Toshiba MT System Description for the WAT2014 Workshop

Satoshi Sonoh, Satoshi Kinoshita, Hiroyuki Tanaka and Satoshi Kamatani

Knowledge Media Laboratory, Corporate Research & Development Center,
Toshiba Corporation.
1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki 212-8582, JAPAN
`satoshi.sonoo@toshiba.co.jp`

Abstract

This paper provides a system description of Toshiba Machine Translation System for WAT2014. We participated in two tasks, namely Japanese-English translation and Japanese-Chinese translation. In each task, we submitted two results; one is a result of a rule-based translation system, and the other is a result which is an output of statistical post editing trained with the ASPEC training corpora. In both tasks, output by statistical post editing shows improvement in machine evaluation, but we obtained different results from human evaluation.

1 Introduction

Toshiba has been developing a Rule-Based Machine Translation (RBMT) system. Recently, open parallel corpora such as ASPEC[1] are available for development of machine translation systems. However, it is extremely high cost to develop a large volume of parallel corpora in wide-domain for commercial use. For this reason, we have developed translation functions by using a monolingual corpus in the target language in order to improve performance. For example, a target word selection is possible based on co-occurrence relationship extracted from a monolingual corpus (Suzuki and Kumano, 2005). Moreover, we tackled a domain adaptation by using a parallel corpus and proposed a system which consists of RBMT and statistical post editing (SPE) (Suzuki, 2011). Furthermore, we have developed a word sense disambiguation based on a monolingual corpus in the target domain, and it has been applied to Japanese-Korean/Korean-Japanese translation system (Kumano, 2012).

In this evaluation task, we submitted two types of results for each Japanese-English translation and Japanese-Chinese translation (Nakazawa et al., 2014). One is our RBMT system which is developed for translating open domain written texts. The other is a combination of RBMT and SPE which is trained by the target domain corpus.

Section 2 describes the overview of our system. The preparation and results for the task are shown in Section 3 and 4. A context-aware translation is discussed in Section 5. Finally, Section 6 concludes this paper.

2 Overview of Toshiba System

2.1 Baseline System

Our system realizes transfer-based machine translation (Izuha et al., 2008). The core framework consists of morphological analysis, syntactic/semantic analysis, target word selection, structural transfer, syntactic generation and morphological generation. Furthermore, huge amount of rules as translation knowledge including word dictionaries can realize both high translation performance and flexibility of customization.

2.2 Statistical Post Editing

Although RBMT may preserve a certain level of adequacy in expected domain such as a manual document, translation knowledge for target word selection is not enough for a specific domain. Moreover, fluency for RBMT output is typically low in comparison with Statistical Machine Translation (SMT).

Statistical post editing by using phrase-based SMT has been proposed and it is an efficient method which is able to adapt translation output to the target domain (Simard et al., 2007). The combination of RBMT and SPE has the potential to improve the translation quality without additional rules (Simard et al., 2007).

---

[1] http://lotus.kuee.kyoto-u.ac.jp/ASPEC/

In the evaluation task, we realized SPE based on phrase-based SMT (Kohen et al., 2003) by using the training corpus (1.0M sentences for English, 0.67M for Chinese) and its translation output by our RBMT system. We used Moses tools (Kohen et al., 2007) for English tokenization, and Kytea (Neubig et al., 2011) for Chinese.

3 Preparation for the Task

3.1 Selecting Technical Term Dictionaries

As a preparation for the evaluation task, we selected technical term dictionaries which improve the baseline system. Toshiba provides customers with technical term dictionaries so that they can get better translation in their target domains. Table 1 shows a list of technical term dictionaries, which contain about 3 million words and 460 thousand words in total for Japanese-English and Japanese-Chinese translation systems respectively. But using all the available dictionaries does not necessarily lead to better translations because even a technical term for one domain should be translated differently in another domain.

To choose the best dictionary sets, we conducted two-step evaluation using the devset of ASPEC corpus. The first step is to find which technical term dictionary is useful or not by comparing the BLEU score with that of the baseline system[2]. The second step is to decide the best combination of dictionaries which are judged useful in the first step by adding them to the baseline system one by one as long as the addition gives a higher BLEU score.

In Japanese-English translation, 10 out of 18 technical term dictionaries provide better translations than the baseline system as shown in Table 2(a). In the second step, we started from adding two dictionaries, "namely Chemistry" and "Medicine/Pharmacology 2", and found that adding 4 technical dictionaries gave the best BLEU score as shown in Table 2(b).

In Japanese-Chinese translation, the first step showed that only two dictionaries provided better translations than the baseline system. We confirmed that this combination provides better result than those obtained in the previous step as shown in Table 2(c) and (d).

JE	JC	dictionary name
x	-	Architecture/Civil Engineering
x	x	Auto
x	x	Biology
x	x	Business
x	x	Chemistry
x	x	Electricity/Electronics
x	x	Finance/Law
x	x	Industrial Technology
x	x	Information/Telecommunication
x	x	Internet
x	-	Japanese Scientific Terms
x	-	JIS Terms
x	-	Math/Physics
x	x	Mechanics
x	-	Medicine (Basic/Clinical）
x	-	Medicine/Pharmacology (Regulatory terms）
x	x	Medicine/Pharmacology 1
x	x	Medicine/Pharmacology 2
x	-	Military/Defense
x	-	Natural Science
x	-	Natural/Social Sciences
x	-	Nichigai/ Science and Technology (Basic Science）
x	-	Nichigai/Medicine (Basic)
x	-	Nichigai/Medicine (Related fields)
x	-	Nichigai/Science and Technology (Engineering）
x	-	Nichigai/Science and Technology (Science/Medicine）
x	-	Nuclear Studies
x	-	Production/Quality
x	x	Proper Nouns (standard)
x	-	Proper Nouns (technical)
x	-	Textile/Garment
x	x	Transportation/Logistics

Table 1: List of technical term dictionaries.

---

[2]The baseline system used technical dictionaries ("Proper noun" and "Internet") as default setting. For Japanese-English, "Natural/Social Sciences" is also used because the target domain is scientific papers.

ID	dictionary name	BLUE
1	Chemistry	16.04
2	Medicine/Pharmacology 2	15.95
3	Nichigai/Science and Technology (Science/Medicine）	15.93
4	Medicine (Basic/Clinical）	15.92
5	Natural/Social Sciences	15.89
6	Nichigai/Medicine (Basic)	15.88
7	Biology	15.88
8	Math/Physics	15.85
9	Information/Telecommunication	15.84
10	Nichigai/Medicine (Related fields)	15.81
	Baseline	15.47

(a) Result of step 1 for JE

dictionary ID	BLUE
1+2	16.11
1+2+3	16.15
1+2+3+4	**16.21**

(b) Result of step 2 for JE

ID	dictionary name	BLUE
1	Chemistry	18.89
2	Biology	18.86
	Baseline	18.73

(c) Result of step 1 for JC

dictionary ID	BLUE
1+2	**18.92**

(d) Result of step 2 for JC

Table 2: BLEU scores after adding technical term dictionaries.

3.2 Building Additional Dictionaries

To improve the performance of the baseline system, we have built additional dictionaries for Japanese-Chinese system by using the given training data. We first built a SMT system by using the standard Moses system, and extracted word pairs whose co-occurrence ratio is more than 0.7 from the phrase table. By using sentence pairs in the training corpus, we got about 10,000 word pairs as dictionary entry candidates. After filtering noisy results, we checked them manually and got about 5,400 terms for this task. By using this dictionary additionally, BLEU score for the development data was improved from 18.92 to 18.98. For Japanese-English system, we have built no additional dictionary.

Because the effect of the additional dictionary was little, finally, we used only selected technical dictionaries, and not use the additional dictionary.

4 Results

This section shows the evaluation results of our systems. We analyzed the relationship between machine evaluation and human evaluation. In addition to the human evaluation provided by WAT 2014, we evaluated the output by relative ranking of 1(best) to 3(worse) for the three systems including RBMT, RBMT with SPE and phrase-based SMT.

4.1 Japanese-English results

Table 3 summarizes the evaluation results of Japanese-English translation systems. In the machine evaluation, BLUE score for our tuned RBMT is 15.69 and 20.61 for RBMT with SPE, respectively. Adopting SPE achieved an improvement of 31.4% in BLUE score.

System	BLUE	RIBES	HUMAN	RANK
RBMT	15.69	0.69	20.25	**1.38**
+SPE	**20.61**	**0.71**	**23.25**	1.45
SMT	18.45	0.65	-	2.24

Table 3: Japanese-English evaluation results.

In the human evaluation, HUMAN score for RBMT with SPE is 3 points higher than RBMT. However, RANK score, which is the average ranks of 200 setences in testset (not same as HUMAN-set), shows a different result from crowdsourcing evaluation. In this regard, we have found out the differences between RBMT and RBMT with SPE as shown in Table 4. Although SPE surely achieved improvements of target word selection, there are some mistranslations such as deletion of word and disagreement of tense in SPE output. As a result, RBMT may get a better result than RBMT with SPE in the RANK evaluation.

4.2 Japanese-Chinese results

Table 4 summarizes the evaluation results of Japanese-Chinese translation systems. In Japanese-Chinese, BLUE score for our tuned RBMT is 19.28 and 27.42 for RBMT with SPE, respectively. A ratio of improvement is larger than JE results, and is 42.2%.

In the human evaluation, both HUMAN and RANK scores show that RBMT with SPE performs better than RBMT. Although results are

System	BLUE	RIBES	HUMAN	RANK
RBMT	19.28	0.76	-5.25	2.13
+SPE	27.42	0.80	**0.75**	1.76
SMT	**27.96**	0.79	-	**1.63**

Table 4: Japanese-Chinese evaluation results.

partially conflicting between machine and human evaluations, these results mean that RBMT with SPE has an equivalent performance of SMT system.

In fact, improvements for target word selection, out of vocabulary words and structural transfer are realized by SPE that complemented lack of rules in the target domain, as shown in Table 6. Meanwhile, mistranslations caused by statistical features also occurred in SPE output.

5 Issues for Context-aware Machine Translation

Our translation system has the following context-aware features, which have been used in our commercial systems (Yoshimura et al., 2011).

- Target word selection based on domain estimation

- Part-of-speech disambiguation and the reuse of translations by using preceding context

Because our system is a rule-based system, semantic dependencies between source words are used as key clues for choosing appropriate target words. Context-aware functions are used when such clue is not available. First is a function to choose a target word based on a domain of the source sentences. Some of the words in translation dictionaries are assigned domain labels, and the translation system tries to decide a domain of input sentence in morphological analysis. For example, in the case of English-Japanese translation, an English word "court" is given a Japanese translation "コート" by default, but it is translated as "法廷" if the current domain is estimated as a legal domain.

Second is used to choose a word of an appropriate part-of-speech, and to output consistent translations for proper nouns. When the translation system finds a proper noun which consists of more than one word, it memorizes a pair of source and target word, each of which is a constituent of the original proper noun. For example,

if "John Snow" is in a translation dictionary and translated as "ジョン・スノー", a memorized word "Snow" will be preferably analyzed as a proper noun, and be translated as "スノー" in later context. If a sequence of words is recognized as a proper noun by named entity recognition, its constituent is treated in the same manner.

We couldn't confirm the effectiveness of the above functions in the translations of the test set due to time constraints. To get better translations, we'd like to utilize the context to make syntactic and semantic analysis more accurate.

6 Conclusion

The overview of Toshiba rule-based machine translation system with statistical post editing for scientific paper translation task is described in this paper.

A combination of RBMT and SPE achieved improvements of BLUE score in both Japanese-English and Japanese-Chinese translation. In contrast, in a part of the human evaluation, RBMT showed better performance than SPE for Japanese-English translation.

References

Hirokazu Suzuki and Akira Kumano. 2005. Learning Translations from Monolingual Corpora. In *Proceedings of MT Summit X*.

Hirokazu Suzuki. 2011. Automatic Post-Editing based on SMT and its selective application by Sentence-Level Automatic Quality Evaluation. In *Proceedings of MT Summit XIII*.

Akira Kumano. 2013. Korean Translation System for Patent Documents. In *Japio YEAR BOOK*, pages 298-301. (In Japanese)

Toshiaki Nakazawa, Hideya Mino, Isao Goto, Sadao Kurohashi and Eiichiro Sumita. 2014. Overview of the 1st Workshop on Asian Translation. IN *Proceedings of the 1st Workshop on Asian Translation (WAT2014)*.

Tatsuya Izuha, Akira Kumano and Yuka Kuroda. 2008. Toshiba Rule-Based Translation System at NTCIR-7 PAT MT. In *Proceedings of NTCIR-7 Workshop Meeting*, pages 430-434.

Michel Simard, Cyril Goutte and Pierre Isabell. 2007. Statistical Phrase-based Post-editing. In *Proceedings of NAACL HLT 2007, ACL*, pages 508-515.

Michel Simard, Nicola Ueffing, Pierre Isabelle and Roland Kuhn. 2007. Rule-based Translation With

Statistical Phrasebased Post-editing. In *Proceedings of the second Workshop on Statistical Machine Translation, ACL*, pages 203-206.

Philipp Kohen, Franz Josef Och and Daniel Marcu. 2003. Statistical Phrase-Based Translation. In *Proceedings of NAACL HLT*, pages 127-133.

Philipp Kohen, Marcell Federuci, Brooke Cowan, Richard Zens, Chris Dyer, Ondrej Bojar Alexandra Constantin and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the ACL*, pages 177-180.

Graham Neubig, Yosuke Nakata and Shinsuke Mori. 2011. Pointwise Prediction for Robust, Adaptable Japanese Morphological Analysis. In *Proceedings of ACL-HLT*.

Yumiko Yoshimura, Satoshi Kinoshita and Miwako Shimazu. 1997. Processing of Proper Nouns and Use of Estimated Subject Area for Web Page Translation. In *Proceedings of TMI-97: The Seventh International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 240-251.

SRC	そこで，流体の性質や条件の違いにより適切なセンサを選択することが必要である。
REF	Then, it is necessary to choose the appropriate sensor in accordance with a difference in the property and condition of the fluid.
RBMT	Then, it is required to choose a suitable sensor by the character of a fluid or the difference among conditions.
+SPE	Then, it is necessary to choose the appropriate sensor by the properties of the fluid and the difference between the conditions.

(a) Improvement of target word selection.

SRC	りん酸基をもつ界面鋳型樹脂によるＣｕ（ＩＩ）の吸着量はｐＨの増大に伴い増大した。
REF	The amount of Cu(II) adsorption by the interface template resin with phosphate group increases with an increase in pH value.
RBMT	The amount of adsorption of Cu(II) by interface mold resin with a phosphoric acid group increased with increase of pH.
+SPE	The adsorption amount of Cu (II) by the mold resin with phosphoric acid group increased with increasing pH.

(b) Mistranslation of noun phrase.

SRC	標題光スイッチングと光記録画像を形成し，その安定性を調べた。
REF	The optical switching of the title and its optical recording image were formed, and the stability was examined.
RBMT	Title optical switching and an optical record picture were formed, and the stability was investigated.
+SPE	Titled optical switching and the optical recording images were formed, and the stability is investigated.

(c) Mistranslation of tense agreement.

Table 5: A comparison between RBMT and RBMT with SPE in Japanese-English translation.

SRC	またＸＭＬの処理の時間も無視できず，実装上の工夫が必要であろう.
REF	此外，不能忽ＸＭＬ的理，在安装上也需要下功夫。
RBMT	外不能也无ＸＭＬ的理的，要需要上的法。
+SPE	外，也不能忽了ＸＭＬ的理，需要安装上下功夫。

(a) Improvement of target word selection.

SRC	ポリエーテルスルホン（ＰＥＳ）非対称ＵＦ膜は高い水浸透率と強度を持つ。
REF	聚（ＰＥＳ）非称ＵＦ膜具有高的水浸透率和度。
RBMT	聚suluhon(PES)不称UF膜有高的水渗透率和度。
+SPE	聚（ＰＥＳ）不称ＵＦ膜具有高的水的渗透率和度。

(b) Improvement of out of vocabulary words.

SRC	図１（ａ）にＪＣＰのチャネル構成を示す.
REF	１（ａ）所示ＪＣＰ的道。
RBMT	在１（ａ）表示ＪＣＰ的成。
+SPE	１（ａ）所示，ＪＣＰ的道成。

(c) Improvement of structural transfer.

SRC	最低メタンロスはメタン産生量の２％へ低下した。
REF	最低甲浪降低到了甲生量的２％以下。
RBMT	最低甲失甲生成量的２％下降了。
+SPE	最低甲失甲生量的２％。

(d) Mistranslation of verb-phrase.

Table 6: A comparison between RBMT and RBMT with SPE in Japanese-Chinese translation.