Handling Ambiguities of Bilingual Predicate-Argument Structures for Statistical Machine Translation

Feifei Zhai, Jiajun Zhang, Yu Zhou and Chengqing Zong
National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, Beijing, China
{ffzhai,jjzhang,yzhou,cqzong}@nlpr.ia.ac.cn

Abstract

Predicate-argument structure (PAS) has been demonstrated to be very effective in improving SMT performance. However, since a source-side PAS might correspond to multiple different target-side PASs, there usually exist many PAS ambiguities during translation. In this paper, we group PAS ambiguities into two types: role ambiguity and gap ambiguity. Then we propose two novel methods to handle the two PAS ambiguities for SMT accordingly: 1) inside context integration; 2) a novel maximum entropy PAS disambiguation (MEPD) model. In this way, we incorporate rich context information of PAS for disambiguation. Then we integrate the two methods into a PAS-based translation framework. Experiments show that our approach helps to achieve significant improvements on translation quality.

1 Introduction

Predicate-argument structure (PAS) depicts the relationship between a predicate and its associated arguments, which indicates the skeleton structure of a sentence on semantic level. Basically, PAS agrees much better between two languages than syntax structure (Fung et al., 2006; Wu and Fung, 2009b). Considering that current syntax-based translation models are always impaired by cross-lingual structure divergence (Eisner, 2003; Zhang et al., 2010), PAS is really a better representation of a sentence pair to model the bilingual structure mapping.

However, since a source-side PAS might correspond to multiple different target-side PASs, there usually exist many PAS ambiguities during translation. For example, in Figure 1, (a) and (b) carry the same source-side PAS $<[A0]_1$ $[Pred(是)]_2$ $[A1]_3>$ for Chinese predicate "是". However, in Figure 1(a), the corresponding target-side-like PAS is $<[X_1]$ $[X_2]$ $[X_3]>$, while in

Figure 1(b), the counterpart target-side-like PAS[1] is $<[X_2]$ $[X_3]$ $[X_1]>$. This is because the two PASs play different roles in their corresponding sentences. Actually, Figure 1(a) is an independent PAS, while Figure 1(b) is a modifier of the noun phrase "中国 和 俄罗斯". We call this kind of PAS ambiguity *role ambiguity*.
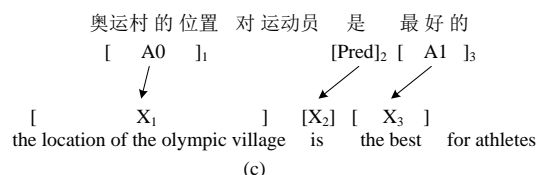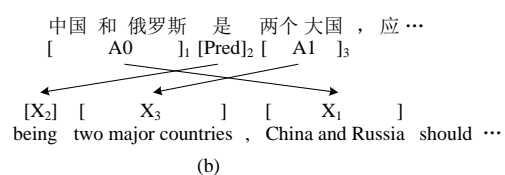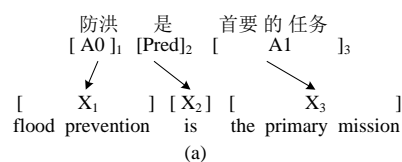

Figure 1. An example of ambiguous PASs.

Meanwhile, Figure 1 also depicts another kind of PAS ambiguity. From Figure 1, we can see that (a) and (c) get the same source-side PAS and target-side-like PAS. However, they are different because in Figure 1(c), there is a gap string "对 运动员" between [A0] and [Pred]. Generally, the gap strings are due to the low recall of automatic semantic role labeling (SRL) or complex sentence structures. For example, in Figure 1(c), the gap string "对 运动员" is actually an argument "AM-PRP" of the PAS, but the SRL system has

---

[1] We use *target-side-like PAS* to refer to a list of general non-terminals in target language order, where a non-terminal aligns to a source argument.

ignored it. We call this kind of PAS ambiguity *gap ambiguity*.

During translation, these PAS ambiguities will greatly affect the PAS-based translation models. Therefore, in order to incorporate the bilingual PAS into machine translation effectively, we need to decide which target-side-like PAS should be chosen for a specific source-side PAS. We call this task *PAS disambiguation*.

In this paper, we propose two novel methods to incorporate rich context information to handle PAS ambiguities. Towards the gap ambiguity, we adopt a method called *inside context integration* to extend PAS to IC-PAS. In terms of IC-PAS, the gap strings are combined effectively to deal with the gap ambiguities. As to the role ambiguity, we design a novel *maximum entropy PAS disambiguation (MEPD)* model to combine various context features, such as context words of PAS. For each ambiguous source-side PAS, we build a specific MEPD model to select appropriate target-side-like PAS for translation. We will detail the two methods in Section 3 and 4 respectively.

Finally, we integrate the above two methods into a PAS-based translation framework (Zhai et al. 2012). Experiments show that the two PAS disambiguation methods significantly improve the baseline translation system. The main contribution of this work can be concluded as follows:

1) We define two kinds of PAS ambiguities: role ambiguity and gap ambiguity. To our best knowledge, we are the first to handle these PAS ambiguities for SMT.

2) Towards the two different ambiguities, we design two specific methods for PAS disambiguation: inside context integration and the novel MEPD model.

2 PAS-based Translation Framework

PAS-based translation framework is to perform translation based on PAS transformation (Zhai et al., 2012). In the framework, a source-side PAS is first converted into target-side-like PASs by PAS transformation rules, and then perform translation based on the obtained target-side-like PASs.

2.1 PAS Transformation Rules

PAS transformation rules (PASTR) are used to convert a source-side PAS into a target one. Formally, a PASTR is a triple *<Pred, SP, TP>*:

- *Pred* means the predicate where the rule is extracted.
- *SP* denotes the list of source elements in source language order.
- *TP* refers to the target-side-like PAS, i.e., a list of general non-terminals in target language order.

For example, Figure 2 shows the PASTR extracted from Figure 1(a). In this PASTR, *Pred* is Chinese verb "是", *SP* is the source element list $<[A0]_1 [Pred]_2 [A1]_3>$, and *TP* is the list of non-terminals $<X_1 X_2 X_3>$. The same subscript in *SP* and *TP* means a one-to-one mapping between a source element and a target non-terminal. Here, we utilize the source element to refer to the predicate or argument of the source-side PAS.
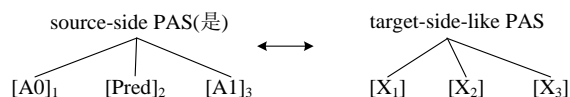
source-side PAS(是)   target-side-like PAS

$[A0]_1$ $[Pred]_2$ $[A1]_3$ $[X_1]$ $[X_2]$ $[X_3]$

Figure 2. An example PASTR.

2.2 PAS Decoding

The PAS decoding process is divided into 3 steps:

(1) *PAS acquisition*: perform semantic role labeling (SRL) on the input sentences to achieve their PASs, i.e., source-side PASs;

(2) *Transformation*: use the PASTR to match the source-side PAS i.e., the predicate *Pred* and the source element list *SP*. Then by the matching PASTRs, transform source-side PASs to target-side-like PASs.

(3) *Translation*: in this step, the decoder first translates each source element respectively, and then a CKY-style decoding algorithm is adopted to combine the translation of each element and get the final translation of the PAS.

2.3 Sentence Decoding with the PAS-based translation framework

Sometimes, the source sentence cannot be fully covered by the PAS, especially when there are several predicates. Thus to translate the whole sentence, Zhai et al. (2012) further designed an algorithm to decode the entire sentence.

In the algorithm, they organized the space of translation candidates into a hypergraph. For the span covered by PAS (PAS span), a multiple-branch hyperedge is employed to connect it to the PAS's elements. For the span not covered by PAS (non-PAS span), the decoder considers all the possible binary segmentations of it and utilizes binary hyperedges to link them.

During translation, the decoder fills the spans with translation candidates in a bottom-up manner. For the PAS span, the PAS-based translation framework is adopted. Otherwise, the BTG system (Xiong et al., 2006) is used. When the span covers the whole sentence, we get the final translation result.

Obviously, PAS ambiguities are not considered in this framework at all. The target-side-like PAS is selected only according to the language model and translation probabilities, without considering any context information of PAS. Consequently, it would be difficult for the decoder to distinguish the source-side PAS from different context. This harms the translation quality. Thus to overcome this problem, we design two novel methods to cope with the PAS ambiguities: inside-context integration and a maximum entropy PAS disambiguation (MEPD) model. They will be detailed in the next two sections.

3 Inside Context Integration

In this section, we integrate the inside context of the PAS into PASTRs to do PAS disambiguation. Basically, a PAS consists of several elements (a predicate and several arguments), which are actually a series of continuous spans. For a specific PAS $<E_1,\ldots, E_n>$, such as the source-side PAS $<[A0][Pred][A1]>$ in Figure 2, its *controlled range* is defined as:

$$range(PAS) = \{s(E_i), \forall i \in [1,n]\}$$

where $s(E_i)$ denotes the span of element E_i. Further, we define the *closure range* of a PAS. It refers to the shortest continuous span covered by the entire PAS:

$$closure\_range = \left[\min_{j \in s(E_0)} j, \max_{j \in s(E_n)} j\right]$$

Here, E_0 and E_n are the leftmost and rightmost element of the PAS respectively. The closure range is introduced here because adjacent source elements in a PAS are usually separated by gap strings in the sentence. We call these gap strings the *inside context* (IC) of the PAS, which satisfy:

$$closure\_range\,(PAS) = \oplus\,(\,IC(PAS) \bigcup range(PAS)\,)$$

The operator $\oplus$ takes a list of neighboring spans as input[2], and returns their combined continuous span. As an example, towards the PAS "<[A0][Pred][A1]>" (the one for Chinese predicate "是 (shi)") in Figure 3, its controlled range is $\{[3,5],[8,8],[9,11]\}$ and its closure range is $[3,11]$. The IC of the PAS is thus $\{[6,7]\}$.

To consider the PAS's IC during PAS transformation process, we incorporate its IC into the extracted PASTR. For each gap string in IC, we abstract it by the sequence of highest node categories (named as s-tag sequence). The s-tag sequence dominates the corresponding syntactic tree fragments in the parse tree. For example, in Figure 3, the s-tag sequence for span $[6,8]$ is "PP VC". Thus, the sequence for the IC (span $[6,7]$) in Figure 3 is "PP". We combine the s-tag sequences with elements of the PAS in order. The resulting PAS is called IC-PAS, just like the left side of Figure 4(b) shows.
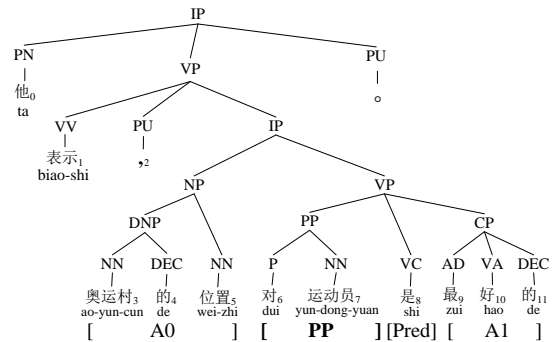

Figure 3. The illustration of inside context (IC). The subscript in each word refers to its position in sentence.

Differently, Zhai et al. (2012) attached the IC to its neighboring elements based on parse trees. For example, in Figure 3, they would attach the gap string "对(dui) 运动员(yun-dong-yuan)" to the PAS's element "Pred", and then the span of "Pred" would become $[6,8]$. Consequently, the span $[6,8]$ will be translated as a whole source element in the decoder. This results in a bad translation because the gap string "对(dui) 运动员 (yun-dong-yuan)" and predicate "是(shi)" should be translated separately, just as Figure 4(a) shows. Therefore, we can see that the attachment decision in (Zhai et al., 2012) is sometimes unreasonable and the IC also cannot be used for PAS disambiguation at all. In contrast, our meth-

---

[2] Here, two spans are neighboring means that the beginning of the latter span is the former span's subsequent word in the sentence. For example, span $[3,6]$ and $[7,10]$ are neighboring spans.

od of inside context integration is much flexible and beneficial for PAS disambiguation.
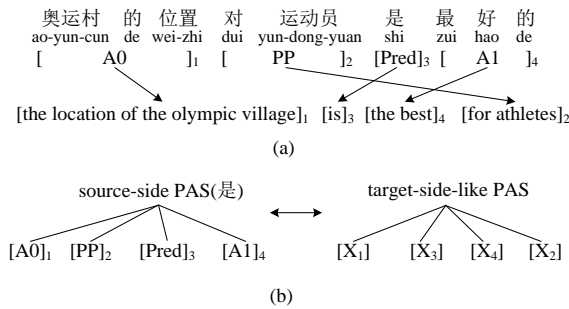

Figure 4. Example of IC-PASTR. (a) The aligned span of each element of the PAS in Figure 3; (b) The extracted IC-PASTR from (a).

Using the IC-PASs, we look for the aligned target span for each element of the IC-PAS. We demand that every element and its corresponding target span must be consistent with word alignment. Otherwise, we discard the IC-PAS. Afterwards, we can easily extract a rule for PAS transformation, which we call IC-PASTR. As an example, Figure 4(b) is the extracted IC-PASTR from Figure 4(a).

Note that we only apply the source-side PAS and word alignment for IC-PASTR extraction. By contrast, Zhai et al. (2012) utilized the result of bilingual SRL (Zhuang and Zong, 2010b). Generally, bilingual SRL could give a better alignment between bilingual elements. However, bilingual SRL usually achieves a really low recall on PASs, about 226,968 entries in our training set while it is 882,702 by using monolingual SRL system. Thus to get a high recall for PASs, we only utilize word alignment instead of capturing the relation between bilingual elements. In addition, to guarantee the accuracy of IC-PASTRs, we only retain rules with more than 5 occurrences.

4 Maximum Entropy PAS Disambiguation (MEPD) Model

In order to handle the role ambiguities, in this section, we concentrate on utilizing a maximum entropy model to incorporate the context information for PAS disambiguation. Actually, the disambiguation problem can be considered as a multi-class classification task. That is to say, for a source-side PAS, every corresponding target-side-like PAS can be considered as a label. For example, in Figure 1, for the source-side PAS "$[A0]_1[Pred]_2[A1]_3$", the target-side-like PAS "$[X_1]$ $[X_2]$ $[X_3]$" in Figure 1(a) is thus a label and "$[X_2]$ $[X_3]$ $[X_1]$" in Figure 1(b) is another label of this classification problem.

The maximum entropy model is the classical way to handle this problem:

$$P_\theta(tp \mid sp, c(sp), c(tp)) = \frac{\exp(\sum_i \theta_i h_i(sp, tp, c(sp), c(tp)))}{\sum_{tp'} \exp(\sum_i \theta_i h_i(sp, tp, c(sp), c(tp)))}$$

where sp and tp refer to the source-side PAS (not including the predicate) and the target-side-like PAS respectively. $c(sp)$ and $c(tp)$ denote the surrounding context of sp and tp. h_i is a binary feature function and θ_i is the weight of h_i.

We train a maximum entropy classifier for each sp via the off-the-shelf MaxEnt toolkit [3]. Note that to avoid sparseness, sp does not include predicate of the PAS. Practically, the predicate serves as a feature of the MEPD model. As an example, for the rule illustrated in Figure 4(b), we build a MEPD model for its source element list sp <[A0] [PP] [Pred] [A1]>, and integrate the predicate "是(shi)" into the MEPD model as a feature.

In detail, we design a list of features for each pair <sp, tp> as follows:

● **Lexical Features**. These features include the words immediately to the left and right of sp, represented as w_{-1} and w_{+1}. Moreover, the head word of each argument also serves as a lexical feature, named as hw(E_i). For example, Figure 3 shows the context of the IC-PASTR in Figure 4(b), and the extracted lexical features of the instance are: w_{-1}=，, w_{+1}=。, hw($[A0]_1$)=位置 (wei-zhi), hw($[A1]_4$)=好(hao).

● **POS Features**. These features are defined as the POS tags of the lexical features, p_{-1}, p_{+1} and phw(E_i) respectively. Thus, the corresponding POS features of Figure 4 (b) are: p_{-1}=PU, p_{+1}=PU, phw($[A0]_1$)=NN, phw($[A1]_4$)=VA.

● **Predicate Feature**. It is the pair of source predicate and its corresponding target predicate. For example, in Figure 4(b), the source and target predicate are "是(shi)" and "is" respectively. The predicate feature is thus "PredF=是(shi)+is". The target predicate is determined by:

$$t\text{-}pred = \underset{j \in t\_range(PAS)}{\arg\max} \ p(t_j \mid s\text{-}pred)$$

where s-$pred$ is the source predicate and t-$pred$ is the corresponding target predicate.

t_range(PAS) refers to the target range covering all the words that are reachable from the PAS via word alignment. t_j refers to the *jth* word in *t_range(PAS)*. The utilized lexical translation probabilities are from the toolkit in Moses (Koehn et al., 2007).

- **Syntax Features**. These features include st(*Ei*), i.e., the highest syntax tag for each argument, and fst(PAS) which is the lowest father node of *sp* in the parse tree. For example, for the rule shown in Figure 4(b), syntax features are st($[A0]_1$)=NP, st($[A1]_4$)=CP, and fst(PAS)=IP respectively.

Using these features, we can train the MEPD model. We set the Gaussian prior to 1.0 and perform 100 iterations of the L-BFGS algorithm for each MEPD model. At last, we build 160 and 215 different MEPD classifiers, respectively, for the PASTRs and IC-PASTRs. Note that since the training procedure of maximum entropy classifier is really fast, it does not take much time to train these classifiers.

5 Integrating into the PAS-based Translation Framework

In this section, we integrate our method of PAS disambiguation into the PAS-based translation framework when translating each test sentence.

For inside context integration, since the format of IC-PASTR is the same to PASTR[4], we can use the IC-PASTR to substitute PASTR for building a PAS-based translation system directly. We use "IC-PASTR" to denote this system. In addition, since our method of rule extraction is different from (Zhai et al., 2012), we also use PASTR to construct a translation system as the baseline system, which we call "PASTR".

On the basis of PASTR and IC-PASTR, we further integrate our MEPD model into translation. Specifically, we take the score of the MEPD model as another informative feature for the decoder to distinguish good target-side-like PASs from bad ones. The weights of the MEPD feature can be tuned by MERT (Och, 2003) together with other translation features, such as language model.

6 Related Work

The method of PAS disambiguation for SMT is relevant to the previous work on context depend-

ent translation.

Carpuat and Wu (2007a, 2007b) and Chan et al. (2007) have integrated word sense disambiguation (WSD) and phrase sense disambiguation (PSD) into SMT systems. They combine rich context information to do disambiguation for words or phrases, and achieve improved translation performance.

Differently, He et al. (2008), Liu et al. (2008) and Cui et al. (2010) designed maximum entropy (ME) classifiers to do better rule section for hierarchical phrase-based model and tree-to-string model respectively. By incorporating the rich context information as features, they chose better rules for translation and yielded stable improvements on translation quality.

Our work differs from the above work in the following two aspects: 1) in our work, we focus on the problem of disambiguates on PAS; 2) we define two kinds of PAS ambiguities: role ambiguity and gap ambiguity. 3) towards the two different ambiguities, we design two specific methods for PAS disambiguation: inside context integration and the novel MEPD model.

In addition, Xiong et al. (2012) proposed an argument reordering model to predict the relative position between predicates and arguments. They also combine the context information in the model. But they only focus on the relation between the predicate and a specific argument, rather than the entire PAS. Different from their work, we incorporate the context information to do PAS disambiguation based on the entire PAS. This is very beneficial for global reordering during translation (Zhai et al., 2012).

7 Experiment

7.1 Experimental Setup

We perform Chinese-to-English translation to demonstrate the effectiveness of our PAS disambiguation method. The training data contains about 260K sentence pairs[5]. To get accurate SRL results, we ensure that the length of each sentence in the training data is among 10 and 30 words. We run GIZA++ and then employ the *grow-diag-final-and (gdfa)* strategy to produce symmetric word alignments. The development set and test set come from the NIST evaluation test data (from 2003 to 2005). Similar to the training set, we also only retain the sentences

---

[4] The only difference between IC-PASTR and PASTR is that there are many syntactic labels in IC-PASTRs.

whose lengths are among 10 and 30 words. Finally, the development set includes 595 sentences from NIST MT03 and the test set contains 1,786 sentences from NIST MT04 and MT05.

We train a 5-gram language model with the Xinhua portion of English Gigaword corpus and target part of the training data. The translation quality is evaluated by case-insensitive BLEU-4 with shortest length penalty. The statistical significance test is performed by the re-sampling approach (Koehn, 2004).

We perform SRL on the source part of the training set, development set and test set by the Chinese SRL system used in (Zhuang and Zong, 2010b). To relieve the negative effect of SRL errors, we get the multiple SRL results by providing the SRL system with 3-best parse trees of Berkeley parser (Petrov and Klein, 2007), 1-best parse tree of Bikel parser (Bikel, 2004) and Stanford parser (Klein and Manning, 2003). Therefore, at last, we can get 5 SRL result for each sentence. For the training set, we use these SRL results to do rule extraction respectively. We combine the obtained rules together to get a combined rule set. We discard the rules with fewer than 5 appearances. Using this set, we can train our MEPD model directly.

As to translation, we match the 5 SRL results with transformation rules respectively, and then apply the resulting target-side-like PASs for decoding. As we mentioned in section 2.3, we use the state-of-the-art BTG system to translate the non-PAS spans.

source-side PAS	counts	number of classes
[A0] [Pred(是)] [A1]	245	6
[A0] [Pred(说)] [A1]	148	6
[A0] [AM-ADV] [Pred(是)] [A1]	68	20
[A0] [Pred(表示)] [A1]	66	6
[A0] [Pred(有)] [A1]	42	6
[A0] [Pred(认为)] [A1]	32	4
[A0] [AM-ADV] [Pred(有)] [A1]	32	19
[A0] [Pred(指出)] [A1]	29	4
[AM-ADV] [Pred(有)] [A1]	26	6
[A2] [Pred(为)] [A1]	16	5

Table 1. The top 10 frequent source-side PASs in the dev and test set.

7.2 Ambiguities in Source-side PASs

We first give Table 1 to show some examples of role ambiguity. In the table, for instance, the second line denotes that the source-side PAS "[A0] [Pred(说)] [A1]" appears 148 times in the devel-

opment and test set all together, and it corresponds to 6 different target-side-like PASs in the training set.

As we can see from Table 1, all the top 10 PASs correspond to several different target-side-like PASs. Moreover, according to our statistics, among all PASs appearing in the development set and test set, 56.7% of them carry gap strings. These statistics demonstrate the importance of handling the role ambiguity and gap ambiguity in the PAS-based translation framework. Therefore, we believe that our PAS disambiguation method would be helpful for translation.

7.3 Translation Result

We compare the translation result using PASTR, IC-PASTR and our MEPD model in this section. The final translation results are shown in Table 2. As we can see, after employing PAS for translation, all systems outperform the baseline BTG system significantly. This comparison verifies the conclusion of (Zhai et al., 2012) and thus also demonstrates the effectiveness of PAS.

MT system	Test set	n-gram precision			
		1	2	3	4
BTG	32.75	74.39	41.91	24.75	14.91
PASTR	33.24*	75.28	42.62	25.18	15.10
PASTR+MEPD	33.78*	75.32	43.08	25.75	15.58
IC-PASTR	33.95*#	75.62	43.36	25.92	15.58
IC-PASTR+MEPD	**34.19*#**	75.66	43.40	26.15	15.92

Table 2. Result of baseline system and the MT systems using our PAS-based disambiguation method. The "*" and "#" denote that the result is significantly better than BTG and PASTR respectively (p<0.01).

Specifically, after integrating the inside context information of PAS into transformation, we can see that system IC-PASTR significantly outperforms system PASTR by 0.71 BLEU points. Moreover, after we import the MEPD model into system PASTR, we get a significant improvement over PASTR (by 0.54 BLEU points). These comparisons indicate that both the inside context integration and our MEPD model are beneficial for the decoder to choose better target-side-like PAS for translation.

On the basis of IC-PASTR, we further add our MEPD model into translation and get system IC-PASTR+MEPD. We can see that this system further achieves a remarkable improvement over system PASTR (0.95 BLEU points).

However, from Table 2, we find that system IC-PASTR+MEPD only outperforms system IC-PASTR slightly (0.24 BLEU points). The result seems to show that our MEPD model is not such

useful after using IC-PASTR. We will explore the reason in section 7.5.

7.4 Effectiveness of Inside Context Integration

The method of inside context integration is used to combine the inside context (gap strings) into PAS for translation, i.e., extend the PASTR to IC-PASTR. In order to demonstrate the effectiveness of inside context integration, we first give Table 3, which illustrates statistics on the matching PASs. The statistics are conducted on the combination of development set and test set.

Transformation Rules	Matching PAS		
	None Gap PAS	Gap PAS	Total
PASTR	1702	1539	3241
IC-PASTR	1546	832	2378

Table 3. Statistics on the matching PAS.

In Table 3, for example, the line for PASTR means that if we use PASTR for the combined set, 3241 PASs (column "Total") can match PASTRs in total. Among these matching PASs, 1539 ones (column "Gap PAS") carry gap strings, while 1702 ones do not (column "None Gap PAS"). Consequently, since PASTR does not consider the inside context during translation, the Gap PASs, which account for 47% (1539/3241) of all matching PASs, might be handled inappropriately in the PAS-based translation framework. Therefore, integrating the inside context into PASTRs, i.e., using the proposed IC-PASTRs, would be helpful for translation. The translation result shown in Table 2 also demonstrates this conclusion.

这 对 经济 复苏 、 尤其是 恢复 投资 信心 是 一个 好 兆头
[A0] [PP] [Pred] [A1]

this is [a good sign] [for economic recovery and the restoration of investors' confidence]
(a) reference

这 对 经济 复苏 、 尤其是 恢复 投资 信心 是 一个 好 兆头
[A0] [Pred] [A1]

this [for economic recovery , especially of investment confidence is] [a good sign]
(b) translation result using PASTR

这 对 经济 复苏 、 尤其是 恢复 投资 信心 是 一个 好 兆头
[A0] [PP] [Pred] [A1]

this is [a good sign] [for economic recovery , especially of investment confidence]
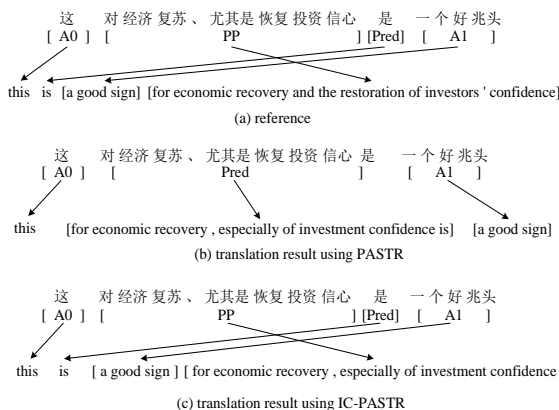(c) translation result using IC-PASTR

Figure 5. Translation examples to verify the effectiveness of inside context.

From Table 3, we can also find that the number of matching PASs decreases after using IC-PASTR. This is because IC-PASTR is more specific than PASTR. Therefore, for a PAS with specific inside context (gap strings), even if the matched PASTR is available, the matched IC-PASTR might not. This indicates that comparing with PASTR, IC-PASTR is more capable of distinguishing different PASs. Therefore, based on this advantage, although the number of matching PASs decreases, IC-PASTR still improves the translation system using PASTR significantly. Of course, we believe that it is also possible to integrate the inside context without decreasing the number of matching PASs and we plan this as our future work.

We further give a translation example in Figure 5 to illustrate the effectiveness of our inside context integration method. In the example, the automatic SRL system ignores the long preposition phrase "对 经济复苏 、 尤其是 恢复 投资信心" for the PAS. Thus, the system using PASTRs can only attach the long phrase to the predicate "是" according to the parse tree, and meanwhile, make use of a transformation rule as follows:

source-side PAS(是)   target-side-like PAS

[A0]$_1$ [Pred]$_2$ [A1]$_3$ [X$_1$]   [X$_2$]   [X$_3$]
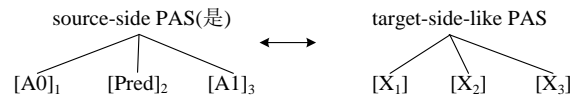
In this way, the translation result is very bad, just as Figure 5(b) shows. The long preposition phrases are wrongly positioned in the translation.

In contrast, after inside context integration, we match the inside context during PAS transformation. As Figure 5(c) shows, the inside context helps to selects a right transformation rule as follows and gets a good translation result finally.

source-side PAS(是)   target-side-like PAS

[A0]$_1$ [PP]$_2$ [Pred]$_3$ [A1]$_4$ [X$_1$] [X$_3$] [X$_4$] [X$_2$]
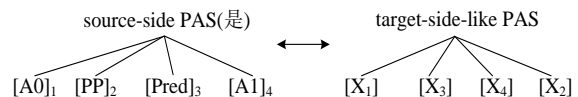
7.5 Effectiveness of the MEPD Model

The MEPD model incorporates various context features to select better target-side-like PAS for translation. On the basis of PASTR and IC-PASTR, we build 160 and 215 different MEPD classifies, respectively, for the frequent source-side PASs.

In Table 2, we have found that our MEPD model improves system IC-PASTR slightly. We conjecture that this phenomenon is due to two possible reasons. On one hand, sometimes, many PAS ambiguities might be resolved by both inside context and the MEPD model. Therefore, the improvement would not be such significant

when we combine these two methods together. On the other hand, as Table 3 shows, the number of matching PASs decreases after using IC-PASTR. Since the MEPD model works on PASs, its effectiveness would also weaken to some extent. Future work will explore this phenomenon more thoroughly.

| Ref	... , [海牙] [是] [其 最后 一站] 。 ... [the hague] [is] [his last stop] .
PASTR	... , [海牙]$_{A0}$ [是]$_{Pred}$ [其 最后 一站]$_{A1}$ 。 ... [is] [his last leg of] [the hague] .
PASTR + MEPD	... , [海牙]$_{A0}$ [是]$_{Pred}$ [其 最后 一站]$_{A1}$ 。 ... [the hague] [is] [the last leg] .

Figure 6. Translation examples to demonstrate the effectiveness of our MEPD model.

Now, we give Figure 6 to demonstrate the effectiveness of our MEPD model. From the Figure, we can see that the system using PASTRs selects an inappropriate transformation rule for translation:

source-side PAS(是)   target-side-like PAS

[A0]$_1$ [Pred]$_2$ [A1]$_3$ [X$_2$]   [X$_3$]   [X$_1$]
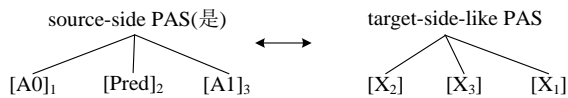
This rule wrongly moves the subject "海牙 (Hague)" to the end of the translation. We do not give the translation result of the BTG system here because it makes the same mistake.

Conversely, by considering the context information, the PASTR+MEPD system chooses a correct rule for translation:

source-side PAS(是)   target-side-like PAS

[A0]$_1$ [Pred]$_2$ [A1]$_3$ [X$_1$]   [X$_2$]   [X$_3$]
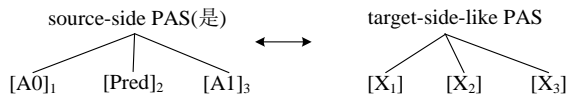
As we can see, the used rule helps to keep the SVO structure unchanged, and gets the correct translation.

8   Conclusion and Future Work

In this paper, we focus on the problem of ambiguities for PASs. We first propose two ambiguities: gap ambiguity and role ambiguity. Accordingly, we design two novel methods to do efficient PAS disambiguation: inside-context integration and a novel MEPD model. For inside context integration, we abstract the inside con-

text and combine them into the PASTRs for PAS transformation. Towards the MEPD model, we design a maximum entropy model for each ambitious source-side PASs. The two methods successfully incorporate the rich context information into the translation process. Experiments show that our PAS disambiguation methods help to improve the translation performance significantly.

In the next step, we will conduct experiments on other language pairs to demonstrate the effectiveness of our PAS disambiguation method. In addition, we also will try to explore more useful and representative features for our MEPD model.

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.

Daniel Bikel. (2004). Intricacies of Collins parsing model. *Computational Linguistics*, 30(4):480-511.

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

Marine Carpuat and Dekai Wu. 2007a. How phrase-sense disambiguation outperforms word sense disambiguation for statistical machine translation. In *11th Conference on Theoretical and Methodological Issues in Machine Translation*, pages 43–52.

Marine Carpuat and Dekai Wu. 2007b. Improving statistical machine translation using word sense disambiguation. In *Proceedings of EMNLP-CoNLL 2007*, pages 61–72.

Yee Seng Chan, Hwee Tou Ng, and David Chiang. 2007. Word sense disambiguation improves statistical machine translation. In *Proc. ACL 2007*, pages 33–40.

Lei Cui, Dongdong Zhang, Mu Li, Ming Zhou and Tiejun Zhao. A Joint Rule Selection Model for Hierarchical Phrase-Based Translation. In *Proc. of ACL 2010*.

1134

Jason Eisner. (2003). Learning non-isomorphic tree mappings for machine translation. In *Proc. of ACL 2003*.

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.

Pascale Fung, Zhaojun Wu, Yongsheng Yang and Dekai Wu. (2007). Learning bilingual semantic frames: shallow semantic sarsing vs. semantic sole projection. In *Proceedings of the 11th Conference on Theoretical and Methodological Issues in Machine Translation*, pages 75-84.

Qin Gao and Stephan Vogel. (2011). Utilizing target-side semantic role labels to assist hierarchical phrase-based machine translation. In *Proceedings of Fifth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 107–115, Portland, Oregon, USA, June 2011. Association for Computational Linguistics

Zhongjun He, Qun Liu, and Shouxun Lin. 2008. Improving statistical machine translation using lexicalized rule selection. In *Proc. of Coling 2008*, pages 321–328.

Franz Josef Och. (2003). Minimum error rate training in statistical machine translation. In *Proc. of ACL 2003*, pages 160‑167.

Franz Josef Och and Hermann Ney. (2004). The alignment template approach to statistical machine translation. *Computational Linguistics*, 30:417‑449.

Dan Klein and Christopher D. Manning. (2003). Accurate unlexicalized parsing. In *Proc. of ACL-2003*, pages 423-430.

Philipp Koehn, Franz Joseph Och, and Daniel Marcu. (2003). Statistical phrase-based translation. In *Proceedings of NAACL 2003*, 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.

P Koehn, H Hoang, A Birch, C Callison-Burch, M Federico, N Bertoldi, B Cowan, W Shen, C Moran and R Zens, (2007). Moses: Open source toolkit for statistical machine translation. In *Proc. of ACL 2007*. pages 177–180, Prague, Czech Republic, June. Association for Computational Linguistics.

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

Ding Liu and Daniel Gildea. (2008). Improved tree-to-string transducer for machine Translation. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 62–69, Columbus, Ohio, USA, June 2008.

Ding Liu and Daniel Gildea. (2010). Semantic role features for machine translation. In *Proc. of Coling 2010*, pages 716–724, Beijing, China, August.

Qun Liu, Zhongjun He, Yang Liu, and Shouxun Lin. Maximum Entropy based Rule Selection Model for Syntax-based Statistical Machine Translation. In *Proc. of EMNLP 2008*.

Yang Liu, Qun Liu and Shouxun Lin. (2006). Tree-to-string alignment template for statistical machine translation. In *Proc. of ACL-COLING 2006*.

Daniel Marcu, Wei Wang, Abdessamad Echihabi and Kevin Knight. (2006). SPMT: Statistical machine translation with syntactified target language phrases. In *Proc. of EMNLP 2006*, pages 44-52.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. (2002). Bleu: a method for automatic evaluation of machine translation. In *Proc. ACL 2002*, 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 modelling toolkit. In *Proceedings of the 7th International Conference on Spoken Language Processing*, pages 901–904, Denver, Colorado, USA, September.

Dekai Wu and Pascale Fung. (2009a). Can semantic role labelling 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 *Proc. NAACL 2009*, pages 13–16, Boulder, Colorado, June.

ShuminWu and Martha Palmer. (2011). Semantic mapping using automatic word alignment and semantic role labelling. In *Proceedings of Fifth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 21–30, Portland, Oregon, USA, June 2011.

Xianchao Wu, Katsuhito Sudoh, Kevin Duh, Hajime Tsukada, and Masaaki Nagata. (2011). Extracting preordering rules from predicate-argument structures. In *Proc. IJCNLP 2011*, pages 29–37, Chiang Mai, Thailand, November.

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.

Deyi Xiong, Min Zhang, and Haizhou Li. (2012). Modelling the translation of predicate-argument structure for smt. In *Proc. of ACL 2012*, pages 902–911, Jeju, Republic of Korea, 8-14 July 2012.

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

Feifei Zhai, Jiajun Zhang, Yu Zhou and Chengqing Zong. Machine Translation by Modeling Predicate-Argument Structure Transformation. In *Proc. of COLING 2012*.

Hui Zhang, Min Zhang, Haizhou Li and Eng Siong Chng. (2010). Non-isomorphic Forest Pair Translation. In *Proceedings of EMNLP 2010*, pages 440-450, Massachusetts, USA, 9-11 October 2010.

Tao Zhuang, and Chengqing Zong. (2010a). A minimum error weighting combination strategy for chinese semantic role labelling. In P*roceedings of COLING-2010*, pages 1362-1370.

Tao Zhuang and Chengqing Zong. (2010b). Joint inference for bilingual semantic role labelling. In *Proceedings of EMNLP 2010*, pages 304–314, Massachusetts, USA, 9-11 October 2010.