Where's the Verb?
Correcting Machine Translation During Question Answering

Wei-Yun Ma, Kathleen McKeown
Department of Computer Science
Columbia University
New York, NY 10027, USA
{ma,kathy}@cs.columbia.edu

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 the main verb and on average, 79% of the modified sentences are judged to be more comprehensible. The QA performance also benefits from the improved MT: 7% of irrelevant response sentences become relevant.

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 source. 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.

We propose a novel method which uses redundant information available at question-answering time to correct 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.

Handling this problem at query time rather than during SMT gives us valuable information which was not available during SMT, 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.

A missing main verb can result in an incomprehensible sentence as seen here where the Chinese verb "被捕" was not translated at all.

MT: On December 13 Saddam .
REF : On December 13 Saddam **was arrested**.
Chinese: 12月13日萨达姆**被捕**。

In other cases, a deleted main verb can result in miscommunication; below the Chinese verb "减退" should have been translated as "reduced". An English native speaker could easily misunderstand the meaning to be "People love classical music every year." which happens to be the opposite of the original intended meaning.

MT: People of classical music loving every year.
REF : People's love for classical music **reduced** every year.
Chinese: 民众对古典音乐的热爱逐年**减退**。

2. Related Work

Post-editing has been used in full MT systems for tasks such as article selection (a, an, the) for

Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, pages 333–336,
Suntec, Singapore, 4 August 2009.

English noun phrases (Knight and Chander 1994). Simard et al in 2007 even developed a statistical phrase based MT system in a post-editing task, which takes the output of a rule-based MT system and produces post-edited target-language text. Zwarts et al. (2008) target selecting the best of a set of outputs from different MT systems through their classification-based approach. Others have also proposed using the question-answering context to detect errors in MT, showing how to correct names (Parton et. al 2008, Ji et. al 2008).

3. 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 any 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 steps: detect a clause with a missing main verb, determine which Chinese verb should have been translated, and find an example sentence in the related documents with an appropriate sentence which can be used to modify the sentence in question. To detect clauses, we first tag the corpus using a Conditional Random Fields (CRF) POS tagger 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". If a clause does not have any POS tag that can serve as a main verb (VB, VBD, VBP, VBZ), it is marked as missing a main verb.

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

In the following sections, we describe how we determine which Chinese verb should have been translated and how that occurs.

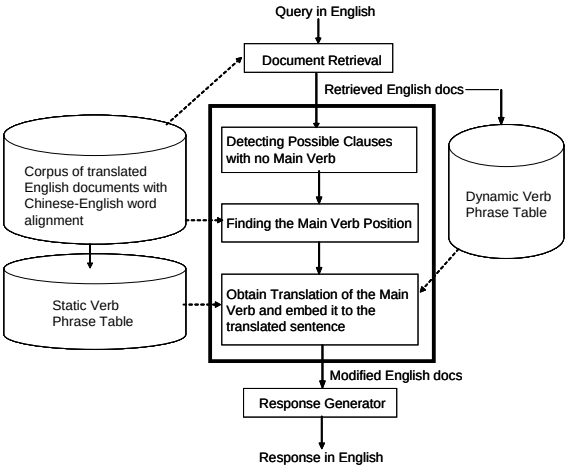

Figure 1. The System Pipeline

4. Finding the Main Verb Position

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 ".". Thus, once we find a VTG, its translation can be inserted into the corresponding position of the target sentence using the alignment.

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 both 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.

5. Obtaining a VTG Translation

We translate VTGs by using verb redundancy in related documents: if the VTG was translated in other places in related documents, the existing translations can be reused. Related documents are likely to use a good translation for a specific VTG 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.

To do this, our system first builds verb phrase tables from relevant documents and then uses the tables to translate the VTG. We use two verb phrase tables: one is built from a collection of MT documents 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".

The construction procedure is the same for both. Given a set of related MT documents and their MT alignments, we collect all Chinese verb phrases and their translations along with their frequencies and contexts.

One key issue is to decide appropriate contextual features of a verb. 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 are the three most useful features. Following their approach, we use the word preceding and the word following a verb 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 the record is not found, the Static one is referenced. If it is not found in either, the given VTG will not be processed. No matter which table is referenced, the following Naive Bayes equation is applied to obtain the translation of a given 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.

6. Experiments

Our test data is drawn from Chinese-English MT results generated by Aachen's 2007 RWTH system (Mauser et al., 2007), a phrase-based SMT system with 38.5% BLEU score on IWSLT 2007 evaluation data.

Newswires and blog articles are retrieved for five queries which served as our experimental test bed. The queries are open-ended and on average, answers were 30 sentences in length.

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
Q4: Provide information on Chen Sui Bian
Q5: What connections are there between World Cup games and stock markets?

We used MT documents retrieved by IR for each query to build the Dynamic Verb Phrase Table. We tested the system on 18,886 MT sentences from the retrieved MT documents for all of the five queries. Among these MT sentences, 1,142 sentences were detected and modified (6 % of all retrieved MT sentences).

6.1 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 better determine if the modified sentence better captured the meaning of the source sentence. We also evaluated relevance of a sentence to a 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 modified. They were asked to decide which sentence is a better translation, after reading the Chinese sentence. An evaluator also had the option of answering "no difference".

6.2 Results and Discussion

We used majority voting (two out of three) to decide the final evaluation of a sentence judged by three people. 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.

Furthermore, we found that for every individual query, the evaluators preferred more of the modified sentences than the original MT. And among these 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.

We also evaluated the impact of post-editing on the 234 sentences returned by 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. With our post-editing technique, 7% of 141 I/T responses become R/PR responses and none of the R/PR responses become I/T responses. This means that R/PR response percentage has an increase of 4%, thus demonstrating that our correction of MT truly improves QA performance. An example of a change from T to PR is:

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***.

6.3 Analysis of Different MT Systems

In order to examine how often missing verbs occur in different recent MT systems, in addition to using Aachen's up-to-date system – "RWTH-PBT"of 2008, we also ran the detection process for another state-of-the-art MT system – "SRI-HPBT" (Hierarchical Phrase-Based System) of 2008 provided by SRI, which uses a grammar on the target side as well as reordering, and focuses on improving grammaticality of the target language. Based on a government 2008 MT evaluation, the systems achieve 30.3% and 30.9% BLEU scores respectively. We used the same test set, which includes 94 written articles (953 sentences).

Overall, 7% of sentences translated by RWTH-PBT are detected with missing verbs while 4% of sentences translated by SRI-HPBT are detected with missing verb. This shows that while MT systems improve every year, missing verbs remain a problem.

7 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. Our approach uses a regular grammar 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. In the future, we plan to extend our approach to tackle other MT error types by using information available at query time.

Acknowledgments

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

References

Clara Cabezas and Philip Resnik. 2005. *Using WSD Techniques for Lexical Selection in Statistical Machine*, Translation Technical report CS-TR-4736

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

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

K. Knight and I. Chander. 1994. *Automated Postediting of Documents*, AAAI

Kristen Parton, Kathleen R. McKeown, James Allan, and Enrique Henestroza. 2008. *Simultaneous multilingual search for translingual information retrieval*, ACM 17th CIKM

Arne Mauser, David Vilar, Gregor Leusch, Yuqi Zhang, and Hermann Ney. 2007. *The RWTH Machine Translation System for IWSLT 2007*, IWSLT

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

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

Kenji Yamada , Kevin Knight. 2001. *A syntax-based statistical translation model*, ACL

S. Zwarts and M. Dras. 2008. *Choosing the Right Translation: A Syntactically Informed Approach*, COLING