Modeling the Translation of Predicate-Argument Structure for SMT

Deyi Xiong, Min Zhang[*], Haizhou Li
Human Language Technology
Institute for Infocomm Research
1 Fusionopolis Way, #21-01 Connexis, Singapore 138632
{dyxiong, mzhang, hli}@i2r.a-star.edu.sg

Abstract

Predicate-argument structure contains rich semantic information of which statistical machine translation hasn't taken full advantage. In this paper, we propose two discriminative, feature-based models to exploit predicate-argument structures for statistical machine translation: 1) a predicate translation model and 2) an argument reordering model. The predicate translation model explores lexical and semantic contexts surrounding a verbal predicate to select desirable translations for the predicate. The argument reordering model automatically predicts the moving direction of an argument relative to its predicate after translation using semantic features. The two models are integrated into a state-of-the-art phrase-based machine translation system and evaluated on Chinese-to-English translation tasks with large-scale training data. Experimental results demonstrate that the two models significantly improve translation accuracy.

1 Introduction

Recent years have witnessed increasing efforts towards integrating predicate-argument structures into statistical machine translation (SMT) (Wu and Fung, 2009b; Liu and Gildea, 2010). In this paper, we take a step forward by introducing a novel approach to incorporate such semantic structures into SMT. Given a source side predicate-argument structure, we attempt to translate each semantic frame (predicate and its associated arguments) into an appropriate target string. We believe that the translation of predicates and reordering of arguments are the two central issues concerning the transfer of predicate-argument structure across languages.

Predicates[1] are essential elements in sentences. Unfortunately they are usually neither correctly translated nor translated at all in many SMT systems according to the error study by Wu and Fung (2009a). This suggests that conventional lexical and phrasal translation models adopted in those SMT systems are not sufficient to correctly translate predicates in source sentences. Thus we propose a discriminative, feature-based **predicate translation model** that captures not only lexical information (i.e., surrounding words) but also high-level semantic contexts to correctly translate predicates.

Arguments contain information for questions of *who*, *what*, *when*, *where*, *why*, and *how* in sentences (Xue, 2008). One common error in translating arguments is about their reorderings: arguments are placed at incorrect positions after translation. In order to reduce such errors, we introduce a discriminative **argument reordering model** that uses the position of a predicate as the reference axis to estimate positions of its associated arguments on the target side. In this way, the model predicts moving directions of arguments relative to their predicates with semantic features.

We integrate these two discriminative models into a state-of-the-art phrase-based system. Experimental results on large-scale Chinese-to-English translation show that both models are able to obtain significant improvements over the baseline. Our analysis on system outputs further reveals that they can indeed help reduce errors in predicate translations and argument reorderings.

[*]Corresponding author

[1]We only consider verbal predicates in this paper.

Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pages 902–911,
Jeju, Republic of Korea, 8-14 July 2012.

The paper is organized as follows. In Section 2, we will introduce related work and show the significant differences between our models and previous work. In Section 3 and 4, we will elaborate the proposed predicate translation model and argument reordering model respectively, including details about modeling, features and training procedure. Section 5 will introduce how to integrate these two models into SMT. Section 6 will describe our experiments and results. Section 7 will empirically discuss how the proposed models improve translation accuracy. Finally we will conclude with future research directions in Section 8.

2 Related Work

Predicate-argument structures (PAS) are explored for SMT on both the source and target side in some previous work. As PAS analysis widely employs global and sentence-wide features, it is computationally expensive to integrate target side predicate-argument structures into the dynamic programming style of SMT decoding (Wu and Fung, 2009b). Therefore they either postpone the integration of target side PASs until the whole decoding procedure is completed (Wu and Fung, 2009b), or directly project semantic roles from the source side to the target side through word alignments during decoding (Liu and Gildea, 2010).

There are other previous studies that explore only source side predicate-argument structures. Komachi and Matsumoto (2006) reorder arguments in source language (Japanese) sentences using heuristic rules defined on source side predicate-argument structures in a pre-processing step. Wu et al. (2011) automate this procedure by automatically extracting reordering rules from predicate-argument structures and applying these rules to reorder source language sentences. Aziz et al. (2011) incorporate source language semantic role labels into a tree-to-string SMT system.

Although we also focus on source side predicate-argument structures, our models differ from the previous work in two main aspects: 1) we propose two separate discriminative models to exploit predicate-argument structures for predicate translation and argument reordering respectively; 2) we consider argument reordering as an argument movement (relative to its predicate) prediction problem and use a discriminatively trained classifier for such predictions.

Our predicate translation model is also related to previous discriminative lexicon translation models (Berger et al., 1996; Venkatapathy and Bangalore, 2007; Mauser et al., 2009). While previous models predict translations for all words in vocabulary, we only focus on verbal predicates. This will tremendously reduce the amount of training data required, which usually is a problem in discriminative lexicon translation models (Mauser et al., 2009). Furthermore, the proposed translation model also differs from previous lexicon translation models in that we use both lexical and semantic features. Our experimental results show that semantic features are able to further improve translation accuracy.

3 Predicate Translation Model

In this section, we present the features and the training process of the predicate translation model.

3.1 Model

Following the context-dependent word models in (Berger et al., 1996), we propose a discriminative predicate translation model. The essential component of our model is a maximum entropy classifier $p_t(e|\mathcal{C}(v))$ that predicts the target translation e for a verbal predicate v given its surrounding context $\mathcal{C}(v)$. The classifier can be formulated as follows.

$$p_t(e|\mathcal{C}(v)) = \frac{exp(\sum_i \theta_i f_i(e, \mathcal{C}(v)))}{\sum_{e'} exp(\sum_i \theta_i f_i(e', \mathcal{C}(v)))} \quad (1)$$

where f_i are binary features, θ_i are weights of these features. Given a source sentence which contains N verbal predicates $\{v_i\}_1^N$, our predicate translation model M_t can be denoted as

$$M_t = \prod_{i=1}^{N} p_t(e_{v_i}|\mathcal{C}(v_i)) \quad (2)$$

Note that we do not restrict the target translation e to be a single word. We allow e to be a phrase of length up to 4 words so as to capture multi-word translations for a verbal predicate. For example, a Chinese verb “发行(issue)” can be translated as “to be issued” or “have issued” with modality words.

This will increase the number of classes to be predicted by the maximum entropy classifier. But according to our observation, it is still computationally tractable (see Section 3.3). If a verbal predicate is not translated, we set e = NULL so that we can also capture null translations for verbal predicates.

3.2 Features

The apparent advantage of discriminative lexicon translation models over generative translation models (e.g., conventional lexical translation model as described in (Koehn et al., 2003)) is that discriminative models allow us to integrate richer contexts (lexical, syntactic or semantic) into target translation prediction. We use two kinds of features to predict translations for verbal predicates: 1) lexical features and 2) semantic features. All features are in the following binary form.

$$f(e, \mathcal{C}(v)) = \begin{cases} 1, & \text{if } e = \clubsuit \text{ and } \mathcal{C}(v).\heartsuit = \spadesuit \\ 0, & \text{else} \end{cases} \quad (3)$$

where the symbol $\clubsuit$ is a placeholder for a possible target translation (up to 4 words), the symbol $\heartsuit$ indicates a contextual (lexical or semantic) element for the verbal predicate v, and the symbol $\spadesuit$ represents the value of $\heartsuit$.

Lexical Features: The lexical element $\heartsuit$ is extracted from the surrounding words of verbal predicate v. We use the preceding 3 words and the succeeding 3 words to define the lexical context for the verbal predicate v. Therefore $\heartsuit \in \{w_{-3}, w_{-2}, w_{-1}, v, w_1, w_2, w_3\}$.

Semantic Features: The semantic element $\heartsuit$ is extracted from the surrounding arguments of verbal predicate v. In particular, we define a semantic window centered at the verbal predicate with 6 arguments $\{A_{-3}, A_{-2}, A_{-1}, A_1, A_2, A_3\}$ where $A_{-3} - A_{-1}$ are arguments on the left side of v while $A_1 - A_3$ are those on the right side. Different verbal predicates have different number of arguments in different linguistic scenarios. We observe on our training data that the number of arguments for 96.5% verbal predicates on each side (left/right) is not larger than 3. Therefore the defined 6-argument semantic window is sufficient to describe argument contexts for predicates.

For each argument A_i in the defined semantic window, we use its semantic role (i.e., ARG0, ARGM-TMP and so on) A_i^r and head word A_i^h to define semantic context elements $\heartsuit$. If an argument A_i does not exist for the verbal predicate v [2], we set the value of both A_i^r and A_i^h to null.

$f(e, \mathcal{C}(v)) = 1$ **if and only if**
e = adjourn and $\mathcal{C}(v).A_{-3}^h$ = 安理会
e = adjourn and $\mathcal{C}(v).A_{-1}^r$ = ARGM-TMP
e = adjourn and $\mathcal{C}(v).A_1^h$ = 天
e = adjourn and $\mathcal{C}(v).A_2^r$ = null
e = adjourn and $\mathcal{C}(v).A_3^h$ = null

Table 1: Semantic feature examples.

Figure 1 shows a Chinese sentence with its predicate-argument structure and English translation. The verbal predicate "休会/adjourn" (in bold) has 4 arguments: one in an ARG0 agent role, one in an ARGM-ADV adverbial modifier role, one in an ARGM-TMP temporal modifier role and the last one in an ARG1 patient role. Table 1 shows several semantic feature examples of this verbal predicate.

3.3 Training

In order to train the discriminative predicate translation model, we first parse source sentences and labeled semantic roles for all verbal predicates (see details in Section 6.1) in our word-aligned bilingual training data. Then we extract all training events for verbal predicates which occur at least 10 times in the training data. A training event for a verbal predicate v consists of all contextual elements $\mathcal{C}(v)$ (e.g., w_1, A_1^h) defined in the last section and the target translation e. Using these events, we train one maximum entropy classifier per verbal predicate (16,121 verbs in total) via the off-the-shelf MaxEnt toolkit[3]. We perform 100 iterations of the L-BFGS algorithm implemented in the training toolkit for each verbal predicate with both Gaussian prior and event cutoff set to 1 to avoid overfitting. After event cutoff, we have an average of 140 classes (target translations) per verbal predicate with the maximum number of classes being 9,226. The training takes an average of 52.6 seconds per verb. In order to expedite the train-

[2] For example, the verb v has only two arguments on its left side. Thus argument A_{-3} doest not exist.

[3] Available at: http://homepages.inf.ed.ac.uk/lzhang10/maxent_toolkit.html

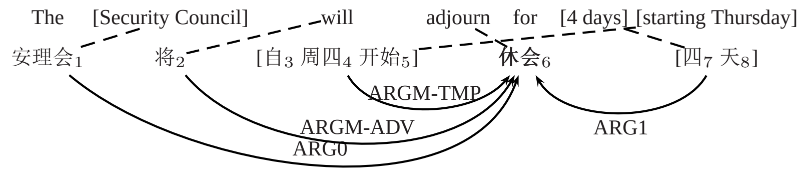

Figure 1: An example of predicate-argument structure in Chinese and its aligned English translation. The bold word in Chinese is the verbal predicate. The subscripts on the Chinese sentence show the indexes of words from left to right.

ing, we run the training toolkit in a parallel manner.

4 Argument Reordering Model

In this section we introduce the discriminative argument reordering model, features and the training procedure.

4.1 Model

Since the predicate determines what arguments are involved in its semantic frame and semantic frames tend to be cohesive across languages (Fung et al., 2006), the movements of predicate and its arguments across translations are like the motions of a planet and its satellites. Therefore we consider the reordering of an argument as the motion of the argument relative to its predicate. In particular, we use the position of the predicate as the reference axis. The motion of associated arguments relative to the reference axis can be roughly divided into 3 categories[4]: 1) no change across languages (NC); 2) moving from the left side of its predicate to the right side of the predicate after translation (L2R); and 3) moving from the right side of its predicate to the left side of the predicate after translation (R2L).

Let's revisit Figure 1. The ARG0, ARGM-ADV and ARG1 are located at the same side of their predicate after being translated into English, therefore the reordering category of these three arguments is assigned as "NC". The ARGM-TMP is moved from the left side of "休会/adjourn" to the right side of "adjourn" after translation, thus its reordering category is L2R.

In order to predict the reordering category for an argument, we propose a discriminative argument reordering model that uses a maximum entropy classifier to calculate the reordering category $m \in \{\text{NC, L2R, R2L}\}$ for an argument A as follows.

$$p_r(m|\mathcal{C}(A)) = \frac{exp(\sum_i \theta_i f_i(m, \mathcal{C}(A)))}{\sum_{m'} exp(\sum_i \theta_i f_i(m', \mathcal{C}(A)))} \quad (4)$$

where $\mathcal{C}(A)$ indicates the surrounding context of A. The features f_i will be introduced in the next section. We assume that motions of arguments are independent on each other. Given a source sentence with labeled arguments $\{A_i\}_1^N$, our discriminative argument reordering model M_r is formulated as

$$M_r = \prod_{i=1}^{N} p_r(m_{A_i}|\mathcal{C}(A_i)) \quad (5)$$

4.2 Features

The features f_i used in the argument reordering model still takes the binary form as in Eq. (3). Table 2 shows the features that are used in the argument reordering model. We extract features from both the source and target side. On the source side, the features include the verbal predicate, the semantic role of the argument, the head word and the boundary words of the argument. On the target side, the translation of the verbal predicate, the translation of the head word of the argument, as well as the boundary words of the translation of the argument are used as features.

4.3 Training

To train the argument reordering model, we first extract features defined in the last section from our bilingual training data where source sentences are annotated with predicate-argument structures. We also study the distribution of argument reordering categories (i.e.,NC, L2R and R2L) in the training data, which is shown in Table 3. Most arguments, accounting for 82.43%, are on the same side of their verbal predicates after translation. The remaining

[4]Here we assume that the translations of arguments are not interrupted by their predicates, other arguments or any words outside the arguments in question. We leave for future research the task of determining whether arguments should be translated as a unit or not.

	Features of an argument A for reordering
src	its verbal predicate A^p
	its semantic role A^r
	its head word A^h
	the leftmost word of A
	the rightmost word of A
tgt	the translation of A^p
	the translation of A^h
	the leftmost word of the translation of A
	the rightmost word of the translation of A

Table 2: Features adopted in the argument reordering model.

Reordering Category	Percent
NC	82.43%
L2R	11.19%
R2L	6.38%

Table 3: Distribution of argument reordering categories in the training data.

arguments (17.57%) are moved either from the left side of their predicates to the right side after translation (accounting for 11.19%) or from the right side to the left side of their translated predicates (accounting for 6.38%).

After all features are extracted, we use the maximum entropy toolkit in Section 3.3 to train the maximum entropy classifier as formulated in Eq. (4). We perform 100 iterations of L-BFGS.

5 Integrating the Two Models into SMT

In this section, we elaborate how to integrate the two models into phrase-based SMT. In particular, we integrate the models into a phrase-based system which uses bracketing transduction grammars (BTG) (Wu, 1997) for phrasal translation (Xiong et al., 2006). Since the system is based on a CKY-style decoder, the integration algorithms introduced here can be easily adapted to other CKY-based decoding systems such as the hierarchical phrasal system (Chiang, 2007).

5.1 Integrating the Predicate Translation Model

It is straightforward to integrate the predicate translation model into phrase-based SMT (Koehn et al., 2003; Xiong et al., 2006). We maintain word alignments for each phrase pair in the phrase table. Given a source sentence with its predicate-argument structure, we detect all verbal predicates and load trained predicate translation classifiers for these verbs. Whenever a hypothesis covers a new verbal predicate v, we find the target translation e for v through word alignments and then calculate its translation probability $p_t(e|\mathcal{C}(v))$ according to Eq. (1).

The predicate translation model (as formulated in Eq. (2)) is integrated into the whole log-linear model just like the conventional lexical translation model in phrase-based SMT (Koehn et al., 2003). The two models are independently estimated but complementary to each other. While the lexical translation model calculates the probability of a verbal predicate being translated given its local lexical context, the discriminative predicate translation model is able to employ both lexical and semantic contexts to predict translations for verbs.

5.2 Integrating the Argument Reordering Model

Before we introduce the integration algorithm for the argument reordering model, we define two functions $\mathcal{A}$ and $\mathcal{N}$ on a source sentence and its predicate-argument structure τ as follows.

- $\mathcal{A}(i, j, \tau)$: from the predicate-argument structure τ, the function finds all predicate-argument pairs which are completely located within the span from source word i to j. For example, in Figure 1, $\mathcal{A}(3, 6, \tau)$ = {(休会, ARGM-TMP)} while $\mathcal{A}(2, 3, \tau)$ = {}, $\mathcal{A}(1, 5, \tau)$ = {} because the verbal predicate "休会" is located outside the span (2,3) and (1,5).

- $\mathcal{N}(i, k, j, \tau)$: the function finds all predicate-argument pairs that cross the two neighboring spans (i, k) and $(k+1, j)$. It can be formulated as $\mathcal{A}(i, j, \tau) - (\mathcal{A}(i, k, \tau) \bigcup \mathcal{A}(k+1, j, \tau))$.

We then define another function $\mathcal{P}_r$ to calculate the argument reordering model probability on all arguments which are found by the previous two functions $\mathcal{A}$ and $\mathcal{N}$ as follows.

$$\mathcal{P}_r(\mathcal{B}) = \prod_{A \in \mathcal{B}} p_r(m_A|\mathcal{C}(A)) \quad (6)$$

where $\mathcal{B}$ denotes either $\mathcal{A}$ or $\mathcal{N}$.

Following (Chiang, 2007), we describe the algorithm in a deductive system. It is shown in Figure 2. The algorithm integrates the argument reordering model into a CKY-style decoder (Xiong et al., 2006). The item $[X, i, j]$ denotes a BTG node X spanning from i to j on the source side. For notational convenience, we only show the argument reordering model probability for each item, ignoring all other sub-model probabilities such as the language model probability. The Eq. (7) shows how we calculate the argument reordering model probability when a lexical rule is applied to translate a source phrase c to a target phrase e. The Eq. (8) shows how we compute the argument reordering model probability for a span (i, j) in a dynamic programming manner when a merging rule is applied to combine its two sub-spans in a straight ($X \rightarrow [X_1, X_2]$) or inverted order ($X \rightarrow \langle X_1, X_2 \rangle$). We directly use the probabilities $\mathcal{P}_r(\mathcal{A}(i, k, \tau))$ and $\mathcal{P}_r(\mathcal{A}(k + 1, j, \tau))$ that have been already obtained for the two sub-spans (i, k) and $(k + 1, j)$. In this way, we only need to calculate the probability $\mathcal{P}_r(\mathcal{N}(i, k, j, \tau))$ for predicate-argument pairs that cross the two sub-spans.

6 Experiments

In this section, we present our experiments on Chinese-to-English translation tasks, which are trained with large-scale data. The experiments are aimed at measuring the effectiveness of the proposed discriminative predicate translation model and argument reordering model.

6.1 Setup

The baseline system is the BTG-based phrasal system (Xiong et al., 2006). Our training corpora[5] consist of 3.8M sentence pairs with 96.9M Chinese words and 109.5M English words. We ran GIZA++ on these corpora in both directions and then applied the "grow-diag-final" refinement rule to obtain word alignments. We then used all these word-aligned corpora to generate our phrase table. Our 5-gram language model was trained on the Xinhua section of the English Gigaword corpus (306 million words) using the SRILM toolkit (Stolcke, 2002) with modified Kneser-Ney smoothing.

To train the proposed predicate translation model and argument reordering model, we first parsed all source sentences using the Berkeley Chinese parser (Petrov et al., 2006) and then ran the Chinese semantic role labeler[6] (Li et al., 2010) on all source parse trees to annotate semantic roles for all verbal predicates. After we obtained semantic roles on the source side, we extracted features as described in Section 3.2 and 4.2 and used these features to train our two models as described in Section 3.3 and 4.3.

We used the NIST MT03 evaluation test data as our development set, and the NIST MT04, MT05 as the test sets. We adopted the case-insensitive BLEU-4 (Papineni et al., 2002) as the evaluation metric. Statistical significance in BLEU differences was tested by paired bootstrap re-sampling (Koehn, 2004).

6.2 Results

Our first group of experiments is to investigate whether the predicate translation model is able to improve translation accuracy in terms of BLEU and whether semantic features are useful. The experimental results are shown in Table 4. From the table, we have the following two observations.

- The proposed predicate translation models achieve an average improvement of 0.57 BLEU points across the two NIST test sets when all features (lex+sem) are used. Such an improvement is statistically significant ($p < 0.01$). According to our statistics, there are 5.07 verbal predicates per sentence in NIST04 and 4.76 verbs per sentence in NIST05, which account for 18.02% and 16.88% of all words in NIST04 and 05 respectively. This shows that not only verbal predicates are semantically important, they also form a major part of the sentences. Therefore, whether verbal predicates are translated correctly or not has a great impact on the translation accuracy of the whole sentence [7].

[5]The corpora include LDC2004E12, LDC2004T08, LDC2005T10, LDC2003E14, LDC2002E18, LDC2005T06, LDC2003E07 and LDC2004T07.

[6]Available at: http://nlp.suda.edu.cn/∼jhli/.

[7]The example in Table 6 shows that the translations of verbs even influences reorderings and translations of neighboring words.

$$\frac{X \to c/e}{[X,i,j] : \mathcal{P}_r(\mathcal{A}(i,j,\tau))} \tag{7}$$

$$\frac{X \to [X_1, X_2] \text{ or } \langle X_1, X_2 \rangle \quad [X_1, i, k] : \mathcal{P}_r(\mathcal{A}(i,k,\tau)) \quad [X_2, k+1, j] : \mathcal{P}_r(\mathcal{A}(k+1,j,\tau))}{[X,i,j] : \mathcal{P}_r(\mathcal{A}(i,k,\tau)) \cdot \mathcal{P}_r(\mathcal{A}(k+1,j,\tau)) \cdot \mathcal{P}_r(\mathcal{N}(i,k,j,\tau))} \tag{8}$$

Figure 2: Integrating the argument reordering model into a BTG-style decoder.

Model	NIST04	NIST05
Base	35.52	33.80
Base+PTM (lex)	35.71+	34.09+
Base+PTM (lex+sem)	36.10++**	34.35++*

Table 4: Effects of the proposed predicate translation model (PTM). PTM (lex): predicate translation model with lexical features; PTM (lex+sem): predicate translation model with both lexical and semantic features; +/++: better than the baseline ($p < 0.05/0.01$). */**: better than Base+PTM (lex) ($p < 0.05/0.01$).

Model	NIST04	NIST05
Base	35.52	33.80
Base+ARM	35.82++	34.29++
Base+ARM+PTM	36.19++	34.72++

Table 5: Effects of the proposed argument reordering model (ARM) and the combination of ARM and PTM. ++: better than the baseline ($p < 0.01$).

- When we integrate both lexical and semantic features (lex+sem) described in Section 3.2, we obtain an improvement of about 0.33 BLEU points over the system where only lexical features (lex) are used. Such a gain, which is statistically significant, confirms the effectiveness of semantic features.

Our second group of experiments is to validate whether the argument reordering model is capable of improving translation quality. Table 5 shows the results. We obtain an average improvement of 0.4 BLEU points on the two test sets over the baseline when we incorporate the proposed argument reordering model into our system. The improvements on the two test sets are both statistically significant ($p < 0.01$).

Finally, we integrate both the predicate translation model and argument reordering model into the final system. The two models collectively achieve an improvement of up to 0.92 BLEU points over the baseline, which is shown in Table 5.

7 Analysis

In this section, we conduct some case studies to show how the proposed models improve translation accuracy by looking into the differences that they make on translation hypotheses.

Table 6 displays a translation example which shows the difference between the baseline and the system enhanced with the predicate translation model. There are two verbal predicates "赶往/head to" and "参加/attend" in the source sentence. In order to get the most appropriate translations for these two verbal predicates, we should adopt different ways to translate them. The former should be translated as a corresponding verb word or phrase while the latter into a preposition word "for". Unfortunately, the baseline incorrectly translates the two verbs. Furthermore, such translation errors even result in undesirable reorderings of neighboring words "伯利恒/Bethlehem and "弥撒/mass". This indicates that verbal predicate translation errors may lead to more errors, such as inappropriate reorderings or lexical choices for neighboring words. On the contrary, we can see that our predicate translation model is able to help select appropriate words for both verbs. The correct translations of these two verbs also avoid incorrect reorderings of neighboring words.

Table 7 shows another example to demonstrate how the argument reordering model improve reorderings. The verbal predicate "进行/carry out" has three arguments, ARG0, ARG-ADV and ARG1. The ARG1 argument should be moved from the right side of the predicate to its left side after translation. The ARG0 argument can either stay on the left side or move to right side of the predicate. Ac-

Base	[数 千] 信徒 **赶往** 伯利恒 **参加** [平安 夜] 弥撒 [thousands of] followers to Mass in Bethlehem [Christmas Eve]
Base+PTM	[数 千] 信徒 **赶往** 伯利恒 **参加** [平安 夜] 弥撒 [thousands of] devotees [rushed to] Bethlehem for [Christmas Eve] mass
Ref	thousands of worshippers head to Bethlehem for Christmas Midnight mass

Table 6: A translation example showing the difference between the baseline and the system with the predicate translation model (PTM). Phrase alignments in the two system outputs are shown with dashed lines. Chinese words in bold are verbal predicates.

PAS	ARG0 ARGM-ADV ARG1 [有关 这 套 灾难 警告 系统] 还要 **进行** [更 多 重要 的 磋商]
Base	[有关 这] 套 灾难 [警告 系统] 还要 [进行 更 多] [重要 的 磋商] the more [important consultations] also set disaster [warning system]
Base+ARM	有关 [这 套] 灾难 [警告 系统] [还要 进行] [更 多] [重要 的 磋商] more [important consultations] on [such a] disaster [warning system] [should be carried out]
Ref	more important discussions will be held on the disaster warning system

Table 7: A translation example showing the difference between the baseline and the system with the argument reordering model (ARM). The predicate-argument structure (PAS) of the source sentence is also displayed in the first row.

cording to the phrase alignments of the baseline, we clearly observe three serious translation errors: 1) the ARG0 argument is translated into separate groups which are not adjacent on the target side; 2) the predicate is not translated at all; and 3) the ARG1 argument is not moved to the left side of the predicate after translation. All of these 3 errors are avoided in the Base+ARM system output as a result of the argument reordering model that correctly identifies arguments and moves them in the right directions.

8 Conclusions and Future Work

We have presented two discriminative models to incorporate source side predicate-argument structures into SMT. The two models have been integrated into a phrase-based SMT system and evaluated on Chinese-to-English translation tasks using large-scale training data. The first model is the predicate translation model which employs both lexical and semantic contexts to translate verbal predicates. The second model is the argument reordering model which estimates the direction of argument movement relative to its predicate after translation. Experimental results show that both models are able to significantly improve translation accuracy in terms of BLEU score.

In the future work, we will extend our predicate translation model to translate both verbal and nominal predicates. Nominal predicates also frequently occur in Chinese sentences and thus accurate translations of them are desirable for SMT. We also want to address another translation issue of arguments as shown in Table 7: arguments are wrongly translated into separate groups instead of a cohesive unit (Wu and Fung, 2009a). We will build an argument segmentation model that follows (Xiong et al., 2011) to determine whether arguments should be translated as a unit or not.

References

Wilker Aziz, Miguel Rios, and Lucia Specia. 2011. Shallow semantic trees for smt. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 316–322, Edinburgh, Scotland, July. Association for Computational Linguistics.

Adam L. Berger, Stephen A. Della Pietra, and Vincent J. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71.

David Chiang. 2007. Hierarchical phrase-based translation. *Computational Linguistics*, 33(2):201–228.

Pascale Fung, Wu Zhaojun, Yang Yongsheng, and Dekai Wu. 2006. Automatic learning of chinese english semantic structure mapping. In *IEEE/ACL 2006 Workshop on Spoken Language Technology (SLT 2006)*, Aruba, December.

Philipp Koehn, Franz Joseph Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 58–54, Edmonton, Canada, May-June.

Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of EMNLP 2004*, pages 388–395, Barcelona, Spain, July.

Mamoru Komachi and Yuji Matsumoto. 2006. Phrase reordering for statistical machine translation based on predicate-argument structure. In *In Proceedings of the International Workshop on Spoken Language Translation: Evaluation Campaign on Spoken Language Translation*, pages 77–82.

Junhui Li, Guodong Zhou, and Hwee Tou Ng. 2010. Joint syntactic and semantic parsing of chinese. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1108–1117, Uppsala, Sweden, July. Association for Computational Linguistics.

Ding Liu and Daniel Gildea. 2010. Semantic role features for machine translation. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 716–724, Beijing, China, August. Coling 2010 Organizing Committee.

Arne Mauser, Saša Hasan, and Hermann Ney. 2009. Extending statistical machine translation with discriminative and trigger-based lexicon models. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 210–218, Singapore, August. Association for Computational Linguistics.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA, July.

Slav Petrov, Leon Barrett, Romain Thibaux, and Dan Klein. 2006. Learning accurate, compact, and interpretable tree annotation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 433–440, Sydney, Australia, July. Association for Computational Linguistics.

Andreas Stolcke. 2002. Srilm–an extensible language modeling toolkit. In *Proceedings of the 7th International Conference on Spoken Language Processing*, pages 901–904, Denver, Colorado, USA, September.

Sriram Venkatapathy and Srinivas Bangalore. 2007. Three models for discriminative machine translation using global lexical selection and sentence reconstruction. In *Proceedings of SSST, NAACL-HLT 2007 / AMTA Workshop on Syntax and Structure in Statistical Translation*, pages 96–102, Rochester, New York, April. Association for Computational Linguistics.

Dekai Wu and Pascale Fung. 2009a. Can semantic role labeling improve smt. In *Proceedings of the 13th Annual Conference of the EAMT*, pages 218–225, Barcelona, May.

Dekai Wu and Pascale Fung. 2009b. Semantic roles for smt: A hybrid two-pass model. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, pages 13–16, Boulder, Colorado, June. Association for Computational Linguistics.

Xianchao Wu, Katsuhito Sudoh, Kevin Duh, Hajime Tsukada, and Masaaki Nagata. 2011. Extracting preordering rules from predicate-argument structures. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 29–37, Chiang Mai, Thailand, November. Asian Federation of Natural Language Processing.

Dekai Wu. 1997. Stochastic inversion transduction grammars and bilingual parsing of parallel corpora. *Computational Linguistics*, 23(3):377–403.

Deyi Xiong, Qun Liu, and Shouxun Lin. 2006. Maximum entropy based phrase reordering model for statistical machine translation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 521–528, Sydney, Australia, July. Association for Computational Linguistics.

Deyi Xiong, Min Zhang, and Haizhou Li. 2011. A maximum-entropy segmentation model for statistical machine translation. *IEEE Transactions on Audio, Speech and Language Processing*, 19(8):2494–2505.

Nianwen Xue. 2008. Labeling chinese predicates with semantic roles. *Computational Linguistics*, 34(2):225–255.