	426	2	0	3	2	7	12	Good	2	E1	0.41	1400	0.33	1742
B_PL_05-2	S2	549	cust-A	T3	Senior translator	1.96	280	1.48	371	4	1	3	2	10	18	Good	1	E1	0.39	1400	0.35	1569
B_PL_05-4	S2	390.25	cust-A	T3	Senior translator	1.39	280	1.07	365	2	0	1	0	3	8	Superior	1	E1	0.28	1400	0.18	2168
B_PL_06-2	S2	595	cust-A	T1	Translator	2.13	280	1.67	356	5	8	5	1	19	32	Mediocre	1	E1	0.43	1400	0.48	1240
B_PL_06-4	S2	417	cust-A	T1	Translator	1.49	280	1.20	348	3	0	1	8	12	29	Good	1	E1	0.3	1400	0.32	1303
B_PL_06-6	S2	517	cust-A	T1	Translator	1.85	280	1.42	364	7	1	5	2	15	29	Good	1	E1	0.37	1400	0.42	1231
B_PL_07-1	S2	612	cust-B	T4	Translator	2.19	280	1.65	371	1	3	0	6	10	16	Good	2	E2	0.44	1400	0.38	1611
B_PL_08-2	S2	668.5	cust-B	T4	Translator	2.39	280	1.85	361	10	2	8	2	22	33	Mediocre	2	E2	0.48	1400	0.52	1286
B_PL_09-2	S2	646	cust-B	T2	Senior translator	2.31	280	2.02	320	4	8	4	1	17	26	Good	1	E2	0.46	1400	0.46	1404
B_PL_09-4	S2	569	cust-B	T2	Senior translator	2.03	280	1.80	316	2	0	3	2	7	12	Good	1	E2	0.41	1400	0.33	1724
B_PL_10-2	S2	533	cust-B	T5	Translator	1.90	280	1.78	299	3	1	6	8	18	34	Mediocre	1	E2	0.38	1400	0.47	1134
B_PL_10-4	S2	491.25	cust-B	T5	Translator	1.75	280	1.65	298	10	3	0	6	19	39	Mediocre	1	E2	0.35	1400	0.42	1170
B_PL_11-2	S2	547.75	cust-C	T4	Translator	1.96	280	1.55	353	2	5	6	4	17	31	Mediocre	2	E2	0.39	1400	0.45	1217
B_PL_11-4	S2	429.5	cust-C	T4	Translator	1.53	280	1.18	364	5	6	4	2	17	40	Mediocre	2	E2	0.31	1400	0.42	1023
B_PL_13-1	S2	426.5	cust-D	T5	Translator	1.52	280	1.38	309	3	2	2	4	11	26	Good	2	E2	0.3	1400	0.35	1219
Total		**9153.75**				**1.92**	**280.00**	**1.52**	**357.35**	**4.06**	**2.53**	**3.12**	**3.18**	**12.88**	**24.24**	**Good**	**1.47**		**0.39**	**1400.00**	**0.39**	**1439.82**

Appendix 10

Detailed results of evaluation for English-Estonian in the second experiment

Task ID (file name)	Scenario	Text size	Trans-lator name	Translator qualification (translator, senior translator)	Esti-mated time	Planned perfor-mance,	Actual time	Actual perfor-mance,	Quality assesment, negative points						Quality total valuation	MT quality feedback where text was not found in TM (Appendix 1, Question 1)	MT quality feedback where text contained formatting or other tags/mark-up (Appendix 1, Question 2)
(in LPS)	(S1, S2)	(adjustted words)			h	(adjusted words/h)	h	(adjusted words/h)	Accu-racy	Lang-uage quality	Style	Termi-nology	Total	Total (per 1000 words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3(best)	Score 1-3(best)
Text 1_1	S1	311.8	T1	Senior translator	0.89 h	350	0.50 h	624	1	0	1	1	3		Good	n/a	n/a
Text 2_1	S1	335	T1	Senior translator	0.96 h	349	0.45 h	744	0	0	0	0	0		Superior	n/a	n/a
Text 3_1	S1	307.1	T1	Senior translator	0.88 h	349	0.65 h	472	1	1	1	3	6		Mediocre	n/a	n/a
Text 4_1	S1	311	T1	Senior translator	0.89 h	349	0.50 h	622	0	4	0	0	4		Good	n/a	n/a
Text 5_1	S1	312.4	T2	Translator	0.89 h	351	0.50 h	625	0	4	0	0	4		Good	n/a	n/a
Text 6_1	S1	330	T2	Translator	0.94 h	351	0.60 h	550	0	2	3	3	8		Mediocre	n/a	n/a
Text 7_1	S1	306	T2	Translator	0.87 h	352	0.60 h	510	3	0	2	1	6		Mediocre	n/a	n/a
Text 8_1	S1	328	T2	Translator	0.94 h	349	0.50 h	656	1	2	0	1	4		Good	n/a	n/a
Text 9_1	S1	296	T3	Translator	0.85 h	348	0.90 h	329	0	0	0	0	0		Superior	n/a	n/a
Text 10_1	S1	311	T3	Translator	0.89 h	349	1.00 h	311	0	1	0	0	1		Superior	n/a	n/a
Text 11_1	S1	320	T3	Translator	0.91 h	352	1.10 h	291	0	0	0	0	0		Superior	n/a	n/a
Text 12_1	S1	271	T3	Translator	0.77 h	352	0.90 h	301	0	2	0	0	2		Superior	n/a	n/a
Text 13_1	S1	306	T4	Translator	0.87 h	352	0.67 h	457	0	5	2	0	7		Good	n/a	n/a
Text 14_1	S1	279.4	T4	Translator	0.80 h	349	0.58 h	482	0	2	4	3	9		Poor	n/a	n/a
Text 15_1	S1	278.2	T4	Translator	0.79 h	352	0.50 h	556	2	5	3	2	12		Very Poor	n/a	n/a
Text 16_1	S1	294.2	T4	Translator	0.84 h	350	0.67 h	439	3	2	3	2	10		Poor	n/a	n/a
Text 17_1	S1	320	T5	Senior translator	0.91 h	352	0.50 h	640	0	0	0	0	0		Superior	n/a	n/a
Text 18_1	S1	245.4	T5	Senior translator	0.70 h	351	0.25 h	982	0	0	0	0	0		Superior	n/a	n/a
Text 19_1	S1	288.2	T5	Senior translator	0.82 h	351	0.33 h	873	0	0	0	0	0		Superior	n/a	n/a
Text 20_1	S1	290.2	T5	Senior translator	0.83 h	350	0.42 h	691	0	0	0	0	0		Superior	n/a	n/a
Text 21_1	S1	325	T5	Senior translator	0.93 h	349	0.75 h	433	0	3	1	0	4		Good	n/a	n/a
Text 22_1	S1	344.2	T5	Senior translator	0.98 h	351	0.50 h	688	0	1	0	0	1		Superior	n/a	n/a
Text 23_1	S1	323	T1	Senior translator	0.92 h	351	0.60 h	538	1	0	1	0	2		Good	n/a	n/a
Text 24_1	S1	296	T1	Senior translator	0.85 h	348	0.50 h	592	0	0	1	0	1		Superior	n/a	n/a
Text 25_1	S1	309	T2	Translator	0.88 h	351	0.60 h	515	1	0	2	0	3		Superior	n/a	n/a
Text 26_1	S1	305	T2	Translator	0.87 h	351	0.40 h	763	1	0	0	2	3		Good	n/a	n/a
Text 27_1	S1	299	T3	Translator	0.85 h	352	1.20 h	249	0	2	0	0	2		Superior	n/a	n/a
Text 28_1	S1	329	T3	Translator	0.94 h	350	1.20 h	274	0	3	0	0	3		Superior	n/a	n/a
Text 29_1	S1	312	T4	Translator	0.89 h	351	0.67 h	466	1	2	4	2	9		Poor	n/a	n/a
Text 30_1	S1	263	T4	Translator	0.75 h	351	0.50 h	526	2	6	0	1	9		Mediocre	n/a	n/a
Text 31_1	S1	318	T2	Translator	0.91 h	349	0.60 h	530	3	2	0	0	5		Mediocre	n/a	n/a
Text 32_1	S1	299.1	T4	Translator	0.85 h	352	0.67 h	446	1	6	1	1	9		Poor	n/a	n/a
Text 33_1	S1	306	T3	Translator	0.87 h	352	1.20 h	255	0	0	0	0	0		Superior	n/a	n/a
Text 34_1	S1	306	T1	Senior translator	0.87 h	352	0.58 h	528	0	1	1	0	2		Superior	n/a	n/a
Text 35_1	S1	306	T5	Senior translator	0.87 h	352	0.50 h	612	0	0	0	0	0		Superior	n/a	n/a
Text 36_1	S1	339	T1	Senior translator	0.97 h	349	0.60 h	565	0	1	1	0	2		Superior	n/a	n/a
Text 37_1	S1	317	T2	Translator	0.91 h	348	0.70 h	453	0	1	2	6	9		Poor	n/a	n/a
Text 38_1	S1	309.8	T1	Senior translator	0.89 h	348	0.40 h	775	0	3	1	1	5		Good	n/a	n/a
Text 39_1	S1	293.8	T2	Translator	0.84 h	350	0.80 h	367	2	4	1	2	9		Poor	n/a	n/a

Task ID (file name)	Scenario	Text size	Trans-lator name	Translator qualification (translator, senior translator)	Esti-mated time	Planned perfor-mance,	Actual time	Actual perfor-mance,	Quality assesment, negative points						Quality total valuation	MT quality feedback where text was not found in TM (Appendix 1, Question 1)	MT quality feedback where text contained formatting or other tags/mark-up (Appendix 1, Question 2)
(in LPS)	(S1, S2)	(adjustted words)			h	(adjusted words/h)	h	(adjusted words/h)	Accu-racy	Lang-uage quality	Style	Termi-nology	Total	Total (per 1000 words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3(best)	Score 1-3(best)
Text 40_1	S1	298.9	T1	Senior translator	0.85 h	352	0.50 h	598	0	2	2	0	4		Good	n/a	n/a
Text 41_1	S1	278.8	T2	Translator	0.80 h	349	0.50 h	558	1	2	0	0	3		Good	n/a	n/a
Text 42_1	S1	294.4	T4	Translator	0.84 h	350	0.50 h	589	2	6	1	2	11		Poor	n/a	n/a
Text 43_1	S1	291	T5	Senior translator	0.83 h	351	0.33 h	882	0	2	0	0	2		Superior	n/a	n/a
Text 44_1	S1	264.2	T3	Translator	0.75 h	352	1.00 h	264	0	0	1	0	1		Superior	n/a	n/a
Text 45_1	S1	306.5	T5	Senior translator	0.88 h	348	0.50 h	613	0	0	1	0	1		Superior	n/a	n/a
Text 46_1	S1	300.6	T3	Translator	0.86 h	350	1.00 h	301	0	0	1	0	1		Superior	n/a	n/a
Text 47_1	S1	276.3	T4	Translator	0.79 h	350	0.50 h	553	0	5	3	0	8		Mediocre	n/a	n/a
Text 48_1	S1	302.2	T3	Translator	0.86 h	351	1.10 h	275	0	0	1	0	1		Superior	n/a	n/a
Text 49_1	S1	303.7	T5	Senior translator	0.87 h	349	0.58 h	524	0	0	0	0	0		Superior	n/a	n/a
Text 50_1	S1	300.8	T4	Translator	0.86 h	350	0.67 h	449	0	4	5	1	10		Mediocre	n/a	n/a
Text 1_2	S2	268.2	T1	Senior translator	0.77 h	348	0.60 h	447	0	1	0	2	3		Good	1	1
Text 2_2	S2	311.4	T1	Senior translator	0.89 h	350	0.50 h	623	0	2	1	2	5		Good	1	1
Text 3_2	S2	316	T1	Senior translator	0.90 h	351	0.50 h	632	0	1	2	1	4		Good	2	2
Text 4_2	S2	305.8	T1	Senior translator	0.87 h	351	0.55 h	556	1	2	0	1	4		Good	1	1
Text 5_2	S2	307	T2	Translator	0.88 h	349	0.50 h	614	0	0	0	0	0		Superior	2	2
Text 6_2	S2	318.5	T2	Translator	0.91 h	350	0.70 h	455	1	0	1	0	2		Superior	2	3
Text 7_2	S2	302	T2	Translator	0.86 h	351	0.40 h	755	0	0	1	0	1		Superior	2	3
Text 8_2	S2	306	T2	Translator	0.87 h	352	0.60 h	510	0	1	0	0	1		Superior	2	3
Text 9_2	S2	275	T3	Translator	0.79 h	348	0.70 h	393	0	0	0	0	0		Superior	1	1
Text 10_2	S2	309.1	T3	Translator	0.88 h	351	1.20 h	258	0	1	1	1	3		Superior	1	1
Text 11_2	S2	253.8	T3	Translator	0.73 h	348	0.80 h	317	0	1	1	0	2		Superior	1	1
Text 12_2	S2	312	T3	Translator	0.89 h	351	1.00 h	312	0	1	0	0	1		Superior	1	1
Text 13_2	S2	299.6	T4	Translator	0.86 h	348	0.58 h	517	0	7	2	2	11		Mediocre	2	2
Text 14_2	S2	272.6	T4	Translator	0.78 h	349	0.58 h	470	1	5	2	1	9		Mediocre	2	1
Text 15_2	S2	303.8	T4	Translator	0.87 h	349	0.67 h	453	0	6	1	4	11		Mediocre	2	1
Text 16_2	S2	293	T4	Translator	0.84 h	349	0.67 h	437	3	7	6	0	16		Poor	2	2
Text 17_2	S2	292	T5	Senior translator	0.83 h	352	0.50 h	584	0	0	0	0	0		Superior	1	n/a
Text 18_2	S2	295.4	T5	Senior translator	0.84 h	352	0.58 h	509	0	1	1	0	2		Good	1	2
Text 19_2	S2	303.4	T5	Senior translator	0.87 h	349	0.50 h	607	0	1	1	0	2		Superior	1	n/a
Text 20_2	S2	308.6	T5	Senior translator	0.88 h	351	0.50 h	617	0	0	0	0	0		Superior	1	n/a
Text 21_2	S2	269	T5	Senior translator	0.77 h	349	0.58 h	464	0	2	0	0	2		Superior	1	3
Text 22_2	S2	281	T5	Senior translator	0.80 h	351	0.33 h	852	0	1	0	0	1		Superior	1	n/a
Text 23_2	S2	301	T1	Senior translator	0.86 h	350	0.55 h	547	0	0	0	0	0		Superior	1	1
Text 24_2	S2	328	T1	Senior translator	0.94 h	349	0.70 h	469	1	1	0	0	2		Good	1	1
Text 25_2	S2	296	T2	Translator	0.85 h	348	0.40 h	740	0	0	1	0	1		Superior	1	2
Text 26_2	S2	315	T2	Translator	0.90 h	350	0.60 h	525	2	1	0	0	3		Good	2	2
Text 27_2	S2	312.1	T3	Translator	0.89 h	351	1.00 h	312	0	0	0	0	0		Superior	1	1
Text 28_2	S2	303	T3	Translator	0.87 h	348	1.10 h	275	0	1	0	0	1		Superior	1	1
Text 29_2	S2	310	T4	Translator	0.89 h	348	0.75 h	413	1	6	1	1	9		Mediocre	1	1
Text 30_2	S2	321.8	T4	Translator	0.92 h	350	0.92 h	350	3	7	1	3	14		Poor	1	1
Text 31_2	S2	293	T2	Translator	0.84 h	349	0.70 h	419	0	0	0	0	0		Superior	1	2
Text 32_2	S2	299.6	T4	Translator	0.86 h	348	0.75 h	399	2	5	4	5	16		Very Poor	1	n/a
Text 33_2	S2	318	T3	Translator	0.91 h	349	1.00 h	318	0	1	1	0	2		Superior	1	1
Text 34_2	S2	293	T1	Senior translator	0.84 h	349	0.50 h	586	0	1	1	1	3		Good	1	1
Text 35_2	S2	313	T5	Senior translator	0.89 h	352	0.75 h	417	0	0	0	0	0		Superior	1	2
Text 36_2	S2	298.9	T1	Senior translator	0.85 h	352	0.55 h	543	0	1	1	0	2		Superior	2	1

Task ID (file name)	Scenario	Text size	Trans-lator name	Translator qualification (translator, senior translator)	Esti-mated time	Planned perfor-mance,	Actual time	Actual perfor-mance,	Quality assesment, negative points						Quality total valuation	MT quality feedback where text was not found in TM (Appendix 1, Question 1)	MT quality feedback where text contained formatting or other tags/mark-up (Appendix 1, Question 2)
(in LPS)	(S1, S2)	(adjustted words)			h	(adjusted words/h)	h	(adjusted words/h)	Accu-racy	Lang-uage quality	Style	Termi-nology	Total	Total (per 1000 words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3(best)	Score 1-3(best)
Text 37_2	S2	323	T2	Translator	0.92 h	351	0.60 h	538	0	0	1	0	1		Superior	2	3
Text 38_2	S2	310.9	T1	Senior translator	0.89 h	349	0.50 h	622	0	0	1	0	1		Superior	1	1
Text 39_2	S2	304.6	T2	Translator	0.87 h	350	0.60 h	508	0	2	2	0	4		Good	2	2
Text 40_2	S2	293.2	T1	Senior translator	0.84 h	349	0.75 h	391	1	0	0	0	1		Good	2	2
Text 41_2	S2	164.1	T2	Translator	0.47 h	349	0.40 h	410	0	0	1	0	1		Superior	2	3
Text 42_2	S2	269.7	T4	Translator	0.77 h	350	0.75 h	360	0	6	0	1	7		Mediocre	1	2
Text 43_2	S2	223.2	T5	Senior translator	0.64 h	349	0.66 h	338	0	0	0	0	0		Superior	1	2
Text 44_2	S2	296.9	T3	Translator	0.85 h	349	0.75 h	396	0	1	1	0	2		Superior	3	2
Text 45_2	S2	290.7	T5	Senior translator	0.83 h	350	0.33 h	881	0	0	1	0	1		Superior	1	2
Text 46_2	S2	280.5	T3	Translator	0.80 h	351	0.80 h	351	0	1	1	0	2		Superior	2	2
Text 47_2	S2	148.4	T4	Translator	0.42 h	353	0.50 h	297	0	5	2	1	8		Poor	1	2
Text 48_2	S2	308	T3	Translator	0.88 h	350	0.75 h	411	0	0	0	0	0		Superior	2	2
Text 49_2	S2	321	T5	Senior translator	0.92 h	349	0.75 h	428	0	1	0	0	1		Superior	1	n/a
Text 50_2	S2	298.2	T4	Translator	0.85 h	351	0.75 h	398	0	8	4	2	14		Poor	2	2

Appendix 11

Detailed results of evaluation for English-Latvian in the second experiment

Task ID (file name)	Scenario,	Text size,	Trans-lator name	Translator qualification (translator, senior translator)	Esti-mated time	Planned perfor-mance,	Actual time	Actual perfor-mance,	Quality assesment, negative points						Quality total valuation	MT quality feedback where text was not found in TM (Appendix 1, Question 1)	MT quality feedback where text contained formatting or other tags/mark-up (Appendix 1, Question 2)
(in LPS)	(S1, S2)	(adjusted words)			h	(adjusted words/h)	h	(adjusted words/h)	Accu-racy	Lang-uage quality	Style	Termi-nology	Total	Total (per 1000 words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3(best)	Score 1-3(best)
Text 1_1	S1	311.80	T1	Translator	0.89 h	350	0.75 h	416	0	5	0	2	7	22	Good	n/a	n/a
Text 2_1	S1	335.00	T1	Translator	0.96 h	349	0.75 h	447	2	3	0	2	7	21	Good	n/a	n/a
Text 3_1	S1	307.10	T1	Translator	0.88 h	349	0.65 h	472	2	5	0	2	9	29	Good	n/a	n/a
Text 4_1	S1	311.00	T1	Translator	0.89 h	349	0.75 h	415	2	5	0	4	11	35	Mediocre	n/a	n/a
Text 5_1	S1	312.40	T2	Senior translator	0.89 h	351	0.57 h	548	6	4	0	4	14	45	Mediocre	n/a	n/a
Text 6_1	S1	330.00	T2	Senior translator	0.94 h	351	0.45 h	733	2	3	0	0	5	15	Good	n/a	n/a
Text 7_1	S1	306.00	T2	Senior translator	0.87 h	352	0.40 h	765	2	3	0	0	5	16	Good	n/a	n/a
Text 8_1	S1	328.00	T2	Senior translator	0.94 h	349	0.35 h	937	6	2	0	4	12	37	Mediocre	n/a	n/a
Text 9_1	S1	296.00	T3	Senior translator	0.85 h	348	0.33 h	897	0	1	0	0	1	3	Superior	n/a	n/a
Text 10_1	S1	311.00	T3	Senior translator	0.89 h	349	0.45 h	691	0	3	0	0	3	10	Superior	n/a	n/a
Text 11_1	S1	320.00	T3	Senior translator	0.91 h	352	0.40 h	800	0	1	0	0	1	3	Superior	n/a	n/a
Text 12_1	S1	271.00	T3	Senior translator	0.77 h	352	0.45 h	602	0	0	0	0	0	0	Superior	n/a	n/a
Text 13_1	S1	306.00	T4	Translator	0.87 h	352	0.46 h	665	5	1	0	8	14	46	Mediocre	n/a	n/a
Text 14_1	S1	279.40	T4	Translator	0.80 h	349	0.57 h	490	4	3	0	8	15	54	Poor	n/a	n/a
Text 15_1	S1	278.20	T4	Translator	0.79 h	352	0.50 h	556	5	4	0	4	13	47	Mediocre	n/a	n/a
Text 16_1	S1	294.20	T4	Translator	0.84 h	350	0.48 h	613	3	4	0	8	15	51	Poor	n/a	n/a
Text 17_1	S1	320.00	T5	Translator	0.91 h	352	0.50 h	640	3	1	0	0	4	13	Good	n/a	n/a
Text 18_1	S1	245.40	T5	Translator	0.70 h	351	0.35 h	701	0	0	0	2	2	8	Superior	n/a	n/a
Text 19_1	S1	288.20	T5	Translator	0.82 h	351	0.35 h	823	4	0	0	2	6	21	Good	n/a	n/a
Text 20_1	S1	290.20	T5	Translator	0.83 h	350	0.35 h	829	5	1	0	2	8	28	Good	n/a	n/a
Text 21_1	S1	325.00	T5	Translator	0.93 h	349	0.75 h	433	0	0	0	0	0	0	Superior	n/a	n/a
Text 22_1	S1	344.20	T5	Translator	0.98 h	351	0.75 h	459	3	0	0	0	3	9	Superior	n/a	n/a
Text 23_1	S1	323.00	T1	Translator	0.92 h	351	0.70 h	461	2	0	0	0	2	6	Superior	n/a	n/a
Text 24_1	S1	296.00	T1	Translator	0.85 h	348	0.65 h	455	0	2	0	0	2	7	Superior	n/a	n/a
Text 25_1	S1	309.00	T2	Senior translator	0.88 h	351	0.67 h	461	0	7	0	0	7	23	Good	n/a	n/a
Text 26_1	S1	305.00	T2	Senior translator	0.87 h	351	0.35 h	871	0	2	0	0	2	7	Superior	n/a	n/a
Text 27_1	S1	299.00	T3	Senior translator	0.85 h	352	0.45 h	664	2	0	0	0	2	7	Superior	n/a	n/a
Text 28_1	S1	329.00	T3	Senior translator	0.94 h	350	0.40 h	823	0	0	0	0	0	0	Superior	n/a	n/a
Text 29_1	S1	312.00	T4	Translator	0.89 h	351	0.47 h	664	2	2	0	6	10	32	Mediocre	n/a	n/a
Text 30_1	S1	263.00	T4	Translator	0.75 h	351	0.57 h	461	0	2	0	4	6	23	Good	n/a	n/a
Text 31_1	S1	318.00	T2	Senior translator	0.91 h	349	0.45 h	707	2	2	0	2	6	19	Good	n/a	n/a
Text 32_1	S1	299.10	T4	Translator	0.85 h	352	0.41 h	730	0	4	0	2	6	20	Good	n/a	n/a
Text 33_1	S1	306.00	T3	Senior translator	0.87 h	352	0.45 h	680	0	2	1	0	3	10	Superior	n/a	n/a
Text 34_1	S1	306.00	T1	Translator	0.87 h	352	0.65 h	471	0	2	0	2	4	13	Good	n/a	n/a
Text 35_1	S1	306.00	T5	Translator	0.87 h	352	0.50 h	612	4	0	1	0	5	16	Good	n/a	n/a
Text 36_1	S1	339.00	T1	Translator	0.97 h	349	0.80 h	424	3	5	0	0	8	24	Good	n/a	n/a
Text 37_1	S1	317.00	T2	Senior translator	0.91 h	348	0.43 h	737	0	2	1	0	3	9	Superior	n/a	n/a
Text 38_1	S1	309.80	T1	Translator	0.89 h	348	0.75 h	413	1	0	0	0	1	3	Superior	n/a	n/a
Text 39_1	S1	293.80	T2	Senior translator	0.84 h	350	0.40 h	735	0	2	0	2	4	14	Good	n/a	n/a

Task ID (file name)	Scenario,	Text size,	Trans-lator name	Translator qualification (translator, senior translator)	Esti-mated time	Planned perfor-mance,	Actual time	Actual perfor-mance,	Quality assesment, negative points						Quality total valuation	MT quality feedback where text was not found in TM (Appendix 1, Question 1)	MT quality feedback where text contained formatting or other tags/mark-up (Appendix 1, Question 2)
(in LPS)	(S1, S2)	(adjusted words)			h	(adjusted words/h)	h	(adjusted words/h)	Accu-racy	Lang-uage quality	Style	Termi-nology	Total	Total (per 1000 words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3(best)	Score 1-3(best)
Text 40_1	S1	298.90	T1	Translator	0.85 h	352	0.65 h	460	0	2	0	0	2	7	Superior	n/a	n/a
Text 41_1	S1	278.80	T2	Senior translator	0.80 h	349	0.27 h	1033	0	3	0	6	9	32	Mediocre	n/a	n/a
Text 42_1	S1	294.40	T4	Translator	0.84 h	350	0.56 h	526	0	2	0	4	6	20	Good	n/a	n/a
Text 43_1	S1	291.00	T5	Translator	0.83 h	351	0.50 h	582	0	1	0	0	1	3	Superior	n/a	n/a
Text 44_1	S1	264.20	T3	Senior translator	0.75 h	352	0.43 h	614	0	5	0	0	5	19	Good	n/a	n/a
Text 45_1	S1	306.50	T5	Translator	0.88 h	348	0.75 h	409	0	2	0	2	4	13	Good	n/a	n/a
Text 46_1	S1	300.60	T3	Senior translator	0.86 h	350	0.41 h	733	0	0	0	0	0	0	Superior	n/a	n/a
Text 47_1	S1	276.30	T4	Translator	0.79 h	350	0.63 h	439	0	7	0	6	13	47	Mediocre	n/a	n/a
Text 48_1	S1	302.20	T3	Senior translator	0.86 h	351	0.33 h	916	0	4	1	0	5	17	Good	n/a	n/a
Text 49_1	S1	303.70	T5	Translator	0.87 h	349	0.75 h	405	0	2	0	2	4	13	Good	n/a	n/a
Text 50_1	S1	300.80	T4	Translator	0.86 h	350	0.58 h	519	0	8	0	0	8	27	Good	n/a	n/a
Text 1_2	S2	268.20	T1	Translator	0.77 h	348	0.60 h	447	5	2	0	2	9	34	Mediocre	1	1
Text 2_2	S2	311.40	T1	Translator	0.89 h	350	0.65 h	479	2	2	0	10	14	45	Mediocre	2	3
Text 3_2	S2	316.00	T1	Translator	0.90 h	351	0.70 h	451	5	2	0	2	9	28	Good	1	1
Text 4_2	S2	305.80	T1	Translator	0.87 h	351	0.60 h	510	4	0	0	4	8	26	Good	2	n/a
Text 5_2	S2	307.00	T2	Senior translator	0.88 h	349	0.37 h	830	1	3	0	4	8	26	Good	2	2
Text 6_2	S2	318.50	T2	Senior translator	0.91 h	350	0.50 h	637	2	4	1	0	7	22	Good	2	2
Text 7_2	S2	302.00	T2	Senior translator	0.86 h	351	0.30 h	1007	2	3	0	2	7	23	Good	2	3
Text 8_2	S2	306.00	T2	Senior translator	0.87 h	352	0.30 h	1020	2	5	0	4	11	36	Mediocre	2	3
Text 9_2	S2	275.00	T3	Senior translator	0.79 h	348	0.35 h	786	0	2	0	0	2	7	Superior	1	2
Text 10_2	S2	309.10	T3	Senior translator	0.88 h	351	0.41 h	754	0	2	0	2	4	13	Good	1	2
Text 11_2	S2	253.80	T3	Senior translator	0.73 h	348	0.35 h	725	0	4	0	2	6	24	Good	1	3
Text 12_2	S2	312.00	T3	Senior translator	0.89 h	351	0.37 h	843	0	5	0	2	7	22	Good	1	1
Text 13_2	S2	299.60	T4	Translator	0.86 h	348	0.70 h	428	2	6	0	6	14	47	Mediocre	2	2
Text 14_2	S2	272.60	T4	Translator	0.78 h	349	0.58 h	470	4	2	0	6	12	44	Mediocre	1	2
Text 15_2	S2	303.80	T4	Translator	0.87 h	349	0.55 h	552	4	1	0	4	9	30	Good	2	2
Text 16_2	S2	293.00	T4	Translator	0.84 h	349	0.53 h	553	0	6	0	6	12	41	Mediocre	1	2
Text 17_2	S2	292.00	T5	Translator	0.83 h	352	0.50 h	584	3	2	0	2	7	24	Good	2	3
Text 18_2	S2	295.40	T5	Translator	0.84 h	352	0.75 h	394	0	1	0	0	1	3	Superior	1	2
Text 19_2	S2	303.40	T5	Translator	0.87 h	349	0.50 h	607	4	0	0	4	8	26	Good	2	2
Text 20_2	S2	308.60	T5	Translator	0.88 h	351	0.75 h	411	0	0	0	8	8	26	Good	1/2	3
Text 21_2	S2	269.00	T5	Translator	0.77 h	349	0.75 h	359	2	0	0	2	4	15	Good	1	2
Text 22_2	S2	281.00	T5	Translator	0.80 h	351	0.75 h	375	0	1	0	2	3	11	Good	1	3
Text 23_2	S2	301.00	T1	Translator	0.86 h	350	0.65 h	463	0	5	0	0	5	17	Good	1	1
Text 24_2	S2	328.00	T1	Translator	0.94 h	349	0.75 h	437	0	0	0	2	2	6	Superior	2	1
Text 25_2	S2	296.00	T2	Senior translator	0.85 h	348	0.51 h	580	0	1	0	0	1	3	Superior	1	1
Text 26_2	S2	315.00	T2	Senior translator	0.90 h	350	0.50 h	630	2	4	2	0	8	25	Good	1	n/a
Text 27_2	S2	312.10	T3	Senior translator	0.89 h	351	0.40 h	780	0	3	0	0	3	10	Superior	1	3
Text 28_2	S2	303.00	T3	Senior translator	0.87 h	348	0.40 h	758	3	2	0	0	5	17	Good	2	N/a
Text 29_2	S2	310.00	T4	Translator	0.89 h	348	0.52 h	596	6	2	1	2	11	35	Mediocre	2	1
Text 30_2	S2	321.80	T4	Translator	0.92 h	350	0.65 h	495	0	2	0	6	8	25	Good	1	1
Text 31_2	S2	293.00	T2	Senior translator	0.84 h	349	0.53 h	553	3	5	0	2	10	34	Mediocre	1	2
Text 32_2	S2	299.60	T4	Translator	0.86 h	348	0.55 h	545	0	3	0	6	9	30	Mediocre	2	1
Text 33_2	S2	318.00	T3	Senior translator	0.91 h	349	0.55 h	578	0	1	0	4	5	16	Good	2	3
Text 34_2	S2	293.00	T1	Translator	0.84 h	349	0.65 h	451	0	0	0	0	0	0	Superior	1	3
Text 35_2	S2	313.00	T5	Translator	0.89 h	352	0.50 h	626	7	1	0	4	12	38	Mediocre	2	2
Text 36_2	S2	298.90	T1	Translator	0.85 h	352	0.60 h	498	0	1	1	2	4	13	Good	1	n/a

Task ID (file name)	Scenario,	Text size,	Trans-lator name	Translator qualification (translator, senior translator)	Esti-mated time	Planned perfor-mance,	Actual time	Actual perfor-mance,	Quality assesment, negative points						Quality total valuation	MT quality feedback where text was not found in TM (Appendix 1, Question 1)	MT quality feedback where text contained formatting or other tags/mark-up (Appendix 1, Question 2)
(in LPS)	(S1, S2)	(adjusted words)			h	(adjusted words/h)	h	(adjusted words/h)	Accu-racy	Lang-uage quality	Style	Termi-nology	Total	Total (per 1000 words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3(best)	Score 1-3(best)
Text 37_2	S2	323.00	T2	Senior translator	0.92 h	351	0.47 h	687	2	3	1	0	6	19	Good	1	n/a
Text 38_2	S2	310.90	T1	Translator	0.89 h	349	0.65 h	478	2	1	0	0	3	10	Superior	2	n/a
Text 39_2	S2	304.60	T2	Senior translator	0.87 h	350	0.37 h	823	0	7	0	0	7	23	Good	2	2
Text 40_2	S2	293.20	T1	Translator	0.84 h	349	0.60 h	489	0	4	0	2	6	20	Good	1	n/a
Text 41_2	S2	164.10	T2	Senior translator	0.47 h	349	0.20 h	821	3	2	0	2	7	43	Mediocre	2	2
Text 42_2	S2	269.70	T4	Translator	0.77 h	350	0.48 h	562	3	2	0	0	5	19	Good	1	1
Text 43_2	S2	223.20	T5	Translator	0.64 h	349	0.50 h	446	0	0	0	6	6	27	Good	2	2
Text 44_2	S2	296.90	T3	Senior translator	0.85 h	349	0.45 h	660	0	3	0	2	5	17	Good	2	1
Text 45_2	S2	290.70	T5	Translator	0.83 h	350	0.50 h	581	3	5	0	2	10	34	Mediocre	2	2
Text 46_2	S2	280.50	T3	Senior translator	0.80 h	351	0.43 h	652	0	2	0	2	4	14	Good	2	2
Text 47_2	S2	148.40	T4	Translator	0.42 h	353	0.32 h	464	3	3	0	4	10	67	Poor	2	3
Text 48_2	S2	308.00	T3	Senior translator	0.88 h	350	0.27 h	1141	0	3	0	2	5	16	Good	2	n/a
Text 49_2	S2	321.00	T5	Translator	0.92 h	349	0.75 h	428	0	0	0	0	0	0	Superior	2	2
Text 50_2	S2	298.20	T4	Translator	0.85 h	351	0.62 h	481	0	6	1	2	9	30	Mediocre	3	2

Appendix 12

Detailed results of evaluation for English-Lithuanian in the second experiment

Task ID (file name)	Scenario,	Text size,	Trans-lator name	Translator qualification (translator, senior translator)	Estimat-ed time	Planned perfor-mance,	Actual time	Actual perfor-mance,	Quality assesment, negative points						Quality total valuation	MT quality feedback where text was not found in TM (Appendix 1, Question 1)	MT quality feedback where text contained formatting or other tags/mark-up (Appendix 1, Question 2)
(in LPS)	(S1, S2)	(adjusted words)			h	(adjusted words/h)	h	(adjusted words/h)	Accu-racy	Lang-uage quality	Style	Termi-nology	Total	Total (per 1000 words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3(best)	Score 1-3(best)
Text 1_1	S1	311.8	T1	senior translator	0.64 h	487	0.25 h	1247	0	1	0	2	3	10	Good	n/a	n/a
Text 2_1	S1	335	T1	senior translator	0.69 h	486	0.20 h	1675	0	0	0	2	2	6	Superior	n/a	n/a
Text 3_1	S1	307.1	T1	senior translator	0.63 h	487	0.25 h	1228	0	3	1	2	6	20	Good	n/a	n/a
Text 4_1	S1	311	T1	senior translator	0.63 h	494	0.30 h	1037	0	0	2	2	4	13	Good	n/a	n/a
Text 5_1	S1	312.4	T2	translator	0.78 h	401	0.49 h	638	0	1	2	2	5	16	Good	n/a	n/a
Text 6_1	S1	330	T2	translator	0.82 h	402	0.43 h	767	0	4	1	0	5	15	Good	n/a	n/a
Text 7_1	S1	306	T2	translator	0.76 h	403	0.34 h	900	0	2	1	0	3	10	Good	n/a	n/a
Text 8_1	S1	328	T2	translator	0.82 h	400	0.57 h	575	0	0	2	0	2	6	Superior	n/a	n/a
Text 9_1	S1	296	T3	freelance translator	0.85 h	348	0.55 h	538	3	14	3	0	20	68	Poor	n/a	n/a
Text 10_1	S1	311	T3	freelance translator	0.89 h	349	0.53 h	587	3	5	2	2	12	39	Mediocre	n/a	n/a
Text 11_1	S1	320	T3	freelance translator	0.91 h	352	0.45 h	711	0	6	2	2	10	31	Mediocre	n/a	n/a
Text 12_1	S1	271	T3	freelance translator	0.77 h	352	0.48 h	565	3	1	2	1	7	26	Good	n/a	n/a
Text 13_1	S1	306	T4	translator	0.97 h	315	0.28 h	1093	2	3	0	0	5	16	Good	n/a	n/a
Text 14_1	S1	279.4	T4	translator	0.88 h	318	0.30 h	931	0	0	0	0	0	0	Superior	n/a	n/a
Text 15_1	S1	278.2	T4	translator	0.88 h	316	0.28 h	994	0	1	0	4	5	18	Good	n/a	n/a
Text 16_1	S1	294.2	T4	translator	0.93 h	316	0.26 h	1132	0	1	0	2	3	10	Good	n/a	n/a
Text 17_1	S1	320	T5	translator	0.91 h	352	0.50 h	640	0	3	0	2	5	16	Good	n/a	n/a
Text 18_1	S1	245.4	T5	translator	0.70 h	351	0.40 h	614	0	5	1	0	6	24	Good	n/a	n/a
Text 19_1	S1	288.2	T5	translator	0.82 h	351	0.50 h	576	0	2	1	2	5	17	Good	n/a	n/a
Text 20_1	S1	290.2	T5	translator	0.83 h	350	0.50 h	580	0	3	2	0	5	17	Good	n/a	n/a
Text 21_1	S1	325	T5	translator	0.93 h	349	0.45 h	722	0	2	2	0	4	12	Good	n/a	n/a
Text 22_1	S1	344.2	T5	translator	0.98 h	351	0.50 h	688	0	4	0	0	4	12	Good	n/a	n/a
Text 23_1	S1	323	T1	senior translator	0.66 h	489	0.25 h	1292	0	3	1	0	4	12	Good	n/a	n/a
Text 24_1	S1	296	T1	senior translator	0.61 h	485	0.25 h	1184	0	0	1	2	3	10	Good	n/a	n/a
Text 25_1	S1	309	T2	translator	0.77 h	401	0.75 h	412	0	2	1	0	3	10	Good	n/a	n/a
Text 26_1	S1	305	T2	translator	0.76 h	401	0.57 h	535	0	2	0	0	2	7	Superior	n/a	n/a
Text 27_1	S1	299	T3	freelance translator	0.85 h	352	0.88 h	340	0	5	6	2	13	43	Mediocre	n/a	n/a
Text 28_1	S1	329	T3	freelance translator	0.94 h	350	0.50 h	658	0	2	2	2	6	18	Good	n/a	n/a
Text 29_1	S1	312	T4	translator	0.99 h	315	0.30 h	1040	0	1	0	2	3	10	Good	n/a	n/a
Text 30_1	S1	263	T4	translator	0.83 h	317	0.25 h	1052	0	1	1	0	2	8	Superior	n/a	n/a
Text 31_1	S1	318	T2	translator	0.79 h	403	0.74 h	430	0	3	0	2	5	16	Good	n/a	n/a
Text 32_1	S1	299.1	T4	translator	0.94 h	318	0.26 h	1150	0	2	1	0	3	10	Good	n/a	n/a
Text 33_1	S1	306	T3	freelance translator	0.87 h	352	0.55 h	556	0	2	5	4	11	36	Mediocre	n/a	n/a
Text 34_1	S1	306	T1	senior translator	0.63 h	486	0.30 h	1020	0	2	0	0	2	7	Superior	n/a	n/a
Text 35_1	S1	306	T5	translator	0.87 h	352	0.35 h	874	0	4	0	2	6	20	Good	n/a	n/a
Text 36_1	S1	339	T1	senior translator	0.70 h	484	0.25 h	1356	0	2	2	0	4	12	Good	n/a	n/a
Text 37_1	S1	317	T2	translator	0.79 h	401	0.71 h	446	0	2	0	2	4	13	Good	n/a	n/a
Text 38_1	S1	309.8	T1	senior translator	0.64 h	484	0.25 h	1239	0	0	2	0	2	6	Superior	n/a	n/a
Text 39_1	S1	293.8	T2	translator	0.73 h	402	0.69 h	426	0	2	2	0	4	14	Good	n/a	n/a

Task ID (file name)	Scenario,	Text size,	Trans-lator name	Translator qualification (translator, senior translator)	Estimat-ed time	Planned perfor-mance,	Actual time	Actual perfor-mance,	Quality assesment, negative points						Quality total valuation	MT quality feedback where text was not found in TM (Appendix 1, Question 1)	MT quality feedback where text contained formatting or other tags/mark-up (Appendix 1, Question 2)
(in LPS)	(S1, S2)	(adjusted words)			h	(adjusted words/h)	h	(adjusted words/h)	Accu-racy	Lang-uage quality	Style	Termi-nology	Total	Total (per 1000 words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3(best)	Score 1-3(best)
Text 40_1	S1	298.9	T1	senior translator	0.61 h	490	0.25 h	1196	0	0	0	0	0	0	Superior	n/a	n/a
Text 41_1	S1	278.8	T2	translator	0.69 h	404	0.34 h	820	0	5	1	2	8	29	Good	n/a	n/a
Text 42_1	S1	294.4	T4	translator	0.94 h	313	0.33 h	892	0	3	0	2	5	17	Good	n/a	n/a
Text 43_1	S1	291	T5	translator	0.83 h	351	0.30 h	970	0	0	0	0	0	0	Superior	n/a	n/a
Text 44_1	S1	264.2	T3	freelance translator	0.75 h	352	0.50 h	528	0	11	2	0	13	49	Mediocre	n/a	n/a
Text 45_1	S1	306.5	T5	translator	0.87 h	352	0.30 h	1022	0	2	1	0	3	10	Good	n/a	n/a
Text 46_1	S1	300.6	T3	freelance translator	0.86 h	350	0.48 h	626	3	6	2	4	15	50	Poor	n/a	n/a
Text 47_1	S1	276.3	T4	translator	0.87 h	318	0.33 h	837	0	1	0	0	1	4	Superior	n/a	n/a
Text 48_1	S1	302.2	T3	freelance translator	0.86 h	351	0.60 h	504	0	4	3	2	9	30	Mediocre	n/a	n/a
Text 49_1	S1	303.7	T5	translator	0.87 h	349	0.33 h	920	0	0	2	0	2	7	Superior	n/a	n/a
Text 50_1	S1	300.8	T4	translator	0.87 h	346	0.41 h	734	0	4	0	0	4	13	Good	n/a	n/a
Text 1_2	S2	268.2	T1	senior translator	0.55 h	488	0.30 h	894	0	0	0	4	4	15	Good	1	3
Text 2_2	S2	311.4	T1	senior translator	0.64 h	487	0.30 h	1038	0	0	0	2	2	6	Superior	1	3
Text 3_2	S2	316	T1	senior translator	0.65 h	486	0.30 h	1053	0	3	0	2	5	16	Good	1	3
Text 4_2	S2	305.8	T1	senior translator	0.63 h	485	0.30 h	1019	0	3	1	0	4	13	Good	1	3
Text 5_2	S2	307	T2	translator	0.76 h	404	0.59 h	520	0	6	0	0	6	20	Good	1	2
Text 6_2	S2	318.5	T2	translator	0.79 h	403	0.54 h	590	5	5	2	4	16	50	Poor	1	2
Text 7_2	S2	302	T2	translator	0.75 h	403	0.49 h	616	0	2	2	0	4	13	Good	1	2
Text 8_2	S2	306	T2	translator	0.76 h	403	0.48 h	638	0	0	2	4	6	20	Good	1	2
Text 9_2	S2	275	T3	freelance translator	0.79 h	348	0.40 h	688	5	1	3	6	15	55	Poor	1	1
Text 10_2	S2	309.1	T3	freelance translator	0.88 h	351	0.78 h	396	2	9	3	6	20	65	Poor	2	2
Text 11_2	S2	253.8	T3	freelance translator	0.73 h	348	0.38 h	668	0	5	5	6	16	63	Poor	2	2
Text 12_2	S2	312	T3	freelance translator	0.89 h	351	0.40 h	780	4	5	3	6	18	58	Poor	2	2
Text 13_2	S2	299.6	T4	translator	0.95 h	315	0.35 h	856	0	0	0	2	2	7	Superior	1	2
Text 14_2	S2	272.6	T4	translator	0.86 h	317	0.23 h	1185	0	0	0	2	2	7	Superior	2	2
Text 15_2	S2	303.8	T4	translator	0.96 h	316	0.30 h	1013	4	0	0	2	6	20	Good	1	1
Text 16_2	S2	293	T4	translator	0.93 h	315	0.26 h	1127	0	0	0	0	0	0	Superior	1	2
Text 17_2	S2	292	T5	translator	0.83 h	352	0.50 h	584	4	4	0	0	8	27	Good	2	1
Text 18_2	S2	295.4	T5	translator	0.84 h	352	0.50 h	591	0	2	0	2	4	14	Good	1	1
Text 19_2	S2	303.4	T5	translator	0.87 h	349	0.50 h	607	0	1	0	0	1	3	Superior	1	1
Text 20_2	S2	308.6	T5	translator	0.88 h	351	0.50 h	617	0	1	0	0	1	3	Superior	1	1
Text 21_2	S2	269	T5	translator	0.77 h	349	0.70 h	384	2	2	1	0	5	19	Good	1	3
Text 22_2	S2	281	T5	translator	0.80 h	351	0.50 h	562	0	2	0	0	2	7	Superior	1	3
Text 23_2	S2	301	T1	senior translator	0.62 h	485	0.30 h	1003	0	5	1	0	6	20	Good	1	1
Text 24_2	S2	328	T1	senior translator	0.67 h	490	0.30 h	1093	0	0	2	0	2	6	Superior	1	1
Text 25_2	S2	296	T2	translator	0.74 h	400	0.69 h	429	0	5	1	0	6	20	Good	1	1
Text 26_2	S2	315	T2	translator	0.78 h	404	0.59 h	534	0	7	0	4	11	35	Mediocre	1	1
Text 27_2	S2	312.1	T3	freelance translator	0.89 h	351	0.56 h	557	2	13	3	0	18	58	Poor	1	1
Text 28_2	S2	303	T3	freelance translator	0.87 h	348	0.48 h	631	0	7	2	4	13	43	Mediocre	1	1
Text 29_2	S2	310	T4	translator	0.98 h	316	0.35 h	886	0	0	0	2	2	6	Superior	1	1
Text 30_2	S2	321.8	T4	translator	1.02 h	315	0.35 h	919	0	0	0	0	0	0	Superior	1	1
Text 31_2	S2	293	T2	translator	0.73 h	401	0.51 h	575	6	0	8	6	20	68	Poor	1	1
Text 32_2	S2	299.6	T4	translator	0.95 h	315	0.28 h	1070	0	2	0	0	2	7	Superior	2	3
Text 33_2	S2	318	T3	freelance translator	0.91 h	349	0.50 h	636	5	5	8	0	18	57	Poor	2	3
Text 34_2	S2	293	T1	senior translator	0.60 h	488	0.30 h	977	0	2	0	2	4	14	Good	1	1
Text 35_2	S2	313	T5	translator	0.89 h	352	0.33 h	948	0	0	0	0	0	0	Superior	2	1
Text 36_2	S2	298.9	T1	senior translator	0.61 h	490	0.30 h	996	0	0	0	0	0	0	Superior	1	1

Task ID (file name)	Scenario,	Text size,	Trans-lator name	Translator qualification (translator, senior translator)	Estimat-ed time	Planned perfor-mance,	Actual time	Actual perfor-mance,	Quality assesment, negative points						Quality total valuation	MT quality feedback where text was not found in TM (Appendix 1, Question 1)	MT quality feedback where text contained formatting or other tags/mark-up (Appendix 1, Question 2)
(in LPS)	(S1, S2)	(adjusted words)			h	(adjusted words/h)	h	(adjusted words/h)	Accu-racy	Lang-uage quality	Style	Termi-nology	Total	Total (per 1000 words)	(Superior, Good, Mediocre, Poor, Very Poor)	Score 1-3(best)	Score 1-3(best)
Text 37_2	S2	323	T2	translator	0.80 h	404	0.63 h	513	0	7	1	0	8	25	Good	1	1
Text 38_2	S2	310.9	T1	senior translator	0.64 h	486	0.30 h	1036	0	0	0	0	0	0	Superior	1	3
Text 39_2	S2	304.6	T2	translator	0.76 h	401	0.62 h	491	0	3	3	2	8	26	Good	1	1
Text 40_2	S2	293.2	T1	senior translator	0.60 h	489	0.30 h	977	0	0	0	2	2	7	Superior	1	2
Text 41_2	S2	164.1	T2	translator	0.41 h	400	0.33 h	497	0	2	1	6	9	55	Poor	1	2
Text 42_2	S2	269.7	T4	translator	0.85 h	317	0.25 h	1079	0	0	0	0	0	0	Superior	1	3
Text 43_2	S2	223.2	T5	translator	0.64 h	349	0.50 h	446	2	1	1	0	4	18	Good	1	3
Text 44_2	S2	296.9	T3	freelance translator	0.85 h	349	0.28 h	1060	2	3	2	0	7	24	Good	2	1
Text 45_2	S2	290.7	T5	translator	0.83 h	350	0.40 h	727	0	2	0	0	2	7	Superior	2	3
Text 46_2	S2	280.5	T3	freelance translator	0.80 h	351	0.41 h	684	0	13	5	4	22	78	Very Poor	2	1
Text 47_2	S2	148.4	T4	translator	0.47 h	316	0.16 h	928	0	1	0	0	1	7	Superior	1	3
Text 48_2	S2	308	T3	freelance translator	0.88 h	350	0.38 h	811	2	4	3	0	9	29	Good	1	1
Text 49_2	S2	321	T5	translator	0.92 h	349	0.60 h	535	1	5	0	0	6	19	Good	1	1
Text 50_2	S2	298.2	T4	translator	0.94 h	317	0.30 h	994	0	2	3	0	5	17	Good	1	1