

The System Combination RWTH Aachen - SYSTRAN for the NTCIR-10 PatentMT Evaluation

Minwei Feng
Lehrstuhl für Informatik 6
RWTH Aachen University
Ahornstr. 55
D-52056 Aachen, Germany
feng@cs.rwth-aachen.de

Markus Freitag
Lehrstuhl für Informatik 6
RWTH Aachen University
Ahornstr. 55
D-52056 Aachen, Germany
freitag@cs.rwth-aachen.de

Hermann Ney
Lehrstuhl für Informatik 6
RWTH Aachen University
Ahornstr. 55
D-52056 Aachen, Germany
ney@cs.rwth-aachen.de

Bianka Buschbeck
SYSTRAN S.A.
5 rue Feydeau
75002 Paris, France
buschbeck@systran.fr

Jean Senellart
SYSTRAN S.A.
5 rue Feydeau
75002 Paris, France
senellart@systran.fr

Jin Yang
SYSTRAN Software, Inc.
4444 Eastgate Mall, Suite 310
San Diego, CA 92121, USA
jyang@systransoft.com

ABSTRACT

This paper describes the joint submission by RWTH Aachen University and SYSTRAN in the Chinese-English Patent Machine Translation Task at the 10th NTCIR Workshop. We specify the statistical systems developed by RWTH Aachen University and the hybrid machine translation systems developed by SYSTRAN. We apply RWTH Aachen's combination techniques to create consensus hypotheses from very different systems: phrase-based and hierarchical SMT, rule-based MT (RBMT) and MT with statistical post-editing (SPE). The system combination was ranked second in BLEU and second in the human adequacy evaluation in this competition.

Categories and Subject Descriptors

I.2.7 [Nature Language Processing]: machine translation

General Terms

Experimentation

Keywords

SMT, Patent Translation, Hybrid Machine Translation, Statistical Post-Editing, Rule-Based Machine Translation

Team Name

RWTH Aachen - SYSTRAN

Subtasks/Languages

Chinese-to-English PatentMT

External Resources Used

SYSTRAN's Enterprise Server 7, GIZA++

1. INTRODUCTION

RWTH Aachen University and SYSTRAN jointly participate in the Chinese-English Patent Translation task of the 10th NTCIR

Workshop [6]. This paper describes the statistical machine translation (SMT) systems developed by RWTH Aachen University and SYSTRAN. The structure of the paper is as follows: in Section 2, we describe the translation systems used in the evaluation, including phrase-based and hierarchical SMT systems, rule-based MT and MT with statistical post-editing. Section specifies the system combination framework developed by RWTH Aachen. Experimental results are given in Section 4. Finally, we draw some conclusions in Section 5.

2. TRANSLATION SYSTEMS

For the NTCIR-10 Patent Translation evaluation we combined RWTH's state-of-the-art phrase-based and hierarchical translation systems with SYSTRAN's rule-based and hybrid machine translation systems. The RWTH Aachen's systems have been described in [5]. In this section, we focus on SYSTRAN's systems. All systems were evaluated using the automatic BLEU [11] and TER [15] metrics.

2.1 SYSTRAN's Rule-Based Machine Translation

This section describes SYSTRAN's Chinese-English rule-based machine translation system that participated in the Patent Machine Translation Task at NTCIR-10 in a system combination with the RWTH Aachen.

The SYSTRAN system is traditionally classified as a *rule-based* system using enormous and diversified linguistic resources. A unified and highly modular architecture applies to all language-pair systems. SYSTRAN's dictionaries and parsers have evolved over a long period of time, have been tested on large amounts of text, and contain extremely detailed linguistic rules and a large terminology database covering various domains.

The development of the SYSTRAN Chinese-English MT system started about a decade ago. Work on lexical development and linguistic analysis have been continuing over the years, with steady improvement. Recent development concentrates on integrating statistical techniques in the various components of the system:

- Corpus-based monolingual and bilingual terminology extraction [4]

Source term	Target term	Frequency	Confidence
实施方案	embodiment	784	0.77
通信系统	communication system	535	0.79
反应混合物	reaction mixture	464	0.89
专利申请	patent application	362	0.85
通信设备	communication device	267	0.72
权利要求	claim	244	0.73
移动终端	mobile terminal	192	0.83
涂料组合物	coating composition	188	0.87
核苷酸序列	nucleotide sequence	161	0.79
数据分组	data packet	157	0.82
数据结构	data structure	149	0.91
移动设备	mobile device	136	0.85

Table 1: Examples of Bilingual Terminology Extraction

- Incorporating corpus evidence in the linguistic rules [13]
- Introducing statistical components to help make difficult linguistic decisions [12]
- Dynamic domain adaptation that goes along with continued extension of domain vocabulary
- Entity recognition and statistical transliteration of named entities

The current RBMT Chinese-English system contains over 2.5 million bilingual words, expressions, and linguistic rules spanning various domains.

2.2 Customization of Rule-Based Machine Translation

Due to the particularities of patent documents their translation is a challenging task for any MT system. Long and complex sentences, large amounts of technical terminology and unknown words, a high degree of ambiguous words and cross references throughout the patent description complicate the translation task.

To better deal with these characteristics we first applied a domain adaptation strategy. SYSTRAN's translation technology is easy to customize. It provides the means to quickly improve the translation quality for a given domain. In particular we applied the following methods:

- Selection of most appropriate SYSTRAN configuration for patent translation (use of domain dictionaries for life science, IT, chemistry, technology and medicine, setting of style parameters for technical documentation)
- Bilingual Terminology Extraction (BTE) on a randomly selected 200K subset of the training data provided for the Chinese-English Patent Translation task, see the Table 1 for some examples. The resulting dictionary is used in the translation process.
- Work on high-frequent patterns containing numeric expressions, brackets and slashes exploited through the SYSTRAN entity module
- Integration of a dictionary treating the translation of sentence-initial adverbial phrases and textual cross-references
- Creation of a word sense disambiguation (WSD) model to improve the translation of highly ambiguous words

Below you find an example that shows some of the improvements coming from the addition of the WSD model (translation of the ambiguous noun 组合物 and verb 包含) and the treatment of text references.

治疗性组合物还可包含上文描述的基因递送系统。

The medical **fabric** may also **encompass** the gene **of preceding text description** to deliver the system.

The medical **composition** may also **include** the gene **described above** to deliver the system.

The methods listed above helped to improve the base-line translation quality (BASE) as shown in Table 2. The dictionary resulting from bilingual terminology extraction (CUST-1) increases BLEU by three points which is quite considerable for a dictionary. Adding the other linguistic components to configuration CUST-2 (word sense disambiguation and treatment of various kinds of patterns) leads to more improvement.

Based on the translations of the customized rule-based systems, described in this section, we performed *statistical post editing* with the provided bilingual and monolingual training corpora, see 2.4.

2.3 SYSTRAN's Hybrid Machine Translation

For the last several years, SYSTRAN has been focusing on the introduction of statistical approaches to its rule-based backbone, leading to *Hybrid Machine Translation*. Our hybrid Chinese-English systems participated in the CWMT 2009 and CWMT 2011 evaluations, ranking first in BLEU at CWMT11 [16] and [17].

The technique of *Statistical Post-Editing* [3] and [14] is used to automatically edit the output of the rule-based system. A Statistical Post-Editing (SPE) module is generated from a bilingual corpus. It is in principal a translation module by itself, but it is trained on rule-based translations and reference data. It performs corrections and adaptations learned from a phrase-based 5-gram language model.

Using this two-step process will implicitly keep long distance relations and other constraints decided by the rule-based system while significantly improving phrasal fluency. It combines the predictability and language consistency of rule-based MT with the fluency and flexibility of statistical MT. It has the advantage that quality improvements can be achieved with very little but targeted bilingual data, thus significantly reducing training time and increasing translation performance. SYSTRAN has incorporated the hybrid technology into its product - SYSTRAN Enterprise Server 7.

RBMT Type	Customization	BLEU	GTM	TER
BASELINE		12.3	37.0	71.0
CUST-1	BTE	15.4	39.9	66.8
CUST-2	BTE + patterns + WSD	16.4	40.4	65.3

Table 2: Improvements of rule-based MT from domain adaption measured on the NTCIR-9 test set

RBMT Type	Method	BLEU	GTM	TER
BASELINE	SPE	26.9	47.5	56.9
CUST-1	SPE	27.4	47.9	56.5
CUST-2	SPE	27.5	48.0	56.3

Table 3: Statistical Post-Editing based on different levels of rule-based MT domain adaption

RBMT Type	Method	BLEU	GTM	TER
CUST-2	SPE	27.5	48.0	56.3
CUST-2	SPE with LM	29.1	48.8	55.6
CUST-2	SPE with LM no pruning	29.5	49.1	54.9

Table 4: Statistical Post-Editing based on best-performing rule-based MT domain adaption

2.4 Statistical Post-Editing

2.4.1 Corpus

For the training of the SPE model we used the 1,000,000 segment Chinese-English patent corpus, for tuning - the 2000-segment development set provided by the organizer for this subtask. As test data the NTCIR-9 evaluation data was used.

2.4.2 System Setup

We trained a bidirectional phrase alignment table and trimmed it [7] to suppress all unique phrase pairs before calculating the probabilities for the final phrase table. The training was conducted in lower-case. The Chinese tokens were segmented by words (not by characters) using the SYSTRAN translation engine [18].

To the target side of the model we added a second monolingual language model built from the monolingual English patent data. We only used the 2005 patents from the United States Patent and Trademark Office *us2005*, segmented them and randomly selected 10 billion sentences. The language model (LM) is a standard 5-gram LM with interpolated modified Kneser-Ney smoothing [8]. Weights for these separate models were tuned by the MERT [10] algorithm provided in the Moses toolkit [9], using the provided development set. The distortion limit for reordering was set to 6 to deal with differences in word order. Moreover, the following measures - limiting unwanted statistical effects - were applied:

- Numeric expressions were replaced by special tokens on both sides. This usually improves word alignment, since the vocabulary size is reduced. After the translation process, the special tokens are replaced by their original values. Entity translation is handled more reliably by the rule-based engine.
- Phrase pairs not containing the same number of these special tokens on the source and the target side were discarded.
- The intersection of both vocabularies (i.e. vocabularies of the rule-based output and the reference translation) was used to produce an additional parallel corpus to help to improve word alignment.

- Singleton phrase pairs were deleted from the phrase table to avoid overfitting.

In the experiments described below, we used Moses for decoding, GIZA++ for word alignment, and the SRILM tool kit for language modeling. The model tuning was done using Minimal Error Rate Training (MERT) with BLEU4-SBP using the development sets provided by the NTCIR-10 organizer.

2.4.3 Results

First, we built SPE systems based on the different quality levels of the rule-based engine, as described in Section 2.2 with an identical training set-up. The result of three of these runs are presented in Table 3. It shows that the SPE system performance is related to the RBMT performance. The smaller the distance between RBMT and the reference translation the higher the gain of the SPE system. This observation is largely supported by our experience with the statistical post-edition technique, especially if little training data is available. However, customization of RBMT is still rather expensive in comparison to purely statistical processing. If we have large bilingual corpora at our disposal RBMT customization effects become less visible, since the statistical model covers most of the rule-based improvements.

In a second step, we continued experimenting only with the CUST-2 RBMT configuration. Table 4 shows the results. The addition of a language model created from 10 billion sentences of the English 2005 patent data helped to improve the performance of the SPE model by 1.6 BLEU points on the NTCIR-9 baseline. Without phrase-table pruning the gain would even be higher (2 BLEU points), but to avoid over-correction on unseen data we used the pruned SPE model for the system combination.

3. SYSTEM COMBINATION

In the field of machine translation, there are many different approaches for generating a translation from a given source sentence. Mainly you can divide the approaches into two categories: statistical based or rule based methods. Within the statistical meth-

ods, there are quite different approaches and each of them has its own advantages and disadvantages. The most important aspect for system combination is: the hypotheses from different systems can differ widely. System combination combines different translations produced by various machine translation systems. In general system combination engines combine different machine translation hypotheses to generate a better translation which has the most overlap with all individual system's hypotheses.

System combination can be divided in two steps. The first step is to align the words of the different system output together, i.e. to create a word alignment between the different hypotheses. The second step is decoding which generates the system combination output with the help of the previously calculated alignments. In the following section we will give a short overview of the two steps.

3.1 Alignment

Different to system selection, in system combination we can produce a sentence combined of different words from different input hypotheses. In order to select words from different hypotheses, a pairwise alignment between all hypotheses is needed. We need to choose one hypothesis as the so-called primary hypothesis. This primary system provides the word order and all remaining hypotheses will be aligned to it. In other words, the primary hypothesis is responsible for the word order. After reordering all other hypotheses to the word order of the primary hypothesis, we build a confusion network of strings in the target language. Now, we can not only extract the original primary hypothesis from the confusion network but also switch words from the primary hypothesis to some words from other hypothesis (also the empty word) or even insert new words or phrases between two words.

To generate a meaningful confusion network, we should generate an alignment which only allows to switch between words which are synonyms, misspellings, stems or on a higher level paraphrases of the primary words. In this work we use METEOR [2] which is designed to reorder a translation for scoring. METEOR has high precision as it only relies on exact words, synonyms, stems or paraphrases. Although the precision is very high, METEOR has a low recall. Synonyms which are not in the METEOR database or punctuation marks like "!" and "?" are not aligned to each other. That is the reason why we add to the METEOR paraphrase table entries like .! .!? or thela to give the decoder the possibility to choose between these options.

We introduce the alignment generation process with the help of one example from the NTCIR-10 dev corpus. The beginning of six individual single system translations of the 470 sentence are given in Table 5.

system 1	will contain isolated cdna
system 2	the isolated cdna library
system 3	the isolated cdna library
system 4	comprising an isolated cdna
system 5	will comprise cdna
system 6	will contain cdna

Table 5: Six different systems (from dev sentence 470).

For the alignment method, we need to choose one primary system as mentioned before. We choose system 1 as primary system and align all other systems to that one. In Table you can see the result after running METEOR. One entry alb means that word a has been aligned to word b of the primary system. As you can see the word order is given by the primary system. Symbol \$ is the empty

word and thus an entry \$lb means that no word could be aligned to the primary system word b.

\$will \$contain the\$ isolatedlisolated cdnalcdna library/\$
\$will \$contain the\$ isolatedlisolated cdnalcdna library/\$
\$will \$contain comprising\$ an\$ isolatedlisolated cdnalcdna
willwill compriselcontain \$isolated cdnalcdna
willwill containlcontain \$isolated cdnalcdna

Table 6: Alignment result after running METEOR. \$ means the empty word.

After getting the pairwise alignment information, we need to build a confusion network to give the decoder the ability to choose between words. In Table 7 the resulting network is given. As you can see, the primary system still provides the word order and the other systems provide translation options (e.g. comprise / contain) or provide new words which could be inserted (e.g. the). Words which have no alignment point to the primary system are inserted as translation option after the last aligned word. This could result in misleading translation options like "the" and "comprising". In our implementation, we include a reordering model which tries to get rid of these wrong pairs and gives these unaligned words better positions. For this example we will skip this reordering model. After building the confusion network, we could extract our first new aligned hypothesis via majority voting. For example "will" has been observed three times and the translation option "\$" (empty word) has been observed four times. The majority voting would extract "\$". As the different single systems have different performance, we assign each system a system weight which is utilized in the decision process. The last line in Table 7 is the majority vote result of our given example with a weight vector (0.1,0.3,0.3,0.1,0.1,0.1).

will	contain	\$	\$	isolated	cdna	\$
\$	\$	the	\$	isolated	cdna	library
\$	\$	the	\$	isolated	cdna	library
\$	\$	comprising	an	isolated	cdna	\$
will	comprise	\$	\$	\$	cdna	\$
will	contain	\$	\$	\$	cdna	\$
\$	\$	the	\$	isolated	cdna	library

Table 7: Majority vote on aligned words. System weights are (0.1,0.3,0.3,0.1,0.1,0.1). The last line is the system combination output.

3.2 Decoding

In this section we will describe the step from building a graph from the given alignment using the open source software openFST [1]. We will stick to the example of Section 3.1 and introduce some additional models as well as the composition of our word graph.

First, we read the alignment and build a word graph. We merge arcs which have the same label as well as the same start and end node. In this example, we use 6 different features each one is belonging to one system. The meaning of one arc labeled with "word/(0,-1,0,-1,0,0)" is that system 2 and 4 want to have "word" on this position. In Figure 1 the first three nodes of our resulting fst is given. In Figure 2 we scored the word graph with system weights (0.1,0.3,0.3,0.1,0.1,0.1) and use the shortest path algorithm to get the final hypothesis with the lowest score. In the end we add three

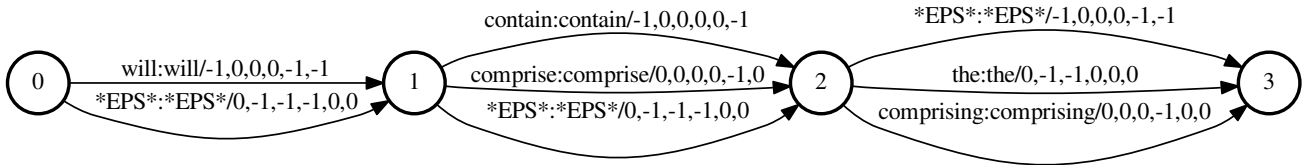


Figure 1: First 3 nodes of the word graph. A weight (a,b,c,d,e,f) means if a = -1, the word occurs in system 1.

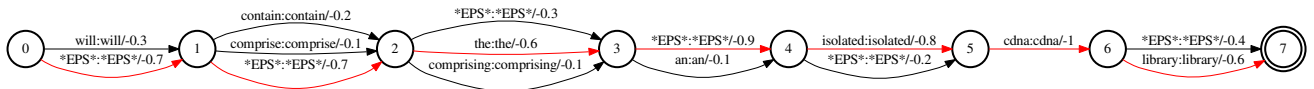


Figure 2: Scored word graph with weights (0.1,0.3,0.3,0.1,0.1,0.1). Red arc is the shortest path.

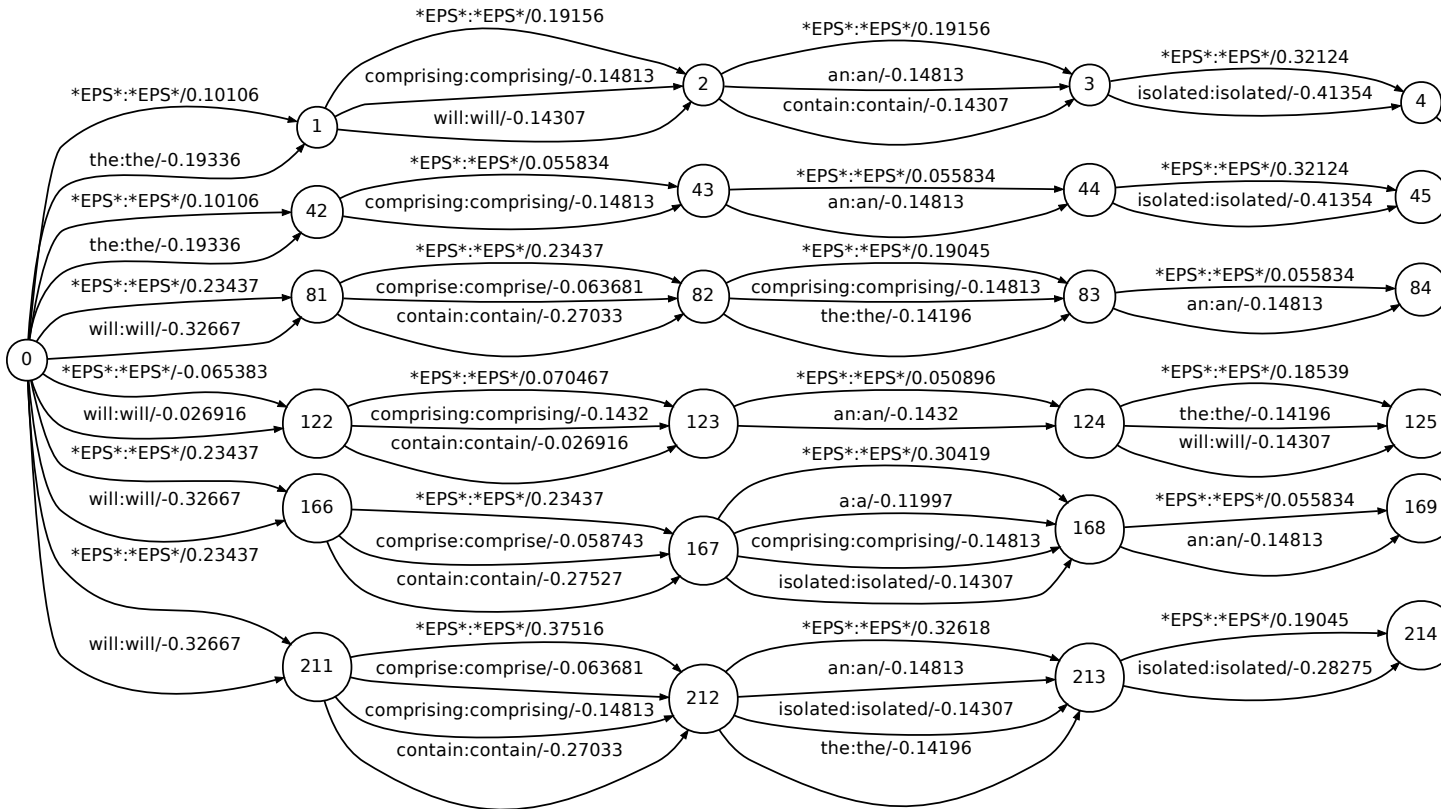


Figure 3: Union of all word graphs with different primary system. System weights are the final optimized weights.

more models. A binary primary feature which marks the arcs from the primary system, a word penalty and a language model. It can be a hard decision to choose one system as the primary one. One option could be the system with the lowest error score, but we could increase the benefit of system combination if we give the decoder the option to choose the primary system for itself. In other words, we build a word graph which includes the word order of all systems. For that we build in our example 6 different word graphs each having a different primary system. At the end we do an union over all the six word graphs. A part of the whole union word graph is given in Figure 3.

To get the best weights for the given features, we use MERT [10] as the optimization algorithm. The nbest size is set to be 200 and MERT converges after 6 or 7 iterations. Although usually the decoding speed is the bottle neck of machine translation system's efficiency, the decoding of our system combination is very fast. Hence the system combination is a fast and easy way to obtain additional improvements of hypotheses from various systems.

4. SYSTEM COMBINATION RESULTS

The results are given in Table 8. **dev** is the develop corpus provided by the NTCIR-10 and **test** is the evaluation corpus of NTCIR-9. Each of them consists of 2000 sentences with single reference. From the scores we can see that the difference between hierarchical phrase-based decoder (HPBT) and phrase-based decoder (PBT) is quite small. For the test corpus, PBT is even 0.1 BLEU and 0.1 TER better than HPBT. The results also show that the inverse hypotheses differs a lot from the normal baseline systems. For hierarchical phrase-based system, the inverse HPBT is 1.1 BLEU and 1.7 TER better than the standard HPBT. For phrase-based system, the inverse PBT is 0.6 BLEU and 0.5 TER worse than the standard PBT. The best single system is the inverse hierarchical phrase-based system (inverse HBPT). For BLEU and TER scores, statistical systems have clear advantage over the RBMT and SPE2 which are systems based on rules. The system combination improves the best single system by 1.2 BLEU and 1.1 TER.

In Table 9 we provide an example of system combination using the dev sentence 470. The table illustrates the TER calculation process. For every system's output, the error information is given like "4/26 (1 shift)" which means in order to edit the output into the reference, four operations needed and one of the operations is shift. 26 represents the reference sentence has 26 tokens. We use different colors to mark the different operations during TER scoring, namely, **substitution**, **deletion**, **insertion** and **shift**. In Table 9, system combination output needs the smallest amount of operations to be edited into the reference which includes only one **substitution**, two **deletions** and no **insertions**. The system combination has better lexicon choice, as most appropriate words already exist in the system combination output.

5. CONCLUSION

This paper describes the joint submission by RWTH Aachen University and SYSTRAN in the Chinese-English Patent Machine Translation Task at the 10th NTCIR Workshop. The submissions is a system combination of different systems by RWTH and SYSTRAN. The RWTH Aachen's phrase-based and hierarchical SMT have been described in [5]. This paper specifies SYSTRAN's rule-based MT (RBMT) and MT with statistical post-editing (SPE). We use system combination technology to unite advantages of different systems. In this work METEOR is used as alignment algorithm and the shortest path algorithm in openFST to extract the hypothesis with the lowest score. With all the efforts, the RWTH Aachen -

SYSTRAN joint submission was ranked second in BLEU and second in the human adequacy evaluation in this competition.

Acknowledgments

This work was achieved as part of the Quaero Programme, funded by OSEO, French State agency for innovation.

Systems	opt criterion	dev		test	
		BLEU	TER	BLEU	TER
HPBT	BLEU	43.0	44.1	39.8	45.3
HPBT inverse	BLEU	43.3	42.8	40.9	43.6
PBT	BLEU	42.3	44.3	39.9	45.2
PBT inverse	BLEU	42.1	44.7	39.3	45.7
RBMT	BLEU	21.4	62.5	19.7	63.5
SPE2	BLEU	30.3	54.3	28.7	55.6
system combination	BLEU	44.6	42.0	42.1	42.7

Table 8: BLEU and TER results. “system combination” is RWTH Aachen - SYSTRAN joint submission.

Single System 1 – 10/26 (0 shifts)	
hyp	the isolated cdna library containing a set of x is used as a template to amplify a polynucleotide encoding lmt , tmlh , tmla , tmabadh and bbh .
shifted hyp	the isolated cdna library containing a set of x is used as a template to amplify a polynucleotide encoding lmt , tmlh , tmla , tmabadh and bbh .
edited hyp	(D) (D) the isolated cdna library containing a set of x is used as a template to amplify a polynucleotide encoding lmt , tmlh , tmla , tmabadh and bbh .
ref	Λ including the separated cdna library pool was used as a template to amplify polynucleotide encoding lmt , tmlh , tmla , tmabadh and bbh .
Single System 2 – 16/26 (3 shifts)	
hyp	the isolated cdna library will contain a set of coding lmt , lambda , was used as template to amplify tmlh , tmla , tmabadh and bbh of the polynucleotide .
shifted hyp	the isolated cdna library will contain set of coding lambda , was used as a template to amplify polynucleotide lmt , tmlh , tmla , tmabadh and bbh of the .
edited hyp	(D) (D) the isolated cdna library will contain set of coding lambda , was used as a template to amplify polynucleotide (D) lmt , tmlh , tmla , tmabadh and bbh of the .
ref	Λ including the separated cdna library pool was used as a template to amplify polynucleotide encoding lmt , tmlh , tmla , tmabadh and bbh .
Single System 3 – 9/26 (2 shifts)	
hyp	will contain isolated cdna lambda library pool was used as a template to amplify coding lmt tmlh tmla , tmabadh and bbh polynucleotide .
shifted hyp	will contain isolated lambda cdna library pool was used as a template to amplify polynucleotide coding lmt tmlh tmla tmabadh and bbh .
edited hyp	will contain isolated lambda cdna library pool was used as a template to amplify polynucleotide coding lmt (D) tmlh (D) tmla , tmabadh and bbh .
ref	Λ including the separated cdna library pool was used as a template to amplify polynucleotide encoding lmt , tmlh , tmla , tmabadh and bbh .
Single System 4 – 8/26 (1 shifts)	
hyp	comprising an isolated cdna library pool was used as template to amplify a polynucleotide encoding lmt tmlh tmla tmabadh and bbh .
shifted hyp	comprising an isolated cdna library pool was used as a template to amplify polynucleotide encoding lmt tmlh tmla tmabadh and bbh .
edited hyp	(D) comprising an isolated cdna library pool was used as a template to amplify polynucleotide encoding lmt (D) tmlh (D) tmla (D) tmabadh and bbh .
ref	Λ including the separated cdna library pool was used as a template to amplify polynucleotide encoding lmt , tmlh , tmla , tmabadh and bbh .
Single System 5 – 13/26 (2 shifts)	
hyp	will comprise cdna library set of separation lamda serves as the template , to increase encoding lmt , tmlh , tmla , tmabadh and bbh polynucleotides .
shifted hyp	will comprise set of cdna library separation lamda serves as the template to , increase encoding lmt , tmlh , tmla , tmabadh and bbh polynucleotides .
edited hyp	will comprise set of cdna library separation lamda serves as the template to , increase encoding lmt , tmlh , tmla , tmabadh and bbh polynucleotides .
ref	Λ including the separated cdna library pool was used as a template to amplify polynucleotide encoding lmt , tmlh , tmla , tmabadh and bbh .
Single System 6 – 11/26 (1 shift)	
hyp	will contain cdna library set of discrete lambda serves as a template to increase encoding lmt , tmlh , tmla , tmabadh and bbh polynucleotides .
shifted hyp	will contain set of cdna library discrete lambda serves as a template to increase encoding lmt , tmlh , tmla , tmabadh and bbh polynucleotides .
edited hyp	will contain set of cdna library discrete lambda serves as a template to (D) increase encoding lmt , tmlh , tmla , tmabadh and bbh polynucleotides .
ref	Λ including the separated cdna library pool was used as a template to amplify polynucleotide encoding lmt , tmlh , tmla , tmabadh and bbh .
System Combination – 4/26 (1 shifts)	
hyp	the isolated cdna library pool was used as template to amplify a polynucleotide encoding lmt , tmlh , tmla , tmabadh and bbh .
shifted hyp	the isolated cdna library pool was used as a template to amplify polynucleotide encoding lmt , tmlh , tmla , tmabadh and bbh .
edited hyp	(D) (D) the isolated cdna library pool was used as a template to amplify polynucleotide encoding lmt , tmlh , tmla , tmabadh and bbh .
ref	Λ including the separated cdna library pool was used as a template to amplify polynucleotide encoding lmt , tmlh , tmla , tmabadh and bbh .

Table 9: System Combination Illustration. We illustrate the TER calculation process. The error information is given like “4/26 (1 shift)” which means 4 operations needed and one of the operations is shift. Operation deletion, insertion and substitution is represented by (D), color and color. To be converted into the reference, the system combination result needs the smallest amount of operations which includes only one substitution, two deletions and no insertions which means most appropriate words are already in the system combination output. So system combination has better lexicon choice.

6. REFERENCES

- [1] C. Allauzen, M. Riley, J. Schalkwyk, W. Skut, and M. Mohri. OpenFst: A General and Efficient Weighted Finite-State Transducer Library. In J. Holub and J. Žďárek, editors, *Implementation and Application of Automata*, volume 4783 of *Lecture Notes in Computer Science*, pages 11–23. Springer Berlin Heidelberg, 2007.
- [2] M. Denkowski and A. Lavie. Meteor 1.3: Automatic Metric for Reliable Optimization and Evaluation of Machine Translation Systems. In *Proceedings of the EMNLP 2011 Workshop on Statistical Machine Translation*, 2011.
- [3] L. Dugast, J. Senellart, and P. Koehn. Statistical Post-Editing on SYSTRAN’s Rule-based Translation System. In *Proceedings of the 2nd Workshop on Statistical Machine Translation*, pages 220–223, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [4] L. Dugast, J. Senellart, and P. Koehn. Selective addition of corpus-extracted phrasal lexical rules to a rule-based machine translation system. In *Proceedings of the twelfth Machine Translation Summit*, Ottawa, Canada, 2009.
- [5] M. Feng, C. Schmidt, J. Wuebker, M. Freitag, and H. Ney. The RWTH Aachen System for NTCIR-10 PatentMT. In *Proceedings of the 10th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access*, NTCIR-10, 2013.
- [6] I. Goto, K. P. Chow, B. Lu, E. Sumita, and B. K. Tsou. Overview of the Patent Machine Translation Task at the NTCIR-10 Workshop. In *Proceedings of the 10th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access*, NTCIR-10, 2013.
- [7] H. Johnson, N. Ueffing, M. Simard, and S. Larkin. NRC’s PORTAGE system for WMT 2007. In *Proceedings of the 2nd Workshop on Statistical Machine Translation*, pages 185–188, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [8] R. Kneser and H. Ney. Improved backing-off for m-gram language modeling. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP’95*, pages 181–184, Detroit, MI, 1995.
- [9] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst. Moses: Open source toolkit for statistical machine translation. In *ACL 2007: Demo and Poster Sessions*, pages 177–180, 2007.
- [10] F. Och. Minimum Error Rate Training for Statistical Machine Translation. In *Proc. Annual Meeting of the Association for Computational Linguistics*, pages 160–167, Sapporo, Japan, July 2003.
- [11] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA, July 2002.
- [12] V. D. P. P.F. Brown and R. Mercer. Word-Sense Disambiguation Using Statistical Methods. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 264–270, Berkeley, CA, 1991.
- [13] J. Senellart. Boosting linguistic rule-based MT systems with corpus-based approaches. In *Global Autonomous Language Exploitation PI Meeting*, Boston, USA, 2006.
- [14] M. Simard, N. Ueffing, P. Isabelle, and R. Kuhn. Rule-based Translation With Statistical Phrase-based Post-Editing. In *Proceedings of the 2nd Workshop on Statistical Machine Translation*, pages 203–206, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [15] M. Snover, B. Dorr, R. Schwartz, L. Micciulla, and J. Makhoul. 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*, pages 223–231, Cambridge, Massachusetts, USA, August 2006.
- [16] J. Yang, S. Enoue, and J. Senellart. SYSTRAN Chinese-English and English-Chinese Hybrid Machine Translation Systems. In *Proceedings of the 5th China Workshop on Machine Translation*, Nanjing, China, 2009.
- [17] J. Yang, S. Enoue, and J. Senellart. SYSTRAN Chinese-English and English-Chinese Hybrid Machine Translation Systems. In *Proceedings of the 7th China Workshop on Machine Translation*, Xiamen, China, 2011.
- [18] J. Yang, J. Senellart, and R. Zajac. SYSTRAN’s Chinese Word Segmentation. In *Proceedings of Second SIGHAN Workshop on Chinese Language Processing*, Sapporo, Japan, 2003.