

Applying Statistical Post-Editing to English-to-Korean Rule-based Machine Translation System

Ki-Young Lee and Young-Gil Kim

Natural Language Processing Team,
Electronics and Telecommunications Research Institute,
138 Gajeongno, Yuseong-gu, Daejeon, Korea
{leeky, kimyk}@etri.re.kr

Abstract

Conventional rule-based machine translation system suffers from its weakness of fluency in the view of target language generation. In particular, when translating English spoken language to Korean, the fluency of translation result is as important as adequacy in the aspect of readability and understanding. This problem is more severe in language pairs such as English-Korean. It's because English and Korean belong to different language family. So they have distinct characteristics. And this issue is very important factor which effects translation quality. This paper describes a statistical post-editing for improving the fluency of rule-based machine translation system. Through various experiments, we examined the effect of statistical post-editing for FromTo-EK¹ system which is a kind of rule-based machine translation system, for spoken language. The experiments showed promising results for translating diverse English spoken language sentences.

¹ FromTo-EK is an English-to-Korean rule-based machine translation system for some various domains (patent, paper, email and messenger).

1 Introduction

There have been many improvements in machine translation from rule-based machine translation (RBMT) to the latest statistical machine translation (SMT). Approaches for machine translation can be typically classified into conventional rule-based approach and statistical approach (Jin et al., 2008). RBMT translates a source sentence to a target sentence through analysis process, transfer process and generation process using analysis rules, dictionaries and transfer rules as its main translation knowledge. On the other hand, SMT system accomplishes translation using translation model and language model obtained from training large parallel corpus composed of source sentences and the corresponding target sentences (Koehn et al. 2003). Comparing two approaches, they have opposite features. That is, rule-based approach is better than statistical approach in the aspect of translation accuracy. However, fluency is contrary to each other. The language pairs that linguistic differences are huge such as English-Korean show these kinds of features apparently.

We aim to improve the translation fluency of RBMT system by introducing SPE. The proposed method is similar to SMT. Difference is the composition of parallel corpus used to build statistical models. To build model for post-editing, parallel corpus should be ready, which is composed of the pairs of the sentence translated by RBMT and the corresponding correct sentence

translated by human translators. Using this parallel corpus, we can build statistical model to post-edit RBMT results. Also, we explain some points to consider when applying SPE to English-to-Korean translation by various experiments.

Our method consists of the following steps:

- Constructing parallel corpus composed of translation results by English-to-Korean RBMT system and translation results by human translator of English source sentences.
- Building translation model and language model for applying SPE (at this phase, SMT toolkits are used).
- Applying decoder for SPE to the output of RBMT system.

The section 2 of this paper presents weakness of conventional rule-based machine translation system. And the overview of our method will be described in the section 3. The section 4 describes experimental parameter, experimental results. In the section 5 and the section 6 we sum up the discussion and show the future research direction.

2 Weakness of RBMT system

Figure 1 shows the configuration of our rule-based machine translation system, FromTo-EK. The flow of machine translation as follows. First, roots of words in an input sentence is restored and part-of-speech (POS) tagging is carried out by morphological analysis and tagging module. Second, syntactic structure is found out by syntactic analysis module (parser). FromTo-EK engine employs full parsing strategy to analyze English source sentences. Third, input sentence structure (parse tree) is transferred to adequate target language structure using transfer patterns. At this step, lexical transfer based on context is conducted using dictionaries and word sense disambiguation knowledge (Yang et al., 2010). Fourth, Korean generator generates final Korean translation sentence.

The advantage of RBMT engine is that it can catch the exact dependency relation between the words in input sentence. It is very helpful to achieve high translation accuracy. In particular, in the case of language pairs which are very similar in the sense of linguistics (for example, Korean-Japanese),

rule-based approach has showed good translation performance.

However, the problem of RBMT system is its poor fluency compared to SMT system. In particular, in translating spoken language sentences, such features are outstanding. Actually, when translating English spoken language sentences, it is found that unnatural, rigid and dried expressions are frequently used. It results from stereotyped language transfer phase, translation knowledge and the limit of translation methodology. We cannot achieve the fluency like human translation by assembling translation knowledge pieces. This is why we propose SPE for RBMT system.

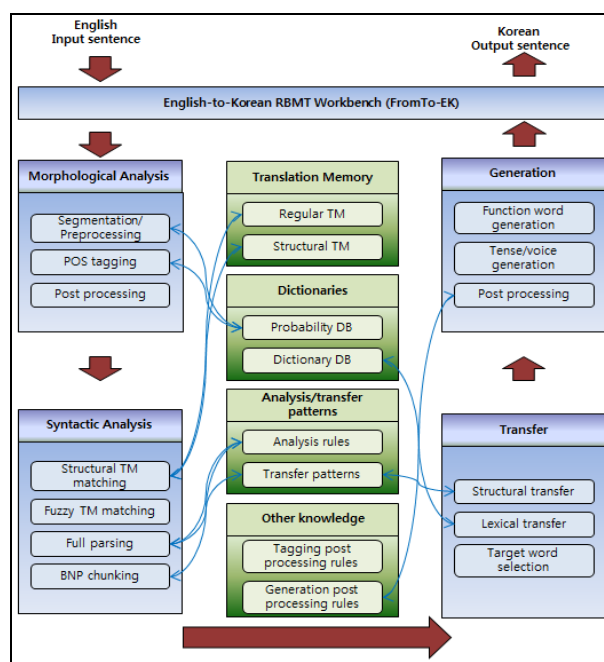


Figure 1. FromTo-EK system configuration

3 Statistical Post-Editing

3.1 Target of post-editing

First of all, to find out the target of applying SPE, we manually evaluated FromTo-EK (English-to-Korean RBMT system) using test set composed of 200 English spoken language sentences.

The measure for human evaluation is showed in Table 1 (Doyon et al., 1998). The score of one sentence in test set is the average of scores assigned by three human translators.

Score	Evaluation categories
4	All meaning expressed in the source fragment appears in the translation fragment
3	Most of the source fragment meaning is expressed in the translation fragment
2	Much of the source fragment meaning is expressed in the translation fragment
1	Little of the source fragment meaning is expressed in the translation fragment
0	None of the meaning expressed in the source fragment is expressed in the translation fragment

Table 1: Accuracy test of machine translation system

Score	Error category	# sent	# errors	Error rate
Score over 2.5	Morphological analysis or POS tagging errors	179	4	2.2%
	Parsing errors		9	5.0%
	Transfer or generation errors		15	8.4%
	Knowledge error		6	3.4%
Score under 2.5	Morphological analysis or POS tagging errors	21	3	14.3%
	Parsing errors		6	28.6%
	Transfer or generation errors		3	14.3%
	Knowledge errors		4	19.0%

Table 2: FromTo-EK evaluation analysis

As is shown in Table 2, analysis of the result of human evaluation shows that the sentences under average 2.5 have errors caused by analysis failure (including POS tagging errors and parsing errors). In this case, input to a transfer module (output from analysis module) is parse tree which already includes incorrect dependency relations. Because of syntactic (or morphological) analysis errors,

these kinds of translation results have difficulty conveying the meaning of the original sentences. So, the sentences under 2.5 have little improvements by SPE. Meanwhile, in the case of the translation results scored over 2.5, it is not difficult to understand the overall meaning of source sentences except one or two words. Sentences over 2.5 have very little of errors from analysis phases. So, misunderstanding brought by analyzing source sentences incorrectly is rarely found. However, the sentences over 2.5 have some other problems. It may be summarized as Table 3. As we know by Table 3, the problem with the sentences over 2.5 can be divided into two categories. The first is intrinsic to English-to-Korean transfer module and Korean generation module. These are because of the limit of RBMT paradigm and the lack of translation knowledge. The second, on the other hand, is some different in that source sentence was well translated to target sentence without the loss of meaning. That is, the only problem is that in the aspect of the command of Korean, RBMT result is not so natural. At the same time, to improve the fluency of RBMT is the target of this paper.

factors	comment
Wrong position of adverb	In Korean, the wrong position of adverb can change the object of modification.
Awkward expression	The lack of ambiguity resolution knowledge generates target words not matched in context.
Stereotyped target word	In most RBMT system, there is perfect strategy to select more natural target word depending on context in the aspect of target language.

Table 3: Factors which cause fluency problems in FromTo-EK

The fluency problem with RBMT system is that translation process has a mechanism with an emphasis on source sentence analysis. In RBMT approach, after input sentence analysis phase, transfer is based on just analysis result and translation knowledge such as dictionaries and transfer patterns. During source-to-target transfer and target sentence generation, the fluency of translation output sentence is not considered. In this paper, we demonstrate the impact of SPE for

RBMT and explain improvements by SPE in the concrete.

3.2 Statistical Post-Editing Architecture

SPE is based on SMT. SPE and SMT differ from the composition of training corpus. For building translation model, SMT uses large parallel corpus composed of the pairs of source sentence and corresponding target sentence. The parallel corpus for SPE training is the pairs of translation sentence by RBMT system and the corresponding correct translation sentence by human translator.

Table 4 shows parallel corpus composed of translation by RBMT and Human respectively. In Table 4, first column is English sentence, second column is Korean translation sentence by RBMT and third column is Korean translation sentence by human translator. In the meaning point, translations by RBMT (second column) have not bad accuracy. However their fluency is not natural. In the other words, the sentences at second column convey right meaning, but are not fluent sentences. We aim to align translation by RBMT into translation by human for getting knowledge for SPE. Through this learning process, the useful data for improving erroneous expressions to correct expressions can be acquired.

	야합니다)	
I'm looking for something for my friend.	Je chingureul wihan eotteon geoseul chatkko itsseoyo (제 친구를 위한 어떤 것을 찾고 있습니다)	Je chinguege julmanhan geoseul chatkko itsseoyo (제 친구에게 줄 만한 것을 찾고 있어요)
Do you have bags made of softer leather?	Dangsineun gabangeul deo budeureoun gajugeuro mandeureojigeha mnikka (당신은 가방을 더 부드러운 가죽으로 만들어지게 합니까)	Deo budeureoun gaguk gabang itnayo (더 부드러운 가죽 가방 있나요)
Please show me some ladies' watches.	Naege yakkanui eoseongui sigereul boyeojuseyo (나에게 약간의 여성의 시계를 보여주세요)	Eoseongyong sige jom boyeojuseyo (여성용 시계 좀 보여주세요)

Table 4: Parallel corpus for SPE

Source sentence	Translation by RBMT	Translation by Human
What time does it leave?	Geugeoseun myeotsie chulbalhamnikka ² ? (그것은 몇 시에 출발합니까)	Myeot sie chulbalhajyo? (몇 시에 출발하죠)
Do you have any other colors?	Dangsineun dareun saegi itsseumnikka? (당신은 다른 색이 있습니까)	Dareun saekkal jom boyeojusigetsseoyo? (다른 색깔 좀 보여주시겠어요)
It's too flashy for me.	Geugeoseun nareul wihae neomu yahamnida. (그것은 나를 위해 너무	Jeohanteneun neomu hwaryeohandeyo (저한테는 너무 화려한데요)

² We follow the Romanization system of Korean, hereafter.

We can build translation model and language model for SPE using this corpus composed of Korean (by RBMT) - Korean (by human). For building models for SPE, the tools for SMT are used at the same way. These models are used to post-edit RBMT results.

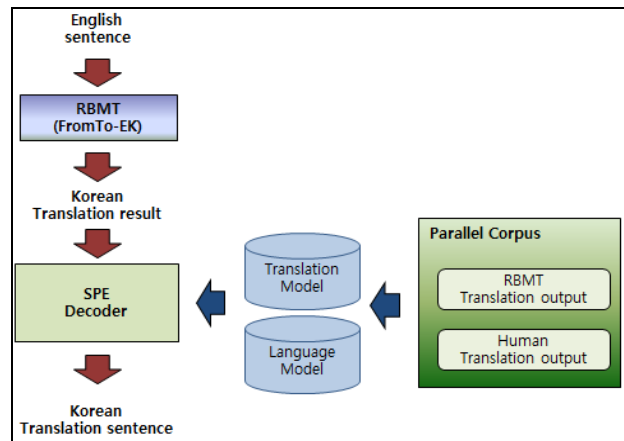


Figure 2. Statistical post-editing architecture

Figure 2 shows the configuration of SPE. We thought that by training based on alignment between incorrect machine translation result and correct human translation result, knowledge to improve the fluency of RBMT can be obtained. In applying SPE to RBMT system, SPE module is located at last place which take the translation result by RBMT as input. In figure 3, translation model and language model means statistical post-editing models which are thought in the concept of statistical machine translation. We use the same tools as SMT for getting SPE models.

4 Experiments

4.1 Setup

Table 5 explains 2 types of training corpora, tuning corpus and test set. The domain for our experiments is tour/travel. For considering the property of Korean, we prepared 2 training corpora. The training_corpus_s is built from Korean surface forms of words in sentences. The training_corpus_m is obtained from Korean morpheme through POS tagging. We tested on translating spoken language test set belonging to tour domain from English to Korean.

Corpus	# Sentences
Training_Corpus_s	1,082 K
Training_Corpus_m	1,082 K
Tuning Set	1 K
Test Set	200

Table 5: Training corpus for evaluation

Table 6 shows the baseline system for building statistical model for post-editing.

Moses ³ (Koehn, 2007)	Revision = “4383” as the baseline system for training and decoding
GIZA++ ⁴ (Och and Ney, 2003)	Version 1.0.5 for alignment between translation result by FromTo-EK and human translation result.
SRI LM ⁵ (Stolcke, 2002)	version 1.5.12 for building a 3-gram language model

Table 6: Baseline system for evaluation

³ <http://www.statmt.org/ Moses/>

⁴ <http://giza-pp.googlecode.com/files/giza-pp-v1.0.5.tar.gz>

⁵ <http://www.speech.sri.com/projects/srilm/>

4.2 Evaluation results

We tested the coverage of SPE using two different training corpora. Table 7 describes the impact of SPE to the translation result of FromTo-EK engine. Regardless of training corpus, many sentences were changed by applying SPE.

	# sent changed	Change rate
Training_Corpus_s	154	77%
Training_Corpus_m	144	72%

Table 7: The number of sentences changed by SPE

Table 8 shows the analysis of the result of applying SPE to the result of FromTo-EK. In table 8, # sent, # imp and # deg means the number of sentences belonging to corresponding score, the number of sentences improved and degraded respectively. Table 8 presents something interesting. First, the characteristic of translation target language should be considered for better SPE performance. In this paper we focused English-to-Korean translation. Korean belonging to agglutinative language has distinct features that function words are well developed and these function words are closely related to fluency. So, when using parallel corpus composed of surface forms of words for training, SPE doesn’t work well. Second, the target of SPE should be defined clearly. Because SPE doesn’t consider source sentence in training phase, translation results under 2.5, including already analysis errors, don’t have the advantage of SPE. This means that it is difficult for SPE to fix errors occurred at analysis phase. Third, translation results over 3.5 are apt to be degraded by SPE, contrary to expectation. This is related to reordering. We explain this with Table 9.

Target sent	# sent	Training_Corpus_s		Training_Corpus_m	
		# imp	# deg	# imp	# deg
over 3.5	96	9	36	7	10
Over 2.5 and under 3.5	83	13	24	29	21
Under 2.5	21	0	0	5	0
Total	200	22	60	41	31

Table 8: The improvement by SPE

Based on experiment with training_corpus_m, we categorized the operation of SPE and had a look at the effect of each operation. Table 9 provides detail figures of SPE for sentences over score 2.5.

Category		# imp	# deg	# imp / # deg
Word change	Function word	9	4	2.25
	Noun	12	11	1.09
	Verb	18	9	2.0
	Adjective	2	0	
	Adverb	1	1	1.0
	Compound word	8	2	4.0
Reordering		2	13	0.15
Subject omission		9	0	

Table 9: The effect of SPE operations for sentences over 2.5

In table 9, “word change” includes insertion, change and deletion. In FromTo-EK engine, function word is generated using just dictionaries and transfer patterns without consideration for rich context. This leads FromTo-EK into its own fluency problems, so does word of other part-of-speech. However, we can know that these kinds of problems can be improved by SPE. In Korean spoken language, “subject omission” is frequent and SPE reflects it well.

However, reordering raises serious side effects. Reordering frequently makes correct dependency relation incorrect. How to minimize the side effects of reordering is also important issue.

There are some examples of the result of applying SPE in Table 10. By SPE, we can see that word change including word insertion and word deletion occurred and fluency is improved. Final translation result by SPE is very good. The literary style expressions appeared in RBMT results are changed to colloquial style by SPE. And many abbreviation forms which are used in spoken language are introduced by SPE. Conventional RBMT cannot get such fluent sentences just using analysis rules, transfer patterns so on.

5 Discussion

We surveyed the effect of statistical post-editing method applied to English-to-Korean rule-based

machine translation system. There still remain some problems to be solved necessarily:

- How to select the sentences and words which could be improved by post-editing? It is needed to devise method to select the target to apply post-editing with minimizing degradation of translation quality. We are considering some language models for resolving these problems.
- Is there a method to apply post-editing for the part of sentence partially? If it is, how to find the phrase or clause to apply post-editing? This problem is similar to first problem to discuss. Up to now, table 9 shows word change category is best target of SPE. However, it is necessary to extend the boundaries of post-editing.
- To improve performance, how to use other information such as source language information? Proposed method considers the alignment data only in target language aspects. If source language clues are used in combination with target language alignment data, the accuracy of post-editing will be improved.
- How to implement RBMT system with SPE module? It is the same as the structure which employs two translation engines sequentially. So its feasibility in application domains should be considered.

6 Conclusion

We proposed how to make use of SPE to improve the fluency of RBMT system. We could see that to maximize the effect of SPE, the consideration on target language has to be preceded. We categorized the factors which can increase the quality of RBMT by SPE. In the case of FromTo-EK system, a kind of RBMT, we knew that SPE works effectively for the sentences in which morphological or syntactic analysis errors did not occur. So, our next research topic is how to decide the target sentences/phrases/words of SPE. And, utilizing the source sentence and its related information (source word, its POS and its syntactic role) is also important research topic to effect SPE performance.

English sentence	FromTo-EK result	score	SPE result	score
What do I have to wrap it with?	Mueoseuro geugeoseul ssayahamnikka? (무엇으로 그것을 싸야 합니까?)	3.83	Mworo pojanghalkkayo? (뭘로 포장할까요?)	4.0
It closes at 11 o'clock at night.	Geugeoseun bame dacheo itsseumnida. (그것은 밤에 11시에 닫혀 있습니다.)	3.33	Bam 11sie kkeunnamnida. (밤 11시에 끝납니다.)	4.0
They depend on the freshness.	Geudeureun sinseonmie uijonhanda. (그들은 신선미 에 의존한다.)	3.0	Geudeureun sinseonmie ddara dareuda. (그들은 신선미 에 따라 다르다.)	3.5
The restroom door on the first floor is locked.	Icheung wiui hwajangsil muneun jamgyeojimnida. (1층 위의 화장실 문은 잠겨집니다.)	2.83	Icheunge itneun hwajangsil muni jamgyeotseumnida. (1층에 있는 화장실 문이 잠겼습니다.)	4.0

Table 10: Example of SPE result

References

- Yun Jin et al. 2008. The Trends of Machine Translation Technology and Case Study. Electronics and Telecommunications Trends, vol. 23, no. 1, pp.89-98.
- Philipp Koehn et al. 2003. Statistical Phrase Based Translation. In Proc. of the HLT/NAACL.
- Seong Il Yang, Young Ae Seo, Young Kil Kim, and Dongyul Ra. 2010. Noun Sense Identification of Korean Nominal Compounds Based on Sentential Form Recovery, ETRI Journal, vol.32, no.5, pp.740-749.
- J. Doyon, K. Taylor and J. White. 1998. The DARPA machine translation evaluation methodology. Proc. of AMTA-98.
- Philipp Koehn, 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, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. the ACL '07 Demo-Poster, pp.177 - 180.
- Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. Computational Linguistics, 2003, pp.19 - 51.
- Andreas Stolcke. 2002. Srlm - an extensible language modeling toolkit. Proc. of ICSLP, pp.901 - 904.