Pre-editing by Forum Users: a Case Study

Pierrette Bouillon[1], Liliana Gaspar[2], Johanna Gerlach[1], Victoria Porro[1], Johann Roturier[2]

[1]Université de Genève FTI/TIM - 40
bvd Du Pont-d'Arve, CH-1211 Genève 4, Suisse
{Pierrette.Bouillon, Johanna.Gerlach, Victoria.Porro}@unige.ch

[2]Symantec Ltd.
Ballycoolin Business Park, Blanchardstown, Dublin 15, Ireland
{Liliana_Gaspar, Johann_Roturier}@symantec.com

Abstract

Previous studies have shown that pre-editing techniques can handle the extreme variability and uneven quality of user-generated content (UGC), improve its machine-translatability and reduce post-editing time. Nevertheless, it seems important to find out whether real users of online communities, which is the real life scenario targeted by the ACCEPT project, are linguistically competent and willing to pre-edit their texts according to specific pre-editing rules. We report the findings from a user study with real French-speaking forum users who were asked to apply pre-editing rules to forum posts using a specific forum plugin. We analyse the interaction of users with pre-editing rules and evaluate the impact of the users' pre-edited versions on translation, as the ultimate goal of the ACCEPT project is to facilitate sharing of knowledge between different language communities.

Keywords: pre-editing, statistical machine translation, user-generated content, language communities

1. Introduction

Since the emergence of the web 2.0 paradigm, forums, blogs and social networks are increasingly used by online communities to share technical information or to exchange problems and solutions to technical issues. User-generated content (UGC) now represents a large share of the informative content available on the web. However, the uneven quality of this content can hinder both readability and machine-translatability, thus preventing sharing of knowledge between language communities (Jiang et al, 2012; Roturier and Bensadoun, 2011).

The ACCEPT project (http://www.accept-project.eu/) aims at solving this issue by improving Statistical Machine Translation (SMT) of community content through minimally-intrusive pre-editing techniques, SMT improvement methods and post-editing strategies, thus allowing users to post questions or benefit from solutions on forums of other language communities. Within this project, the forums used are those of Symantec, one of the partners in the project. Pre-editing and post-editing are done using the technology of another project partner, the Acrolinx IQ engine (Bredenkamp et al, 2000). This rule-based engine uses a combination of NLP components and enables the development of declarative rules, which are written in a formalism similar to regular expressions, based on the syntactic tagging of the text.

Within the project, we used the Acrolinx engine to develop different types of pre-editing rules for French, specifically designed for the Symantec forums. Primarily, the aim of pre-editing in this context is to obtain a better translation quality in English without retraining the system with new data. In previous work, we have found that the application of these rules significantly improves MT output quality, where improvement was assessed through human comparative evaluation (Gerlach et al, 2013a; Seretan et al, to appear). Another study suggested that for specific phenomena, for example for the register mismatch between community content and training data, pre-editing produces comparable if not better results than retraining with new data (Rayner et al, 2012). Further work (Gerlach et al, 2013b) has shown that pre-editing rules that improve the output quality of SMT also have a positive impact on bilingual post-editing time, reducing it almost by half.

However, it is still unclear whether pre-editing can successfully be implemented in a forum, which is the real life scenario targeted by the ACCEPT project. In the previous studies, the pre-editing rules were applied by native speakers with a translation background, i.e., with excellent language skills. In contrast, in the targeted scenario, the pre-editing task will have to be accomplished by the community members themselves. Although the task was simplified as much as possible for the forum users, by integration of a checking tool in the forum interface, it still involves choosing among one or multiple suggestions, or even correcting the text manually, following instructions when no reliable suggestions can be given. Applying these changes might prove difficult for users with varied linguistic knowledge, as it can involve quite complex modifications, for example restructuring a sentence to avoid a present participle. Another aspect to consider is the motivation of the users: if pre-editing requires too much time or effort, users will be less inclined to complete this step. Additionally, as users probably have little knowledge of the functioning of an SMT engine or the consequences of pre-editing, the importance of making certain changes to the source will not be obvious to them.

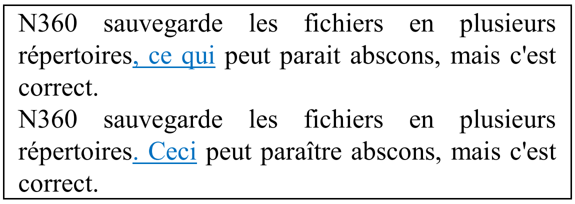
The aim of this study is therefore to ascertain whether light pre-editing rules which were developed using the Acrolinx formalism and which have proved to be useful for SMT can be applied successfully by forum users.

In the rest of the paper, Section 2 provides more details about the French Acrolinx pre-editing rules developed for the Symantec forums. Section 3 describes the experimental setup and provides details about the experiments conducted for evaluating the rules with forum users. In Section 4, we discuss the results obtained in these experiments and, finally, conclusions and directions for future work are provided in Section 5.

2. Pre-editing in ACCEPT

Pre-editing can take different forms: spelling and grammar checking; lexical normalisation (e.g. Han & Baldwin, 2011, Banerjee et al., 2012); Controlled Natural Language (CNL) (O'Brien, 2003; Kuhn, 2013); or reordering (e.g. Wang et al, 2007; Genzel, 2010). However, few pre-editing scenarios combine these different approaches. For partially historical reasons, CNL was mostly associated with rule based machine translation (RBMT) (Pym, 1988; Bernth & Gdaniec, 2002; O'Brien & Roturier, 2007; Temnikova, 2011, etc. (one exception is (Aikawa et al, 2007)). On the contrary, spellchecking, normalisation and reordering were frequently used as pre-processing steps for SMT. In this work, the particularities of community content have led us to choose an eclectic approach. We developed rules of all the types mentioned above which answer the following criteria:

- The rules focus on specificities of community content that hinder SMT, namely informal and familiar style (not well covered by available training data), word confusion (related to homophones) and divergences between French and English.
- As we cannot reasonably ask forum users, whose main objective is obtaining or providing solutions to technical issues, to painstakingly study pre-editing guidelines, compliance with the rules must be checked automatically. Therefore rules must be implemented within a checking tool, in our case Acrolinx. This entails some restrictions, especially due to the nature of the Acrolinx formalism, which is for example not well suited to detect non local phenomena. On the positive side, it also means that rules are easily portable to other similar tools since they don't require a lot of linguistic resources.
- Another condition for successful rule application by forum users is that suggestions are provided, since we cannot expect forum users to reformulate based only on linguistic instructions (such as "avoid the present participle", "avoid direct questions", "avoid long sentences", etc). For this reason, common CNL rules like "avoid long sentences" were replaced by more specific rules, accompanied by an explanation which appears on a tooltip. A good example is the rule which replaces ", ce qui", by a full stop followed by a pronoun: ". Ceci" (see **Figure 1**).



N360 sauvegarde les fichiers en plusieurs répertoires, ce qui peut parait abscons, mais c'est correct.

N360 sauvegarde les fichiers en plusieurs répertoires. Ceci peut paraître abscons, mais c'est correct.

Figure 1. Example of pre-editing rule used to substitute traditional CNL rules like "avoid long sentences"

In the absence of forum post-edited data that would have allowed identification of badly translated phrases or phenomena, the rules were developed mainly using a corpus-oriented approach. Two specific resources proved to be particularly useful: the out-of-vocabulary (OOV) items, which are a good indicator of the data that is not covered in the training set (see Banerjee et al, 2012), and the list of frequent trigrams and bigrams, present in the development data but absent from the training corpus.

Three sets of rules were developed intended to be used in sequence. A first distinction is made between rules for humans (which also improve source quality) and rules for the machine (which can degrade it or change it considerably since the only aim is to improve MT output) (Hujisen, 1998). The rules for humans were split up into two sets, according to the pre-editing effort they require.

A first set (Set1) contains rules that can be applied automatically. This set includes rules that treat unambiguous cases and have unique suggestions. It contains rules for homophones, word confusion, tense confusion, elision and punctuation. While the precision of the rules included in this set is reasonably high, it is not perfect. The automatic application of this set does therefore produce some errors that might be avoided if the rules were applied manually instead. Examples of rules contained in this set are given in **Table 1**.

Rule	**Raw**	**Pre-edited**
Confusion of the homophones "sa" and "ça"	oups j'ai oublié, j'ai **sa** aussi.	oups j'ai oublié, j'ai **ça** aussi.
Missing or incorrect elision	Lancez Liveupdate et regardez **si il** y a un code d'erreur.	Lancez Liveupdate et regardez **s'il** y a un code d'erreur.
Missing hyphenation	Il est **peut être** infecté, ce qui serait bien dommage.	Il est **peut-être** infecté, ce qui serait bien dommage.

Table 1. Examples for Set1

A second set (Set2) contains rules that have to be applied manually as they have either multiple suggestions or no suggestions at all. The rules correct agreement (subject-verb, noun phrase, verb form) and style (cleft sentences, direct questions, use of present participle, incomplete negation, abbreviations), mainly related to informal/familiar language. The human intervention required to apply these rules can vary from a simple selection between two suggestions, to manual changes, for example for checking a bad sequence of words. Examples of rules contained in this set are given in **Table 2**.

Rule	Raw	Pre-edited
Avoid direct questions *Avoid abbreviations*	**Tu as** lu le **tuto** sur le forum?	**As-tu** lu le **tutoriel** sur le forum?
Avoid the present participle	Certains jeux **utilisant** Internet ne fonctionnent plus.	Certains jeux **qui utilisent** Internet ne fonctionnent plus.
Avoid letters between brackets	Regarde le**(s)** barre**(s)** que tu as téléchargées et surtout le**(s)** site**(s)** web où tu les as récupérés.	Regarde les barres que tu as téléchargées et surtout les sites web où tu les as récupérés.

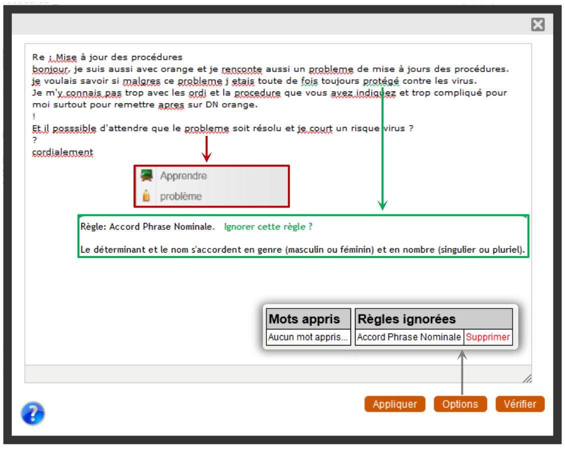
Table 2. Examples for Set2

Finally, the rules for the machine were grouped in a third set (Set3) that is applied automatically and will not be visible to end-users. These rules modify word order and frequent badly translated words or expressions to produce variants better suited to SMT. The rules developed in this framework are specific to the French-English combination and to the technical forum domain. Examples of rules contained in this set are given in **Table 3**.

Rule	Raw	Pre-edited
Avoid informal 2^{nd} person	J'ai apporté une modification dans le titre de **ton** sujet.	J'ai apporté une modification dans le titre de **votre** sujet
Replace pronoun by "ça"	Il est recommandé de **la** tester sur une machine dédiée.	Il est recommandé de tester **ça** sur une machine dédiée.
Avoid "merci de"	**Merci de** nous tenir au courant.	**Veuillez** nous tenir au courant.

Table 3. Examples for Set3

In ACCEPT, pre-editing is completed through the ACCEPT plugin directly in the Symantec forum. This plugin was developed using Acrolinx's technologies and specifically conceived to check the compliance with the rules directly where content is created (ACCEPT Deliverable D5.2, 2013). This plugin "flags" potential errors or structures by underlining them in the text. Depending on the rules, when hovering with the mouse cursor over the underlined words or phrases, the user receives different feedback to help him apply the rule correction (**Figure 2**). For rules with suggestions, a contextual menu provides a list of potential replacements, which can be accepted with a mouse click. For rules without suggestions, a tool-tip comes up with the description of the error but no list of potential replacement is provided. Modifications then have to be done directly by editing the text. Besides these two main interactions, users can also choose to "learn words", i.e. add a given token to the system so that it will not be flagged again, or "ignore rules", i.e. completely deactivate a given rule. Both actions are stored within the user profile and remain active for all subsequent checking sessions. By means of a properties window, users can view learned words and ignored rules, which can be reverted at any time. **Figure 2** shows the plugin in action.



Re : Mise à jour des procédures
bonjour, je suis aussi avec orange et je renconte aussi un probleme de mise à jours des procédures.
je voulais savoir si malgres ce probleme j etais toute de fois toujours protégé contre les virus.
Je m'y connais pas trop avec les ordi et la procedure que vous avez indiquez et trop compliqué pour moi surtout pour remettre apres sur DN orange.
!
Et il posssible d'attendre que le probleme soit résolu et je court un risque virus ?
?
cordialement

Règle: Accord Phrase Nominale. Ignorer cette règle ?

Le déterminant et le nom s'accordent en genre (masculin ou féminin) et en nombre (singulier ou pluriel).

Mots appris	Règles ignorées	
Aucun mot appris...	Accord Phrase Nominale	Supprimer

Figure 2. ACCEPT pre-editing plugin used for this study

In this study, our aim is twofold. In a first step, we want to compare rule application by forum users and experts. In a second step, we wish to determine if it is preferable to have a semi-automatic, yet not entirely reliable process (where Set1 is applied automatically), or a manual process where all the rules from Set1 and Set2 are checked manually. This last approach will strongly depend on the motivation and skills of the users. These different scenarios (user vs expert, manual vs automatic) will be compared in terms of pre-editing activity (number of changes made in the source and the target) and in terms of the impact of changes on translation output. This impact will be evaluated using human comparative evaluation. In the next section, we will describe the experimental setup for the scenarios mentioned above.

3. Experimental Setup

3.1 Pre-editing

In order to compare the different pre-editing scenarios, we collected the following pre-edited versions of our corpus:

UserSemiAuto: Rules from Set1 were applied automatically. Then, the corpus was submitted to the forum users, who applied the rules from Set2 manually using the ACCEPT plugin.

UserAllManual: The raw corpus was submitted to the forum users, who applied the rules from Set1 and Set2 manually using the ACCEPT plugin. This version was produced at one week interval from UserSemiAuto.

Expert: Rules from Set1 were applied automatically. Then, the corpus was submitted to a native French speaking language professional, who applied the rules from Set2 manually.

Oracle: This version is the result of manual post-processing of the Expert version by a native French speaker. All remaining grammar, punctuation and spelling issues were corrected. No style improvements were made in this step.

For the User scenarios, the pre-editing activity was recorded using the ACCEPT plugin. This included recording the number and type of errors flagged by the rules and the actions performed during the process (accepted suggestions, displayed tooltips, ignored rules and words learned). The output data was collected in a JSON format.

To complete the pre-editing process as designed for ACCEPT, once all manual pre-editing steps were performed, we applied the rules from Set3 automatically to all pre-edited versions. All versions 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). We then set up five human comparative evaluations on Amazon Mechanical Turk and measured the pre-editing activity as explained in the following section.

3.2 Evaluation

3.2.1 MT output

For the comparative evaluations, the test data was split into sentences. We presented three bilingual judges with sentence pairs in randomised order. These sentences are translations of different pre-edited versions of the same source sentence. Sentences with identical translations were not included in the evaluation. The judges were asked to assign a judgement to each pair on a five-point scale {first clearly better, first slightly better, about equal, second slightly better, second clearly better}. The majority judgement for each sentence was calculated.

The evaluations were performed on Amazon Mechanical Turk, using the same setup as in previous studies (Rayner et al, 2012; Gerlach et al, 2013a). Tasks were restricted to workers residing in Canada and having a reliable work history on AMT. We chose to use AMT workers for this evaluation because we have found that for simple tasks like these, the results obtained are reliable and can be obtained fast.

We first compared the translations of the Raw with the translations of the version pre-edited by the Expert (**Raw vs Expert**). The result was used as a baseline for the evaluations of the User versions and allowed us to corroborate the positive impact of our rules on translation and validate the results obtained in previous studies (Gerlach et al, 2013a, 2013b).

In a second evaluation, we compared the translations of the different User versions with the translations of the Raw (**Raw vs User**), to evaluate the impact of our rules when applied by users.

In a third evaluation, we compared the translations of the different User versions against the Expert version (**Users vs Expert**), in order to complement results obtained with the second evaluation.

A fourth evaluation was designed to determine the impact of applying some of the rules automatically, as opposed to performing an entirely manual application. For this evaluation, we asked judges to compare the translations produced in each scenario, UserSemiAuto and UserAllManual, for a same user (**UserSemiAuto vs UserAllManual**).

Finally, we compared the translations of the Raw with the translations of the Oracle version (**Raw vs Oracle**). This allowed us to assess the potential of correcting all grammar, punctuation and spelling issues that are not covered by our rules.

3.2.2 Pre-editing activity

In order to gain more insight into the effort required for applying pre-editing rules, we performed a quantitative analysis of the activities logged by the plugin during the pre-editing process. We looked at the number of flagged errors (errors found) and the total number of actions performed by users. We also investigated the acceptance rate of suggestions as well as the rules and words which had been ignored and/or learned. Additionally, we calculated the Levenshtein distance between the raw and the pre-edited User versions to quantify the total tokens changed during pre-editing. We compared results per scenario and per user.

3.3 Data selection

The amount of data we could reasonably expect volunteer forum users to process being limited, we chose to create a corpus of about 2500 words for this study. From an initial corpus of 10000 forums posts, only posts of 250 words or less were selected to ensure that the final corpus would contain posts with a diversity of writers and topics. Among these, we then chose to select posts with a relatively high occurrence of errors and structures to pre-edit. Focussing on posts with many errors allowed us to cover a larger number of pre-editing rules, and thereby increase the chances that users would treat or reflect upon a diversity of rules, giving us more insight into the difficulties encountered with each rule category. To this end, we processed our corpus with the Acrolinx Batch Checker, which produces reports that summarise all the errors found for each rule. In Acrolinx, rules are grouped in three categories: grammar, style and spelling. For this study, we chose to focus on grammar and style rules, as the application of these is more likely to cause difficulties to our participants, as opposed to spelling, which works like any other spelling checker that most users are familiar with. Therefore, we kept only posts with at least 3 grammar and 3 style errors (mean number of errors per post: 5.7). Among these, we selected the posts with the highest error/words ratio, resulting in a set of 25 posts. These posts were made available to users of the French Norton forum[1] in the forum itself to maximize the ecological validity of the study. Specific forum sections were created for each participant and automatically populated with the selected posts using the Lithium API.[2] In this study users were asked to edit texts that they had not necessarily authored,

[1] http://fr.community.norton.com

[2] http://www.lithium.com/products/technology/integration

which would not be the case in a real-life scenario.

3.4 User selection

To recruit users willing to participate in our study, we made an open call for participation in the French-Speaking Norton forum. We did not look for any specific profile. The only prerequisite was to be a French native speaker. 7 users showed their willingness to participate and were contacted, but only 2 had completed all tasks at the time of this study.

4. Results

In this section we present the results of the evaluations for the two main research questions (Users vs Expert and SemiAuto vs AllManual) we seek to answer both in terms of translation quality and pre-editing activity.

4.1 Users vs Expert

4.1.1 Translation quality

The results obtained for the Expert version through a comparative evaluation confirm those of previous studies, namely that correct application of the pre-editing rules has a significant positive impact on translation quality. **Table 4** shows that for 52% of sentences, the translation of the pre-edited version is better, while the translation is degraded for only 6% of sentences. A McNemar test showed that the difference of cases in which pre-editing had a positive vs a negative impact is statistically significant (p<0.001).

	identical	raw better	about the same	pre-edited better	no majority judgement
Expert	32%	6%	4%	52%	5%
Oracle	29%	6%	2%	60%	3%

Table 4. Raw against Expert pre-edited and Oracle

The Oracle version only produces slightly better results (60%) than the Expert version. This suggests that our light pre-editing rules, in their current state, can produce high-quality results not far from those obtained with the Oracle.

Table 5 presents the results for the User scenarios. We observe that they are very close to those obtained with Expert pre-editing.

	identical	raw better	about equal	user better	no majority judgement
SemiAuto					
user1	42%	7%	2%	45%	4%
user2	41%	4%	1%	50%	3%
AllManual					
user1	43%	6%	2%	47%	3%
user2	44%	2%	2%	50%	2%

Table 5. Raw against User pre-edited

For both scenarios and users, the translations of nearly half of the sentences are improved by pre-editing. As in the case of the Expert, the difference between improved and degraded sentences is statistically significant (p<0.001).

However, while the number of improved sentences is similar, these results do not tell us if pre-editing by the users produced as good a result as pre-editing by the Expert. It cannot be excluded that, while they were judged as better than the Raw version, some of the improved sentences are still of lesser quality than the Expert version. For this reason, we decided to compare the User versions against the Expert version. Results are shown in **Table 6**.

	identical	user better	about equal	expert better	no majority judgement
SemiAuto					
user1	65%	5%	2%	25%	3%
user2	60%	13%	4%	19%	3%
AllManual					
user1	65%	10%	3%	19%	3%
user2	57%	12%	4%	24%	3%

Table 6. User against Expert

In all scenarios, flag application performed by the users and the Expert produced identical translations for more than half of the sentences (65%-60%/65%-57%). In all scenarios, the Expert version is considered better than the Users version in less than a quarter of the sentences (19% to 25%). In some cases, the User version is considered better than the Expert. Globally, in three out of four cases the differences are statistically significant (p<0.0001) but small, which suggests that users are not far from the Expert.

4.1.2 Pre-editing activity

In terms of activity performed, the users and the Expert are also close. The comparison of the Levenshtein distance for all versions against Raw (2274 original tokens) shows that users made less changes than the Expert in both scenarios, but again the difference is small. In average, the Expert changed 5% more tokens than the users. This may also be due to the incomplete application of rules. The additional changes made in the Oracle version amount only to 5%. **Table 7** displays the Levenshtein distance from Raw for all scenarios.

	User SemiAuto	**User AllManual**	**Expert**	**Oracle**
Tokens	449 (user1)	465 (user1)	582	694
	527 (user2)	480 (user2)		
% of total	20% (user1)	20% (user1)	26%	31%
	23% (user2)	21% (user2)		

Table 7. Levenshtein distance from Raw - All scenarios

From Section 4.1 we can then conclude that both users and experts can reach a good pre-editing performance, with a significant impact on SMT.

4.2 UserSemiAuto vs UserAllManual

4.2.1 Translation quality

For each user, version for scenario 1 (SemiAuto) was compared with version for scenario 2 (AllManual).

	identical	semi-auto better	about equal	all manual better	no majority judgement
user1	72%	8%	6%	13%	0%
user2	58%	18%	6%	16%	2%

Table 8. UserSemiAuto against UserAllManual

Table 8 shows that for more than half of the sentences, there is no difference between the two versions. The difference between UserSemiAutoBetter and UserAllManualBetter is relatively small and is not statistically significant (McNemar test, p>0.05).

4.2.2 Pre-editing activity

The data logged using the ACCEPT plugin provided information about number of flags and actions performed to correct the text in both User scenarios (UserSemiAuto vs UserAllManual).

As expected, users had to deal with more flags in the UserAllManual scenario than in the UserSemiAuto because they had to apply both sets (1 and 2) manually (430 vs 642). This fact required more attention from users, as evidenced by the higher number of actions performed in the UserAllManual scenario (347 and 327 in UserSemiAuto vs 501 vs 512 in UserAllManual). A summary of actions and flags is provided in **Table 9**.

	UserSemiAuto		**UserAllManual**	
	user1	**user2**	**user1**	**user2**
totalFlags	430		642	
total actions performed	347	327	501	512
of which accepted suggestions (%)	213 (61%)	211 (65%)	431 (86%)	375 (73%)
total available suggestions	333		539	
% of accepted suggestions over total available	64%	63%	80%	70%

Table 9. Flags and actions logged by the ACCEPT plugin

In both scenarios, suggestions are among the most frequent type of performed actions. They represent 61%-86% of actions for user1 and 65%-73% of actions for user2 (UserSemiAuto and UserAllManual respectively). Moreover, suggestions have a high acceptance rate for both users in both scenarios (64%-80% for user1 and 63%-70% for user2 over the total available suggestions), which suggests that the suggestions provided are considered useful.

The Levenshtein distance for the two user scenarios (UserSemiAuto and UserAllManual) revealed information about the number of edits performed by users in each scenario (see **Table 10** below). In the UserSemiAuto scenario, 141 tokens were changed after the automatic application of Set1 to the raw original corpus. This scenario then required 326 more changes from user1 when applying Set2 manually, and 407 from user2. Conversely, more tokens were changed when applying both Set1 and Set2 manually in the UserAllManual scenario, which shows that more edit activity was required in this scenario: 465 tokens were changed by user1 (+ 39%) and 480 by user2 (+ 17%).

Scenario			**Changed tokens**
Auto application of Set1 to Raw			141
User SemiAuto	manual set2	**user1**	326
		user2	407
User AllManual	manual set1&set2	**user1**	465
		user2	480

Table 10. Levenshtein distance - User scenarios

The conclusion from Section 4.2 is therefore that the high-precision (yet not perfect) rules from Set1 can be safely automatically applied with less effort from users.

4.3 Learned words and ignored rules

Considering that we had only two participants and a relatively small amount of data, results presented in this section are too scarce to perform a significant quantitative analysis, but they still provide insights into user preferences. As we suspect that the distinction between “learn word” and “ignore rule” might not have been entirely clear for the users, we have chosen to regroup both cases. In the following, we will call these “rejected flags”.

In both scenarios, both users chose to reject a certain number of flags, as shown in **Table 11**.

	semiAuto	**allManual**
user1	6	22
user2	22	21

Table 11. Rejected flags per user

A closer investigation shows that by far the most frequently rejected are spelling flags (14, counted over both users and both scenarios). Among these, only 5 are “real” spelling issues such as missing accents or typos, while the others are either proper nouns, anglicisms or abbreviations, all very common on a technical forum, and not always incorrect. Three of these flags were also rejected by the Expert. Unsurprisingly, the next rule that was rejected frequently is "avoid anglicisms" (13 flags, counted over both users and both scenarios). Words such as “boot”, “Trojan” or “software” are very common in French techie speak, and users might not see the use of replacing them with less common French equivalents. The remaining ignored flags are mostly style rules, such as "avoid conjunctions at Beginning of Sentence" and "avoid present participle".

We also examined the impact of flag rejection on translation. However, due to the experimental setup it is not possible to draw direct conclusions, as the evaluation is sentence-based and most of the sentences had several flags. It is therefore not possible to determine whether omission of one flag was the determining change that influenced the evaluation of an entire sentence. We did however find that for 17% of sentences where a flag was rejected, the translation was identical to that obtained with the Expert version where the flags had effectively been applied. It must be noted that in 6 cases, users corrected the flagged word or phrase, despite choosing to ignore the rule or learn the word. This might be due to manipulation errors.

5. Conclusion

In this paper, we ascertained that pre-editing rules developed with a light formalism (regular expressions) are sufficient to produce significant improvement on SMT and can be applied successfully by some forum users. In particular, we have found that:

- The two users who participated in this study are close to experts in terms of pre-editing activity and produce significant impact on SMT.

- The semi-automatic process can be safely applied without degrading the quality of the results. Besides, it saves time and effort from users, as less edits and actions are required when Set1 is applied automatically.

- The analysis of interaction with rules allowed us to discriminate between rules that users might be willing to apply from those rules perceived as incorrect or purely stylistic, and thus not essential and time-consuming. This can help in the future to filter out unnecessary rules or to decide which rules to place in an automatic set (a decision which implies increasing precision in detriment of coverage). For example, some rules rejected by users but with a high impact on SMT, as "avoid present participle" could be restricted to be automatic. Further research will be needed in this sense.

6. 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.

7. References

ACCEPT Deliverable D4.1 (2012), http://www.accept.unige.ch/Products/

ACCEPT Deliverable D5.2 (2013), http://www.accept.unige.ch/Products/

Aikawa, T., Schwartz, L., King, R., Corston-Oliver, M., and Lozano, C. (2007). Impact of controlled language on translation quality and post-editing in a statistical machine translation environment. In *Proceedings of the MT Summit XI*, Copenhagen, Denmark, pp. 10-14,

Banerjee, P., Naskar, S. K., Roturier, J., Way, A. and Van Genabith, J. (2012). Domain Adaptation in SMT of User-Generated Forum Content Guided by OOV Word Reduction: Normalization and/or Supplementary Data? In *Proceedings of EAMT*, Trento.

Bernth, A. and Gdaniec, C. (2002). MTranslatability. In *Machine Translation 16*, pp. 175-218.

Bredenkamp, A., Crysmann B., and Petrea, M. (2000). Looking for errors: A declarative formalism for resource-adaptive language checking. In *Proceedings of LREC 2000*. Athens, Greece.

Genzel, D. (2010). Automatically learning source-side reordering rules for large scale machine translation. In *Proceedings of the 23rd International Conference on Computational Linguistics*, Beijing, China.

Gerlach, J., Porro, V., Bouillon, P., and Lehmann, S. (2013a). La préédition avec des règles peu coûteuses, utile pour la TA statistique des forums ? *In Proceedings of TALN/RECITAL 2013*. Sables d'Olonne, France.

Gerlach, J., Porro, V., Bouillon, P., and Lehmann, S. (2013b). Combining pre-editing and post-editing to improve SMT of user-generated content. In *Proceedings of MT Summit XIV Workshop on Post-editing Technology and Practice*. Nice, France.

Han, B. and Baldwin, T. (2011). Lexical normalisation of short text messages: Makn sens a #twitter, In *ACL 2011*, Portland, OR, USA, pp. 368–378.

Hujisen, W. O. (1998). Controlled Language: An introduction. In *Proceedings of CLAW 98,* Pittsburg, Pennsylvania, pp. 1–15.

Jiang, J., Way, A., and Haque, R. (2012). Translating User-Generated Content in the Social Networking Space. In *Proceedings of AMTA 2012*, San Diego, CA, United States.

Kuhn, T. (2013) A survey and classification of controlled natural languages. *Computational Linguistics*. Early Access publication: June 26, 2013. doi: 10.1162/COLI_a_00168.

O'Brien, S. (2003). Controlling controlled English: An Analysis of Several Controlled Language Rule Sets. In *EAMT-CLAW-03*, Dublin, pp. 105-114.

O'Brien, S. and Roturier, J. (2007). How Portable are Controlled Languages Rules? A Comparison of Two Empirical MT Studies. In *MT Summit XI*, Copenhagen, Denmark, pp. 105-114.

Pym, P. J. (1988). Pre-editing and the use of simplified writing for MT: an engineer's experience of operating an MT system. In *Translating and the Computer 10*.

Rayner, M., Bouillon P. and Haddow B. (2012). Using Source-Language Transformations to Address Register Mismatches in SMT. In *Proceedings of AMTA*, San Diego, CA, United States.

Roturier, J., and Bensadoun, A. (2011). Evaluation of MT Systems to Translate User Generated Content. In *Proceedings of the MT Summit XIII*, p. 244-251.

Roturier, J., Mitchell, L., and Silva, D. (2013). The ACCEPT Post-Editing Environment: a Flexible and Customisable Online Tool to Perform and Analyse Machine Translation Post-Editing. In *Proceedings of MT Summit XIV Workshop on Post-Editing Technology and Practice*. Nice, France.

Ruffino, J.R. (1982). Coping with machine translation. In:

Veronica Lawson (ed.) *Practical Experience of Machine Translation*: *Proceedings of a Conference*, pp. 57-60.

Seretan, V., Bouillon P. and Gerlach J. (to appear). A Large-Scale Evaluation of Pre-editing Strategies for Improving User-Generated Content Translation. In *LREC 2014*.

Streiff, A. A. (1985). New developments in TITUS 4. In: Veronica Lawson (ed.) *Tools for the Trade: Translation and the Computer Aslib*, London, United Kingdom, pp. 185-192.

Temnikova, I (2011). Establishing Implementation Priorities in Aiding Writers of Controlled Crisis Management Texts. In *Recent Advances in Natural Language Processing (RANLP 2011),* Hissar, Bulgaria, pp. 654-659.

Wang, C., Collins, M. and Koehn, P. (2007). Chinese Syntactic Reordering for Statistical Machine Translation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (EMNLP-CoNLL), pp. 737-745.