

Combining pre-editing and post-editing to improve SMT of user-generated content

Johanna Gerlach¹, Victoria Porro¹, Pierrette Bouillon¹, Sabine Lehmann²

(1) Université de Genève FTI/TIM - 40, bvd du Pont-d'Arve, 1211 Genève 4, Switzerland

(2) Acrolinx GmbH, Friedrichstr. 100, 10117 Berlin, Germany

Johanna.Gerlach@unige.ch, Victoria.Porro@unige.ch,

Pierrette.Bouillon@unige.ch, Sabine.Lehmann@acrolinx.com

Abstract

The poor quality of user-generated content (UGC) found in forums hinders both readability and machine-translatability. To improve these two aspects, we have developed human- and machine-oriented pre-editing rules, which correct or reformulate this content. In this paper we present the results of a study which investigates whether pre-editing rules that improve the quality of statistical machine translation (SMT) output also have a positive impact on post-editing productivity. For this study, pre-editing rules were applied to a set of French sentences extracted from a technical forum. After SMT, the post-editing temporal effort and final quality are compared for translations of the raw source and its pre-edited version. Results obtained suggest that pre-editing speeds up post-editing and that the combination of the two processes is worthy of further investigation.

1 Introduction and Background

User-generated content (UGC) such as can be found on forums, blogs and social networks is increasingly used by the online community to share technical information or to exchange problems and solutions to technical issues. Since the users contributing to the content are mainly domain specialists but not professional writers, the text quality cannot be compared with usual publishable content. In the context of a forum, where the focus is on solving problems, linguistic accuracy is often not a priority. Spelling, grammar and punctuation conventions are not always respected (cf. Figure 1). The language used is closer to spoken language, using informal syntax, colloquial vocabulary, abbreviations and tech-

nical terms (Jiang et al, 2012; Roturier and Bensadoun, 2011). Correcting or reformulating UGC is therefore not only interesting to improve readability, but also needed to improve machine-translatability.

J'ai redémarrer l'ordi (apparition de la croix rouge) mais pas besoin de restaurer le système; Toute ces mises à jour on été faite le 2013-03-13

Figure 1. Example from a forum post showing errors (agreement, word confusions) and word usage (abbreviations) typical for technical UGC

The work presented in this paper is part of the Automated Community Content Editing PorTal (ACCEPT) research project and focusses on the relationship between pre-editing and post-editing. The ACCEPT project aims at improving Statistical Machine Translation (SMT) of community content by investigating minimally-intrusive pre-editing techniques, SMT improvement methods and post-editing strategies. Within this project, the forums used are those of Symantec, one of the partners in the project. Pre-edition is carried out through the Acrolinx IQ engine and translation is done with a phrase-based Moses system.

Although several studies have explored the potential of MT of forum and user-generated content (Carrera et al, 2009; Roturier and Bensadoun, 2011; Jiang et al, 2012), few of them have looked into the role of pre- and post-editing as MT complementary modules (Aikawa et al, 2007).

In previous work (Gerlach et al., 2013), we have shown that it is possible to develop pre-editing rules that significantly improve MT output quality, where improvement was assessed through comparative evaluation. In this paper we intend to investigate whether pre-editing rules that have a positive impact on the raw SMT out-

put also have an impact on post-editing temporal effort, which is generally considered one of the most important factors in post-editing evaluations (Krings, 2001). It could be that even though the quality of raw MT output is improved, this does not facilitate the post-editor's task. We will also compare the time required for pre-editing and post-editing tasks and investigate whether time can be gained by combining both activities. Furthermore, we will analyse the final translation quality and look at the satisfaction of the post-editors.

Our aim in this study is twofold, namely: 1) ascertain whether pre-editing rules that improve MT can reduce post-editing effort, and 2) confirm that comparative human evaluation is a valid method to evaluate and select such rules, thus justifying the use of this evaluation method for the ACCEPT project.

In the next sections (2 and 3), we briefly describe the pre-editing approach used in the ACCEPT project. In section 3 we describe the experimental setup and the methodology followed. The data obtained for each experiment is analysed in section 4. Conclusions and future work are presented in section 5.

2 Pre-edition in ACCEPT

In ACCEPT, pre-edition is carried out through the Acrolinx IQ engine, which supports spelling, grammar, style and terminology checking (Brendenkamp et al, 2000). This rule-based engine follows a phenomena-oriented approach to language checking, using a combination of NLP components such as a morphological analyser and a POS tagger to obtain linguistic annotations which can be used to define complex linguistic objects. These are then used in declarative rules written in a formalism similar to regular expressions that marks phenomena that should be pre-edited. Rules can also include correction suggestions, making the pre-editing process semi-automatic, where users only have to accept suggestions provided by the system.

The Symantec community will have access to the Acrolinx engine through a browser plugin, allowing the users to check their text and apply the rules directly in the browser window when writing a forum post (Accept Deliverable D5.2, 2013). The interface of the pre-editing plugin is shown in Figure 2.

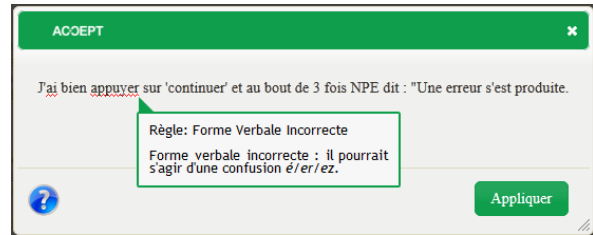


Figure 2. ACCEPT pre-editing plugin. Example of a rule which detects incorrect verb forms.

2.1 Pre-editing rules

During the first period of the project, a stable set of rules with significant positive impact was developed from scratch for French technical UGC. The rules focus mainly on four phenomena, which were proven troublesome for SMT: word confusion (due to homophones), informal and familiar French, punctuation, and structural divergences between French and English. The main criteria for their definition have been precision and impact on translation into English. Impact on translation has been assessed through human comparative evaluation, performed by advanced translation students as well as Amazon Mechanical Turk judges (Gerlach et al., 2013).

The rules are grouped into three sets. Besides the obvious separation of rules for humans and rules for the machine (Hujisen, 1998), they are grouped according to the pre-editing effort they require. Indeed, considering the end-users of the rules, namely forum users who might not be inclined to invest much time in pre-edition, we intended to offer several pre-editing options that would require different amount of involvement.

Some of the rules treat unambiguous cases and have unique suggestions. These are therefore grouped in a set (Set 1) which can be applied automatically with no human intervention. This contains rules for homophones, word confusion, tense confusion, elision and punctuation. Examples are shown in Table 1.

	Raw	Pre-edited
Source	oups j'ai oublié , j'ai sa aussi.	oups j'ai oublié, j'ai ça aussi.
SMT output	Oops I forgot, I have its also.	I have forgotten, I have this too.
Source	avez vous des explications ou astuces pour que cela fonctionne?	Avez-vous des explications ou astuces pour que cela fonctionne?
SMT output	Have you explanations or tips for it to work?	Do you have any explanations or tips for it to work?

Table 1. Examples for rule set 1

The remainder of the rules for humans have either multiple suggestions or no suggestions, thus requiring human intervention. These are grouped in a second set (Set 2), which contains rules for agreement (subject-verb, noun phrase, verb form) and style (cleft sentences, direct questions, use of present participle, incomplete negation, abbreviations), mainly for correcting informal/familiar language. An example is shown in Table 2.

	Raw	Pre-edited
Source	Tu as lu le tuto sur le forum?	As-tu lu le tutoriel sur le forum?
SMT output	You have read the Tuto on the forum?	Have you read the tutorial on the forum?

Table 2. Example for rule set 2

Finally, a third set (Set 3) contains the rules for the machine that should not be visible to end-users. The rules in this set modify word order and frequent badly translated words or expressions to produce variants better suited to MT. One important rule converts the informal second person (*Tu as compilé?*) into its formal correspondent (*Vous avez compilé?*), more frequent in the training data (Rayner et al, 2012). Another rule deals with French clitics that are easily confused with definite articles, replacing them with less ambiguous structures. Examples are shown in Table 3.

	Raw	Pre-edited
Source	J'ai apporté une modification dans le titre de ton sujet.	J'ai apporté une modification dans le titre de votre sujet
SMT output	I have made a change in the title of tone subject	I have made a change in the title of your issue
Source	Il est recommandé de la tester sur une machine dédiée.	Il est recommandé de tester ça sur une machine dédiée.
SMT output	It is recommended to the test on a dedicated machine.	It is recommended to test it on a dedicated machine.

Table 3. Example for rule set 3

In the rest of the paper we describe the experimental setup with the different tasks, the evaluation methodology and the results obtained.

3 Experiment Setup and Methodology

3.1 Corpus

The data used for this study is extracted from the French Symantec forums, where users discuss technical problems with anti-virus and other security software.

In order to create a representative corpus, we selected 684 sentences from the data provided by Symantec, based on bigram frequency, keeping the same proportion of sentences of each length. Sentence lengths range from 6 to 35 words. As a result of this selection process, all sentences were out of context.

Due to the characteristics of UGC, the segmentation of forum data into sentences is not always straightforward. Consequently, some of the automatically extracted sentences are in fact only fragments of the sentences as intended by their authors and can be difficult to understand out of context. We chose not to remove these at this stage, as we did not want to alter the data.

3.2 Participants

For both the pre-editing and post-editing tasks, we recruited translation students in the second year of the MA program at the Faculty of Translation and Interpreting (FTI) of the University of Geneva. For the pre-editing task, we recruited a native French speaker. For the post-editing task, we recruited three native English speakers who had French as a working language. None of the participants had any specific technical knowledge.

3.3 Pre-editing Task

The pre-editing task was divided in three steps. First, we applied the rules from Set 1 automatically, using Acrolinx's AutoApply Client, which replaces each flag (marked phenomena) with the first suggestion available. Since the precision of the rules is not perfect, this step can induce minor deterioration of some sentences, which we did not correct. In a second step, we had the French translator manually apply the rules from Set 2 using Acrolinx's MSWord plugin. This plugin marks all incorrect words in colour, provides information about the error in a contextual menu and, if suggestions are available, allows the user to select a correction from a list. The translator also corrected spelling errors flagged by the Acrolinx spelling module. The pre-editor was asked to treat all correct flags. During this process, we logged the keystrokes, mouse clicks and time. In a third step, we applied Set 3 automatically, using the same method as for Set 1. 456 of the original 684 sentences were affected by pre-editing, i.e. had one or more changes. The flags reported at each step are summarized in Table 4.

Set	grammar, punctuation	style, reformulations	spelling
1	87	7	-
2	74	115	362
3	-	191	-
total	161	313	362

Table 4. Flags for each step

3.4 Translation and Data Selection

The 456 sentences affected by pre-edition were then translated into English using the project's baseline system, a phrase-based Moses system, trained on translation memory data supplied by Symantec, europarl and news-commentary (Accept Deliverable D4.1, 2012).

For 319 sentences, the translation of the pre-edited version was different from that of the raw version.

In order to retain only those sentences where pre-edition had a positive impact on MT output, the translation results (319) were submitted to a comparative evaluation, on the same principle as what was done in previous works (Gerlach et al, 2013). This evaluation was performed by three bilingual judges, using a five-point scale {raw better, raw slightly better, about equal, pre-edited

slightly better, pre-edited better}. The "better" and "slightly better" judgments for each category (raw and pre-edited) were regrouped and the majority judgement for each sentence pair was calculated. The results of the comparative evaluations are shown in Table 5. When considering the majority judgements, the pre-editing rules have a significant positive impact on translation quality. In 65% of cases, translation was improved, while degradation was only observed in 11% of cases. For this specific work, we only considered unanimous judgements. Only those sentences where all three judges considered that pre-editing had had a positive impact on the translation were retained for the post-editing task. This selection had the additional benefit of removing problematic sentences, as we had noticed that judges often fail to reach a unanimous judgement when the presented sentences are difficult to understand, due to bad segmentation or very poor language quality. This final selection resulted in a set of 158 sentences, which added up to 2524 words.

Total sentences	Raw better	About equal	Pre-edited better	No majority judgement	p-value
Majority judgements					
319	34 (11%)	63 (20%)	209 (65%)	13 (4%)	<0.0001
Unanimous judgements					
193	11 (6%)	24 (12%)	158 (82%)	-	<0.0001

Table 5. Comparative evaluation

3.5 Post-editing Task

The resulting set of 158 sentences was used to investigate bilingual post-editing productivity as well as the impact of pre-edition on the quality of the final output after post-editing. Translators were asked to post-edit the machine translation output both of the raw source and of its pre-edited counterpart. This added up to a total of 316 sentences, which were randomly distributed in 71 sets of 20 pairs each.

The post-editing task was performed using the project's post-editing portal (<http://www.accept-portal.eu>, Accept Deliverable D5.2, 2013; cf. Figure 3). The portal logs editing time as well as

keystrokes for each source-target pair. This data can be exported in XLIFF format.

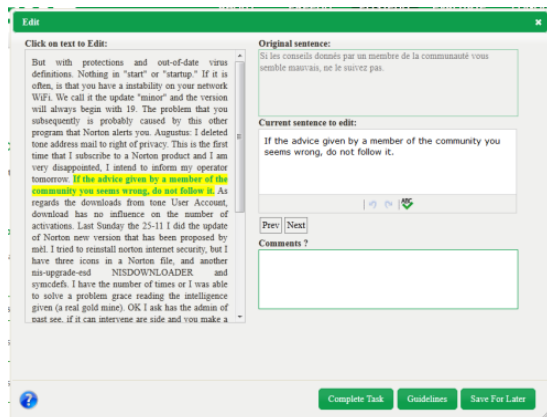


Figure 3. Post-editing Portal Interface

Post-editors were presented with a source-target pair, where the target was the machine translation of either the raw or the pre-edited sentence. Post-editing guidelines and a glossary for the domain covered by the data were provided. Post-editors were asked to render a grammatically correct target sentence, which should convey the same meaning as the original, while using as much of the raw MT output as possible. Terminology and style were not given priority. No time limit was given and all participants were paid.

At the end of the task, the participants were asked to complete a short questionnaire, which was designed to gather information about the post-editors' profile, their previous experience with MT and post-editing, and their feelings towards it.

In this experimental setup, post-editors processed each sentence twice: once the translation of the raw source and once the translation of its pre-edited counterpart. As the sentences were presented in a random order, in some cases the translation of the raw source was treated before that of the pre-edited source and vice-versa. It is logical to expect the post-editor to spend more time reading and post-editing the first instance of a pair of sentences. When the second instance appears, the post-editor has at least already read and processed the meaning of the source and thus will probably spend less time in post-editing. Since the order randomisation of our data produced an unfair distribution (69 pre-edited first vs 89 raw first), we chose to remove 20 sentences where the translation of the raw source had been processed in the first place, in order to balance the impact of processing order.

The quality of the final translations was evaluated using the LISA QA Model. The errors in all 276 sentences for each of the three post-editors were annotated by two bilingual persons, whose annotations were then put in common and discussed to resolve ambiguities and disagreements.

In the next section, we will present the results for all tasks.

4 Results

4.1 Pre-editing Effort

The pre-editor spent 53 minutes processing the entire corpus (684 sentences) using the MSWord Plugin, making 334 keystrokes, 576 left-clicks and 542 right-clicks. This process changed 567 tokens in the corpus and affected 456 sentences (cf. Table 6).

The pre-editor found the rules straightforward to apply and the pre-editing process globally quite easy, except for some terminology issues related to the unfamiliar domain.

Pre-editing task : 456 sentences	
Total time (mins)	53
Total keys	334
Total mouse-clicks	1118

Table 6. Pre-editing effort

4.2 Post-editing Effort

The post-editing effort in terms of time and keystrokes is clearly lower for the translations of pre-edited sentences. While the post-editing speed differs strongly among post-editors, the relative time gain is very similar for all three. On average, the total post-editing time for all 138 sentences is reduced by 47% with $sd=4\%$. The one-tailed t-test shows that the difference is highly significant for all three post-editors ($p<0.0025$, $t=4.581/3.094/3.635$). The results for the three post-editors are shown in Table 7.

Post-editing task : 2*138 sentences (2*2194 words)						
	PE 1		PE 2		PE 3	
	Raw	Pre-edited	Raw	Pre-edited	Raw	Pre-edited
Total time (mins)	53	29	98	56	109	54
Total keys	3492	1763	3907	2181	5579	3263
Processing speed (w/mins)	41	76	22	39	20	40

Table 7. Pre- and post-editing effort

Table 8 shows an example of a sentence before and after pre-editing, with its corresponding MT output and the post-editing times for each post-editor (in seconds).

	Raw	Pre-edited
Source	quelqu'un a t'il déjà rencontré se problème?.....	quelqu'un a-t-il déjà rencontré ce problème?.....
SMT output	Someone has it already you encountered is problem?.....	Has anyone had this problem?.....
Post-editing time (PE1/PE2/PE3)	7.7s/14.2s/16.1s	5.1s/0s/6.5s

Table 8. Examples of MT output with corresponding post-editing time

The histogram in Figure 4 illustrates the results presented in table 7. It represents the frequency distribution of time gain percentages from raw to pre-edited for each of the post-editors, which were calculated per sentence, in relation to the time used to post-edit MT output of the raw sentence. The data range is distributed into “bins” of equal size on the x axis and the frequency within each bin is shown vertically on the y axis. 25 outliers¹ were removed.

Although the post-editing time for the pre-edited sentence is not always lower than the time for the raw sentence, we observe that the cases where pre-editing reduces the post-editing time are more frequent. 312 of 389 sentences are plotted on the right-hand side of the histogram.

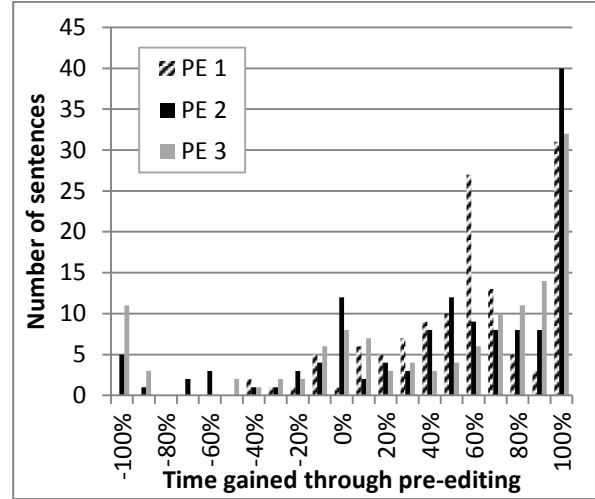


Figure 4. Distribution of relative time gained for each post-editor

While the absolute pre- and post-editing times may not be directly comparable, due to the different number of sentences processed and to the possibly artificially low post-editing times caused by the double processing mentioned in section 3.5, it remains interesting to combine these times. As not all pre-edited sentences have been post-edited, we have estimated the pre-edition time for the effectively post-edited sentences proportionally to the number of sentences, based on the data shown in Table 6, resulting in an approximate pre-editing time of 16 minutes for 138 sentences.

We observe that, for our set of sentences where pre-editing had a positive impact on MT output, the post-editing time gained by using a pre-edited source (respectively 24/42/55 minutes for each of the post-editors, cf. Table 7) outweighs the time invested in the pre-editing process itself. Combined results are shown in Table 9. Furthermore, it can be argued that for an equal time investment, the pre-editing effort is “cheaper” than the post-editing effort, as 1) it is a monolingual process, thus requiring less qualification from the user, and 2) it is semi-automatic, as most of the rules have suggestions and can be applied by selecting an item in a list.

¹ We apply one of the common definitions of outliers using the interquartile range (IQR): lower than the 1st quartile minus 1.5*IQR or greater than the 3rd quartile plus 1.5*IQR.

Combined time for 138 sentences (mins)						
	PE1		PE2		PE3	
	Raw	Pre-edited	Raw	Pre-edited	Raw	Pre-edited
Pre-editing	-	16	-	16	-	16
Post-editing	53	29	98	56	109	54
Total	53	46	98	81	139	78

Table 9. Combined pre- and post-editing times

As another indicator of post-editing effort in terms of number of edit operations, we computed the Translation Error Rate (TER) (Snover et al., 2006) for each of the two MT outputs (raw and pre-edited) using the three corresponding post-edited versions as reference. The case sensitive TER score for the translation of the raw source is 20.17, the score for the translation of the pre-edited source is 10.76, indicating a lower number of edits for the pre-edited version.

4.3 Quality Evaluation

On the whole, pre-edition seems to slightly reduce the number of errors in the final output, but the number of errors is insufficient to determine whether the difference is significant (cf. Table 10). A similar number of errors was found for all three post-editors in both versions, although far less time was spent post-editing the pre-edited version. We can therefore assume that the increase in processing speed does not entail an increase in the number of errors.

Total errors			
	Raw	Pre-edited	Reduction
PE 1	44	37	7
PE 2	28	29	-1
PE 3	41	35	6

Table 10. Error counts for each post-editor

A closer examination of the individual annotated errors does not indicate a clear relation between the errors and the output that was post-edited (MT of raw sentence or MT of pre-edited sentence). However, we have observed that there are proportionally more sentences with errors among those with longer edit distances (Levenshtein) between the raw MT output and the post-edited version. This supports the assumption that post-editors will make fewer errors when presented with a relatively clean MT output needing only few edits (rather than an output that requires heavy reformulation and corrections at

many places). While our data is insufficient to quantify this claim, this observation suggests that pre-editing can also have a positive impact on final post-edited translation quality.

Table 11 shows the error counts by category, averaged over the three post-editors. Mistranslations are the most frequent type of error, which was to be expected considering that 1) the sentences were out of context and sometimes badly segmented, making them difficult to understand, 2) the post-editors were not familiar with the domain, 3) the post-editors, not being native French speakers, might have had difficulties understanding the colloquial French used on the forums. The only category where we observe no improvement is terminology, but the number of errors is too small to be significant. The most important reduction can be observed for language errors, which include spelling, punctuation, grammar and semantics.

Final Quality Evaluation (LISA QA)			
Average per category	Raw	Pre-edited	% error reduction
Mistranslation	17.3	16.7	-4%
Accuracy	6.0	5.7	-6%
Terminology	1.3	2.3	75%
Language	9.7	6.0	-38%
Style	3.3	3.0	-10%
TOTAL	37.7	33.7	-11%

Table 11. Average error counts by error category

Most of the errors observed in our data can be attributed to typos, lack of attention and hesitation to seriously reformulate the MT output, which can at least partially be explained by the participants profiles and insights described in the next section.

4.4 Questionnaire. Insights from participants.

After the post-editing task, we asked participants to complete an anonymous questionnaire to establish their profiles and gather their insights about the post-editing task. This questionnaire was based on the questionnaire used in another experiment performed at FTI, also involving translation students, texts from the same forum and the same MT system (Morado Vázquez et al., 2013), where globally feedback was very positive. From the analysis of the answers provided,

we gathered the following information. All participants claimed to translate about 250 words per hour on an average 8-hour day of work, but had little experience as professional translators (only one claimed to have been working as a freelance for 2 years) and had hardly ever post-edited MT-output before. As for CAT tools, one only uses them when required to do so and the other two have tried them but do not use them on a daily basis.

Participants were not familiar with the topic or with Symantec products. Two found the task difficult from a terminology point of view and one indicated she had mainly experienced linguistic-related doubts.

More interestingly, when asked about the helpfulness of MT proposals to produce a final translation, two seemed sceptical (they responded 3 on a 6-point scale, where 6 stood for “Not at all, I would have preferred working from scratch”) and the third was negative (she responded 5). Nonetheless, we observed that their attitude towards post-editing itself was quite positive: they considered that post-editing was “definitely needed [...] and can help a lot” (PE1) and “useful” (PE2), except for the third participant, who found post-editing harder than translating from scratch. Despite this, they all agreed in saying that if more context was provided and if they mastered the domain or topic of the texts, they would find post-editing machine translations more useful and interesting.

5 Conclusion and Future Work

We have observed that pre-editing rules that have a significant positive impact on translation output also have a significant positive impact on post-editing time, reducing it almost by half. The combination of pre-editing and post-editing to process user-generated content seems promising, as easy monolingual pre-editing effort effectively reduces the more tedious bilingual post-editing effort. Based on the fact that a translation judged as being better is also faster to post-edit, we conclude that comparative evaluation is a valid method to select pre-editing rules for a workflow such as envisaged in the ACCEPT project. We plan to extend our investigations to examine whether pre-editing that does not directly improve translation quality also has an impact on post-editing effort.

While pre-editing does not significantly improve the quality of the final post-edited transla-

tions, there is no loss of quality linked to the time gain. The most frequent errors in the final translations are mistranslations. While the bad segmentation and lack of context are probably responsible for many of these, we suspect that the lack of experience and insufficient domain knowledge of the MA students have also influenced the results. In order to refine these results, we plan to perform in-context tests, processing entire forum posts, using both professional translators and savvy real users. This would give us more information about the causes of the mistranslations and might point to phenomena that could be corrected by pre-editing.

Finally, regarding the pre-editing task, we would like to see how pre-editors apply the rules, i.e. if, in non-controlled circumstances, they will apply all rules systematically or choose only those they consider useful.

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 288769.

References

- Accept Deliverable D4.1 (2012), <http://www.accept.unige.ch/Products/>
- Accept Deliverable D5.2 (2013), <http://www.accept.unige.ch/Products/>
- Aikawa, Takako, Schwartz, Lee, King, Ronit, Corston-Oliver, Mo, Lozano, Carmen. 2007. Impact of Controlled Language on Translation Quality and Post-editing in a Statistical Machine Translation Environment. In *Proceedings of the MT Summit XI*, 10-14 September, Copenhagen, Denmark, pp.1-7.
- Allen, Jeffrey. 2003. “Post-editing”, in Somers, Harold: *Computers and Translation. A Translator's Guide*, John Benjamins Publishing Company, Amsterdam/Philadelphia, p. 297-317.
- Bredenkamp, A., B. Cysmann and M. Petrea. 2000. Looking for errors: A declarative formalism for resource-adaptive language checking. In *Proceedings of LREC*. Athens, Greece.
- Carrera, Jordi, Olga Beregovaya and Alex Yanishevsky. 2009. Machine Translation for Cross-Language Social Media, available: http://www.promt.com/company/technology/pdf/machine_translation_for_cross_language_social_media.pdf [accessed May 23rd 2013]

- Gerlach, Johanna, Victoria Porro and Pierrette Bouillon. 2013. La prédiction avec des règles peu coûteuses, utile pour la TA statistique des forums ? In *Proceedings of TALN/RECITAL 2013*. Sables d'Olonne, France.
- Hujisen, W. O. 1998. Controlled Language: An introduction. In *Proceedings of CLAW 98* (pp. 1–15). Pittsburg, Pennsylvania: Language Technologies Institute, Carnegie Mellon University
- Jiang, Jie, Andy Way and Rejwanul Haque. 2012. Translating User-Generated Content in the Social Networking Space. In *Proceedings of AMTA 2012*, San Diego, CA, United States.
- Krings, Hans P. 2001. Repairing texts: Empirical investigations of machine translation post-editing process. The Kent State University Press, Kent, OH.
- Morado Vázquez, Lucía, Silvia Rodríguez Vázquez and Pierrette Bouillon. 2013. Comparing forum data post-editing performance using translation memory and machine translation output: a pilot study. In *Proceedings of 14th Machine Translation Summit*, 2013, Nice, France.
- O'Brien, Sharon and Johann Roturier. 2007. How Portable are Controlled Languages Rules? A Comparison of Two Empirical MT Studies. In *Proceedings of the MT Summit XI*, Copenhagen, pages 105–114.
- Rayner, Manny, Pierrette Bouillon and Barry Haddow. 2012. Using Source-Language Transformations to Address Register Mismatches in SMT. In *Proceedings of AMTA*, San Diego, CA, United States.
- Roturier, Johann, and Anthony Bensadoun. 2011. Evaluation of MT Systems to Translate User Generated Content. In *Proceedings of the Thirteenth Machine Translation Summit*, 244–251.
- Snover, M., B. Dorr, R. Schwartz, L. Micciulla and J. Makhoul. 2006. A Study of Translation Edit Rate with Targeted Human Annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation of the Americas*. Cambridge, Massachusetts.

