Correcting Machine Translation During Question Answering

Wei-Yun Ma, Kathleen R. KcKeown *

Abstract

When a multi-lingual question-answering (QA) system provides an answer that has been incorrectly translated, it is very likely to be regarded as irrelevant. In this paper, we propose a novel method for correcting a deletion error that affects overall understanding of the sentence. Our post-editing technique uses information available at query time: examples drawn from related documents determined to be relevant to the query. Our results show that 4%-7% of MT sentences are missing a main verb and on average, 79% of the sentences modified by our system are judged to be more comprehensible. QA performance also benefits from the improved MT: 7% of irrelevant response sentences become relevant.

Keywords: machine translation, multi-lingual question-answering, main verb, post-editing.

1 Introduction

We are developing a multi-lingual question-answering (QA) system that must provide relevant English answers for a given query, drawing pieces of the answer from translated foreign sources. In the multi-lingual QA scenario, relevance and translation quality are usually inseparable: an incorrectly translated sentence in the answer is very likely to be regarded as irrelevant even when the corresponding source language sentence is actually relevant. We use a phrase-based statistical machine translation system for the MT component and thus, for us, MT serves as a black box that produces the translated documents in our corpus; we cannot change the MT system itself. As MT is used in more and more multi-lingual applications, this situation will become quite common.

* Columbia University Department of Computer Science
E-mail: wm2174@columbia.edu; kathy@cs.columbia.edu

We propose a novel method which uses redundant information available at question-answering time to correct translation errors. We present a post-editing mechanism to both detect and correct errors in translated documents determined to be relevant for the response. In this paper, we focus on cases where the main verb of a Chinese sentence has not been translated. The main verb usually plays a crucial role in conveying the meaning of a sentence. In cases where only the main verb is missing, an MT score relying on edit distance (e.g., TER or Bleu) may be high, but the sentence may nonetheless be incomprehensible, further affecting the performance of a QA system in terms of the relevance of its response answers.

Handling this problem at query time rather than during MT gives us valuable information which was not available during MT, namely, a set of related sentences and their translations which may contain the missing verb. By using translation examples of verb phrases and alignment information in the related documents, we are able to find an appropriate English verb and embed it in the right position as the main verb in order to improve MT quality. This post-editing framework is applicable across a variety of MT systems, since it only needs word alignments.

In the following sections, we first illustrate the impact of a missing main verb in a response sentence in Section 2 and then discuss how this is a universal problem across different MT systems in Section 3. We then introduce related post-editing work on MT in Section 4 and provide a system overview in Section 5. In Section 6 we describe how we identify clauses and detect those clauses which are missing a main verb. Section 7 shows how to determine which Chinese verb phrase should have been translated while Section 8 describes how to find the appropriate translation of this Chinese verb phrase in order to embed it into the target sentence. And finally, we present experimental results and summarize our conclusions in Section 9 and Section 10 respectively.

2 Impact of Missing Main Verb

MT systems usually have two major goals:

- Accuracy: the target language must convey the correct meaning of the source language;
- Fluency: the target language must be easy to read.

For a practical MT system, in many cases syntactic errors not only influence fluency of a target sentence but also affect accuracy in overall comprehension. Correcting these ungrammaticalities is inevitably necessary if we are to ensure the translation accuracy. A sentence that is missing a main verb belongs in this category. If a sentence is missing a verb, in addition to poor fluency, there are three possible consequences in comprehension for readers: the sentence can be easily understood, the sentence is incomprehensible, or the sentence may create false impressions, resulting in miscommunication.

The first case is usually caused by incorrect lexical substitution. For example,

Chinese: 国际社会也普遍**期望**伊政治进程能向前推进

REF: The international community **expects** that the political process will move forward.

MT: The international community **expectations** that the political process to move forward.

(Note: MT means machine translation result. REF means correct translation.)

In this example, the Chinese verb “期望” is wrongly translated as “expectations” instead of “expects”. For an English native speaker, however, this error does not interfere with his understanding of the MT sentence. Therefore, in this paper we do not deal with cases where the missing main verb resulted from substitution of a verb with a noun or other part of speech.

Instead, we focus on the problem where the missing main verb was caused by inappropriate deletions, which often make the translated sentence incomprehensible. For example,

Chinese: 12月13日萨达姆**被捕**。

REF : On December 13 Saddam **was arrested**.

MT: On December 13 Saddam .

In the above example, the Chinese verb “被捕” is not translated at all; it should be translated as “was arrested”. Even for an English native speaker with background knowledge about Saddam’s arrest, it is impossible for him/her to understand the original meaning of the Chinese sentence just by reading this MT sentence.

When a sentence is missing a main verb, this can also cause miscommunication and this is also a focus of this paper. For example,

Chinese: 民众对古典音乐的热爱逐年**减退**。

REF : People’s love for classical music **reduced** every year.

MT: People of classical music loving every year.

In the above example, the Chinese verb “减退” is not translated, but should be translated as “reduced”. An English native speaker could easily misunderstand the meaning to be “People love classical music every year.” or “People love classical music of every year”, which both happen to be the opposite of the original intended meaning.

3 Different MT Systems and the Missing Verb Problem

One of the recent trends in statistic machine translation (SMT) is to move from words to phrases as the basic unit of translation, allowing the systems to learn local reordering and translation operations that are sensitive to local context (Och and Ney, 2004). And for capturing more global translation operations and constraints, in recent years, syntax-driven phrase-based SMT (Chiang 2005, Liang Huang et al., 2006, Lavie 2008) models have also been developed. However, global constraints, such as the constraint of main verbs, which must consider the syntactic structure of the whole sentence, are still not ensured in these models.

In order to examine how often missing verbs occur in different up-to-date MT systems, we examined Aachen's phrased-based SMT system – "RWTH-PBT", and another state-of-the-art SMT system – "SRI-HPBT" (Hierarchical Phrase-Based Translation) provided by SRI, which uses a synchronous context-free grammar learned from a bitext, and focuses on phrase reordering and improving grammaticality of the target language.

Based on a 2008 government MT evaluation, "RWTH-PBT" and "SRI-HPBT" achieve 30.3% and 30.9% BLEU score respectively on newswire articles, and 25.1% and 24.9% BLEU score respectively on blog articles, We use the same testing set, which includes 57 newswires (480 sentences) and 37 blog articles (473 sentences), as our experimental test bed for missing verb detection. (The detection technique is illustrated in section 6.)

	RWTH-PBT	SRI-HPBT
newswire	7%	4%
blog	7%	3%
total	7%	4%

Table 1. Detection of missing main verbs for two systems

A sentence is regarded as missing a main verb if at least one of its clauses (main clause, subordinate clause or conjunct clause) or the sentence itself is detected as missing a main verb. Overall, 7% of sentences translated by RWTH-PBT are missing a main verb while 4% of sentences translated by SRI-HPBT are missing a main verb. This shows that missing a main verb is a common problem across up-to-date MT systems.

4 Related Post-Editing Work

Human post-editing is commonly used in commercial applications of MT, to correct errors in MT output. Knight and Chander (1994) introduced the idea of an automatic post-editing system that would be "portable across MT systems" and worked on article selection (a, an, the) for English noun phrases. More recently, Elming (2006) applied transformation-based learning to learn substitution rules for post-editing an MT system. Ji and Grishman (2007) used joint inference over information extraction and entity translation to improve name translation. Simard et al. (2007) developed a statistical machine translation system to learn to post-edit the output of a rule-based MT system.

Others have also proposed post-editing MT errors at query time. Parton et. al (2008) focused on the task of correcting named entity translation at query time for their multi-lingual QA framework. They utilized Wikipedia and word alignment to match and replace the incorrect name translation with the name from the given query. Ji et. al (2008) proposed a method of fuzzy phone matching between query names and transliterated/translated names in retrieved relevant documents and wanted to further correct errors in automatic speech recognition and MT. Like their work, our approach corrects MT errors at query time. It is novel in that we focus on correcting verb deletion errors by utilizing QA context, i.e. retrieved relevant documents for finding verb translations in our post-editing mechanism.

5 System Overview

The architecture of our QA system is shown in Figure 1. Our MT post-editing system (the bold block in Figure 1) runs after document retrieval has retrieved all potentially relevant documents and before the response generator selects sentences for the answer. It modifies all MT documents retrieved by the embedded information retrieval system that are missing a main verb. All MT results are provided by a phrase-based SMT system.

Post-editing includes three subtasks: detect a clause with a missing main verb, determine which Chinese verb phrase should have been translated, and find an appropriate translation for the Chinese verb phrase from related translated documents retrieved for a given query. Each subtask is described in the following sections.

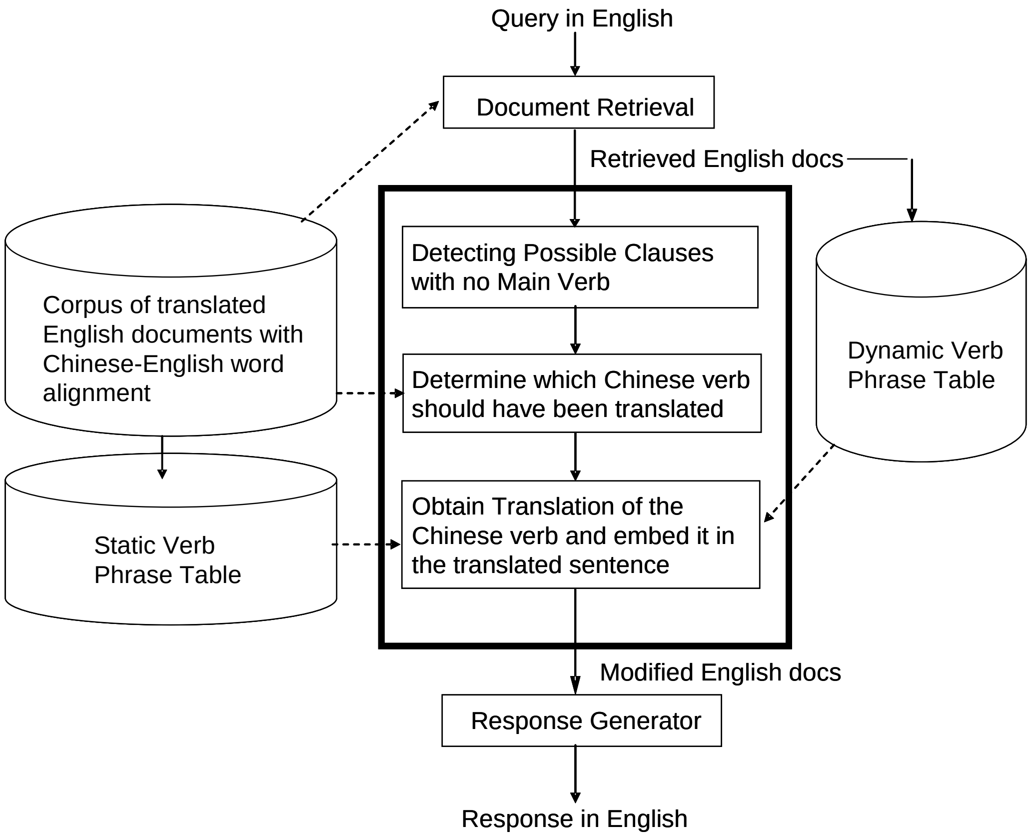


Figure 1. The System Pipeline

6 Detecting Possible Clauses with no Main Verb

Our first step is to identify clauses of a target sentence. We first tag the corpus using a Conditional Random Fields (CRF) POS tagger (Phan et al., 2007) and then use manually designed regular expressions to identify main clauses of the sentence, subordinate clauses (i.e., clauses which are arguments to a verb) and conjunct clauses in a sentence with conjunction. We do not handle adjunct clauses. Hereafter, we simply refer to all of these as "clause". The grammar is applied to a POS-tagged corpus; in the result, the scope of clauses is tagged:

"<c>I hope </c> <c>that the defence counsel will be held. </c>"

"After Saddam is captured, <c> he was in prison . </c>".

"<c>Saddam is captured, </c> and <c> he was in prison <c >"

After identifying clauses, we aim to detect those clauses without a main verb. Intuitively, we might think of utilizing a syntactic parser to parse a sentence or clause to judge whether there is a main verb within it. However, for MT sentences with serious grammatical mistakes, parser results are often unpredictable. For those sentences with no main verbs, parsers tend to incorrectly interpret a noun as a main verb.

Compared with a parser, a POS tagger is relatively reliable in processing MT results. Our system scans the output of a POS tagger to determine if there is an error. If a clause does not have any POS tag that can serve as a main verb, it is marked as missing a main verb. Based on the PENN Treebank standard tagset, a verb's POS tag can be one of the six types: "VB", "VBD", "VBG", "VBN", "VBP" and "VBZ". And according to the PENN definition, "VB" can represent either "base verb form" or "infinitive verb". We modified its definition as only representing "base verb form" and create a new POS tag – "VBT" to represent "infinitive verb".

We determine only "VB", "VBD", "VBP" and "VBZ" qualifies to serve as a main verb by their definitions, as shown in the following table.

POS	Definition	Can be a main verb's POS？
VB	base verb form	Y
VBT	infinitive verb	N
VBD	past tense	Y
VBG	gerund or present participle	N
VBN	past participle	N
VBP	non-3rd person singular present	Y
VBZ	3rd person singular present	Y

Table 2. A main verb's POS

MT alignment information is used to further ensure that the resulting marked clauses are really missing main verbs. We segment and tag the Chinese source sentence using the Stanford Chinese segmenter (Tseng et al., 2005) and the CRF Chinese POS tagger developed by Purdue University. If we find any verb in the Chinese source sentence that was not aligned with any English words in the SMT alignment tables, then we confirm that the marking was correct.

Recall the previous example- "On December 13 Saddam. (12月13日萨达姆被捕。)", shown below. The translation sentence is marked as missing a main verb because its POS-tagged result – "On/IN December/NNP 13/CD Saddam/NNP ./." does not include any POS tag which can serve as a main verb. In addition, the word “捕” in Chinese, which is POS-tagged with “VV”, is not aligned with any English words, so we further confirm that the translation sentence- “On December 13 Saddam” is missing a main verb.

Chinese: 12月13日萨达姆**被捕**。

POS-tagged Chinese: 12月/NN 13日/NN 萨达姆/NR **被/SB 捕/VV** 。/PU

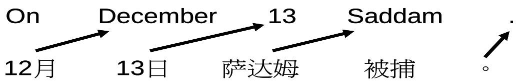
REF : On December 13 Saddam **was arrested**.

MT: On December 13 Saddam .

POS-tagged MT: On/IN December/NNP 13/CD Saddam/NNP ./.

7 Determine Non-translated Chinese Verb Phrase

Once a clause is identified as missing a main verb, the next step is to determine which Chinese verb phrase should have been translated to serve in that role. Here we define the term “Verb Translation Gap (VTG)” as a non-translated string of successive Chinese characters where the string comprises at least one Chinese verb and the string’s preceding phrase and following phrase are translated. Recalling the previous example- “On December 13 Saddam. (12月13日萨达姆被捕。)”. Its alignment can be visualized as follows:



“被捕” is a VTG because it comprises the verb “捕” and its preceding phrase-“萨达姆” and following phrase-“。” are translated but the verb phrase itself is not.

Once we find a VTG, the system will try to embed its translation into the target sentence. This assumes, however, that there is only one VTG found within a clause. In practice, more than one VTG may be found in a clause. If we choose one of them, we risk making the wrong choice. Instead, we insert the translations of VTGs simultaneously. This strategy could result in more than one main verb in a clause, but it is more helpful than having no verb at all.

8 Obtaining and Utilizing the Translation of a VTG

We translate VTGs by using verb redundancy in related documents: if the VTG was translated in other places in related documents, those translations can be reused. Related documents are likely to use a good translation for a specific VTG or provide appropriate lexical choices as it is used in a similar context. A verb’s aspect and tense can be directly determined by referencing the corresponding MT examples and their contexts. If, unfortunately, a given VTG did not have any other translation record, then the VTG will not be processed.

8.1 Building Verb Phrase Tables

To record the QA context for providing necessary information, we build verb phrase tables from relevant documents and then use the tables to translate a VTG. Two verb phrase tables are then designed: one is built from a collection of MT documents of the same genre before any query and is called the “Static Verb Phrase Table”, and the other one is dynamically built from the retrieved relevant MT documents for each query and is called the “Dynamic Verb Phrase Table”. Construction of the verb phrase tables is the same for both. Given a set of MT documents and their MT alignments, we collect all Chinese verb phrases and their translations along with their frequencies and contexts. Chinese verb phrases here are any string composed of three successive Chinese words where at least one of them is a verb.

Another key issue is to decide what kind of contextual information for a given verb phrase is appropriate for our task. A number of researchers (Cabezas and Resnik 2005, Carpuat and Wu 2007) provide abundant evidence that rich context features are useful in MT tasks. Carpuat and Wu (2007) tried to integrate a Phrase Sense Disambiguation (PSD) model into their Chinese-English SMT system and they found that the POS tag preceding a given phrase, the POS tag following the phrase and bag-of-words of the full sentence are the three most useful features. Following their approach, we use the word preceding and the word following a verb phrase as the context features.

The Static and Dynamic Verb Phrase Tables provide us with MT examples to translate a VTG. The system first references the Dynamic Verb Phrase Table as it is more likely to yield a good translation. If a record for the VTG is not found, the Static Table is referenced. If it is not found in either, the given VTG will not be processed.

8.2 Translating a VTG by Referencing Verb Phrase Tables

Regardless of which table is referenced, the following Naive Bayes equation is applied to obtain the translation of a given VTG. Given the VTG and its contextual features – preceding and following source words - our goal is to find the most appropriate translation of the VTG, which is formalized as finding the highest $P(t_k \mid pw, fw)$ among all possible translation t_k for VTG.

$$t' = \arg\max_{t_k} P(t_k \mid pw, fw) = \arg\max_{t_k} (\log P(t_k) + \log P(pw \mid t_k) + \log P(fw \mid t_k))$$

pw, fw and t_k respectively represent the preceding source word, the following source word and a translation candidate of a VTG.

8.3 Determine the Embedded Position in the Target Sentence

Finally, once we have selected an English translation, we use word alignment information to insert the translation back into the translated sentence. Chinese ordering differs from English mainly in clause ordering (Wang et al., 2007) and within the noun phrase. But within a clause centered by a verb, Chinese mostly uses a SVO or SV structure, like English (Yamada and Knight 2001) and we can assume the local alignment centered by a verb between Chinese and English is a linear mapping relation. Under this assumption, the translation of “被捕” in the above example should be placed in the position between “Saddam” and “.”

9 Experiment

Our test data is drawn from Chinese-English MT results generated by Aachen’s RWTH system (Mauser et al., 2007), a phrase-based SMT system with 38.5% BLEU score on IWSLT 2007 evaluation data. For us, the translation system serves as a black box that produces the translated documents in our corpus.

We adopted the QA test set from a government multi-lingual QA program (DARPA Global Autonomous Language Exploitation Year 2) as our experimental test bed. This test set consists of open-ended queries in English and require providing relevant answers in English translated from the Chinese source. On average, the answers were 30 sentences in length. We ran our post-editing module on five queries from the QA test set.

Q1: Who/What is involved in Saddam Hussein's trial

Q2: Produce a biography of Jacques Rene Chirac

Q3: Describe arrests of person from Salafist Group for Preaching and Combat
and give their role in organization

Q4: Provide information on Chen Sui Bian

Q5: What connections are there between World Cup games and stock markets?

We used an IR engine toolkit – “Indri”[1], a language model-based search engine for complex queries, developed by UMass Amherst and CMU. We tested our approach on 18,886 translation sentences retrieved by Indri for all of the five queries. Among these translation sentences, 1,142

[1] http://ciir.cs.umass.edu/research/indri/index.html

sentences were detected as missing a verb and thus modified by our system (6 % of all retrieved translation sentences).

	sentence # of retrieved documents	modified sentence #
Q1 (newswire)	734	24 (3%)
Q2 (newswire)	1348	56 (3%)
Q3 (newswire)	819	35 (4%)
Q4 (blog)	15093	953 (6%)
Q5 (blog)	892	74 (8%)
Total	**18886**	**1142** (6%)

Table 3. Size of retrieved translation sentences

9.1 Evaluation Methodology

For evaluation, we used human judgments of the modified and original MT. We did not have reference translations for the data used by our question-answering system and thus, could not use metrics such as TER or Bleu. Moreover, at best, TER or Bleu score would increase by a small amount and that is only if we select the same main verb in the same position as the reference. Critically, we also know that a missing main verb can cause major problems with comprehension. Thus, readers could more easily determine if the modified sentence better captured the meaning of the source sentence. We also evaluated relevance of a sentence to the query before and after modification.

We recruited 13 Chinese native speakers who are also proficient in English to judge MT quality. Native English speakers cannot tell which translation is better since they do not understand the meaning of the original Chinese. To judge relevance to the query, we used native English speakers.

Each modified sentence was evaluated by three people. They were shown the Chinese sentence and two translations, the original MT and the modified one. Evaluators did not know which MT sentence was the modified one. They were asked to decide which sentence is a better translation after reading the Chinese sentence. An evaluator had the option of answering “no difference”. An example follows:

Chinese Sentence ： 12月13日萨达姆被捕。

( ) On December 13 Saddam .

(✓) On December 13 Saddam was arrested.

( ) no difference

Because our aim is to assist question-answering systems, a translated sentence is better if it accurately represents the meaning of the corresponding Chinese sentence. Therefore “a better machine translation” was described to evaluators as “it represents more accurate, clear or complete meaning of the original Chinese.” Evaluators were asked to ignore any grammar mistake of both MT versions in their decision. We used the majority vote (two out of three) to decide the final evaluation of a sentence judged by three people.

We also evaluated the impact of post-editing on our response generator. In our QA task, response sentences were judged as “Relevant(R)”, “Partially Relevant(PR)”, “Irrelevant(I)” and “Too little information to judge(T)” sentences by native English speakers.

9.2 Results and Discussions

The results are shown in Table 4 and Table 5.

	Modified sentence #	Better	Same	Worse	Totally disagree
Q1 (newswire)	24	92 %	0 %	8 %	0 %
Q2 (newswire)	56	86 %	4 %	5 %	5 %
Q3 (newswire)	35	89 %	0 %	11 %	0 %
Q4 (blog)	953	77 %	6 %	11 %	7 %
Q5 (blog)	74	91 %	1 %	5 %	3 %
Total	**1142**	**79 %**	**5 %**	**10 %**	**6 %**

Table 4. Modification Performance

Better			Same			Worse			Totally disagree
900 (79%)			56 (5%)			116 (10%)			70 (6%)
BBB#	BBS#	BBW#	SSS#	BSS	WSS	BWW	WWS	WWW	BSW
629	122	149	2	33	21	58	26	32	70

Note: BBS represents a modified sentence that is judged as “Better” by two evaluators, and as “Same” by one evaluator, and so on.

Table 5. Cases of all 1142 modified sentences

On average, 900 (79%) of the 1142 modified sentences, which comprise 5% of all 18,886 retrieved MT sentences, are better than the original sentences based on majority voting. And for

629 (70%) of these 900 better modified sentences all three evaluators agreed that the modified sentence is better.

Among these 900 truly improved sentences, 81% sentences reference the Dynamic Verb Phrase Table, while only 19% sentences had to draw from the Static Verb Phrase Table, thus demonstrating that the question answering context is quite helpful in improving MT.

Chinese	印度25日组织全国几百万儿童**服用**脊髓灰质炎疫苗。
Ref.	In India on 25 th, millions of children **received** polio vaccine .
Original MT	India 25 th National millions children polio vaccine .
Modified MT	India 25 th National millions children **received** polio vaccine .
Chinese	另一名分析家**也同意**。
Ref.	Another analyst **would also agree with that**.
Original MT	Another analyst.
Modified MT	Another analyst **would also agree that**.
Chinese	超过七十岁者,**可以由**代理发言。
Ref.	Elderly who are more than 70 years **could be taken up by** Mr Deputy .
Original MT	More than 70 years , Mr Deputy .
Modified MT	More than 70 years , **could be taken up by** Mr Deputy .
Chinese	大陆**也是**台商对外投资的首选。
Ref.	Mainland China **is also** Taiwan businessmen's first choice of investment.
Original MT	mainland China, Taiwan businessmen's investment of first elections.
Modified MT	mainland China, **is also** Taiwan businessmen's investment of first elections.

Table 6. Examples where the modified MT is better

Chinese	古巴国家合唱团星期六结束了对加拿大为期10天的访问**演出**。
Ref.	Cuban State chorus on Saturday ended ten-day visit of performances in Canada.
Original MT	Cuban State chorus on Saturday Canada for a ten-day visit .
Modified MT	Cuban State chorus on Saturday Canada for a ten-day visit **performed** .
Chinese	想進來的包括日本,還有歐洲幾個國家,還**有**韓國等。
Ref.	The countries that want to come in include Japan , several European countries and

	South Korea .
Original MT	The coming in , including Japan , several European countries and South Korea .
Modified MT	The coming in , including Japan , several European countries and **there are** South Korea .

Table 7. Examples where original MT was better

With our post-editing technique, 7% of I/T responses become R/PR responses and none of the R/PR responses become I/T responses, yielding an increase in accuracy of 4%, thus demonstrating that our correction of MT truly improves QA performance.

	Response Sentence # (answer sentence #)	R/PR Response Sentence #	I/T Response Sentence #
Q1	**161**	**73**	**88**
Q2	27	10	17
Q3	**45**	**10**	**35**
Q4	0	0	0
Q5	1	0	1
Total	**234**	**93 (40%)**	**141 (60%)**

Table 8. QA performance without post-editing

	Response Sentence # (answer sentence #)	R/PR Response Sentence #	I/T Response Sentence #
Q1	**161**	**81**	**80**
Q2	27	10	17
Q3	**45**	**12**	**33**
Q4	0	0	0
Q5	1	0	1
Total	**234**	**103 (44%)**	**131 (56%)**

Table 9. QA performance with post-editing

Three examples of a change from I/T to R/PR are shown in the following:

Question: Who/What is involved in Saddam Hussein's trial?
Original QA answer: On May 8 , Saddam defence counsel in Jordan 's capital Amman .
Modified QA answer: On May 8 , Saddam defence counsel in Jordan 's capital Amman ***formally established*** .

Question: Who/What is involved in Saddam Hussein's trial?
Original QA answer: According to the Iraqi transitional government announced that Saddam would on October 19.
Modified QA answer: According to the Iraqi transitional government announced that Saddam would on October 19 ***to be tried***.

Question: What connections are there between World Cup games and stock markets?
Original QA answer: But if winning the ball, not necessarily in the stock market .
Modified QA answer: But if winning the ball, not necessarily in the stock market ***increased*** .

The original QA answers in the three examples are all missing main verbs and thus hard to read, but once verbs are inserted back, the fluency and the translation accuracy are both improved, enabling readers to more easily understand the provided QA answers. In the first example, although “Saddam defence counsel” in the original QA answer is already able to respond to “Who/What” of the query, the answer would become clearer and more complete when “formally established” is added in that we know how the “Saddam defence counsel” is involved in Saddam Hussein's trial. The second example is similar: “to be tried” enables readers to get a clearer idea of how “the Iraqi transitional government” is involved. And for the third example, once “increased” is inserted back, readers can get the clear idea that winning the ball not necessarily causes an increase of stock markets and then consider the provided QA answer to be relevant for the given query.

10 Conclusions

In this paper, we have presented a technique for detecting and correcting deletion errors in translated Chinese answers as part of a multi-lingual QA system, which builds on previous work in automatic post-editing of machine translation to correct errors in MT, with the goal of returning more readable, relevant answers. Our approach utilizes part-of-speech tagging and alignment information to detect missing verbs and draws from examples in documents determined to be relevant to the query to insert a new verb translation. Our evaluation demonstrates that MT quality and QA performance are both improved due to our post-editing module.

While our approach has only been tested on Chinese-English translation so far, we expect that it would be applicable to other language pairs via the adjustment with consideration of different language-specific verb structures, such as VSO structure in Arabic or SOV structure in Japanese.

We are also interested in extending our verb-post-editing framework to a sentence-rewriting procedure: given one MT system and a number of different retrieved sentences with similar or even the same meaning, some are translated better than others due to different representations in source language. We are investigating how to detect that at query time and utilize those better translations of some sentences to rewrite poorer translations of the other sentences with similar or the same meaning on the sentence level.

Acknowledgments

This material is based upon work supported by the U.S. Defense Advanced Research Projects Agency under Contract No. HR0011-06-C-0023

Reference

David Chiang. "A Hierarchical Phrase-Based for statistical machine translation," 2005, In *Proceedings of ACL, pages 263-270.*

Clara Cabezas and Philip Resnik. "Using WSD Techniques for Lexical Selection in Statistical Machine," 2005, Translation Technical report CS-TR-4736 / LAMP-TR-124 / UMIACS-TR-2005-42

Marine Carpuat and Dekai Wu. "Context-Dependent Phrasal Translation Lexicons for Statistical Machine Translation," 2007, In *Proceedings of Machine Translation Summit XI, Copenhagen*

J. Elming. "Transformation-based correction of rule-based MT," 2006, In *Proceedings of EAMT.*

Liang Huang, Kevin Knight, and Aravind Joshi. "Statistical Syntax-Directed Translation with Extended Domain of Locality," 2006, In *Proceedings of the 7th Biennial Conference of the Association for Machine Translation in the Americas (AMTA)*

Heng Ji, Ralph Grishman and Wen Wang. "Phonetic Name Matching For Cross-lingual Spoken Sentence Retrieval," 2008, *Proc. IEEE-ACL SLT08.* Goa, India

K. Knight and I. Chander. "Automated Postediting of Documents," 1994, In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*

Alon Lavie. "A general search-based syntax-driven framework for machine translation," 2008, In *Invited paper in Proceedings of CICLing-2008, pages 362-375. Computational Linguistics and Intelligent Text Processing, LNCS 4919, Springer.*

Arne Mauser, David Vilar, Gregor Leusch, Yuqi Zhang, and Hermann Ney. “The RWTH Machine Translation System for IWSLT 2007,” In *Proceedings of the International Workshop on Spoken Language Translation*, 2007, Trento, Italy.

F. J. Och, and H. Ney. “The Alignment Template Approach to Statistical Machine Translation,” 2004, *Computational Linguistics, volume 30, number 4, pages 417-449,* Cambridge, MA, USA

Xuan-Hieu Phan, Le-Minh Nguyen, Yasushi Inoguchi, and Susumu Horiguchi. “High Performance Training Conditional Random Fields for Large-scale Applications of Labeling Sequence Data,” 2007, *IEICE Transactions on Information and Systems, Vol.E90D, No.1, pp.13-21*

Kristen Parton, Kathleen R. McKeown, James Allan, and Enrique Henestroza. “Simultaneous multilingual search for translingual information retrieval,” 2008, *In Proceedings of ACM 17th Conference on Information and Knowledge Management (CIKM),* Napa Valley, CA.

Michel Simard, Cyril Goutte and Pierre Isabelle. “Statistical Phrase-based Post-Editing,” 2007, *Proceedings of NAACL-HLT*

Huihsin Tseng, Pichuan Chang, Galen Andrew, Daniel Jurafsky and Christopher Manning. "A Conditional Random Field Word Segmenter." 2005, *In Fourth SIGHAN Workshop on Chinese Language Processing.*

Chao Wang, Michael Collins, and Philipp Koehn. “Chinese Syntactic Reordering for Statistical Machine Translation,” 2007, In *proceedings of EMNLP-CoNLL.*

Kenji Yamada , Kevin Knight. “A syntax-based statistical translation model,” 2001, In *Proceedings of ACL, pages 523-530.*