Feedback Selecting of Manually Acquired Rules Using Automatic Evaluation

Xianhua Li[†], Yajuan Lü[‡], Yao Meng[†], Qun Liu[‡], Hao Yu[†]

[†]Fujitsu R&D Centre CO., LTD, Beijing, China

{lixianhua, mengyao, yu}@cn.fujitsu.com

[‡]Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

{lvyajuan, liuqun}@ict.ac.cn

Abstract

As the number of manually acquired rules in a patent translation system increases, conflicts between rules inevitably exist. Meanwhile, lacking the matched cases of manually acquired rules in real bilingual corpus, people may conceive rules which cause implausible impact on machine translation quality. In this paper, we propose a feedback selecting algorithm for manually acquired rules in patent translation using automatic evaluation, which picks out manually acquired rules that benefit machine translation quality. Experiments show that we achieve significant improvement in terms of BLEU (+5.23 points over baseline).

1 Introduction

Transfer rules, which are very important resources in rule-based machine translation (RBMT) systems, have shown their effects in statistical machine translation (SMT) recently. There're two ways to get translation rules: automatically acquired and manually acquired. **Automatically acquired rules** are automatically extracted from bilingual corpus. Automatic rule extraction usually generates a large amount of rules which contain incorrect or redundant rules. **Manually acquired rules** are accumulated by human being, which could summarize common linguistic phenomena and should be of high quality ideally.

Sentences in patent documents are usually long and have neat structures. Manually acquired rules can capture sentence structures and have been proved to be effective in translating patent information. Shu Cai (2009) employed 104 manually acquired rules in a patent machine translation system trained on a small data set and achieved an improvement of 1.84 points in terms of BLEU. However, with the increase of manually acquired rules, conflict problem becomes serious. Meanwhile, some manually acquired rules may be low quality and generate incorrect results.

- How to select manually acquired rules that benefit translation results?
- How to integrate manually acquired rules into statistical machine translation system to improve machine translation (MT) quality?

Recently there have been researches focusing on combining the advantages of both SMT systems and rule-based machine translation systems. Ahsan et al. (2010) couple the RBMT and SMT systems at different stages in the RBMT pipeline and report significant improvement. Eisele et al. (2008) describe an architecture that combines multiple rule-based MT engines with a statistical machine translation system in a hybrid system. It uses SMT technologies to align translation outputs from RBMT systems with the source text and extract phrases from the alignments. These phrases are incorporated into the phrase table of the SMT system so that the SMT system can find good combinations of phrases from SMT phrase table and phrases derived from RBMT systems. Dugast et al.

Proceedings of the 4th Workshop on Patent Translation, pages 52-59,
MT Summit XIII, Xiamen, China, September 2011

(2007) combine an RBMT system and a SMT system in a similar way and report improvements for four European language pairs. They also provide qualitative analysis on the contributions of the statistical post editing layer. Ehara (2010) describes a system architecture combining RBMT technique and statistical post-editing techniques. Instead of heavy combination of RBMT and SMT techniques, they adopt light combination for the easiness of system construction.

These approaches all take advantages of both empirical translation rules from RBMT systems and the corpus power of SMT systems. Meanwhile, all these above approaches need to build both SMT system(s) and RBMT system(s). Different from their work, we build our system on the foundation that we already have obtained manually acquired rules for a patent translation system. In other words, we don't have to build an RBMT system to generate such rules. Our work focuses on how to make good use of these rules.

In this paper, we propose a feedback selecting algorithm for manually acquired rules in patent translation, using automatic evaluation of MT quality. Based on a hierarchical phrase-based (HPB) statistical machine translation system (Section 2.1) and an accumulation of manually acquired rules (Section 2.2), we utilize BLEU (Section 2.3) to select manually acquired rules and develop a decoder which employs both automatically acquired rules and manually acquired rules (Section 3).

We evaluated our refined decoder which integrates manually acquired rules into a HPB system (Chiang, 2005; Chiang, 2007) on a test set from patent documents. Experimental results show that the refined HPB decoder with manually acquired rules achieves an absolute improvement of 5.23 BLEU points over traditional HPB decoder (Section 4).

2 Backgrounds

2.1 Hierarchical Phrase-based SMT

Hierarchical phrase-based model (Chiang, 2005; Chiang, 2007) is the state-of-the-art SMT model. By utilizing hierarchical phrases consisting of both words and variables, it captures both short and long distance reorderings and outperforms previous phrase-based models (Keohn et al., 2003; Och and Ney, 2004). We employ the traditional HPB system as our baseline system.

In traditional HPB system, the most important translation resources are automatically acquired rules:

$$X \rightarrow \langle \alpha, \gamma, \sim \rangle \qquad (1)$$

where X is a nonterminal, α and γ are source and target strings respectively. $\sim$ is the one-to-one correspondence between nonterminals in α and γ.

To limit the quantity of automatically acquired rules, there're constraints during rule extraction (Chiang 2005; Chiang 2007). These constraints make automatically acquired rules not sufficient to solve reordering problems in patent MT, as sentences in patent information may be especially long. Manually acquired rules can capture long sentence structures and make up for this shortcoming.

Chiang (2007) utilized the following features in the traditional HPB model: phrase translation probability $p(\alpha|\gamma)$, inverse phrase translation probability $p(\gamma|\alpha)$, lexical translation probability $p_w(\alpha|\gamma)$, inverse lexical translation probability $p_w(\gamma|\alpha)$, word penalty, rule penalty and a target n-gram language model. The log-linear model (Och and Ney, 2003) is used to combine different features:

$$P\gamma(e|f) \propto \sum_{i} \lambda_i h_i(\alpha, \gamma) \qquad (2)$$

where $h_i(\alpha, \gamma)$ is a feature function, λ_i is the weight of h_i. In this paper, we add two new features for automatically acquired rules and manually acquired rules into the log-linear model to distinguish different rules. The feature weights are optimized by minimum error rate training (Och, 2003).

2.2 Manually acquired Rules

Manually acquired rules are accumulated by human being. Compared with automatically acquired rules, manually acquired rules are of better quality and can capture long sentence structures. Also, there can be constraints for nonterminals in manually acquired rules so that they could match more accurately. Examples of manually acquired rules in our Chinese-to-English patent translation system are shown in Table 1.

具有##1[0]{-；　，}?的功效 ||| has effects of ##1
与##1[0]{-，}?相比 ||| compared with ##1
把##1[0]{-，}?倒进 ||| pour ##1 into

Table 1: Examples of Chinese-to-English manually acquired rules.

As we can see from Table 1, each manually acquired rule consists of two parts: source side and target side, which are separated by "|||". Each part is composed of terminals and nonterminals.

The full definition of nonterminals in Chinese part is:

$$\#\#\mathrm{N}[\mathrm{m},\mathrm{n}]\{+/-\mathrm{words}\}?$$

where $\#\#\mathrm{N}$ denotes the nonterminal, $[\mathrm{m},\mathrm{n}]$ indicates the length constraint of the nonterminal (0 means no length constraint), $\{+/-\mathrm{words}\}$ indicates the nonterminal must/must-not contain these words according to the symbol $+/-$. ? indicates whether to match the first valid position or the last one in case there're several valid positions.

The full definition of nonterminals in English part is simply $\#\#\mathrm{N}$. The number of nonterminals in source side and target side should be the same.

With these constraints, a manually acquired rule should be more confident than an automatically acquired rule with the same source side without constraints, when both of them can be applied to a source phrase with the same pattern.

A case of rule-matching is shown in Figure 1.

In Figure 1, the manually acquired rule can match sentence 1, as all the nonterminal constraints are satisfied, while it fails to match sentence 2, because the comma in sentence 2 is not allowed for the nonterminal according to its constraints. As for sentence 3, the manually acquired rule would match the first valid position for the existence of **?** constraint.

With the increase of manually acquired rules, some rules conflict with others and cause implausible influence on SMT quality. Meanwhile, as people are not clear about how the rules they've conceived work in real big corpus, some manually acquired rules may not play their original roles. These all become serious problems when applying manually acquired rules to patent machine translation system.

Manually acquired rule:
具有##1[0]{-；　，}?**的功效** ||| has effects of ##1

Sentence 1:
阿司匹林是一种血液稀释剂,它**具有**预防突发心脏病**的功效**。

Sentence 2:
该产品**具有**净化，抗菌，除臭，释放远红外线**的功效**。

Sentence 3:
乌贼骨**具有**制止胃酸分泌**的功效**及帮助伤口愈合**的功效**。

Figure 1: A case of rule-matching

2.3 Automatic Evaluation of MT Quality

We evaluate the MT quality in our experiments using the BLEU automatic evaluation metric (Papineni et al., 2002). BLEU measures the similarity between machine translation results and human-made translation results (called references) by N-gram precision scores and allows multiple reference translations to model the variety of possible translations. BLEU aims to replace subjective evaluation and speed up the development cycle of MT systems. It is now used not only to aid developers but also for automatically tuning of MT systems (e.g., (Och, 2003; Su et al., 1992; Imamura et al., 2003)).

In this paper, we use case-insensitive BLEU-4 and there's only one reference for each input sentence.

3 Feedback Selecting Using Automatic Evaluation

Based on the above discussion, we propose an algorithm called feedback selecting of manually acquired rules using automatic evaluation. In this section, we'll introduce the procedure of feedback selecting and the strategies we employ to reduce time-cost.

3.1 Feedback Selecting Algorithm

Generally, tuning set is much smaller than training set. Therefore, only part of manually acquired rules can be tested by tuning set. In order to avoid this problem, we apply the feedback selecting algorithm to training set and test all the manually acquired rules. To avoid confusion, we call the corpus to test manually acquired rules "evaluation corpus". The algorithm is shown in Figure 2.

static: B_{base}, the BLEU score of traditional HPB system with no manually acquired rules
C_{eval}, the evaluation corpus
$R_{man\text{-}acqd}$, the manually acquired rule set
$R_{auto\text{-}acquired}$, the automatically acquired rule set

function FEEDBACK-SELECTING () **returns** selected manually acquired rule set
  $R_{selected} \leftarrow \Phi$
  $R_{current} \leftarrow R_{auto\text{-}acquired}$
  $B_{iter} \leftarrow B_{base}$
  repeat
    $SelectedRule_{iter} \leftarrow \Phi$
    $Set\text{-}bleu \leftarrow$ GET-CONTRIBUTION-OF-R$_{\text{MAN-ACQD}}$ ($R_{MAN\text{-}ACQD}$, $R_{current}$)
    sort *Set-bleu* by ΔB_i in decrease order
    $< R_{i\text{-}max}, \Delta B_{i\text{-}max}, Set_{sentences\text{-}max} > \leftarrow$ first element in *Set-bleu*
    $Set_{sentences\text{-}pre} \leftarrow Set_{sentences\text{-}max}$
    for each $<R_i, \Delta B_i, Set_{sentences\text{-}i}>$ **in** *Set-bleu* **do**
      if $\Delta B_i < 0$ **then do continue;**
      if $Set_{sentences\text{-}i} \cap Set_{sentences\text{-}pre} = \Phi$ **and** $\Delta B_i >= 0$ **then do**
        add R_i to $SelectedRule_{iter}$
        $Set_{sentences\text{-}pre} \leftarrow Set_{sentences\text{-}i} \cup Set_{sentences\text{-}pre}$
    end
    $R_{selected} \leftarrow SelectedRule_{iter} \cup R_{selected}$
    $R_{man\text{-}acqd} \leftarrow R_{man\text{-}acqd} - SelectedRule_{iter}$
    $R_{current} \leftarrow R_{selected} \cup R_{current}$
    $B_{iter} \leftarrow$ GET-BLEU(C_{eval}, $R_{current}$)
  until $SelectedRule_{iter} = \Phi$
  return $R_{selected}$

function GET-CONTRIBUTION-OF-R$_{\text{MAN-ACQD}}$ ($R_{MAN\text{-}ACQD}$, $R_{current}$) **returns** a set of $<R_i, \Delta B_i, Set_{sentences}>$
  $Set\text{-}bleu \leftarrow \Phi$
  for each R_i **in** $R_{man\text{-}acqd}$ **do**
    $R_{mixed} \leftarrow$ add R_i to $R_{current}$
    $Set_{sentences} \leftarrow$ sentences **in** C_{eval} which matches R_i
    $B_i \leftarrow$ GET-BLEU(C_{eval}, R_{mixed})
    $\Delta B_i \leftarrow B_i - B_{iter}$
    add $< R_i, \Delta B_i, Set_{sentences} >$ to *Set-bleu*
  end
  return *Set-bleu*

Figure 2: Feedback selecting algorithm

The algorithm can be summarized as follows.

- Translate the evaluation corpus with no manually acquired rules and get a baseline BLEU score
- For each unselected manually acquired rule, get the BLEU score after integrating this rule with the selected rules into the refined HPB model and obtain the difference between baseline BLEU score and the new score. We call the difference **rule contribution**.
- Sort rule contribution in decrease order. Add the first manually acquired rule into selected manually acquired rule set and delete it from unselected manually acquired rule set. Create a set S to store the sentences it matches. For other rules in unselected set, if its rule contribution is not negative and its matched sentences do not exist in the sentence set S, add it into the selected rule set and add its matched sentences into S.

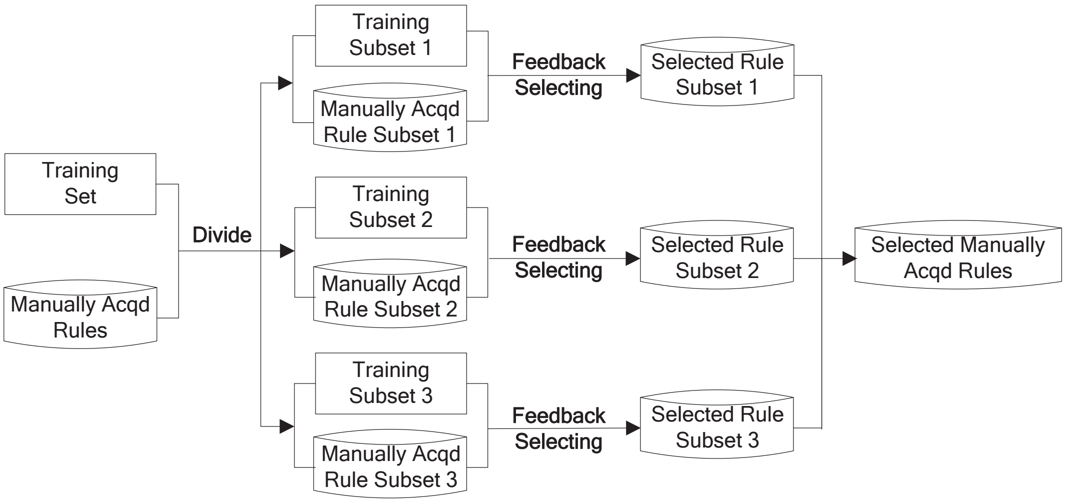

Figure 3: Structure of feedback selecting on training set (in the case that training set and manually acquired rules can be divided into three parts)

- Employ the manually acquired rules in selected rule set and translate the evaluation corpus. Update the baseline BLEU score with this new one.
- Repeat the above three steps until the selected rule set no longer expands.

Automatically acquired rules are utilized in the whole process. To enhance the competitiveness of manually acquired rules during decoding, here we set all the four probabilities introduced in Section 2.1 to the highest probability 1. Besides, nonterminal constraints for manually acquired rules are checked during decoding.

3.2 Computational Cost Reduction

In feedback selecting, the most time-consuming procedure is to get rule contribution of each unselected manually acquired rule. We have to translate the whole evaluation corpus each time. For example, in the first iteration to get rule contribution of each manually acquired rule, we have to translate $C \times R$ sentences, where C denotes the evaluation corpus size and R denotes the manually acquired rule size. In this paper, the training set contains 1000000 sentences and the manually acquired rule size is 9283. So the number of sentences to translate in the first iteration is nearly one billion, not to mention the sentences to translate in later iterations.

In order to reduce computational cost and make feedback selecting practicable, we go further to reduce sentences to translate. As the sentences a manually acquired rule may be applied to are definite, only the results of these sentences may change. Therefore, we only translate the sentences the current rule may be applied to, rather than the whole corpus. By replacing the corresponding results in baseline results, we get the result for current manually acquired rule. Assume that one manually acquired rule may be applied to three sentences on average, sentences to translate in the first iteration become $3 \times R$. Compared with $C \times R$, this obviously reduces computational cost.

In the worst case, we can only pick out one manually acquired rule in each iteration, and all the manually acquired rules should be selected. Therefore, we have to translate $3 \times (R-N+1)$ sentences in the Nth iteration. The total sentence number to translate is $3 \times R + (3 \times R-1) + \dots + 3 \times 1 = 3/2 \times (R+1)^2$. As R = 9283 in this paper, total sentences to translate become 129 million. That’s still beyond endurance. Thus we divide evaluation corpus and manually acquired rules into several self-contained parts and apply feedback selecting algorithm to each part simultaneously to speed up the selecting procedure. Here by self-contained we mean that sentences/manually acquired rules in one part can only match manually acquired rules/sentences in the same part. Structure of feedback selecting on training set is shown in Figure 3.

corpus	TCM	CI	M	PE	total
training set	200,000	400,000	200,000	200,000	1,000,000
tuning set	200	400	200	200	1,000
test set	2,000	4,000	2,000	2,000	10,000

Table 2: Corpus details

4 Experiments

4.1 Datasets

Our experiments were carried on Chinese-to-English patent translation. The bilingual corpus used in the following experiments was from patent domain bilingual corpus. It consisted of four parts: Traditional Chinese Medicine (TCM), Chemical Industry (CI), Machinery (M) and Physical Electronics (PE). We used the SRI Language Modeling Toolkit (Stolcke, 2002) to train a 7-gram model on training set. We evaluated the translation results using case-insensitive BLEU metric (Papineni et al., 2002). The corpus details (the number of sentences) are shown in Table 2.

The manually acquired rules employed in this paper were accumulated for a patent information translation system (Lü et al., 2007). There're totally 9283 manually acquired rules. Each of them has at least one matched sentence in the training set.

The goal of our experiments was to pick out high quality manually acquired rules which could benefit translation quality.

4.2 Description of Experiments

Baseline Our baseline system was a traditional hierarchical phrase-based system. We obtained word alignments of training data by running GIZA++ (Och and Ney, 2003) and then applied the refinement rule "grow-diag-and-final" (Koehn et al., 2003). After that, we extracted automatically acquired rules according to Chiang 2007. We tuned the feature weights with minimum error rate training (Och, 2003). The traditional hierarchical phrase-based system achieved a BLEU score of 30.55 on the test set.

Applying all manually acquired rules Our refined decoder is able to integrate manually acquired rules into the traditional HPB machine translation system. We combined manually acquired rules with automatically acquired rules by assigning manually acquired rules probabilities so that these two types of rules could compete during decoding. Also, the decoder would check the constraints of nonterminals in manually acquired rules so that they would match the right sentences.

We first employed all the manually acquired rules and set all the four probabilities (Section 2.1) 1. Assigning the highest probabilities to manually acquired rules enhanced their competitiveness during decoding. To our surprise, the BLEU score on test set decreased to 29.82, indicating the manually acquired rules should not be imposed on the system. We examined the translation results and found some short manually acquired rules caused great confusion. Examples of such manually acquired rules are shown in Table 3.

```
于##1[0]{-，　于}?后 ||| after ##1
并对##1[0]{-，} ||| and for ##1
第##1[0]{-，}?个 ||| No.##1
经##1[0]{-，}?后 ||| after ##1
呈##1[0]{-呈，}?状 ||| in ##1 form
对##1[0]{-对，}?施##2[0]{-，} ||| Apply ##2 to ##1
```

Table 3: Examples of short manually acquired rules which caused problems

The main problem of such rules was no limits on nonterminal length, so that they would match sentences which they shouldn't match. By modifying the length limitation of nonterminals, these rules would play their due roles.

Applying feedback selected manually acquired rules Then we adopted feedback selecting of manually acquired rules using automatic evaluation. As introduced before, we divided the training set and manually acquired rules into 28 parts to reduce time cost. We applied feedback selecting algorithm to each part. As a result, 2983 manually acquired rules were picked out. We integrated these 2983 rules into the refined HPB system with their probabilities set to 1 and the BLEU score achieved 30.87.

Adding two new features for all rules To the rest 6300 manually acquired rules, they either conflicted with the selected 2938 rules, or might have implausible effects on MT quality. We modified those short rules by editing their length constrains for nonterminals to avoid incorrect matches.

For all rules (including automatically acquired rules and manually acquired rules), we added two new features to distinguish them: one feature to indicate whether the rule is a manually acquired rule or not, the other to indicate whether the rule is a selected manually acquired rule.

Afterwards, we applied all the manually acquired rules in the refined HPB system and adopted minimum error rate training to tune the feature weights. As a result, the BLEU score achieved 35.78. By adding two additional features for all rules, we achieved great improvements over baseline.

Details of experiments are shown in Table 4.

	# of manually acquired rules	BLEU
Traditional HPB system	0	30.55
+All ManAcqd rules	9283	29.82
+Selected ManAcqd rules	2983	30.87
+ All ManAcqd (with two more features)	9283	35.78

Table 4: BLEU scores with different manually acquired rules

5 Discussions and Analysis

5.1 Domain Adaptation

In this paper, the training corpus and manually acquired rules we used are from different patent domains. As a result, we have selected manually acquired rules which are feasible for different domains. When translating documents in a certain patent domain, we can take the corpus of this domain as evaluation corpus and carry out feedback selecting algorithm on it. Then the selected manually acquired rule set will be adapted to this domain.

5.2 Automatic Evaluation Metrics

Recently there have been researches on automatic evaluation of MT quality, such as Akiba et al. (2001) and Yasuda et al. (2001). The feedback selecting algorithm is independent of BLEU. In other words, it is feasible to adopt other evaluation methods which output scores of MT quality instead of BLEU. Actually, we have observed manually acquired rules which are suitable for some sentences not selected by feedback selecting because they don't coincide with the reference. Different evaluation metrics may lead to different results. The evaluation metrics for feedback selecting remains an interesting future work.

5.3 Other Translation Resources

The feedback selecting algorithm is applicable to select manually acquired rules. It can also be employed to test other translation resources such as dictionaries and so on.

6 Conclusions and Future Work

In this paper, we proposed a feedback selecting algorithm for manually acquired rules employed in a patent machine translation system. We measured the contribution of each manually acquired rule and utilized BLEU to evaluate the translation quality. We also introduced strategies to reduce computational cost. After feedback selecting, we integrated manually acquired rules into a refined HPB machine translation system. We added two additional features into the log-linear model and optimized the feature weights by MERT. As a result, we observed significant improvement of SMT quality by 17.12% and by 5.23 in terms of BLEU. This is an absolute improvement over baseline.

Our future efforts would focus on the automatically acquisition of promising long distance reordering rules, especially the rules which capture patent sentence structures. We also plan to work on rule recommendation to reduce the manual work and relieve rule conflict problem.

Acknowledgement

This work was partly supported by National Natural Science Foundation of China Contract 60873167 and 60736014. We thank the anonymous reviewers for their insightful comments.

References

Arafat Ahsan, Prasanth Kolachina, Sudheer Kolachina, Dipti Misra Sharma, Rajeev Sangal. 2010. *Coupling Statistical Machine Translation with Rule-based Transfer and Generation.* In Proc. of the 9th Conference of the Association for Machine Translation in the Americas.

Yasuhiro Akiba, Kenji Imamura, Eiichiro Sumita. 2001. *Using multiple edit distance to automatically rank machine translation output.* In Proc. of Machine Translation Summit VIII. pages 15-20.

Shu Cai, Yajuan Lü, Qun Liu. 2009. *Application of Linguistic Translation Templates in Statistical Machine Translation Systems.* In Proc. of the 5th China Workshop on Machine Translation.

David Chiang. 2005. *A Hierarchical Phrase-based Model for Statistical Machine Translation.* In Proc. of the 43th Annual Meeting on Association for Computational Linguistics, pages 263-270.

David Chiang. 2007. *Hierarchical Phrase Based Translation.* Computational Linguistics, Vol 33, pages 201-228.

Loïc Dugast, Jean Senellart, Philipp Koehn. 2007. *Statistical post-editing on SYSTRAN's rule-based translation system.* In Proc. of the Second Workshop on Statistical Machine Translation , pages 220-223.

Terumasa Ehara. 2010. *Machine translation for patent documents combining rule-based translation and statistical post-editing.* In Proc. of the 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies Information Retrieval Question Answering and CrossLingual Information Access, pages 384-386

Andreas Eisele, Christian Federmann, Herve Saint-Amand, Michael Jellinghaus, Teresa Herrmann, Yu Chen. 2008. *Using Moses to Integrate Multiple Rule-Based Machine Translation Engines into a Hybrid System.* In Proc. of the Third Workshop on Statistical Machine Translation, pages 179-182.

Kenji Imamura, Eiichiro Sumita, Yuji Matsumoto. 2003. *Feedback cleaning of machine translation rules using automatic evaluation.* In Proc. of the 41st Annual Meeting on Association for Computational Linguistics. Vol 1, pages 447-454.

Phillip Koehn, Franz Josef Och, Daniel Marcu. 2003. *Statistical phrase-based translation.* In Proc. of NAACL03.

A L Lagarda, V Alabau, Francisco Casacuberta, R Silva, E Díaz De Liaño. 2009. *Statistical Post-Editing of a Rule-Based Machine Translation System.* In Proc. of NAACL HLT, pages 217-220.

Yajuan Lü, Lei Fu, Jin Huang, Zhongjun He, Qun liu. 2007. *Design and implementation of a statistical machine translation system for patent translation.* Information Technology Letter., pages 33-42.

Franz Josef Och. 2003. *Minimum Error Rate Training for Statistical Machine Translation.* In Proc. of the 41st Annual Meeting of the ACL. Vol 1, pages 160-167.

Franz Josef Och, Hermann Ney. 2003. *A systematic comparison of various statistical alignment models.* Computational Linguistics, 29(1).

Franz Josef Och, Hermann Ney. 2004. *The alignment template approach to statistical machine translation.* Computational Linguistics, pages 417-449

Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu. 2002. *Bleu: a Method for Automatic Evaluation of Machine Translation.* In Proc. of the 40th Annual Meeting on Association for Computational Linguistics, pages 311-318.

Keh-Yih Su, Ming-Wen Wu, Jing-Shin Chang. 1992. *A new quantitative quality measure for machine translation systems.* In Proc. of COLING-92, pages 433-439.

Andreas Stolcke. 2002. *Srilm-an Extensible Language Modeling Toolkit.* In Proc. of the International Conference on Spoken Language Processing. Vol 2, pages 901-904.

Ehara Terumasa. 2007. *Rule Based Machine Translation Combined with Statistical Post Editor for Japanese to English Patent Translation.* In Proc. of Machine Translation Summit XI (2007), pages 13-18

Kenji Yasuda. Fumiaki Sugaya, Toshiyuki Takezawa, Seiichi Yamamoto, Masuzo Yanagida. 2001. *An automatic evaluation method of translation quality using translation answer candidates queries from a parallel corpus.* In Proc. of Machine Translation Summit VIII, pages 373-378.