D9.1: 1ST REPORT ON SYSTEM VALIDATION & EVALUATION

Grant Agreement number	ICT-248307
Project acronym	**PRESEMT**
Project title	**P**attern **RE**cognition-based **S**tatistically **E**nhanced **MT**
Funding Scheme	Small or medium-scale focused research project – STREP – CP-FP-INFSO
Deliverable title	**D9.1: 1st Report on system validation & evaluation**
Version	**6**
Responsible partner	**ILSP**
Dissemination level	Public
Due delivery date	30.11.2011 *(+ 60 days)*
Actual delivery date	20.1.2012

Project coordinator name & title	**Dr. George Tambouratzis**
Project coordinator organisation	**Institute for Language and Speech Processing / RC 'Athena'**
	+30 210 6875411
Fax	+30 210 6854270
E-mail	giorg_t@ilsp.gr
Project website address	www.presemt.eu

Contents

Tables

Figures

1. Executive summary

The current deliverable, falling within Tasks *T9.1 and T9.2* of **WP9: Validation & Evaluation**, provides an outline of the validation and evaluation activities that were carried out within the PRESEMT project after the release of the 1st system prototype (1st Validation & Evaluation cycle). These activities concern the assessment of the system in terms of performance & conformance to the system design principles (validation), and is a consortium-internal process, and of translation quality (evaluation).

The validation process, on which this deliverable reports, concerned the testing of two system functionalities, (a) *translation process* and (b) *post-processing*. It was performed consortium-internally at each partner's site by personnel members not belonging in the PRESEMT development team and it followed a concrete plan and time schedule. Validators experimented with both system functionalities and documented their experience on purpose-built validation forms.

The evaluation of the translation output, using data compiled for development purposes, involved eight (8) language pairs, those covered by the 1st system prototype, and is was also performed consortium-internally based on automatic evaluation metrics.

Source Language	Target Language
English	German
German	English
Greek	German
Greek	English
Czech	German
Czech	English
Norwegian	German
Norwegian	English

The deliverable has the following structure: Section 2 is dedicated to the validation process and provides a unified account of the validators' comments and suggestions. Section 3 describes the evaluation data used and reports on the results obtained. Finally, in a series of appendices more details on the validation process are given, namely validation forms & schedule and a comprehensive presentation of the results obtained.

2. Validation activities

Validation within PRESEMT involves testing of PRESEMT modules and functionalities and is aimed at ascertaining that they function in accordance to the general system design principles and those of the individual modules.

According to deliverable *D2.2: Evaluation Set-up*, which outlined the validation and evaluation plan of the project, 3 validation sessions have been foreseen, the first one of which had been estimated to take place around M20, following the release of the 1st PRESEMT prototype. During the first validation session the following four system functionalities were scheduled to be tested:

1. **Functionality 1: Translation process for an already created language pair**

   The aim of this activity is to ensure that the PRESEMT prototype can perform the translation of given sentences or given pieces of text, the main concern here being to ensure that a non-trivial working translation is generated and in a reasonable amount of time.

2. **Functionality 2: Optimisation of the translation system**

   In this case, the system optimisation process will be examined by utilising a set of reference translations provided by the user in order to automatically modify the translation system parameters.

3. **Functionality 3: Post-processing of translations using the PRESEMT GUI**

   In this case, the aim is to ensure that the PRESEMT GUI allows the user to modify the system-generated translation in an effective manner according to their preferences.

4. **Functionality 4: Adaptation of the translation system**

   The aim here is to test whether the system is able to be adapted towards the user-specified corrections.

Within the aforementioned timeframe, only functionalities 1 & 3 underwent a validation process, since the Optimisation module (functionality 2) had not yet been finalised when the validation was initiated, while the User adaptation module (functionality 4) was still under development.

When testing functionality 1, the aim was naturally to check whether the system produced a translation, but additional aspects were also of interest such as the system behaviour when handling long texts, operation time, display features, relation of text size to the system performance time etc.

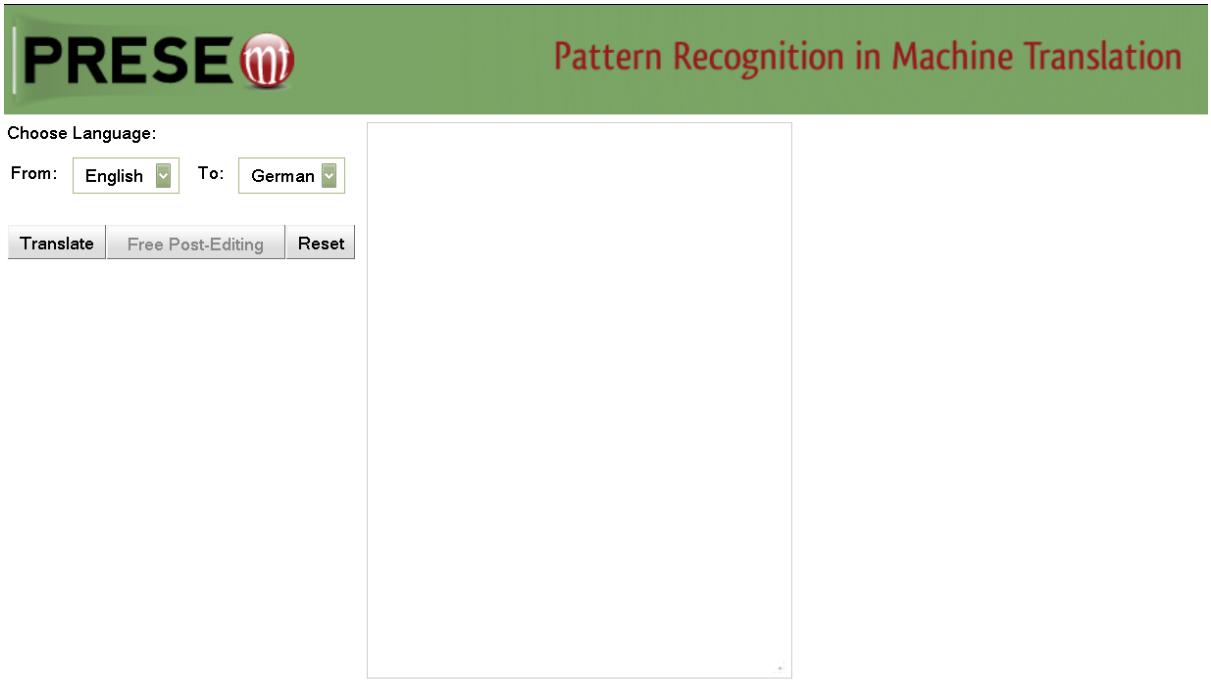
For functionality 3, we wanted to test whether user-oriented post-processing provisions implemented were functional, such as the lexical substitution and the free-post-editing. Display features were also of interest as well as the validators' opinion on the post-processing process as a whole.

2.1 Description of the validation process

ILSP was responsible for coordinating the validation process, which took place at each partner's site. A relevant schedule was drawn up (see Appendix II: Validation schedule), according to which validators, by definition not belonging to the development teams of the project, assessed the performance of two functionalities, i.e. the **translation process** and the **post-processing**[1], which are available via the PRESEMT GUI [http://147.102.3.151:8080/presemt_interface/].

It should be noted that when those functionalities were tested, only two language pairs had been integrated into the main system platform, namely German → English and English → German. Hence, all validators used these language pairs.

Figure 1: PRESEMT GUI



The validators were requested to document their experimentation with the system and report on any problems by filling in the appropriate validation forms (see Appendix I: Validation forms), which have been compiled for this purpose.

The validators' profile included almost exclusively computer analysts and linguists, as expected, since the process was a consortium-internal one involving personnel members of the partners' sites.

2.2 Validation results

The comprehensive results of the validation, as these were depicted in the corresponding forms, are to be found in Appendix II: Validation results – Translation & Appendix IV: Validation results – Post-processing.

The comments of the validators, highlighting problems they have encountered during the validation and including suggestions for improvement, relate to the GUI layout and the function of the translation server, to potential incompatibilities with specific browsers and the text formatting. The comments are summarised as follows:

[1] It should be noted that the same validation pattern is to be followed in the future for the residual system functionalities.

TRANSLATION

Translation server & GUI

Almost all validators noted that the server crashed after a few minutes of continuous use, thus forcing them to restart the browser and reinitiate the whole process.

There is a general consensus about the system being slow; the **suggestion** was expressed that there should be a progress indicator.

One validator noted that when opening a new tab, while waiting for the translation of a long text to be completed in the first tab, then the second one got broken, when the translation completed.

Furthermore, a few comments related to the interface layout and the positioning and size of the buttons, while it was noted that sometimes the interface buttons were disabled, thus preventing the user from launching the translation process.

Finally, it was **suggested** that the source text should remain intact and not be cleared, when changing language.

Browser

A few validators observed that the text rendering was faulty, when using Google Chrome, or that the interface did not work at all with that browser. So the validators turned to either Internet Explorer or Mozilla Firefox.

Text formatting & character rendering

Almost all validators pointed out the fact that, when trying to copy and paste the system translation output, each word appeared in a new line with multiple empty lines in between.

In a similar vein, a few validators noticed that the first letter of sentences was not capitalised.

Similarly, it was pointed out that some characters (e.g. double quote characters [**“ ”**] or the hyphen [-]) were replaced by a question mark in the translated text.

POST-PROCESSING

Translation server & GUI

It was noted that the small size of the input box makes the free-editing of long text inconvenient; so it was **suggested** that the text-area element should be used.

It was noted that it is possible to press the "*Free Post-Editing*" button before the completion of the translation process, thus resulting in a post-editing GUI without text. So, the **suggestion** was expressed that the “*Free Post-editing*” button should be disabled until the translation process is terminated.

All the aforementioned comments have been forwarded to the development team for revising the technical and design characteristcis of the prototype as appropriate.

3. Evaluation activities

Evaluation within PRESEMT involved assessing the quality of the system translation output. Within the reporting period, the results evaluated were obtained by the 1st PRESEMT prototype, which handles the following eight (8) language pairs:

Table 1: Language pairs evaluated

Source Language	Target Language
English	German
German	English
Greek	German
Greek	English
Czech	German
Czech	English
Norwegian	German
Norwegian	English

At the current development phase, the evaluation of the translation output was performed consortium-internally and relied solely on automatic evaluation metrics, using data compiled from material drawn from the web.

3.1 Compiling the evaluation data

Before compiling the evaluation data it has been decided to collect two sets of data: (a) **development** data and (b) **test** data.

The development data would be evaluated with automatic metrics and used consortium-internally to study the system's performance. In other words, this data would be utilised for discovering possible problems in the translation engine. In a similar vein, this set is planned to be used as input to the Optimisation module for optimising the system parameters.

The second category of data involves a sentence set, which is planned to be used both consortium-internally and consortium-externally and will be evaluated on the basis of automatic metrics as well as assessed by humans[2].

The process of creating both data categories (up to this point only the development data have been compiled) was subject to appropriately defined specifications (cf. Table 2). All data originate from the web. More specifically, the web was crawled over for retrieving 1,000 sentences of specific length for each project source language. Thus, five (5) corpora were collected, one per source language.

[2] At this point it should be noted that it is intended to use primarily benchmark data for consortium-external evaluation (e.g. data sets compiled for MT competition purposes). However, the lack of such data for some project languages, namely Greek and Norwegian, necessitates the creation of these data sets.

Subsequently, 200 sentences were randomly chosen out of each corpus, these sentences constituting the development set, and manually translated into the project target languages, namely English and German. The correctness of these translations, which would serve as reference ones, was next checked by native speakers.[3]

Table 2 summarises the particulars of the evaluation data.

Table 2: Evaluation data details

Features	Development	Testing
Source languages	Czech, English, German, Greek, Norwegian	Czech, English, German, Greek, Norwegian
Corpora per language	1	1
Total number of corpora	5	5
Number of sentences per corpus	1,000	1,000
Sentence size	7 – 40 tokens	7 – 40 tokens
Sets per language	1	---
Total number of sets	5	---
Number of selected sentences per set *(approximately)*	180 – 200	---
Number of reference translations	1[4]	---

3.2 Automatic evaluation metrics used

For the current evaluation phase four (4) automatic evaluation metrics were employed, i.e. BLEU, NIST, Meteor and TER.

BLEU (**Bil**ingual **E**valuation **U**nderstudy)[5] metric was developed by IBM (Papineni et al., 2002) and currently is one of the most widely used metrics in the MT field, although primarily designed for assessing the translation quality of statistical MT systems. Its basic function is to calculate the number of common n-grams between a translation produced by the system (candidate translation) and the whole of the reference translations provided. The BLEU score may range between [0 – 1], with 1 denoting a perfect match, i.e. a perfect translation.

NIST (NIST 2002)[6], developed by the National Institute for Standards and Technology, encompasses a similar philosophy to that of BLEU, in that it also counts the matching n-grams between candidate and reference translations. However, it additionally introduces information weights for less frequently occurring, hence more informative, n-grams. The score range is [0 – ∞), where a higher score signifies a better translation quality.

[3] The same process is planned to be followed for compiling the test set.
[4] The number of reference translations will be increased in the future.
[5] ftp://jaguar.ncsl.nist.gov/mt/resources/mteval-v13a-20091001.tar.gz
[6] http://www.nist.gov/speech/tests/mt/

Meteor (**M**etric for **E**valuation of **T**ranslation with **E**xplicit **OR**dering) was developed at CMU (Banerjee & Lavie (2005) and Denkowski & Lavie (2011)), with the aim of explicitly addressing weaknesses in BLEU such as the lack of recall (Banerjee & Lavie 2005: 3), hoping to achieve a higher correlation with human judgements. It "*evaluates a machine translation hypothesis against a reference translation by calculating a similarity score based on an alignment between the two strings. When multiple references are provided, the hypothesis is scored against each and the reference producing the highest score is used.*" It additionally offers various options (such as stemming or paraphrasing) for achieving a better matching. Its score range is [0 – 1], where 1 signifies a perfect translation.

TER (**T**ranslation **E**rror **R**ate)[7], developed at the University of Maryland, resembles the philosophy of Levenshtein distance, in that it calculates the minimum number of edits needed to change a hypothesis (i.e. candidate translation) so that it exactly matches one of the reference translations, normalised by the average length of the references (Snover et al., 2006: 3). In case of more than one references, then only the reference translation closest to the hypothesis is taken into account, since this entails the minimum number of edits. The calculated score, with a range of [0 – ∞), derives from the total number of edits, namely insertion, deletion and substitution of single words as well as shifts of word sequences. Hence a zero score (number of edits = 0) denotes a perfect translation. Another variant of this metric, TER-Plus (TERp), additionally provides more options (paraphrasing, stemming etc.).

3.3 Evaluation results

The following table illustrates the scores obtained per metric and language pair.

Table 3: Evaluation results

Language pair		Sentence set		Reference translations	Metrics			
SL	TL	Number	Source		BLEU	NIST	Meteor	TER
English	German	189	web	1	0.1052	3.8433	0.1789	83.233
German	English	195	web	1	0.1305	4.5401	0.2324	74.804
Greek	German	200	web	1				
Greek	English	200	web	1				
Czech	German	183	web	1	0.0168	2.1878	0.1007	99.383
Czech	English	183	web	1	0.0424	2.5880	0.1739	99.798
Norwegian	German	200	web	1	0.0604	3.2351	0.1484	84.728
Norwegian	English	200	web	1	0.0942	3.6830	0.2110	78.078

According to the results summarised in Table 3, it can be seen that the best results are obtained for the German-to-English and English-to-German corpora, both for NIST and BLEU. For these two languages, the BLEU scores are approximately 0.10 to 0.13, while NIST scores are in the range of 3.8 to 4.3. Similarly, the METEOR results are around the 0.20 mark, while TER results are above 70.0.

[7] http://www.cs.umd.edu/~snover/tercom/

Since the development of the PRESEMT translation system started with these two language pairs, it may be expected that these results are better than those achieved for instance for language pairs involving Norwegian and Czech. Still, it is very promising that by using the same modules, it was possible to build the MT systems in a short period of time. As indicated by the BLEU results for the language pairs involving Czech and Norwegian, there is definitely scope for further improvement for these language pairs. The same applies of course to the pairs German-to-English and English-to-German.

3.3.1 Analysis of the evaluation results

In the present section, the aim is to visualise the evaluation results. In Figure 2, the BLEU results are indicated in a scatter plot, as a function of the sentence size. As can be seen, there does not seem to be a dominant relation between the size in tokens and the BLEU score. Even by grouping together different sizes to create fewer classes (where the first bin is generated by grouping together sentences with between 1 and 5 tokens, the second contains sentences from 6 to 10 tokens etc.) no trend is clearly shown.

Figure 2: Scatter plot of Bleu results for the EN-DE language pair

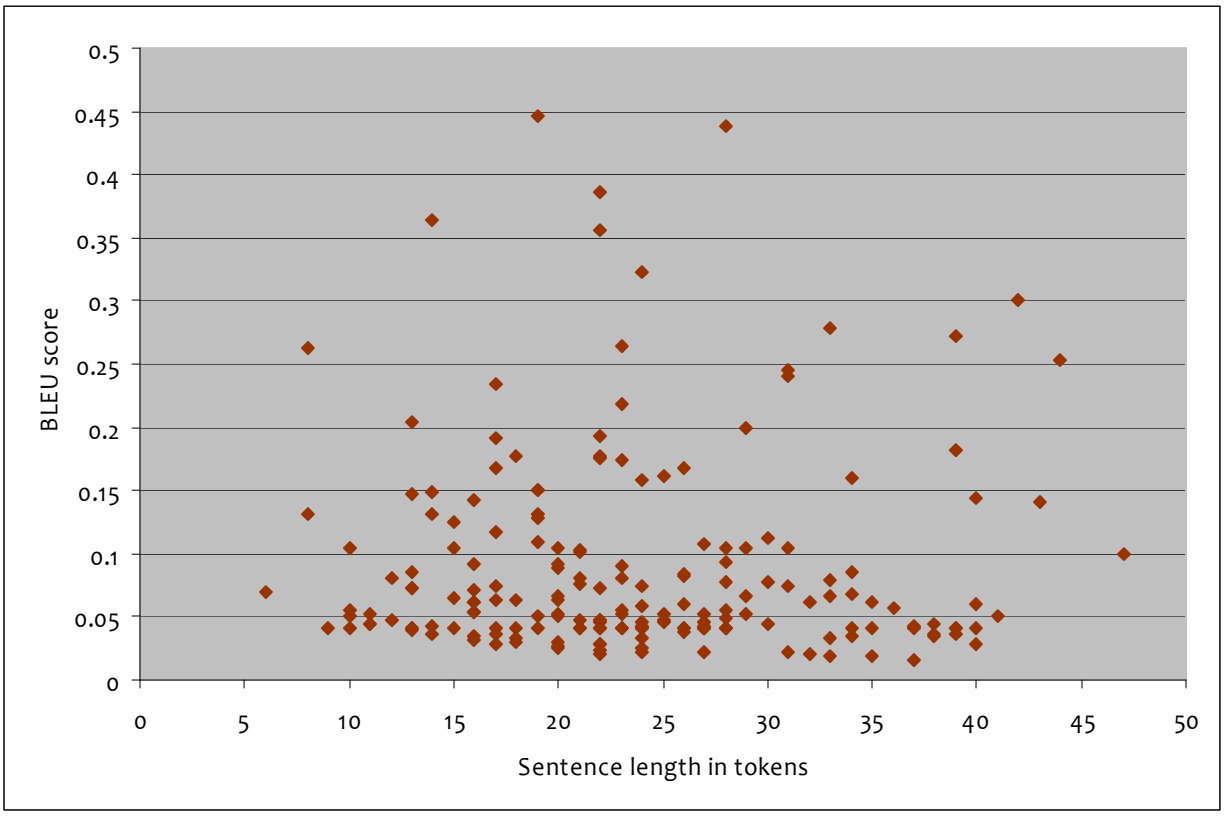
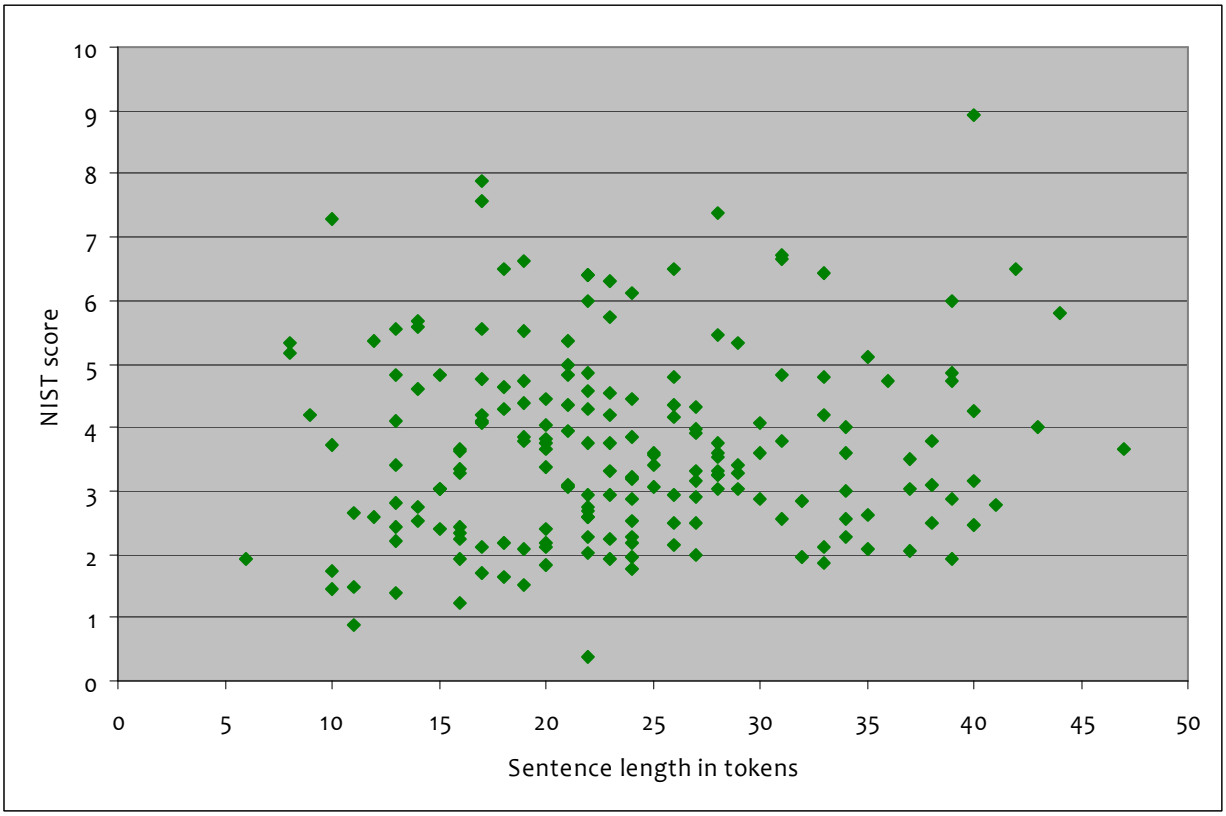
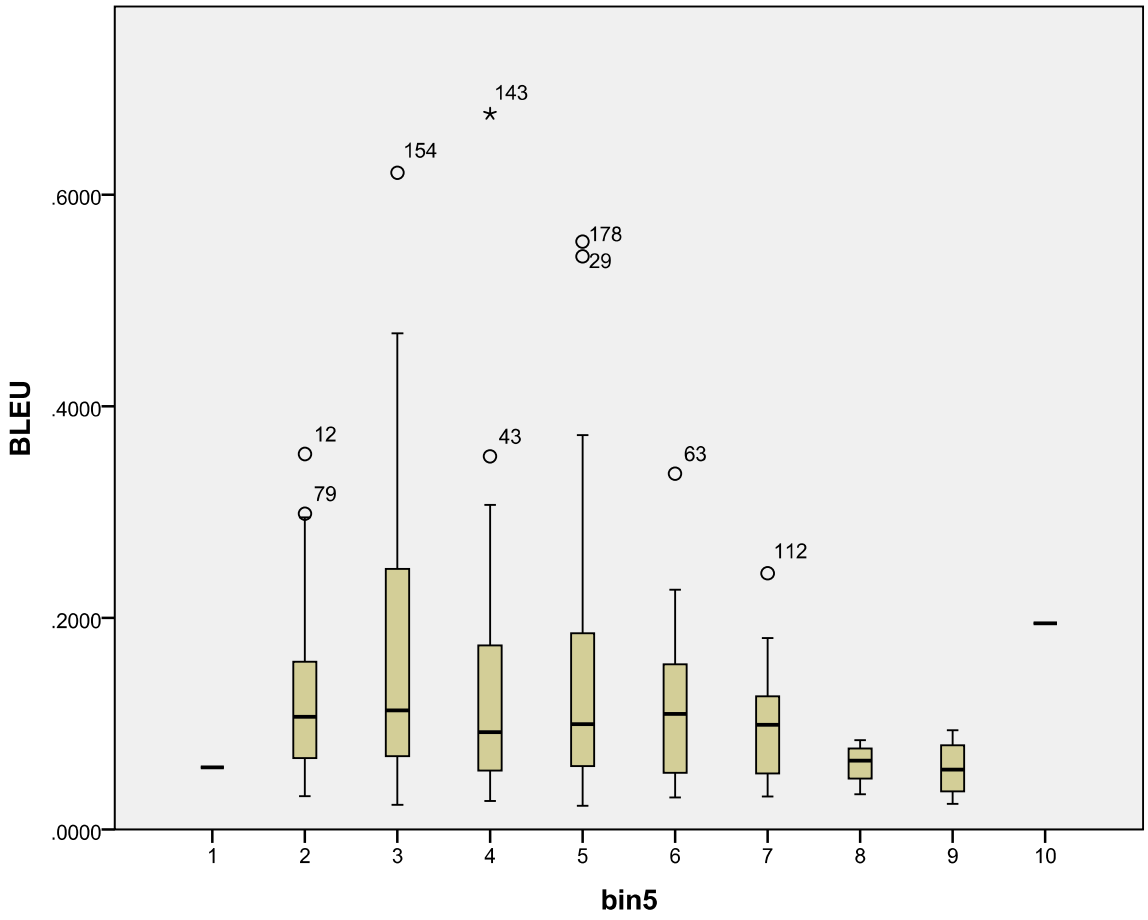


Figure 3: Scatter plot of NIST results for the EN-DE language pair



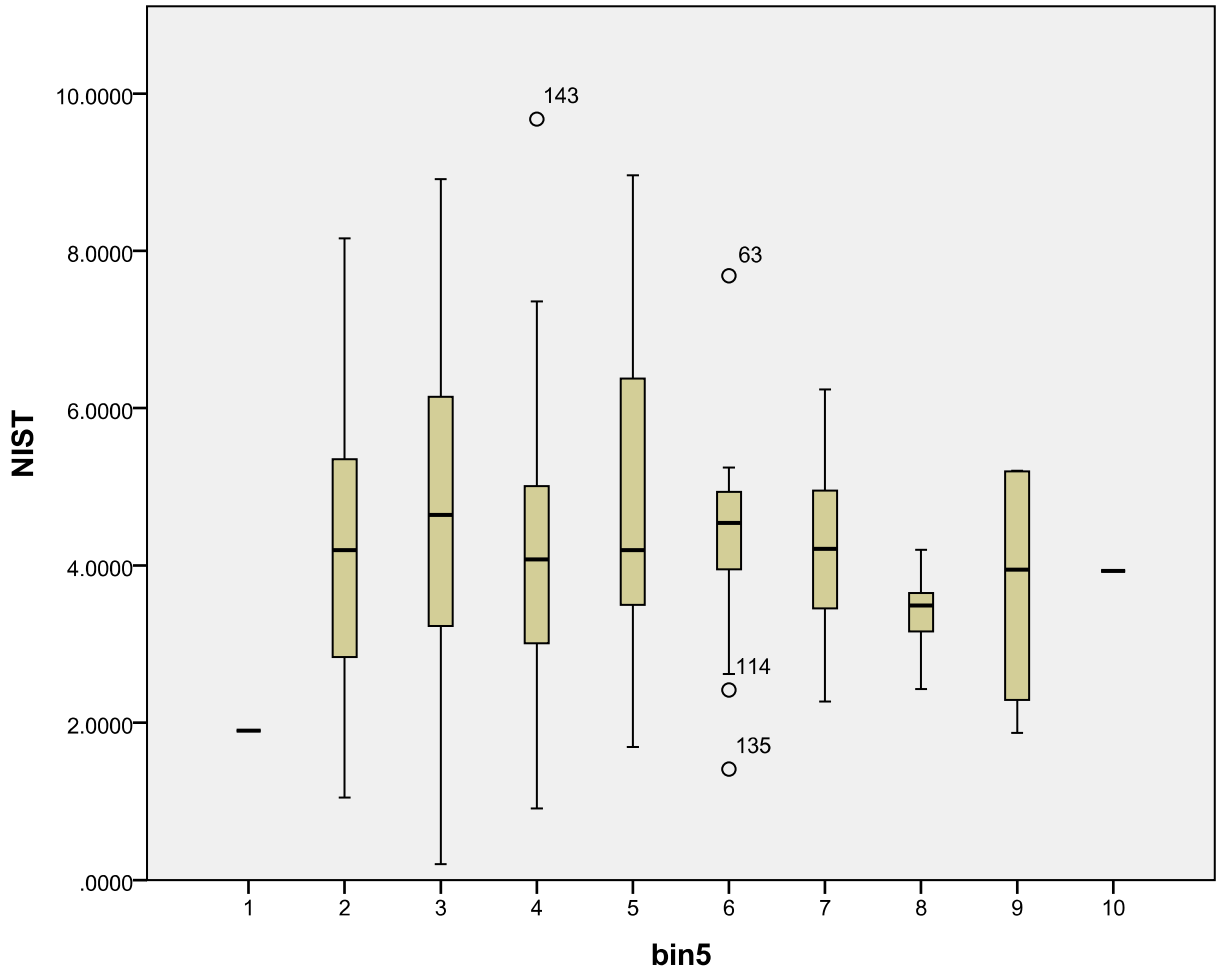
Furthermore, a boxplot diagram is used to indicate for each of the aforementioned bins the characteristics of BLEU scores, as shown in Figure 4. It can be seen that the average BLEU score does not vary too much for bins 2 to 7, indicating that the BLEU score is affected for larger sentences, at least up to a size of 35 tokens (bin 7). The variance is largest for bin 3, whiel a few outliers appear.

Figure 4: Box plot of BLEU results for the EN-DE language pair



Finally, in Figure 5, the same diagram is created for the NIST metric. In this case, the best translation accuracy seems to be obtained for bin 3, though again similar results are obtained for sizes up to 35 tokens. It is only for bin 8 and thereafter (i.e. for sentences with more than 35 tokens) that the NIST score is reduced. Of course these observations are to be verified by extending to other language pairs.

Figure 5: Box plot of NIST results for the EN-DE language pair



3.3.2 Comments and Future Work

Even, however, if the scores obtained are not particularly high, there are a number of factors that need to be taken into account, as listed below:

* One of them is the trade-off between translation accuracy and ease of development of new language pairs. For instance a higher accuracy could result in more demanding specifications regarding the linguistic resources to be provided as well as linguistic knowledge. At least the proposed methodology is easily applicable to other language pairs while it should be noted that PRESEMT aims to provide a translation quality suitable for gisting purposes.
* The second one concerns the chain of modules responsible for the translation. Currently, for a new language pair this involves the phrase alignment of the parallel corpus, the PGM-derived parser for the input sentence, the first translation phase and the second translation phase. All of these of course probably introduce small errors in comparison to dedicated resources, for a selected language pair, and it is likely that these errors multiply. Thus, the final accuracy may be reduced quite considerably. On the other hand, by improving the accuracy of even a single stage, the actual improvement may be substantial.

* The linguistic resources may provide only limited coverage. For instance, the lexica used for most language pairs are not particularly large. In addition, by design the small bilingual corpus from which the TL structure is extracted is limited in size. On the contrary, the monolingual corpus is sufficiently large as it stands. Therefore, it is intended to investigate the effect of each linguistic resource in more detail to provide coverage information. This shall be reported in the next evaluation report.
* Also, it may be that the reference translations are not sufficient (only one reference translation is provided per sentence, currently).
* Therefore, it has been decided to perform a more detailed evaluation of the aforementioned results. This will include a study to indicate the main sources of errors. For the relevant translation stages that cause the largest problems a specific study will be performed. The time to provide the present deliverable has been limited due to the constraints of the review date, so relevant work will continue along the lines described above.

NOTE: In the next version of this deliverable, for objective measures (such as BLEU, NIST METEOR and TER), it is planned to also test other systems, to provide reference values. Candidates to serve as reference systems include commercial systems as well as freely available ones (indicatively, one can mention GoogleTranslate, Systran and Moses).

4. References

Banerjee, S. & Lavie, A., 2005. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43rd Annual Meeting of the Association of Computational Linguistics (ACL-2005), Ann Arbor, Michigan, pp. 65-72

Denkowski, M. & Lavie, A., 2011. Meteor 1.3: Automatic Metric for Reliable Optimization and Evaluation of Machine Translation Systems. Proceedings of the EMNLP 2011 Workshop on Statistical Machine Translation, Edinburgh, Scotland, pp. 85-91

Levenshtein, V.I. 1966). Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady 10: 707–10.

NIST (2002). Automatic Evaluation of Machine Translation Quality Using n-gram Co-occurrences Statistics

Papineni, K., Roukos, S., Ward, T. & Zhu, W.J., 2002. BLEU: A Method for Automatic Evaluation of Machine Translation. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, U.S.A., pp. 311-318

Snover, M., Dorr, B., Schwartz, R., Micciulla, L. & Makhoul, J., 2006. A Study of Translation Edit Rate with Targeted Human Annotation. Proceedings of Association for Machine Translation in the Americas

5. Appendix I: Validation forms

Validation form					
Functionality 1: Translation process					
Date		**Experiment number**		**Site**	

Name				
Profile	Computer analyst	☐	Linguist	☐
	Other	☐	Please specify:	

Input	Sentence	☐	**Number of words**	
	Text	☐	**Number of sentences**	
	LANGUAGE PAIR			
	Source language:		**Target language:**	
	Source text:		**Translation:**	

Can you select the language pair?	Yes	☐	No	☐
Does the system produce a translation?	Yes	☐	No	☐
Does the system display the source text & its translation next to each other?	Yes	☐	No	☐
Translation time *(approximately)*				
Problems with longer texts	Yes	☐	No	☐
If yes, please explain				
Does the '*Reset*' button clear the screen?	Yes	☐	No	☐
Process	Successful	☐	Unsuccessful	☐
If unsuccessful, please explain				
Comments				

Help notes

1. You should complete the form above and save a different copy for each new experiment. Please use the following naming: **Translation-ValidForm_ExperXX.doc**, where '*XX*' stands for the number of a given experiment.
2. Please fill in the **date**, the **serial number** of the experiment and the **site** you work at in the respective fields.
3. Next proceed with your personal details.
4. In the '*Input*' section you should state whether you input a sentence or text for translation and specify the number of words, in case of sentences, and the number of sentences, in case of texts.
5. Next use the drop-down lists for selecting the source and target languages of the experiment.
6. The fields '*Source text*' and '*Translation*' should be filled with the text that you input to the system and the system translation respectively.
7. Please describe any possible problems that the system may have encountered with the size of the input text.
8. If the overall process was unsuccessful, please state so and describe the problem.
9. Finally, add any comments.

Validation form					
Functionality 3: Post-processing					
Date		**Experiment number**		**Site**	

Name				
Profile	Computer analyst	☐	Linguist	☐
	Other	☐	Please specify:	

Input	**LANGUAGE PAIR**	
	Source language:	**Target language:**
	Source text:	**Translation:**

Are the words highlighted when moving the cursor over them?	Yes	☐	No	☐
Does the system provide lexical alternatives?	Yes	☐	No	☐
Can you substitute a word with a lexical alternative?	Yes	☐	No	☐
Can you freely post-edit the text?	Yes	☐	No	☐
If no, please explain				
Process	Successful	☐	Unsuccessful	☐
If unsuccessful, please explain				
Comments				

Help notes

1. You should complete the form above and save a different copy for each new experiment. Please use the following naming: **PostProcessing-ValidForm_ExperXX.doc**, where '*XX*' stands for the number of a given experiment.
2. Please fill in the **date**, the **serial number** of the experiment and the **site** you work at in the respective fields.
3. Next proceed with your personal details.
4. In the '*Input*' section use the drop-down lists for selecting the source and target languages of the experiment.
5. The fields '*Source text*' and '*Translation*' should be filled with the text that you input to the system and the system translation respectively.
6. If the overall process was unsuccessful, please state so and describe the problem.
7. Finally, add any comments.

6. Appendix II: Validation schedule

All partners will ask members of their personnel not belonging to the development teams to validate two system functionalities, (a) the translation process and (b) the post-processing. The whole process should have been completed by **early December**.

The validators will access the PRESEMT web interface for performing the corresponding activity. The interface version tested will be the one implemented by the 10th of November 2011.

Before the validation process

Before the actual process the validators should preferably read the user manual (**Deliverable D7.3.1**) or receive the corresponding guidance by the partner.

Besides, every validation form includes accompanying help notes, which guide the validators.

Validation process details

The validators will be asked to document the whole process by filling in the corresponding validation form[8]. Every form is in *.doc* format and should be completed electronically. A different copy of the form should be completed for each new experiment.

The following form naming convention should be used:

* **Translation-ValidForm_ExperXX.doc** [where '*XX*' stands for the number of a given experiment]
* **PostProcessing-ValidForm_ExperXX.doc** [where '*XX*' stands for the number of a given experiment]

After the validation process

After the validation process is over, the completed forms should be uploaded on the PRESEMT website, in the Archive under the folder **'Validation'**, where each partner will create their own folder.

The validation process is summarised in the following table:

Validation activity	Partner	Validator profile	Number of validators	Submission deadline
Translation process	All	Partner staff; non-member of the development team	At least 2 per partner	5.12.2011
Post-processing	All	Partner staff; non-member of the development team	At least 2 per partner	5.12.2011

[8] The validation forms can be found in the Archive under the folder **'Validation'.**

7. Appendix II: Validation results – Translation process

In this section the validation results for the translation functionality are presented. Table 4 contains the responses of the validators and is followed by their comments, as these were recorded in the corresponding forms. The comments are presented per partner. The numbers enclosed in brackets denote the form from which the comments originate.

Table 4: Validators' responses for the translation functionality

s/n	Experiment	Site number	Site	Profile	Input	Number	SL-TL	LP selection	Translation	Display	Time	Long Text	Reset	Process	Comments
1	1	1	ILSP	Linguist	Text	2	EN-DE	Yes	No	No	---	No	No	Unsuccessful	Yes
2	2	1	ILSP	Linguist	Sentence	18	DE-EN	Yes	Yes	Yes	2	No	Yes	Successful	No
3	3	1	ILSP	Linguist	Sentence	26	EN-DE	Yes	Yes	Yes	2	No	Yes	Successful	Yes
4	4	1	ILSP	Linguist	Text	5	DE-EN	Yes	No	No	---	No	Yes	Unsuccessful	No
5	5	1	ILSP	Linguist	Text	2	DE-EN	Yes	Yes	Yes	2	No	Yes	Successful	Yes
6	6	1	ILSP	Linguist	Text	2	DE-EN	Yes	Yes	Yes	2	No	Yes	Successful	Yes
7	7	1	ILSP	Linguist	Text	3	DE-EN	Yes	Yes	Yes	10	No	Yes	Successful	Yes
8	8	1	ILSP	Linguist	Sentence	17	DE-EN	Yes	Yes	Yes	10	No	Yes	Successful	Yes
9	9	1	ILSP	Linguist	Sentence	19	DE-EN	Yes	Yes	Yes	2	No	Yes	Successful	Yes
10	10	1	ILSP	Linguist	Sentence	25	DE-EN	Yes	Yes	Yes	2	No	Yes	Successful	Yes
11	1	1	ILSP	Computer analyst	Sentence	5	DE-EN	Yes	Yes	Yes	4	No	Yes	Successful	Yes
12	2	1	ILSP	Computer analyst	Sentence	7	EN-DE	Yes	Yes	Yes	35	No	Yes	Successful	Yes
13	3	1	ILSP	Computer analyst	Sentence	6	DE-EN	Yes	Yes	Yes	4	No	Yes	Successful	Yes
14	4	1	ILSP	Computer analyst	Sentence	5	EN-DE	Yes	Yes	Yes	3	No	Yes	Successful	Yes
15	5	1	ILSP	Computer analyst	Text	3	DE-EN	Yes	Yes	Yes	10	No	Yes	Successful	Yes
16	6	1	ILSP	Computer analyst	Sentence	8	DE-EN	Yes	Yes	Yes	4	No	Yes	Successful	Yes
17	7	1	ILSP	Computer analyst	Text	2	DE-EN	Yes	Yes	Yes	5	No	Yes	Successful	Yes
18	8	1	ILSP	Computer analyst	Text	4	DE-EN	Yes	Yes	Yes	15	No	Yes	Successful	Yes
19	9	1	ILSP	Computer analyst	Text	3	DE-EN	Yes	Yes	Yes	5	No	Yes	Successful	No

s/n	Experiment	Site number	Site	Profile	Input	Number	SL-TL	LP selection	Translation	Display	Time	Long Text	Reset	Process	Comments
20	10	1	ILSP	Computer analyst	Text	3	DE-EN	Yes	Yes	Yes	20	Yes	Yes	Successful	Yes
21	1	2	GFAI	Linguist	Sentence	6	DE-EN	Yes	Yes	Yes	3	No	Yes	Successful	No
22	2	2	GFAI	Linguist	Sentence	5	DE-EN	Yes	Yes	Yes	3	No	Yes	Successful	No
23	3	2	GFAI	Linguist	Sentence	3	DE-EN	Yes	Yes	Yes	1	No	Yes	Successful	No
24	4	2	GFAI	Linguist	Sentence	7	DE-EN	Yes	Yes	Yes	3	No	Yes	Successful	No
25	5	2	GFAI	Linguist	Sentence	8	DE-EN	Yes	Yes	Yes	3	No	Yes	Successful	No
26	6	2	GFAI	Linguist	Text	7	DE-EN	Yes	Yes	Yes	60	No	Yes	Successful	No
27	7	2	GFAI	Linguist	Text	6	DE-EN	Yes	Yes	Yes	70	No	Yes	Successful	No
28	8	2	GFAI	Linguist	Text	27	DE-EN	Yes	No	No	---	Yes	Yes	Unsuccessful	Yes
29	9	2	GFAI	Linguist	Text	19	DE-EN	Yes	Yes	Yes	480	No	Yes	Successful	No
30	10	2	GFAI	Linguist	Text	5	DE-EN	Yes	Yes	Yes	90	No	Yes	Successful	No
31	1	2	GFAI	Translator	Text	4	EN-DE	Yes	Yes	Yes	300	No	Yes	Successful	Yes
32	2	2	GFAI	Translator	Text	2	EN-DE	Yes	Yes	Yes	15	No	Yes	Successful	Yes
33	3	2	GFAI	Translator	Text	3	EN-DE	Yes	Yes	Yes	10	No	Yes	Successful	Yes
34	4	2	GFAI	Translator	Text	5	EN-DE	Yes	Yes	Yes	7	No	Yes	Successful	Yes
35	5	2	GFAI	Translator	Text	10	EN-DE	Yes	Yes	Yes	240	No	Yes	Successful	Yes
36	6	2	GFAI	Translator	Sentence	7	EN-DE	Yes	Yes	Yes	2	No	Yes	Successful	Yes
37	7	2	GFAI	Translator	Text	3	EN-DE	Yes	Yes	Yes	60	No	Yes	Successful	Yes
38	8	2	GFAI	Translator	Sentence	36	EN-DE	Yes	No	No	---	Yes	No	Unsuccessful	Yes
39	9	2	GFAI	Translator	Text	4	EN-DE	Yes	No	No	---	Yes	No	Unsuccessful	Yes
40	10	2	GFAI	Translator	Sentence	19	EN-DE	Yes	Yes	Yes	15	No	Yes	Successful	Yes
41	1	3	NTNU	Computer analyst	Sentence	1	EN-DE	Yes	Yes	Yes	3	No	Yes	Successful	Yes
42	2	3	NTNU	Computer analyst	Sentence	3	EN-DE	Yes	Yes	Yes	1	No	Yes	Successful	No
43	3	3	NTNU	Computer analyst	Sentence	18	EN-DE	Yes	Yes	Yes	2	No	Yes	Successful	Yes
44	4	3	NTNU	Computer analyst	Text	3	DE-EN	Yes	Yes	Yes	10	No	Yes	Successful	No
45	5	3	NTNU	Computer analyst	Text	29	DE-EN	Yes	Yes	Yes	600	No	Yes	Successful	Yes
46	1	3	NTNU	Computer analyst-2	Sentence	8	EN-DE	Yes	Yes	Yes	1	No	Yes	Successful	No
47	2	3	NTNU	Computer analyst-2	Text	2	EN-DE	Yes	Yes	Yes	22	No	Yes	Successful	No

s/n	Experiment	Site number	Site	Profile	Input	Number	SL-TL	LP selection	Translation	Display	Time	Long Text	Reset	Process	Comments
48	3	3	NTNU	Computer analyst-2	Text	7	EN-DE	Yes	Yes	Yes	1,200	Yes	Yes	Successful	Yes
49	4	3	NTNU	Computer analyst-2	Text	4	EN-DE	Yes	Yes	Yes	45	No	Yes	Successful	Yes
50	5	3	NTNU	Computer analyst-2	Text	8	DE-EN	Yes	Yes	Yes	45	No	Yes	Successful	Yes
51	1	4	ICCS	Secretary	Text	2	EN-DE	Yes	Yes	Yes	---	No	Yes	Successful	No
52	2	4	ICCS	Secretary	Text	1	EN-DE	Yes	Yes	Yes	few	No	Yes	Successful	No
53	3	4	ICCS	Secretary	Sentence	6	EN-DE	Yes	Yes	Yes	3	No	Yes	Successful	No
54	4	4	ICCS	Secretary	Text	2	EN-DE	Yes	Yes	Yes	22	No	Yes	Successful	No
55	5	4	ICCS	Secretary	Text	2	EN-DE	Yes	Yes	Yes	20	No	Yes	Successful	No
56	6	4	ICCS	Secretary	Sentence	15	EN-DE	Yes	Yes	Yes	8	No	Yes	Successful	No
57	7	4	ICCS	Secretary	Sentence	24	EN-DE	Yes	Yes	Yes	10	No	Yes	Successful	No
58	8	4	ICCS	Secretary	Sentence	19	EN-DE	Yes	Yes	Yes	8	No	Yes	Successful	No
59	9	4	ICCS	Secretary	Sentence	4	EN-DE	Yes	Yes	Yes	0	No	Yes	Successful	No
60	10	4	ICCS	Secretary	Text	4	EN-DE	Yes	Yes	Yes	120	No	Yes	Successful	No
61	11	4	ICCS	Secretary	Sentence	4	EN-DE	Yes	Yes	Yes	0	No	Yes	Successful	No
62	12	4	ICCS	Secretary	Sentence	20	EN-DE	Yes	Yes	Yes	3	No	Yes	Successful	No
63	13	4	ICCS	Secretary	Sentence	4	EN-DE	Yes	Yes	Yes	0	No	Yes	Successful	No
64	14	4	ICCS	Secretary	Sentence	15	EN-DE	Yes	Yes	Yes	0	No	Yes	Successful	No
65	15	4	ICCS	Secretary	Sentence	14	EN-DE	Yes	Yes	Yes	3	No	Yes	Successful	No
66	16	4	ICCS	Linguist	Sentence	8	EN-DE	Yes	Yes	Yes	3	No	Yes	Successful	No
67	17	4	ICCS	Linguist	Sentence	11	EN-DE	Yes	Yes	Yes	2	No	Yes	Successful	No
68	18	4	ICCS	Linguist	Sentence	8	EN-DE	Yes	Yes	Yes	2	No	Yes	Successful	No
69	19	4	ICCS	Linguist	Sentence	6	EN-DE	Yes	Yes	Yes	2	No	Yes	Successful	No
70	20	4	ICCS	Linguist	Sentence	6	EN-DE	Yes	Yes	Yes	2	No	Yes	Successful	No
71	21	4	ICCS	Linguist	Sentence	9	EN-DE	Yes	Yes	Yes	4	No	Yes	Successful	No
72	22	4	ICCS	Linguist	Sentence	6	EN-DE	Yes	Yes	Yes	2	No	Yes	Successful	No
73	23	4	ICCS	Linguist	Sentence	5	EN-DE	Yes	Yes	Yes	2	No	Yes	Successful	No
74	24	4	ICCS	Linguist	Sentence	6	EN-DE	Yes	Yes	Yes	2	No	Yes	Successful	No
75	25	4	ICCS	Linguist	Text	2	EN-DE	Yes	Yes	Yes	50	No	Yes	Successful	No

s/n	Experiment	Site number	Site	Profile	Input	Number	SL-TL	LP selection	Translation	Display	Time	Long Text	Reset	Process	Comments
76	**26**	**4**	ICCS	Linguist	Text	3	EN-DE	Yes	Yes	Yes	52	No	Yes	Successful	No
77	**27**	**4**	ICCS	Linguist	Text	3	EN-DE	Yes	Yes	Yes	18	No	Yes	Successful	No
78	**28**	**4**	ICCS	Linguist	Text	3	EN-DE	Yes	Yes	Yes	5	No	Yes	Successful	No
79	**29**	**4**	ICCS	Linguist	Text	4	EN-DE	Yes	Yes	Yes	70	No	Yes	Successful	No
80	**30**	**4**	ICCS	Linguist	Text	3	EN-DE	Yes	Yes	Yes	65	No	Yes	Successful	No
81	**1**	**5**	MU	Computer analyst	Sentence	29	EN-DE	Yes	Yes	Yes	5	No	Yes	Successful	Yes
82	**2**	**5**	MU	Computer analyst	Sentence	13	EN-DE	Yes	Yes	Yes	3	No	Yes	Successful	Yes
83	**3**	**5**	MU	Computer analyst	Sentence	18	EN-DE	Yes	Yes	Yes	4	No	Yes	Successful	Yes
84	**4**	**5**	MU	Computer analyst	Sentence	38	EN-DE	Yes	Yes	Yes	8	No	Yes	Successful	Yes
85	**5**	**5**	MU	Computer analyst	Sentence	23	EN-DE	Yes	Yes	Yes	5	No	Yes	Successful	Yes
86	**6**	**5**	MU	Computer analyst	Sentence	10	DE-EN	Yes	Yes	Yes	8	No	Yes	Successful	Yes
87	**7**	**5**	MU	Computer analyst	Sentence	20	DE-EN	Yes	Yes	Yes	5	No	Yes	Successful	Yes
88	**8**	**5**	MU	Computer analyst	Sentence	40	DE-EN	Yes	Yes	Yes	130	No	Yes	Successful	Yes
89	**9**	**5**	MU	Computer analyst	Sentence	14	DE-EN	Yes	Yes	Yes	5	No	Yes	Successful	Yes
90	**10**	**5**	MU	Computer analyst	Sentence	21	DE-EN	Yes	Yes	Yes	5	No	Yes	Successful	Yes
91	**1**	**5**	MU	other	Sentence	6	EN-DE	Yes	Yes	Yes	2	No	Yes	Successful	No
92	**5**	**5**	MU	other	Sentence	17	EN-DE	Yes	Yes	Yes	5	No	Yes	Successful	No
93	**8**	**5**	MU	other	Sentence	22	EN-DE	Yes	Yes	Yes	5	No	Yes	Successful	No
94	**9**	**5**	MU	other	Sentence	18	EN-DE	Yes	Yes	Yes	10	No	Yes	Successful	No
95	**10**	**5**	MU	other	Sentence	24	EN-DE	Yes	Yes	Yes	5	No	Yes	Successful	No
96	**1**	**6**	LCL	Computer analyst	Sentence	3	EN-DE	Yes	Yes	Yes	1	No	Yes	Successful	No
97	**2**	**6**	LCL	Computer analyst	Sentence	4	DE-EN	Yes	Yes	Yes	2	No	Yes	Successful	No
98	**3**	**6**	LCL	Computer analyst	Sentence	10	EN-DE	Yes	Yes	Yes	3	No	Yes	Successful	No
99	**4**	**6**	LCL	Computer analyst	Sentence	9	DE-EN	Yes	Yes	Yes	3	No	Yes	Successful	No
100	**5**	**6**	LCL	Computer analyst	Sentence	11	EN-DE	Yes	Yes	Yes	6	No	Yes	Successful	No
101	**6**	**6**	LCL	Computer analyst	Sentence	11	DE-EN	Yes	Yes	Yes	5	No	Yes	Successful	No
102	**7**	**6**	LCL	Computer analyst	Sentence	37	EN-DE	Yes	Yes	Yes	84	Yes	Yes	Successful	No
103	**8**	**6**	LCL	Computer analyst	Sentence	34	DE-EN	Yes	Yes	Yes	250	Yes	Yes	Successful	No

s/n	Experiment	Site number	Site	Profile	Input	Number	SL-TL	LP selection	Translation	Display	Time	Long Text	Reset	Process	Comments
104	9	6	LCL	Computer analyst	Text	5	EN-DE	Yes	Yes	Yes	55	No	Yes	Successful	No
105	10	6	LCL	Computer analyst	Text	3	DE-EN	Yes	Yes	Yes	5	No	Yes	Successful	No
106	1	6	LCL	Computer analyst-2	Sentence	7	DE-EN	Yes	Yes	Yes	2	No	Yes	Successful	No
107	2	6	LCL	Computer analyst-2	Sentence	6	DE-EN	Yes	Yes	Yes	3	No	Yes	Successful	No
108	3	6	LCL	Computer analyst-2	Sentence	13	DE-EN	Yes	No	No	---	No	Yes	Unsuccessful	Yes
109	4	6	LCL	Computer analyst-2	Sentence	14	DE-EN	Yes	Yes	Yes	10	No	Yes	Successful	No
110	5	6	LCL	Computer analyst-2	Sentence	44	DE-EN	Yes	Yes	Yes	75	No	Yes	Successful	No
111	1	6	LCL	Computer analyst-2	Sentence	8	EN-DE	Yes	Yes	Yes	2	No	Yes	Successful	No
112	2	6	LCL	Computer analyst-2	Sentence	9	EN-DE	Yes	Yes	Yes	2	No	Yes	Successful	No
113	3	6	LCL	Computer analyst-2	Sentence	16	EN-DE	Yes	Yes	Yes	5	No	Yes	Successful	No
114	4	6	LCL	Computer analyst-2	Sentence	19	EN-DE	Yes	Yes	Yes	8	No	Yes	Successful	No
115	5	6	LCL	Computer analyst-2	Sentence	40	EN-DE	Yes	Yes	Yes	120	No	Yes	Successful	No

Notes

* **Experiment**: The given experiment's serial number
* **Site:** The partner responsible for the corresponding experiment
* **Profile**: The validator's profile
* **Input**: The type of text input for translation
* **Number**: The number of words or sentences constituting the input. When the input is a sentence, the number refers to words; when the input is text, the number refers to sentences.
* **SL-TL**: The language pair selected for a given experiment
* **LP selection**: It corresponds to the question: "Can you select the language pair?"
* **Translation**: It corresponds to the question: "Does the system produce a translation?"
* **Display**: It corresponds to the question: "Does the system display the source text & its translation next to each other?"
* **Time**: It corresponds to the question: "Translation time *(approximately)*". The time is measured in seconds.
* **Long Text**: It corresponds to the question: "Problems with longer texts".
* **Reset**: It corresponds to the question: "Does the *'Reset'* button clear the screen".
* **Process**: It indicates whether the whole experiment was successful or not.
* **Comments**: It indicates whether there were comments inserted by the validator.

Table 5: Validators' comments for the translation functionality

ILSP	
Validator 1	
[1]	"Client cannot be found" is the message that appears
[3]	the output is in column format
[5]	the output is in column format
[6]	when I paste the output, it is in column format
[7]	when I paste the output, it is in column format
[8]	when I paste the output, it turns into column format
[9]	when I paste the output, it turns into column format
[10]	when I paste the output, it turns into column format
Validator 2	
[1]	Not the right translation, though With Google chrome the translation results are presented to the user in a top to bottom way. With Mozilla Firefox, the results are presented as they should.
[2]	Wrong translation
[3]	Wrong translation
[4]	Wrong translation
[5]	Bad quality in translation
[6]	Bad translation
[7]	Bad quality in translation
[8]	Not every word can be selected. New sentences do not start with capital letter.
[10]	Process successful but wrong translation. Bad translation

GFAI	
Validator 1	
[8]	The source language text consists of many paragraphs separated by one or more blank lines. Server crashes, an error message is displayed. Client has to be restarted.
Validator 2	
[1]	System often chooses the wrong translation for the current context, but provides the correct translation in the list of lexical alternatives, e.g.: source: civilizations, translation: Kulturen --> lexical alternative: Zivilisation; source: scale, translation: Dimensionen --> lex. alternatives: Ausmaß
[2]	System often chooses the wrong translation for the current context, but provides the correct translation in the list of lexical alternatives, e.g.: source: low, translation: geringes --> lexical alternative: Tiefdruckgebiet; source: 24 hours, translation: 24 Zeiten --> lexical alternative: Stunden System delivers a word-by-word translation most of the time: source, e.g.: A deep low pressure system, translation: ein tiefes geringes Belastung System
[3]	System didn't translate the questions correctly, but delivered a word-by-word translation.
[4]	System often chooses the wrong translation in the current context, but provides the correct translation in the list of lexical alternatives System translated names (despite upper case writing): BBC Travel, Lonely Planet System doesn't recognise imperative sentences: source: Verify critical information before travel. , translation:
[5]	System often chooses the wrong translation for the current context, but provides the correct translation in the list of lexical alternatives, e.g.: source: nuclear reactors, translation: nukleare Apparaten --> lex. alternatives: nukleare Reaktor word-by-word translations, e.g.: source: This containment absorbs radiation and prevents radioactive material from being released into the environment . translation: diese Begrenzung absorbiert Strahlung und verhindert radioaktives Material aus lautesten gelöst in die Umgebung . no adaption for the genetive case: source: reactor core's heat, translation: Reaktorkern's Lauf
[6]	word-by-word translation. However, system recognised the superlative form correctly.
[7]	System often chooses the wrong translation for the current context, but provides the correct translation in the list of lexical alternatives, e.g.: source: translation tool, translation: Umsetzung Gerät --> lex. alternative: Übersetzung Tool word-by-word translation: source: We suggest that you print this tutorial manual as you follow the step-by-step instructions to complete the various exercises. translation: wir vorstellen dass euch Eindruck diese Anleitung Anleitung wie folgen ihrer die schrittweisen Anleitungen den vielfältigen Aufgaben zu erledigen
[8]	system crashed after 5 min (1. try) / after 10 min (2. try)
[9]	system crashed after 5 min (3 times)
[10]	System often chooses the wrong translation for the current context, but provides the correct translation in the list of lexical alternatives: source: growth, translation: Entwicklung --> lex. alternatives: Wachstum word-by-word translation, but a good result in this case

NTNU

Validator 1

[1]	The layout could be much nicer (look at Google Translate UI) No automatic language detection. When changing languages a textfield is cleared, which is inconvinient when a user have already typed in or pasted a text into the textfield. Behaviour of buttons is inconsistent, sometimes translation is impossible because the Translate button is disabled. Doesn't work in Google Chrome, had to switch to Internet Explorer. Very slow with no indicator of the progress. When copying the translated text and pasting it, each word appears in a newline with multiple empty lines in between.
[3]	Had to repeat this experiment 2 times. The first time when I pasted a text into the textfield Translate button (as well as all other buttons) remained disabled.
[5]	Very slow

Validator 2

[3]	20 minutes felt a bit long, given the size of the text. I am not sure whether or not this can be characterised as a problem though. It took some time, but I got a result in the end.
[4]	This is two times the text from Expero2.
[5]	When changing languages, I did at one point got error message every time I tried to change languages. The message said something like "unknown", without any additional information. I am not able to recreate the situation at will, so I am guessing it have something to do with the page/GUI and its communication with the webservice. It got fixed after I refreshed the page.

MU

Validator 1

[1]	There's no indicator that the system actually does something once you hit “Translate”. In another preliminary experiment, I encountered pop-up error message “The client could not be find”. I needed to reload the page several times to fix it. I cannot reproduce the problem now.
[2]	There's no indicator that the system actually does something once you hit “Translate”. I was able to reproduce the problem with the error pop-up. When the browser (Firefox) with the PRESEmt interface open is left idle for an hour or so, it will reject any input to translate with the “The client could not be found” error message. Reloading the page solves the problem.
[3]	There's no indicator that the system actually does something once you hit “Translate”. I was able to reproduce the problem with the error pop-up. When the browser (Firefox) with the PRESEmt interface open is left idle for an hour or so, it will reject any input to translate with the “The client could not be found” error message. Reloading the page solves the problem.

MU	
[4]	There's no indicator that the system actually does something once you hit “Translate”. I was able to reproduce the problem with the error pop-up. When the browser (Firefox) with the PRESEmt interface open is left idle for an hour or so, it will reject any input to translate with the “The client could not be found” error message. Reloading the page solves the problem.
[5]	There's no indicator that the system actually does something once you hit “Translate”. I was able to reproduce the problem with the error pop-up. When the browser (Firefox) with the PRESEmt interface open is left idle for an hour or so, it will reject any input to translate with the “The client could not be found” error message. Reloading the page solves the problem.
[6]	There's no indicator that the system actually does something once you hit “Translate”.
[7]	There's no indicator that the system actually does something once you hit “Translate”.
[8]	The process got too long for the input sentence. No result produced after 2 minutes of waiting, no feedback provided, see comment below. When I gave up and reset the form, the translation finally appeared... There's no indicator that the system actually does something once you hit “Translate”.
[9]	There's no indicator that the system actually does something once you hit “Translate”.
[10]	There's no indicator that the system actually does something once you hit “Translate”.

LCL	
Validator 1	
o	first letter of the translated sentences were often in lower case.
o	the "Translate" button was of different height compared to other buttons and the location of buttons were changing during filling the text.
o	changing language causes clearing the source text
o	when user opens a new tab while translating a long text in the first one, the second tab gets broken, when the translation arrives into the first tab.
Validator 2	
[3]	A popup message appeared "Client could not be found"

8. Appendix IV: Validation results – Post-processing

In this section the validation results for the post-processing functionality are presented. Table 6 contains the responses of the validators and is followed by their comments, as these were recorded in the corresponding forms. The comments are presented per partner. The numbers enclosed in brackets denote the form from which the comments originate.

Table 6: Validators' responses for the post-processing functionality

s/n	Experiment	Site number	Site	Profile	SL-TL	Highlight	Lexical alternatives	Substitution	Post-editing	Process	Comments
1	1	1	ILSP	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
2	2	1	ILSP	Linguist	EN-DE	No	No	No	No	Unsuccessful	Yes
3	3	1	ILSP	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
4	4	1	ILSP	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	Yes
5	5	1	ILSP	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	Yes
6	6	1	ILSP	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	Yes
7	7	1	ILSP	Linguist	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
8	8	1	ILSP	Linguist	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
9	9	1	ILSP	Linguist	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
10	10	1	ILSP	Linguist	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
11	1	1	ILSP	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
12	2	1	ILSP	Computer analyst	EN-DE	Yes	Yes	Yes	Yes	Successful	Yes
13	3	1	ILSP	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
14	4	1	ILSP	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
15	5	1	ILSP	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
16	6	1	ILSP	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
17	7	1	ILSP	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
18	8	1	ILSP	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
19	9	1	ILSP	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
20	10	1	ILSP	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes

s/n	Experiment	Site number	Site	Profile	SL-TL	Highlight	Lexical alternatives	Substitution	Post-editing	Process	Comments
21	1	2	GFAI	Linguist	DE-EN	Yes	Yes	Yes	Yes	Successful	No
22	2	2	GFAI	Linguist	DE-EN	Yes	Yes	Yes	Yes	Successful	No
23	3	2	GFAI	Linguist	DE-EN	Yes	Yes	Yes	Yes	Successful	No
24	4	2	GFAI	Linguist	DE-EN	Yes	Yes	Yes	Yes	Successful	No
25	5	2	GFAI	Linguist	DE-EN	Yes	Yes	Yes	Yes	Successful	No
26	1	2	GFAI	Translator	EN-DE	Yes	Yes	Yes	Yes	Successful	No
27	2	2	GFAI	Translator	EN-DE	Yes	Yes	Yes	Yes	Successful	No
28	3	2	GFAI	Translator	EN-DE	Yes	Yes	Yes	Yes	Successful	No
29	4	2	GFAI	Translator	EN-DE	Yes	Yes	Yes	Yes	Successful	No
30	5	2	GFAI	Translator	EN-DE	Yes	Yes	Yes	Yes	Successful	No
31	6	3	NTNU	Computer analyst	EN-DE	Yes	Yes	Yes	Yes	Successful	Yes
32	7	3	NTNU	Computer analyst	EN-DE	Yes	Yes	Yes	Yes	Successful	No
33	8	3	NTNU	Computer analyst	EN-DE	Yes	Yes	Yes	Yes	Successful	No
34	9	3	NTNU	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	No
35	10	3	NTNU	Computer analyst	DE-EN	No	No	No	No	Unsuccessful	Yes
36	1	3	NTNU	Computer analyst-2	EN-DE	Yes	Yes	Yes	Yes	Successful	No
37	2	3	NTNU	Computer analyst-2	EN-DE	Yes	Yes	Yes	Yes	Successful	No
38	3	3	NTNU	Computer analyst-2	EN-DE	Yes	Yes	Yes	Yes	Successful	No
39	4	3	NTNU	Computer analyst-2	EN-DE	Yes	Yes	Yes	Yes	Successful	No
40	5	3	NTNU	Computer analyst-2	EN-DE	Yes	Yes	Yes	Yes	Successful	Yes
41	1	4	ICCS	Secretary	EN-DE	Yes	Yes	Yes	Yes	Successful	No
42	2	4	ICCS	Secretary	EN-DE	Yes	Yes	Yes	Yes	Successful	No
43	3	4	ICCS	Secretary	EN-DE	Yes	Yes	Yes	Yes	Successful	No
44	4	4	ICCS	Secretary	EN-DE	Yes	Yes	Yes	Yes	Successful	No
45	5	4	ICCS	Secretary	EN-DE	Yes	Yes	Yes	Yes	Successful	No
46	6	4	ICCS	Secretary	EN-DE	Yes	Yes	Yes	Yes	Successful	No
47	7	4	ICCS	Secretary	EN-DE	Yes	Yes	Yes	Yes	Successful	No
48	8	4	ICCS	Secretary	EN-DE	Yes	Yes	Yes	Yes	Successful	No

s/n	Experiment	Site number	Site	Profile	SL-TL	Highlight	Lexical alternatives	Substitution	Post-editing	Process	Comments
49	9	4	ICCS	Secretary	EN-DE	Yes	Yes	Yes	Yes	Successful	No
50	10	4	ICCS	Secretary	EN-DE	Yes	Yes	Yes	Yes	Successful	No
51	11	4	ICCS	Secretary	EN-DE	Yes	Yes	Yes	Yes	Successful	No
52	12	4	ICCS	Secretary	EN-DE	Yes	Yes	Yes	Yes	Successful	No
53	13	4	ICCS	Secretary	EN-DE	Yes	Yes	Yes	Yes	Successful	No
54	14	4	ICCS	Secretary	EN-DE	Yes	Yes	Yes	Yes	Successful	No
55	15	4	ICCS	Secretary	EN-DE	Yes	Yes	Yes	Yes	Successful	No
56	16	4	ICCS	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
57	17	4	ICCS	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
58	18	4	ICCS	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
59	19	4	ICCS	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
60	20	4	ICCS	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
61	21	4	ICCS	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
62	22	4	ICCS	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
63	23	4	ICCS	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
64	24	4	ICCS	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
65	25	4	ICCS	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
66	26	4	ICCS	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
67	27	4	ICCS	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
68	28	4	ICCS	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
69	29	4	ICCS	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
70	30	4	ICCS	Linguist	EN-DE	Yes	Yes	Yes	Yes	Successful	No
71	1	5	MU	Computer analyst	EN-DE	Yes	Yes	Yes	Yes	Successful	Yes
72	2	5	MU	Computer analyst	EN-DE	Yes	Yes	Yes	Yes	Successful	Yes
73	3	5	MU	Computer analyst	EN-DE	Yes	Yes	Yes	Yes	Successful	Yes
74	4	5	MU	Computer analyst	EN-DE	Yes	Yes	Yes	Yes	Successful	Yes
75	5	5	MU	Computer analyst	EN-DE	Yes	Yes	Yes	Yes	Successful	Yes
76	6	5	MU	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes

s/n	Experiment	Site number	Site	Profile	SL-TL	Highlight	Lexical alternatives	Substitution	Post-editing	Process	Comments
77	7	5	MU	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
78	8	5	MU	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
79	9	5	MU	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
80	10	5	MU	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
81	1	5	MU	other	EN-DE	Yes	Yes	Yes	Yes	Successful	No
82	3	5	MU	other	EN-DE	Yes	Yes	Yes	Yes	Successful	No
83	4	5	MU	other	EN-DE	Yes	Yes	Yes	Yes	Successful	No
84	5	5	MU	other	EN-DE	Yes	Yes	Yes	Yes	Successful	No
85	6	5	MU	other	EN-DE	Yes	Yes	Yes	Yes	Successful	No
86	8	5	MU	other	EN-DE	Yes	Yes	Yes	Yes	Successful	No
87	9	5	MU	other	EN-DE	Yes	Yes	Yes	Yes	Successful	No
88	10	5	MU	other	EN-DE	Yes	Yes	Yes	Yes	Successful	No
89	1	6	LCL	Computer analyst	EN-DE	Yes	Yes	Yes	Yes	Successful	No
90	2	6	LCL	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
91	3	6	LCL	Computer analyst	EN-DE	Yes	Yes	Yes	Yes	Successful	Yes
92	4	6	LCL	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	Yes
93	5	6	LCL	Computer analyst	EN-DE	Yes	Yes	Yes	Yes	Successful	No
94	6	6	LCL	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	No
95	7	6	LCL	Computer analyst	EN-DE	Yes	Yes	Yes	Yes	Successful	No
96	8	6	LCL	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	No
97	9	6	LCL	Computer analyst	EN-DE	Yes	Yes	Yes	Yes	Successful	No
98	10	6	LCL	Computer analyst	DE-EN	Yes	Yes	Yes	Yes	Successful	No
99	1	6	LCL	Computer analyst-2	DE-EN	Yes	Yes	Yes	Yes	Successful	No
100	2	6	LCL	Computer analyst-2	DE-EN	Yes	Yes	Yes	Yes	Successful	No
101	3	6	LCL	Computer analyst-2	DE-EN	Yes	Yes	Yes	Yes	Successful	No
102	4	6	LCL	Computer analyst-2	DE-EN	Yes	Yes	Yes	Yes	Successful	No
103	5	6	LCL	Computer analyst-2	DE-EN	Yes	Yes	Yes	Yes	Successful	No
104	1	6	LCL	Computer analyst-2	EN-DE	Yes	Yes	Yes	Yes	Successful	No

s/n	Experiment	Site number	Site	Profile	SL-TL	Highlight	Lexical alternatives	Substitution	Post-editing	Process	Comments
105	2	6	LCL	Computer analyst-2	EN-DE	Yes	Yes	Yes	Yes	Successful	No
106	3	6	LCL	Computer analyst-2	EN-DE	Yes	Yes	Yes	Yes	Successful	No
107	4	6	LCL	Computer analyst-2	EN-DE	Yes	Yes	Yes	Yes	Successful	No
108	5	6	LCL	Computer analyst-2	EN-DE	Yes	Yes	Yes	Yes	Successful	No

Notes

* **Experiment:** The given experiment's serial number
* **Site:** The partner responsible for the corresponding experiment
* **Profile:** The validator's profile
* **SL-TL:** The language pair selected for a given experiment
* **Highlight:** It corresponds to the question: "Are the words highlighted when moving the cursor over them?"
* **Lexical alternatives:** It corresponds to the question: "Does the system provide lexical alternatives?"
* **Substitution:** It corresponds to the question: "Can you substitute a word with a lexical alternative?"
* **Post-editing:** It corresponds to the question: "Can you freely post-edit the text".
* **Process:** It indicates whether the whole experiment was successful or not
* **Comments:** It indicates comments inserted by the validator.

Table 7: Validators' comments for the post-processing functionality

ILSP	
Validator 1	
[2]	client could not be found
[4]	the output is in column format
[5]	the output is in column format
[6]	the output is in column format
[7]	the output is in column format
[8]	the output is in column format
[9]	the output is in column format
[10]	the output is in column format
Validator 2	
[1]	Not every word is highlighted, when moving the cursor over them
[2]	Bad translation
[3]	Process successful, bad translation though.
[4]	Process successful, bad translation though. Sentences do not start with capital letters. System faces problems in recognising compound words.
[5]	Process successful, bad translation though. Sentences do not start with capital letters. System faces problems in recognising compound words. The system often use "that" instead of "the"
[6]	Process successful, bad translation though. Not able to recognise compound words or not able to translate correct the compound words from source language.
[7]	Process successful, bad translation though. Tense is not really translated as it should. Not every translated word is highlighted when hovering the mouse over.
[8]	Process successful, bad translation though. Compound words of source language are not correctly translated. Not every word is highlighted when hovering the mouse over.
[9]	Process successful, bad translation though. Compound words of source language are not correctly translated. Not every word is highlighted when hovering the mouse over.
[10]	Process successful, bad translation though. Compound words of source language are not correctly translated. Not every word is highlighted when hovering the mouse over.

NTNU

Validator 1

[6]	Goggle Translate provides better user experience (UX) for word substitution. Post editing user UX can be improved by preserving the formatting of text in several lines rather than presnting it in a textbox in one line. It is possible to press "Free Post-Editing" button before the translation is completed which results in a post-editing GUI without text. It is better to disable this button before the translation process is completed.

Validator 2

[5]	I noticed that after translation, all "-" with space on each side were replaced by question marks in both texts.

MU

Validator 1

[1]	I cannot go back from the Free Post-Editing view to the view with lexical alternatives. The translated text is technically a vertical text, for viewing and editing it might be useful to convert it into a paragraph. When free-editing the translated text, only a single-line input field is available which is inconvenient even for a longer sentence. Please use the textarea element for editing.
[2]	I cannot go back from the Free Post-Editing view to the view with lexical alternatives. The translated text is technically a vertical text, for viewing and editing it might be useful to convert it into a paragraph. When free-editing the translated text, only a single-line input field is available which is inconvenient even for a longer sentence. Please use the textarea element for editing.
[3]	I cannot go back from the Free Post-Editing view to the view with lexical alternatives. The translated text is technically a vertical text, for viewing and editing it might be useful to convert it into a paragraph. When free-editing the translated text, only a single-line input field is available which is inconvenient even for a longer sentence. Please use the textarea element for editing.
[4]	I cannot go back from the Free Post-Editing view to the view with lexical alternatives. The translated text is technically a vertical text, for viewing and editing it might be useful to convert it into a paragraph. When free-editing the translated text, only a single-line input field is available which is inconvenient even for a longer sentence. Please use the textarea element for editing.
[5]	Double quote characters (“”) were wrongly converted in to question marks, thus messing up the sentence borders in the Free Post-Edit mode. I cannot go back from the Free Post-Editing view to the view with lexical alternatives. The translated text is technically a vertical text, for viewing and editing it might be useful to convert it into a paragraph. When free-editing the translated text, only a single-line input field is available which is inconvenient even for a longer sentence. Please use the textarea element for editing.

MU	
[6]	I cannot go back from the Free Post-Editing view to the view with lexical alternatives. The translated text is technically a vertical text, for viewing and editing it might be useful to convert it into a paragraph. When free-editing the translated text, only a single-line input field is available which is inconvenient even for a longer sentence. Please use the textarea element for editing.
[7]	I cannot go back from the Free Post-Editing view to the view with lexical alternatives. The translated text is technically a vertical text, for viewing and editing it might be useful to convert it into a paragraph. When free-editing the translated text, only a single-line input field is available which is inconvenient even for a longer sentence. Please use the textarea element for editing.
[8]	I cannot go back from the Free Post-Editing view to the view with lexical alternatives. The translated text is technically a vertical text, for viewing and editing it might be useful to convert it into a paragraph. When free-editing the translated text, only a single-line input field is available which is inconvenient even for a longer sentence. Please use the textarea element for editing.
[9]	I cannot go back from the Free Post-Editing view to the view with lexical alternatives. The translated text is technically a vertical text, for viewing and editing it might be useful to convert it into a paragraph. When free-editing the translated text, only a single-line input field is available which is inconvenient even for a longer sentence. Please use the textarea element for editing.
[10]	I cannot go back from the Free Post-Editing view to the view with lexical alternatives. The translated text is technically a vertical text, for viewing and editing it might be useful to convert it into a paragraph. When free-editing the translated text, only a single-line input field is available which is inconvenient even for a longer sentence. Please use the textarea element for editing.

LCL	
Validator 1	
o when free-editing the longer sentences do not fit in the input box o when user manages to press "free-editing" during long computation - the form gets filled with previous results.	
[2]	the translation isn't really helpful
[3]	the sentence during free-editing does not fit in the input box
[4]	the text does not fit into input field when free-editting