
Simplify Sentence Structure for Improving Human Post-editing Efficiency on Chinese-to-English Patent Machine Translation

Xiaona Ren
Yongpeng Wei
Rile Hu

renxiaona@lingosail.com
weiyongpeng@lingosail.com
hurile@lingosail.com

Beijing Lingosail Tech Co., Ltd, Qingyun Dangdai Building, No.43, West Road of North 3rd Ring Road, Haidian District, Beijing, P. R. China

Abstract

In this paper, we propose a new approach to improve human post-editing efficiency by simplifying the sentence structure on Chinese-to-English patent machine translation (PMT). To simplify the structure of a patent sentence, we use a recognizer to recognize the max noun phrases (MNPs)¹ in a Chinese sentence before translating the sentence. The MNPs are replaced with their head words in the sentence, which makes the sentence structure simpler to be translated. Therefore, the task of translating a complicated sentence is transformed into two subtasks: one is the translation task of MNPs and the other is the translation task of the simplified sentence. And then, the translation results of two subtasks are combined to get the final translation result of the input Chinese sentence. This method outperforms NTCIR-10 official baseline by approximate 2 BLEU points. Moreover, the translation results are beneficial for human post-editing, which can save human post-editing time and improve the quality of translation.

1 Introduction

Patent information is important for communities all around the world, and there is a significant practical need for translations in order to understand patent information written in foreign languages and to apply for patents in foreign countries (Goto et al., 2013). PMT is a significant and important task.

Patent document belongs the category of scientific literature. Patent constitute one of the challenging domains for machine translation because patent sentences can be quite long and contain complex structures, and the presentation of sentence is rigorous and contains a lot of professional terms. Moreover, the structure of Chinese sentence is complex, and the word order is rather arbitrary. Therefore it is difficult to translate a sentence in Chinese patent with general phrase-based statistical machine translation (SMT) method.

In this paper, we propose a novel method for Chinese-to-English PMT, which uses a MNP recognizer to recognize the MNPs in Chinese sentence for simplifying the sentence structure. That is to say, the task of translating a long Chinese sentence can be divided into two subtasks, one subtask is the translation of MNP, and the other subtask is the translation of simplified

¹Maximal noun phrase is the noun phrase which is not contained by any other noun phrases.

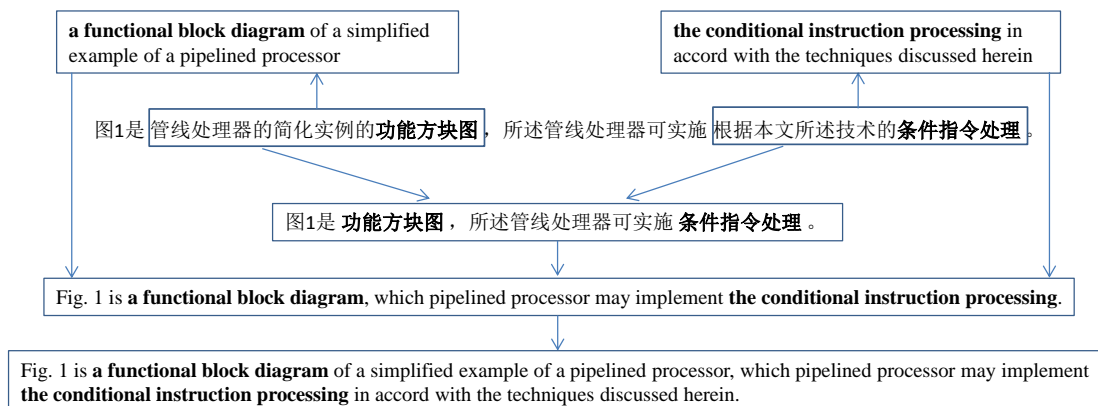


Figure 1: The translation process of a example sentence.

sentence. And then the translation results of two subtasks can be combine to get the final translation result of input Chinese sentence. Our method can be illustrated by the following example sentence as shown in Figure 1. "管线处理器的简化实例的功能方块图" and "根据本文所述技术的条件指令处理" are the MNPs of the example sentence, which be replaced with their core heads to simplify the sentence structure. Therefore, the simplified sentence is "图1是功能方块图, 所述管线处理器可实施条件指令处理。", and then combine the translation results of MNPs and the simplified sentence to get the final translation of the input sentence. Different from the previous works, our approach has two advantages. One is that our approach can improve the translation performance of long sentences on Chinese-to-English PMT, and the other is that the translation result is convenient to be post-edited by translators.

The rest of the paper is organized as follows: Section 2 describes the related works, and Section 3 introduces our translation system. We will present the evaluation results in Section 4. Finally, we conclude this paper and discuss future work in Section 5.

2 Related Works

Most recent researches on Chinese-to-English PMT are based on phrase-based translation model (Fujita and Carpuat, 2013; Zhao et al., 2013b), and some people (Zhao et al., 2013a) use a combined output from two types of grammars supported in their SMT engine (Zheng, 2007), with two different word segmentations. Huang et al. (2013) investigated a sentence-level language model adaptation approach to take advantage of the characteristics of patent documents, and developed SMT system based on the string-to-dependency translation model (Shen et al., 2010) for a variety of languages and contributed substantially to the improvement of translation quality.

One of the characteristics of patent sentences is long, complicated modification. A modification is identified by the presence of a head word. Yokoyama (2013) constructed a modifier correcting system using head words extracted from about 1 million patent sentences, which corrected 60% of the errors. However, our system used a MNP recognizer to recognize the MNPs of patent sentence for simplifying the structure of patent sentence, which can improve the performance of PMT and the efficiency of human post-editing.

Post-editing (PE) is not a new topic in MT-related research, it was studied quite eagerly already in the 1980s especially in Europe where international organizations had a need to share information rapidly in many languages. The initial motivation was to develop effective strate-

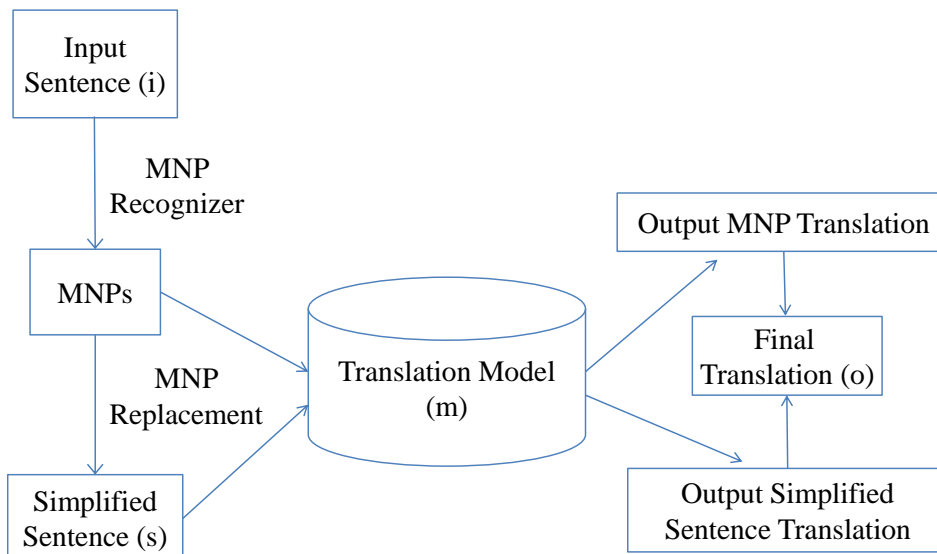


Figure 2: The framework of our machine translation system.

gies and methodologies of human PE in order to make the most efficient use of MT output (Tatsumi, 2010). Sometimes only accuracy is needed, but sometimes stylistic refinement is required (McElhaney and Vasconcellos, 1988; Austerlind, 2001; Allen, 2001; TAUS, 2010). Ehara (2007) reports on their system which is specifically developed to work on English to Japanese patent translation. Based on the experiment and evaluation, he concluded that the rule-based part of the system is good at handling structural transfer, and the statistical part of the system is good at lexical transfer of technical terms. Our approach is inspired by a similar idea used in post-edit process. We can simplify the structure of patent sentence with a MNP recognizer, which is advantage for translators to handle the structure of sentence. And we use the method of statistical machine translation, which is advantage for translators to translate the patent terms.

3 Overview of the translation system

As illustrated in Figure 2, our translation system consists of the following five steps:

- Recognize the MNPs of a input Chinese sentence i using a MNP recognizer;
- Replace the MNPs with their head words, and simplify the input sentence i , and generate the simplified sentence s ;
- Use the translation model m for translating each MNP;
- Use the translation model m for translating the simplified sentence s ;
- Combine the translation of MNPs and the simplified sentence s , and generate the final translation result o of the input sentence i .

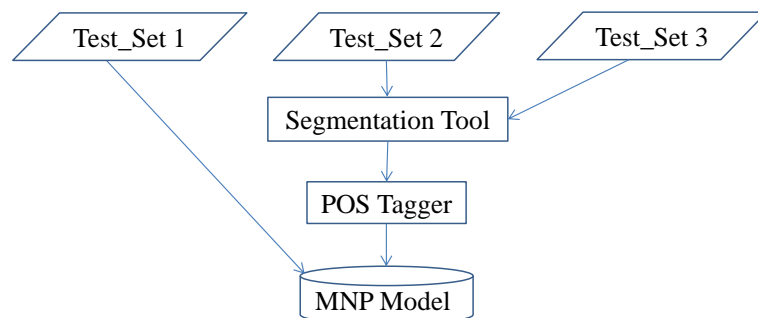


Figure 3: The flow charts of three experiments.

3.1 MNP Recognizer

In this work, we use a statistical method to recognize MNPs of a sentence, which is similar to the related work (Ren et al., 2010). We use the same feature template, but not include the post-process step. The Chinese Treebank (CTB) 5.1 is used in this experiment, and is split into three partitions for training, developing and testing, respectively, following its conventional split in most previous works in the field, as shown in Table 1.

	Sections	Sentences
Train	others	18,104
Dev	271-300	350
Test	301-325	348

Table 1: The split of CTB 5.1 for experiments.

We conducted three experiments to show the performance of the MNP recognizer. Figure 3 shows the flow charts of three experiments. The first experiment is to evaluate the performance of the MNP recognizer based on the CTB 5.1 test corpus that word segmentation and part-of-speech tagging are full correct (test_set1). In the second experiment, we used a Chinese word segmentation tool (i.e. Institute of Computing Technology, Chinese Lexical Analysis System, namely, ICT-CLAS) to implement word segmentation and the Stanford POS tagger to tag CTB 5.1 test corpus (test_set2). The third experiment is to evaluate the performance using the same method as the second experiment just the test sentences are different. In the third experiment, we used the NTCIR-10 Chinese-English bilingual patent text corpus (test_set3) as our experimental test set. The results of three comparative experiments are shown in Table 2.

	Precision	Recall	F1 score
Experiment 1	82.59%	84.09%	83.33%
Experiment 2	71.43%	72.29%	71.86%
Experiment 3	61.18%	63.41%	62.28%

Table 2: The results of three experiments.

However, as can be seen from Table 2, the performance of MNP recognizer is relatively poor on patent domain. One reason is that the word segmentation and POS tagging errors affect the MNP recognition, and the other important reason is that the model for MNP recognition was

trained by Chinese Treebank rather than the patent corpus. Because there is not the Treebank of patent corpus, and manual tagging corpus takes a lot of time and hard work. Although the performance of MNP recognition is relatively poor on patent domain, the MNP recognizer still is favorable for simplifying the sentence structure for machine translation.

3.2 MNP Analysis

In this part, we analysis the influence of using a MNP recognizer on the machine translation. The boundaries of MNPs may be recognized wrong, but not all of them are harmful for translation, which can be seen in the following example.

Input Chinese Sentence: 为了更清楚地说明本发明[实施例或现有技术中的技术方案].
Output English Sentence: to more clearly illustrate the present invention [embodiments or in the prior art technical solution].

In the example, the phrases in the brackets are the recognized MNP of the input Chinese sentence and the corresponding English translation result in output English sentence. We can see that the left boundary of Chinese MNP is wrong. The correct MNP of input Chinese sentence should be recognized as the phrase "本发明实施例或现有技术中的技术方案". But, the translation result is acceptable for the post-editing translators, which does not affect the post-editing translators to understand the sentence structure. Therefore, we conducted the following experiments to analysis in detail the influence of the MNP recognizer.

We randomly extract 300 sentences from patent test set, and use our MNP recognizer to recognize MNPs from them. There are 305 MNPs be recognized, which contain 192 MNPs are recognized to be right, 113 MNPs are recognized to be wrong. And then we define the MNPs that are beneficial for translators to post-edit translation as good MNPs, and otherwise as bad MNPs. In the Table 3 we present the results of our experiments.

	Good	Bad	Percentage
Right MNPs (192)	188	4	97.92%
Wrong MNPs (113)	44	69	38.94%

Table 3: The experimental results of MNP analysis.

From the experimental results, we can see that 97.92 percent of right MNPs are convenient to simplify the structure of sentence and post-edit the machine translation results. Moreover, 38.94 percent of wrong MNPs also are convenient for translators to understand the structure of sentence and post-edit the translation results. Therefore, the method of simplifying sentence structure using a MNP recognizer is beneficial for translators to understand the overall structure of sentence and improve the human post-editing efficiency.

3.3 Translation System

Our machine translation system is a phrase-based system based on two translation models. Like many other MT systems, we use two phrase-based SMT models with some factored translation features for the MNP and simplified sentence translation tasks. However, the two translation models were trained by standard methods. The only difference between them is that the distortion limit parameter value of translating simplified sentence is set as 6 while the distortion limit parameter value of translating MNP is set as 3, which is the best value proved by our experiments.

3.4 Post-Editing Process

Depalma and Kelly (2009) state that even when the MT output needs human Post-editing (PE), it is generally faster and cheaper than human translation, and when the cost is the same, MT



Figure 4: The interface of our machine translation system.

plus PE achieves faster turnaround. Also, some studies have shown that the quality of the final product of MT plus PE can in some cases exceed the quality of human translation (Fiederer and O'Brien, 2009; Koehn, 2009), which is especially obvious in the domain of patents. Therefore, it may further justify the increasing employment of this workflow.

In this paper, we develop a platform for Chinese-English machine translation. Firstly, the imputed Chinese sentence can be translated to English by our MT method, and then be correct into right English translation by translators with the post-editing interface of our platform. This workflow of post-editing includes some basic operation, such as add, delete, remove, modify, etc. And we highlight the MNPs of sentence for translators more convenient to translate. The machine translation system is available on our web site, and the interface is shown as the following Figure 4.

4 Experiments

The experiment process is divided into two steps. The first step is to evaluate the performance of our translation system, which recognizes the MNPs with a MNP recognizer for simplifying the sentence structure. The second step is to evaluate whether the translation of our translation system is more advantageous for post-editing translators than other statistical translation systems.

4.1 Data Sets

We used the experimental corpus from the NTCIR-10 Patent MT Chinese-to-English task, which contains 1 million bilingual sentences pairs for training and 2,000 sentence pairs for development data, and monolingual patent corpus in English covering a span of 13 years (1993-2005).

To measure the overall performance of our method (i.e. translation system and post-editing strategy), we used two metrics: automatic evaluation scores (sentence-level) and human evaluation scores. We then applied some currently available automatic evaluation methods BLEU (Papineni et al., 2002) and NIST (Doddington, 2002) metric to evaluate the performance of our translation system. BLEU and NIST scores were calculated using NIST's mteval-v13a.pl². The

²<http://www.itl.nist.gov/iad/mig/tools/>

human evaluation method is used to evaluate the efficiency of post-editing, which include three indicators: degrees of fit and fluent, the number of operation steps and time. The detail information of evaluation will be shown on subsection 4.3.

4.2 Automatic Evaluation

We trained standard phrase-based SMT system (baseline) on training set and tuned model parameters on development set using Moses (Koehn et al., 2007). A trigram language model (LM) was trained on the English side of the training set by SRILM Toolkit for both systems. GIZA++ (Och and Ney, 2003) was used to obtain symmetric word alignment model. We used a Chinese word segmentation tool (ICT-CLAS) to implement word segmentation. The automatic evaluation results including the baseline system and our system are shown in Table 4.

	Baseline	Our system
BLEU	33.22	33.84 (+0.62)
NIST	8.2389	8.2763 (+0.0374)

Table 4: The automatic evaluation results of two systems.

From the results in Table 4, it can be seen that although our system achieves slight improvements over the baseline system, there is still much to be done to improve it. However, our system is more beneficial to translate complex sentences. We randomly extract 1500 sentences from the test set (Auto_test_set1) as a new test set (Auto_test_set2) that contains at least 70 Chinese words. Table 5 is shown the automatic evaluation results based on Auto_test_set2.

	Baseline	Our system
BLEU	32.76	34.46 (+1.7)
NIST	8.2233	8.4171 (+0.1938)

Table 5: The automatic evaluation results of two systems based on Auto_test_set2.

Obviously, the experimental results reveal that our method is superior to the baseline system based on the new test set. It can be seen that the accuracy rate of MNP recognizer is not high, but our method improved the performance by 1.7 BLEU points and 0.1685 NIST points. Therefore, our translation system is more beneficial to translate long sentences.

4.3 Human Evaluation

In this subsection, we use manual evaluation method to evaluate the improvement of efficiency for post-editing translators. The manual evaluation method contains the following three metrics:

- (1) Semantic accuracy and fluency (SA and SF), which follows the general evaluation criterion of machine translation system (King, 1996).
- (2) Translation error rate (TER), which is the ratio of the number of edits incurred to the total number of words in the reference text (Przybocki et al., 2006; Snover et al., 2006).
- (3) Time, which is the individual productivity in words per hour based on the results of machine translation system (Plitt and Masselot, 2010).

We used 300 sentences with average 230 characters from set2 as the test set of this subsection. Three translators for patent participated in the human evaluation work. Each translator

	SA	Allocation of scores				
		5	4	3	2	1
Baseline	2.98	6	75	132	82	5
Our system	3.16	8	89	149	52	2

	SF	Allocation of scores				
		5	4	3	2	1
Baseline	2.78	3	52	126	113	6
Our system	2.95	7	60	145	87	1

Table 6: The human evaluation results of semantic accuracy and fluency.

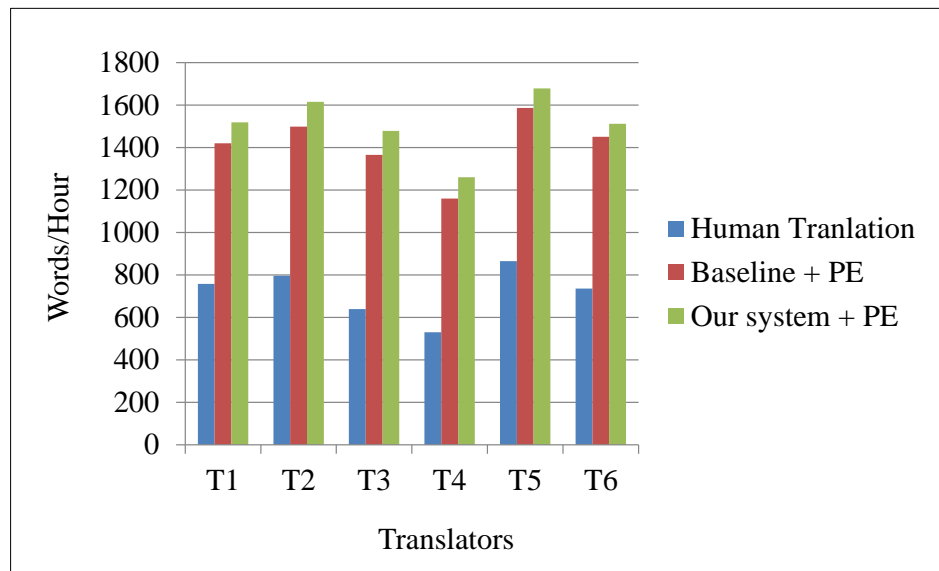


Figure 5: Individual productivity in words per hour.

evaluated the semantic accuracy and fluency of the translation of baseline system and our system. The average values and the allocation of scores of semantic accuracy and fluency are shown in Table 6.

TER is an error metric for machine translation that measures the number of edits required to change a system output into one of the references. We define the edit distance similar to the reference (Przybocki et al., 2006), which is the number of insertions, deletions, and substitutions that are required in order to make a system translation equivalent in meaning to that of a reference translation, using understandable English. We also use publicly available software³ developed by Snover (Snover et al., 2005) to calculate edit distance. We divided the test sentences into three groups, and there were a total of six test participants. Each two participants are the same group, and they translate the same test sentences based on two different machine translation results. Figure 5 shows the individual productivity of each test participant in words per hour, and the translation error rates of each participant are shown in Figure 6.

From the above two Figures and Table 6, we can see that the results of manual evaluation are consistent with the results of automatic evaluation as shown in Table 5. Our translation

³<http://www.cs.umd.edu/snover/tercom/>

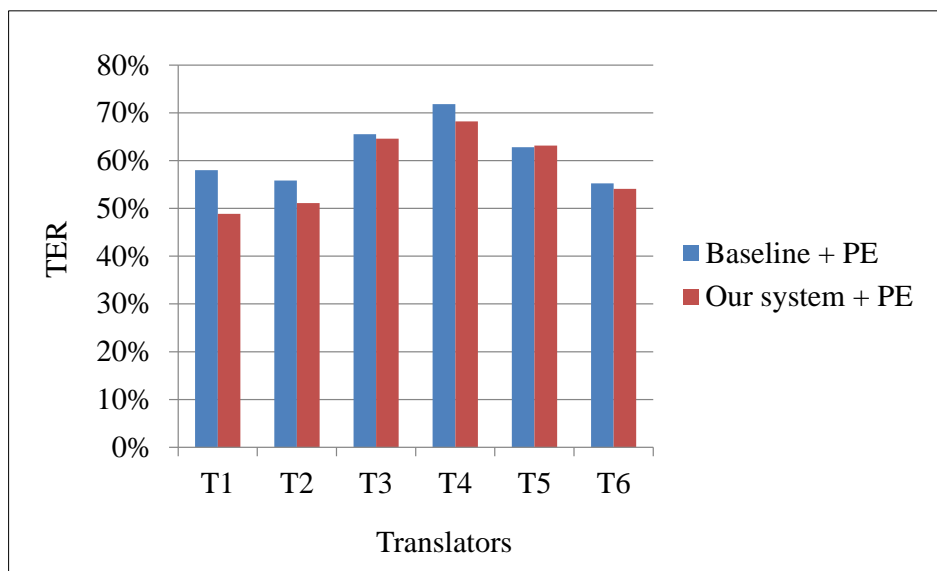


Figure 6: Translation error rate of each test participant.

system performs somewhat better than the baseline system, though there is still much to be explored and improved. Figure 5 shows that post-editing time is shorter based on the translation provided by our system than baseline system, and we can see from the Figure 6 the translation error rates are relatively lower based on the translation provided by our system than baseline system. Therefore, our translation system is more suitable for translators to post-editing.

5 Conclusions and Future Works

The experiments show the efficiency of the proposed method of simplifying sentence structure in the field of PMT, and this proposed method is more efficient to improve the speed and quality of translation for post-editing translators.

In future work, we would like to explore further performance improvement of our translation system. Accurately, there is a big space for our system to improve the performance of translation. On the one hand, we can improve the accuracy of MNP recognizer by increasing the size of the training corpus. We build an extended patent training set automatically using a Chinese word segmentation tool (ICT-CLAS), Stanford POS tagger and our MNP recognizer, and then correct the high frequency of MNPs. The extended model is more suitable for the MNP recognition in the domain of patents. On the other hand, the training corpus for training MNP translation model also can be extended, which be added 1000 pairs of bilingual MNPs. We make a preliminary experiment using the extended training corpus to train a new MNP translation model. The preliminary experiments indicate that the method achieves some improvements in the BLEU score on the same test data.

References

- Allen, J. (2001). Posting: an Integrated Part of a Translation Software Program. *Language International Magazine*, pages 26–29.
- Austermuhl, F. (2001). *Electronic Tools for Translators*. Routledge.

- Depalma, D. and Kelly, N. (2009). The Business Case for Machine Translation: How Organizations Justify and Adopt Automated Translation. *Common Sense Advisory, Inc.*
- Doddington, G. (2002). Automatic Evaluation of Machine Translation Quality using N-gram Co-occurrence Statistics. In *Proceedings of the Second International Conference on Human Language Technology Research*, pages 138–145.
- Ehara, T. (2007). *A Proposal for A New Automatic Machine Translation Evaluation Metrics: NMG*. Japio.
- Fiederer, R. and O'Brien, S. (2009). Quality and Machine Translation: a Realistic Objective? *Specialized Translation*, 11:52–74.
- Fujita, A. and Carpuat, M. (2013). FUN-NRC: Paraphrase-augmented Phrase-based SMT System for NTCIR-10 PatentMT. In *Proceedings of the 10th NII Testbeds and Community for Information access Research Conference*, pages 327–334, Tokyo, Japan.
- Goto, I., Chow, K. P., Lu, B., sumita, E., and Tsou, B. K. (2013). Overview of the Patent Machine Translation Task at the NTCIR-10 Workshop. In *Proceedings of the 10th NII Testbeds and Community for Information access Research Conference*, pages 260–286, Tokyo, Japan.
- Huang, Z., Devlin, J., and Matsoukas, S. (2013). BBN's Systems for the Chinese-English Sub-task of the NTCIR-10 PatentMT Evaluation. In *Proceedings of the 10th NII Testbeds and Community for Information access Research Conference*, pages 287–293, Tokyo, Japan.
- King, M. (1996). Evaluating Natural Language Processing Systems. *Communications of the ACM*, 39(1):73–79.
- Koehn, P. (2009). A Process Study of Computer-aided Translation. *Machine Translation*, 23(4):241–263.
- Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., and Zens, R. (2007). Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceeding of Annual Meeting-Association for Computational Linguistics*, page 2.
- McElhaney, T. and Vasconcellos, M. (1988). The Translator and the Postediting Experience. *State University of New York at Binghamton*.
- Och, F. and Ney, H. (2003). A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19–51.
- Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318.
- Plitt, M. and Masselot, F. (2010). A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context. *The Prague Bulletin of Mathematical Linguistics*, pages 7–16.
- Przybocki, M., Sanders, G., and Le, A. (2006). Edit Distance: A Metric for Machine Translation Evaluation. In *Proceedings of the fifth international conference on language resources and evaluation*, pages 2038–2043.
- Ren, X., Zhou, Q., Kit, C., and Cai, D. (2010). Automatic Identification of Predicate Heads in Chinese Sentences. In *Proceedings of the First CIPS-SIGHAN Joint Conference on Chinese Language Processing*, Beijing.
- Shen, L., Xu, J., and Weischedel, R. (2010). String-to-dependency Statistical Machine Translation. *Computational Linguistics*.

- Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. (2005). A Study of Translation Edit Rate with Targeted Human Annotation. *LAMP-TR-126, CS-TR-4755, UMIACS-TR-2005-58*.
- Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. (2006). A Study of Translation Edit Rate with Targeted Human Annotation. In *Proceedings of the 7th conference of the association for machine translation in the Americas*, Cambridge, Massachusetts, USA.
- Tatsumi, M. (2010). Post-Editing Machine Translated Text in a Commercial Setting: Observation and Statistical Analysis. *PHD Thesis*.
- TAUS (2010). Post-editing in Practice. *A TAUS Report*.
- Yokoyama, S. (2013). Analysis of Parallel Structures in Patent Sentences, Focusing on the Head Words. In *Proceedings of the 5th Workshop on Patent Translation*, pages 24–27, Nice, France.
- Zhao, B., Zheng, J., and Wang, W. (2013a). SRI's Submissions to Chinese-English PatentMT NTCIR10 Evaluation. In *Proceedings of the 10th NII Testbeds and Community for Information access Research Conference*, pages 315–319, Tokyo, Japan.
- Zhao, H., Zhang, J., Utiyama, M., and Sumita, E. (2013b). An Improved Patent Machine Translation System Using Adaptive Enhancement for NTCIR-10 PatentMT Task. In *Proceedings of the 10th NII Testbeds and Community for Information access Research Conference*, pages 376–379, Tokyo, Japan.
- Zheng, J. (2007). Srinterp: Sri's Scalable Multipurpose SMT Engine. *Technical Report*.