A Novel Statistical Pre-Processing Model for Rule-Based Machine Translation System

Yanli Sun
SALIS, Dublin City University
yanli.sun2@mail.dcu.ie

Sharon O'Brien
SALIS, Dublin City University
Sharon.obrien@dcu.ie

Minako O'Hagan
SALIS, Dublin City University
Minako.ohagan@dcu.ie

Fred Hollowood
Research & Deployment (SES), Symantec Corporation, Ireland
FHollowood@symantec.com

Abstract

This paper introduces a new statistical pre-processing model for Rule-Based Machine Translation (RBMT) systems. We train a Statistical Machine Translation (SMT) system using monolingual corpora. This model can transform a source input to an RBMT system into a more target-language friendly or RBMT-system friendly "pivot" language. We apply this proposed model to translation from English to Chinese in a pilot project. Automatic evaluation scores (BLEU, TER and GTM) show that this pre-processing model can increase the quality of the output of the RBMT system, especially with an increase in the size of the training corpus. This model is applicable to language pairs which differ in grammar and language structures.

1 Introduction

Recently, a new pre-processing approach that suggests changing the source language to be closer to the structure of the target language has been reported. Wang et al. (2007) reported that transforming Chinese sentences by using hand-coded linguistic rules to be closer to English in terms of syntactic structure, could increase the scores of the final translation by an MT System. Xu and Seneff (2008) transformed English texts into "Zhonglish" (English words in Chinese structure) before translating them by an MT system and found that human evaluations prefer the translation of "Zhonglish" to the translation of the original English texts. A number of other researchers have also described their pre-processing methods on other language pairs. Xia and MacCord (2004) reported the effect of automatically learnt rewrite patterns in improving English and French translation. Crego and Marino (2007) proposed an approach to coupling reordering and decoding in statistical machine translation and reported significant improvements in translation quality between English and Spanish. These papers focused on incorporating syntactic information into Statistical MT (SMT) systems with rules either hand-crafted or automatically extracted. Babych et al. (2009) conducted a similar study for an RBMT system. They applied "construction-level human evaluation" to discover systematically mistranslated contexts and then to "create automatic pre-editing rules to make the constructions more tractable for an RBMT system" (p36). Their study concentrated on some of the most frequently occurring light verb constructions ("verb phrases with a semantically depleted verb and its objects, such as take part, put pressure" etc.) In addition, they still needed to compose the pre-editing rules manually.

Another common pre-processing approach to improve the performance of RBMT systems is Controlled Language (CL) authoring. CL rules can reduce post-editing effort and improve comprehensibility of the final translation (O'Brien, 2003; O'Brien and Roturier, 2007). However, again CL rules have to be manually crafted. Technical writers usually have to manually correct the sentences that violate certain CL rules. Besides, it is difficult to define specific rules and some general rules such as "discourage the use of dangling prepositions" (O'Brien, 2003), could be hard for technical writers to implement.

This paper proposes a new statistical pre-processing model for an RBMT system. The design of the current model differs from the previous ones in the following ways: firstly, the pre-processing model is designed for an RBMT system while most of the previous work focuses on SMT systems; secondly, the transformation process is automated without any hand-coded rules; thirdly, the translation direction is from English to Chinese which is less studied compared to Chinese to English translation.

The remainder of the paper is organized as follows. Sections Two and Three explain the

rationale of our pre-processing model. Section Two presents a new test on back and forward translation and Section Three reports the linguistic analysis of the new test. Section Four presents the general methodology of the pre-processing model and Section Five demonstrates the experimental set-up. Some preliminary evaluation results are reported in Section Six. Section Seven concludes the paper and proposes some future research questions.

2 New Test on Back- and Forward-Translation

Our method was inspired by a pilot project related to "round-trip translation" (Somers, 2005), one intuitive evaluation approach usually (and especially) used by lay people to determine the quality of an MT system. "Round-trip translation" includes translating a text in one language into a second language (Forward-Translation); and then translating it back into the original language (Back-Translation). In cases where the evaluators do not know the target language or no target language reference is available, "round-trip" translation seems to be an intuitive and easy solution for judging the performance of an MT system based on the assumption that the Back-Translation can represent the quality of the Forward-Translation. However, by comparing the BLEU scores (Papineni et al., 2002) of the Forward-Translation and the Back-Translation, Somers (2005) claimed that overall "round-trip" translation was not suitable for MT evaluation as Back-translations tend to get higher scores than Forward-translations. However, it could be useful at sentence level evaluation according to Rapp (2009).

Whether "round-trip" translation could or could not be used as a means of MT evaluation is not the focus of this paper. Instead, Forward-Translation and Back-Translation are defined differently in this paper from their traditional definitions. Generally speaking, Forward- and Back-Translation occur across two different languages, with Forward-Translation into the target language and Back-Translation into the source language. In this paper, we compared a new pair of "Forward-" and "Back-Translation" which are in the same language (in this paper, both are in Chinese). To avoid confusion with the traditional definition of Forward-Translation and Back-Translation, a new set of symbols are used. Procedure 1 below explains how to obtain this new pair of translations for comparison.

Procedure 1: Steps to Obtain New Back Translation and Forward Translation

(i) Input a source English text ($\mathbf{E^O}$, **O**riginal **E**nglish) into an RBMT system and get the target language translation (in this paper, Chinese). Name this translation as $\mathbf{ZH^{MTF}}$ (it can be regarded as a "Forward-Translation" from English to Chinese by the MT system);

(ii) Input the Chinese reference **ZH** (which is human translation of the above English text) into the same RBMT system and get an English translation output. Name this English translation as $\mathbf{E^{MT}}$ (it can also be regarded as a "Forward-Translation" from Chinese to English by the MT system);

(iii) Input $\mathbf{E^{MT}}$ from the above step into the same RBMT system and get the final Chinese translation output. Name this Chinese translation as $\mathbf{ZH^{MTB}}$ (it can be regarded as a "Back-Translation" of the Chinese reference mentioned in the second step. The whole process is $\mathbf{ZH->E^{MT}->ZH^{MTB}}$ (translate the Chinese reference into English and then translate back into Chinese by the MT system)).

To see which translation is better, the Chinese "Back-Translation" (ZH^{MTB}, from step 3 in Procedure 1) or the Chinese "Forward-Translation" (ZH^{MTF}, from step 1 in Procedure 1), two samples were randomly selected from a technical knowledge base (a corpus contains various types of technical documents) of a software company. The Chinese reference was extracted from the in-house Translation Memory, which is the human translation of the English corpus. The entire Chinese corpus in this test was segmented into words. The statistics of the two samples are shown in Table 1.

Table 1: Statistics of the Two Samples

Corpus (#Sentences)	# English Words	# Chinese Words
Sample 1 (500)	9830	10703
Sample 2 (1000)	15915	17257

Samples 1 and 2 were processed according to the three steps in Procedure 1 and two pairs of Chinese translations (ZH^{MTB} and ZH^{MTF}) were generated. Only one automatic evaluation metric

based on precision, recall and F-measure (GTM, Turian et al., 2003) was applied to get the scores of the translations by comparing them to the Chinese reference. GTM is one of the most commonly used automatic evaluation metrics. According to Tatsumi (2009) GTM scores correlate best with human post-editing speed (one of the human evaluation measurements of MT quality). Table 2 reports the GTM scores of ZH^{MTF} and ZH^{MTB} of the two samples.

Table 2: GTM Scores of the Translations

Samples	ZH^{MTF}	ZH^{MTB}
Sample 1	0.65	0.74
Sample 2	0.67	0.75

The scores in Table 2 show that Chinese "Back-Translation" (ZH^{MTB}) is better than Chinese "Forward-Translation" (ZH^{MTF}) in terms of GTM scores for both samples. The next section compares the two translations in detail and reveals one key reason for their differences. And finally a new pre-processing model is proposed based on that key reason.

3 Qualitative Comparisons

One possible reason for the differences between the GTM scores of ZH^{MTF} and ZH^{MTB} relates to what Somers (2005) mentioned about the difference between Forward-Translation and Back-Translation in his tests:

> Although systems perform source-text analysis to a certain extent, when all else fails they resort to word-to-word translation, and where there is a choice of target word they would go for the most general translation. Clearly, when the input to the process is difficult to analyse, the word-for-word translation will deliver pretty much the same words in the BT as featured in the original text. (p130)

Hence, in our test when the Chinese reference was translated into English (the E^{MT} in Procedure 1) due to some failed source analysis the system generated some word-to-word translation in English with some Chinese flavoured structures. When this English translation was translated back to Chinese, a second round of word-to-word translation generated some translations that were the same as the original Chinese reference. In other words, one assumption about E^{MT} is that it contains target-language friendly or at least MT-friendly structures and that is why its translation (ZH^{MTB}) is better than the translation of the source English text (ZH^{MTF}). This assumption arose after comparing the E^{MT} and the E^{O} text. The following example shows their differences.

- E^{O}: A proactive threat scan looks at *the behaviour of active processes* **at the time that the scan runs**.
- E^{MT}: The Proactive Threat Scan **will be scanning the runtime**, checks *the active process the act*.

The two major differences between these two English sentences are marked by bold font and italics. In English, the adverbial phrase (in this example, "**at the time that the scan runs")** is placed after the main verb of the sentence ("looks at") while in Chinese, it is usually placed before the main verb. The E^{MT} sentence shows this characteristic by moving the phrase (which is "**will be scanning the runtime**" in E^{MT}) in front of the main verb ("checks"). Another difference is the position of the modifier and the modified. In the source English sentence, modifiers follow the modified in an attributive clause (such as "**the time that the scan runs**") or in prepositional phrases (such as "*the behaviour of active processes"*). However, in Chinese, the modifiers appear before the modified. Again, the E^{MT} sentence exhibits this grammatical characteristic: "**the time that the scan runs**" was changed to "**scanning the runtime**" and the prepositional phrase "*the behaviour of active processes"* was changed to "*the active process the act"*, both of which put the original modifier before the modified. The differences between the source English sentence (E^{O}) and the English sentence E^{MT} are most likely the reason why the ZH^{MTB} gets a higher GTM score than ZH^{MTF}.

The fact that ZH^{MTB} receives higher scores than ZH^{MTF} also reflects one of the drawbacks of most of the automatic evaluation metrics, i.e. scores of translations are based on the similarity between the machine translation output and the provided reference even though other alternatives of the translation are also acceptable. However, from the automatic evaluation scores, one hypothesis which can be derived is that if an English source sample can be pre-processed into the structures of E^{MT}, its Chinese translation could be better than the direct Chinese translation of this English sample. Therefore, in the next section, we introduce a statistical model to automatically pre-process the source texts which will be translated by an RBMT system into the structures similar to that of E^{MT}. As we mentioned in the introduction, there are already studies showing that changing a source text to be

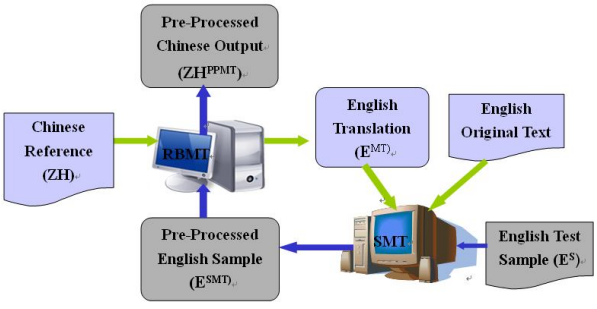
closer to the target language could improve the translation output (Wang et al. 2007; Xu and Seneff 2008).

4 Statistical Pre-Processing

To test the hypothesis that "If we pre-process an English sample into the structure of E^{MT}, the final translation should be better", we need a model that can learn the structures of E^{MT} and automatically transform a new English sample into similar structures. An SMT system, which is trained using two parallel corpora (a source language corpus and a target language corpus) and some statistical methods to try to generate the best target translation for a source sentence, is a good candidate to conduct this transformation. Recently, SMT systems have been applied to post-edit the output of RBMT systems (this process is called Statistical Post-Editing (SPE)) and has been reported to be effective in improving the MT output in many studies (Simard et al. 2007 etc.). The SPE process includes the following steps: first, a corpus is translated using an RBMT system from one language (let us continue with the example of English) into a target language (Chinese). Secondly, an SMT system is trained using this Chinese translation as the "source language" and the Chinese reference as the "target language". The SMT system will learn how to post-edit raw Chinese RBMT output into the corresponding Chinese reference translation. Thirdly, once a new English text is translated using the same RBMT system into Chinese, the translation can be input into the trained SMT system to be post-edited into a revised translation.

Our proposal combines an SMT system and an RBMT system in a similar but a novel manner, i.e. using an SMT system to pre-process the source for the RBMT system instead of post-editing the output. The process is described in Procedure 2 below (and illustrated in Figure 1):

Figure 1: Statistical Pre-Processing Model



Procedure 2: Statistical Pre-Processing

(i) Input a Chinese reference corpus into an RBMT system and get the English translation output. As in Procedure 1, name it $\mathbf{E}^{MT}$ (a translation from Chinese into English by an MT system). It will function as a "pivot" English with some Chinese characteristics or the RBMT-system friendly structures;

(ii) Train an SMT system using the $\mathbf{E}^{MT}$ corpus as the "target" text and the source English corpus $\mathbf{E}^{O}$ as the "source" text. Let the SMT system learn how to translate or pre-process the source English into $\mathbf{E}^{MT}$ style English (a kind of pseudo English);

(iii) Input a new English sample $\mathbf{E}^{S}$ (with no sentences that have appeared in the training corpus) into the trained SMT system. The output will be an English text with $\mathbf{E}^{MT}$ style or flavour. Name the new English sample as $\mathbf{E}^{SMT}$.

(iv) Translate $\mathbf{E}^{SMT}$ (the English text generated from the last step) into Chinese using the RBMT system used in the above steps. The final output is **P**re-**p**rocessed **Ch**inese **M**achine **T**ranslation output ($\mathbf{ZH}^{PPMT}$, for short).

5 Experiment Set-Up

The RBMT system employed is Systran (version 6, customized with a special user dictionary), a well-established and widely used commercial RBMT system. Together with the corpus, this RBMT system is provided by the software company. The SMT system used is Moses, which is an open resource toolkit (Koehn et al. 2007). There were two corpora used in our experiment. The first corpus is a small preposition corpus provided following a related study on improving English to Chinese preposition translation. This corpus contains sentences with at least one preposition per sentence. In order to check how effective this pre-processing model is in improving the translation of prepositions compared to the other methods that have been tested, in the current test, the same test set and training corpus as in previous tests are used. The test set was randomly selected from this preposition corpus. The remaining sentences were used as the first training corpus for the pre-processing experiment. The training corpus and the test set are described

in Table 3. As this training corpus is prepared in a controlled manner, i.e. a preposition must be present in each sentence; it is called the controlled corpus (Cont_Corp, for short).

Table 3: Statistics of Controlled Corpora

	#sentence	#English words	#Chinese words
Cont_Corp	5439	77268	85501
Test set	944	14839	16100

In order to scale the experiment, besides this controlled corpus, we also used another larger corpus which is the in-house English-to-Chinese translation memory knowledge base (from which the above controlled corpus was extracted). Three training corpora were randomly selected from the larger corpus. The first random corpus contains the same number of sentences as the controlled corpus. However, this random corpus is less similar to the test set compared to the controlled corpus in the sense that the random corpus was not filtered to contain sentences with prepositions. The controlled corpus and the test set were all, as mentioned above, restricted to contain sentences with prepositions. The effect of the similarity between training data and test set on the final translation quality could be partly revealed through the comparison of these two corpora. The second and third corpora are larger randomly extracted corpora. The purpose of using another two larger corpora is to see if the size of the training corpus will affect the performance of the pre-processing model. Besides the training corpus, a tuning set was set up for tuning and optimizing the model.

The three random training corpora and the tuning set are listed in Table 4. For all the English corpora, the corresponding Chinese references were extracted again from the in-house Translation Memory of the company. As these corpora are randomly selected, they are called random corpora (Rand_Corp, for short).

Table 4: Statistics of Random Corpora

	#sentence	#English words	#Chinese words
Rand_Corp 1	5439	55846	69410
Rand_Corp 2	9934	106457	119480
Rand_Corp 3	269913	2787175	3382309
Tuning set	903	10677	10764

For each of the training corpora, the four steps listed in Procedure 2 (or Figure 1 in Section 4) were repeated to get the final Chinese translations. Four different translations of the test set were generated from four pre-processing models trained using the four different training corpora. To assess the effectiveness of this pre-processing model a baseline translation was obtained by translating the test set using the default Systran settings of the company (the customized dictionary was employed in all the translation processes) without any other pre-processing process. The final five translations, namely, the baseline translation (Baseline), the translations of the three random training corpora respectively (Ran1_ZH^{PPMT}, Ran2_ZH^{PPMT} and Ran3_ZH^{PPMT}) and the translation using the controlled training corpus (Con_ZH^{PPMT}) are scored by comparing them to the reference using automatic evaluation metrics. The next section reports the scores of these translations and gives a brief analysis of the translation results.

6 Results

In addition to GTM, two more automatic evaluation metrics were used to compare the translations, i.e. BLEU and TER (Snover et al, 2006). They are also among the most commonly used metrics in the field. Besides, using more metrics can reflect the difference between the translations more confidently. Table 5 below reports the final automatic scores of the five translations with and without the source pre-processing model.

Table 5: Automatic Scores of Translations

	GTM	BLEU	TER
Baseline	0.6565	0.2490	0.5249
Ran1_ZH^{PPMT} (Rand_Corp 1)	0.6553	0.2229	0.5499
Ran2_ZH^{PPMT} (Rand_Corp 2)	0.6567	0.2303	0.5436
Ran3_ZH^{PPMT} (Rand_Corp 3)	0.6836 *	0.2746 *	0.5058 *
Con_ZH^{PPMT} (Cont_Corp)	0.6751 *	0.2646 **	0.5261

We performed significance tests on the improvement of the automatic scores compared to the baseline translation using approximate randomization (Noreen, 1989). Scores with * are significantly better than the score of the baseline translation at $p<0.01$ and scores with ** are significant at $p<0.1$. The first model

(Ran1_ZH^{PPMT}) failed to show better scores than the baseline translation. However, the score of the third model (Ran3_ZH^{PPMT}) is quite promising, which is significantly better than the baseline translation. The difference between the training corpus and the test set and the size of the training corpus are the major reasons for the lower scores of the first two models (Ran1_ZH^{PPMT} and Ran2_ZH^{PPMT}). With bigger or more similar corpora, the pre-processing model can render a better translation (Ran3_ZH^{PPMT} and Con_ZH^{PPMT}) than the baseline translation. The results reflect one important criterion in SMT training data selection. While the more the better is still true, it should also be the more similar the better. Although the controlled corpus is much smaller than the biggest random training corpus (Rand_Corp 3), the two models trained using these two corpora work almost as well as each other. Therefore, we can hypothesize that if the biggest random corpus (Rand_Corp 3) was also more similar to the test corpus, the translation can get much higher scores. To sum up, the pre-processing model can improve the output of the RBMT system, especially when the pre-processing model is trained with a bigger training corpus or similar corpus.

To give an example of the improvement introduced by the pre-processing model, the author selected one sentence whose translation from the pre-processing model (Ran3_ZH^{PPMT} to be specific) received higher scores than the baseline translation. Table 6 compares the baseline translation with the reference and Table 7 compares the translation generated after pre-processing (Ran3_ZH^{PPMT}) with the same reference. The shaded blocks indicate where the translations are the same as the reference. The source English sentences are put at the top of the tables. The English sentences at the bottom of the two tables are the glosses of the translations.

From Tables 6 and 7, we can see that although both translations share the same number of correct translations with the reference, their orders are different. Besides the missing word, the translation in Table 7 has the same order as the reference while the baseline translation (in Table 6) has a different order. The glosses show that Ran3_ZH^{PPMT} has almost the same meaning as the source English sentence but the baseline translation has a totally different meaning from the original one.

Table 6: Baseline Translation

English Source	About the processes that proactive threat scans detect							
Ref / MT	关于	主动型	威胁	扫描	所	检测	的	进程
关于	■							
主动型		■						
威胁			■					
扫描				■				
的							■	
进程								■
请								
检测						■		
Gloss	About the processes of proactive threat scans please detect							

Table 7: Ran3_ZH^{PPMT} Translation

English	About the processes that proactive threat scans detect							
Ref / MT	关于	主动型	威胁	扫描	所	检测	的	进程
关于	■							
主动型		■						
威胁			■					
扫描				■				
检测						■		
的							■	
进程								■
Gloss	About the process that proactive threat scans detect							

The original English sentences and the English sentences after pre-processing were compared at sentence level to reveal what changes made by the pre-processing model to the English sentences. The following example exhibits some of the changes that the pre-processing model made to the source English sentence:

- E^{O}: Allows other users in your network to browse files and folders on your computer.
- E^{MT} (Pre-Processed English): Permits other user in your network to browse for the file and folder on your machine.

"Allows" and "computer" in the original English sentence are changed into "permits" and "machine" after pre-processing. "for" and "the" are two new additions found in the pre-processed English sentence. "files and folders" become singular form "file and folder". Further qualitative assessment of these changes is necessary to reveal why, or if, these changes are leading to better translation.

Using TER, we extracted at word level the list of deletions, insertions and substitutions made by the pre-processed model compared to the original English text. Table 8 reports the total number of insertions, deletions and substitutions as well as the top five most frequent changes in each category.

Table 8: No. and Examples of Insertions, Deletions and Substitutions

Category (# occurred)	Example	Frequency	
Insertion (1158)	the		248
	will		46
	,		41
	”		39
	to		36
Deletion (992)	the		102
	of		85
	a		65
	that		59
	you		49
Substitution (5307)	a	the	166
	can	may	150
	computer	machine	64
	that	which	58
	click	clicks	49

Table 8 shows that most of the changes are function words, for example, “the” is both the most frequently inserted and deleted word. Besides the word level changes, we also examined the changes at sentence level. For the 944 English sentences in the test set, the statistical pre-processing model modified 942 sentences (99.8% of the corpus), with only 2 sentences remaining unchanged. We divided all the “pseudo” English sentences (944 sentences) into three groups: group1 (242 sentences) contains sentence with correct English grammar and easily understandable meaning; group 2 (243 sentences) consists of sentences with minor problems in English grammar and understandable meaning; group 3 (456 sentences) contains sentences with ungrammatical grammar and unclear meaning. It is found that some “pseudo” English sentences have different meaning from the original English sentence. This may account for the degradations of the pre-processing model. Although overall the translation of the pre-processed English is better than the original English source text in terms of automatic scores, more valid human evaluation is necessary before classifying whether these changes are improvements or degradations.

7 Future Work

This paper proposes a new pre-processing model for RBMT systems. It also opens up a new perspective on combining SMT and RBMT systems. The new model makes use of the grammatical difference between English and Chinese and the inefficiency of the RBMT system in dealing with the difference. One advantage of the proposed new pre-processing model is that it is not only language independent but also system independent. Overall, this statistical pre-processing model shows promise in terms of automatic evaluation. However, brief examination of the translations suggests there are both improvements and degradations generated in the translations, for example, the position of prepositional phrases. While in some cases moving the prepositional phrase in front of the word it modifies can lead to better translation, it is also a major reason for degradation. Human evaluation with multiple evaluators will be conducted next to reveal more about the degradations and improvements of the pre-processing model. Besides, how to optimize the model by regulating or configuring the translation process of either the RBMT system or the SMT system or both is another topic worth exploring, for example, training corpus cleaning or selection. Another advantage of this model is that it is compatible with other pre- and post-processing approaches. As was mentioned in part four, SMT systems have been used to post-edit the output of the RBMT system and have been shown to be effective in improving the output. Further experiments combining statistical pre-processing and statistical post-editing is currently in progress.

Acknowledgement

This work was financed by Enterprise Ireland and Symantec Corporation (Ireland). The author would like to thank Dr. Johann Roturier and Dr. Yanjun Ma for their support and help on training and decoding of SMT system and other technical problems. Thanks also to the anonymous reviewers for their insightful comments.

Reference

Babych, Bogdan, Anthony Hartley and Serge Sharoff. 2009. “Evaluation-guided pre-editing of source text: improving MT-tractability of light verb constructions”. In P*roceedings of the 13th Annual Conference of the European Association for Machine Translation*, May, Barcelona, pp.36-43.

Crego, Josep Maria and Jose Bernardo Marino. 2006. “Integration of POStag-based source reordering into SMT decoding by an extended search graph”. In *Proceedings of the 7th Biennial Conference of the Association for Machine Translation in the Americas*, August, Boston, pp.29-36.

Koehn, Philipp, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, Evan Herbst. 2007. "Moses: open source toolkit for statistical machine translation". In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL)*, demonstration session, June, Prague, pp.177-180.

Midori, Tatsumi. 2009. “Correlation between automatic evaluation metric scores, post-editing speed, and some other factors”. In *Proceedings of the Twelfth Machine Translation Summit*, August, Ottawa, no page number.

Noreen, Eric W. 1989. *Computer intensive methods for testing hypotheses: An introduction*. New York: John Wiley & Sons.

O'Brien, Sharon. 2003. “Controlling controlled English: an analysis of several controlled language rules sets”. In *Proceedings of EAMT-CLAW-03*, Dublin city university, Dublin, Ireland, pp.105-114.

O'Brien, Sharon and Johann Roturier. 2007. “How portable are controlled language rules? a comparison of two empirical MT studies”. In *Proceeding of MT Summit XI, September, Copenhagen*, pp.345-352.

Papineni, Kishore, Salim Roukos, Todd Ward and Zhu Wei-jing. 2002. “BLEU: a method for automatic evaluation of machine translation”. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), July,* Philadelphia, pp.311-318.

Rapp, Reinhard. 2009. “The back-translation score: automatic MT evaluation at the sentence level without reference translations”. In P*roceeding of the Association for Computational Linguistics – International Joint Conference on Natural Language Processing 2009 Conference Short Papers*, August, Suntec, Singapore, pp. 133-136.

Simard, Michel, Nicola Ueffing, Pierre Isabelle and Roland Kuhn. 2007. “Rule-based translation with statistical phrase-based post-editing”. In *Proceedings of the Second Workshop on Statistical Machine Translation*, June, Prague, pp. 203-206.

Snover, Matthew, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. 2006. "A study of translation edit rate with targeted human annotation". In *Proceedings of Association for Machine Translation in the Americas*, August, Massachusetts, pp. 223-231.

Somers, Harold. 2005 “Round-trip translation: what is it good for?” In *Proceedings of the Australasian Language Technology Workshop 2005*, December, Sydney, pp. 127-133.

Turian, Joseph P., Luke Shen and I. Dan Melamed. 2003. “Evaluation of Machine Translation and its Evaluation”. In *Proceedings of the MT Summit IX*, New Orleans, pp.386-393.

Wang, Chao, Michael Collins and Philipp Koehn. 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*, pp. 737-745.

Xia, Fei and Michael McCord. 2004. “Improving a statistical MT system with automatically learned rewrite patterns”. In *Proceedings of the 20th international conference on Computational Linguistics*, August, Geneva, pp. 508-514.

Xu, Yushi and Stephanie Seneff. 2008. “Two-stage translation: a combined linguistic and statistical machine translation framework”. In *Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas*, October, Hawaii, pp. 222-231