Frontiers of Information Technology & Electronic Engineering
www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Incorporating target language semantic roles into a string-to-tree translation model*

Chao SU[†1,3], Yu-hang GUO[1], He-yan HUANG[†‡1,2], Shu-min SHI[1,2], Chong FENG[1,2]
([1]School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China)
([2]Beijing Engineering Research Center of High Volume Language Information Processing and Cloud Computing Applications, Beijing 100081, China)
([3]Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing 100048, China)
[†]E-mail: suchao@bit.edu.cn; hhy63@bit.edu.cn

Received June 18, 2016; Revision accepted Nov. 30, 2016; Crosschecked Nov. 3, 2017

Abstract: The string-to-tree model is one of the most successful syntax-based statistical machine translation (SMT) models. It models the grammaticality of the output via target-side syntax. However, it does not use any semantic information and tends to produce translations containing semantic role confusions and error chunk sequences. In this paper, we propose two methods to use semantic roles to improve the performance of the string-to-tree translation model: (1) adding role labels in the syntax tree; (2) constructing a semantic role tree, and then incorporating the syntax information into it. We then perform string-to-tree machine translation using the newly generated trees. Our methods enable the system to train and choose better translation rules using semantic information. Our experiments showed significant improvements over the state-of-the-art string-to-tree translation system on both spoken and news corpora, and the two proposed methods surpass the phrase-based system on large-scale training data.

Key words: Machine translation; Semantic role; Syntax tree; String-to-tree
https://doi.org/10.1631/FITEE.1601349 **CLC number:** TP391

1 Introduction

Machine translation (MT) is one of the most promising applications of natural language processing (NLP) technology, and one of the most difficult to accomplish well. Statistical machine translation (SMT) techniques (Brown *et al.*, 1990), learning statistical models from large amounts of data, have experienced significant progress including word-based (Brown *et al.*, 1993), phrase-based (Koehn *et al.*, 2003; Och and Ney, 2004), and tree-based methods, including both formal grammar based (Wu, 1995; 1996; Chiang, 2005) and linguistic syntax based (Yamada and Knight, 2001; Galley *et al.*, 2004; Liu *et al.*, 2006; Marcu *et al.*, 2006; Mi *et al.*, 2008) methods. The syntax-based methods have been research hotspots due to their modeling of structural or syntactic aspect of language.

The string-to-tree model (Galley *et al.*, 2004; Marcu *et al.*, 2006) is one of the most successful syntax-based models. It employs translation rules that represent the source side as strings and the target side as syntactic structures. One of the limitations in the string-to-tree model is that it does not use any useful semantic information. As a result, it tends to produce error translations containing confusion with respect to semantic roles and error sequences of chunks. Those errors will mislead readers and cause their misunderstanding of the essential meaning of the original sentences: who did what to whom, how, where, when,

[†] Corresponding author
[*] Project supported by the National Natural Science Foundation of China (Nos. 61132009, 61201352, 61502035, and 61201351), the National Basic Research Program (973) of China (No. 2013CB329303), and the Beijing Advanced Innovation Center for Imaging Technology (No. BAICIT-2016007)
ORCID: Chao SU, http://orcid.org/0000-0001-6771-329X; He-yan HUANG, http://orcid.org/0000-0002-0320-7520

and why. Our goal is to take advantage of semantic role labeling to improve the translation quality of string-to-tree translation models.

We replace the target-language syntax tree with our proposed two kinds of trees in the string-to-tree translation model. These two kinds of trees are the syntax-role tree (SRT) and role-syntax tree (RST). In SRT, we incorporate semantic roles in the syntax tree; in RST, we transform semantic role labeling results into a tree, and then incorporate syntactic and lexical information. In the experimental results, SRT shows an improvement of 0.56 BiLingual Evaluation Understudy (BLEU) points on the Broad Operational Language Translation (BOLT) spoken corpus, and an improvement of 1.33 BLEU points on the Foreign Broadcast Information Service (FBIS) news corpus on average. The result of RST is poorer than that of SRT on the BOLT and FBIS corpora, but better on large-scale training data (an improvement of 1.13 BLEU points).

1.1 Syntax-based machine translation

Galley *et al.* (2004) described their whole process of transformation from a string of source symbols to a target syntactic tree, and proposed algorithms for extracting syntactically motivated translation rules. Liu *et al.* (2006) employed a parser on source language and extracted tree-to-string alignment templates from source-side parsed parallel texts. Liu and Gildea (2008) improved the tree-to-string (TTS) transducer through normalizing the TTS templates, building a syntax-based word alignment model, and modeling the tree decomposition.

In another direction, to alleviate the problem that the tree-based systems use only 1-best parse to direct the translation and tend to make mistakes due to parsing errors, Huang and Chiang (2005) conducted an experiment using *k*-best parsing. Mi *et al.* (2008) proposed a forest-based approach that encodes many more alternatives than the standard *k*-best lists. Liu and Liu (2010) proposed a joint decoder that produces simultaneously a parse tree on the source side and a translation on the target side.

1.2 Semantic role labeling

SRL identifies the semantic relationships filled by constituents of a sentence within a semantic frame (Gildea and Jurafsky, 2002). It is a process to assign WHO did WHAT to WHOM, WHEN, WHERE, WHY, HOW, etc. structures to sentences. It has been used in many NLP applications, such as information extraction, question answering, and summarization.

Thanks to hand-crafted resources, such as FrameNet (Baker *et al.*, 1998), PropBank (Palmer *et al.*, 2005), and NomBank (Meyers *et al.*, 2004), we have seen considerable achievements in SRL tasks in the last decade. Pradhan *et al.* (2004) improved the performance using an algorithm based on support vector machines.

Semantic roles are well known to be helpful in improving translation accuracy, because they tend to agree better between two languages than syntactic structures and constitute the skeleton of a sentence (Liu and Gildea, 2010). There is also a significant difference caused by the roles of the noun phrase (NP) between different languages such as English and Chinese (Liu and Gildea, 2008). Inspired by Liu and Gildea (2008) and Wu and Fung (2009), we try to merge semantic roles with syntax information in one string-to-tree model.

1.3 Motivations for SRT and RST

Because of a lack of semantic information, the string-to-tree model often creates translation with semantic role confusion. Adding the semantic role labeling results to it can help change the situation. A syntactic structure often undertakes a semantic role. For example, a noun phrase may be a role of ARG0 (agent) or ARG1 (patient). We argue that the syntactic structure can be the child node of a semantic argument in a tree. That is to say, the translation system can learn which syntactic structure often takes which semantic role from the data, and how to order the semantic chunks.

Furthermore, the semantic roles show some hierarchical structure. That is to say, a predicate and its arguments may become one argument of another predicate. For example, in Fig. 1, the predicate 'get' and its arguments, [ARG0 you] and [ARG1 it], constitute the [ARGM-TMP] argument of the predicate 'run'. This hierarchy inspires us to build a structure-like parse tree, which can help us understand the semantic relations among the chunks and improve translation performance.

Our SRT and RST retain all syntactic information and incorporate semantic role information. If there is overlap among role labels, some labels will be dropped. RST also drops the overlap labels and the syntactic information in RST is sometimes incomplete.

(a) When [ARG0 you] [TARGET get] [ARG1 it], run at top speed to buy a SLR.

(b) [ARGM-TMP When you get it], [TARGET run] [ARGM-MNR at top speed] [ARGM-PRP to buy a SLR].

(c) When you get it, run at top speed to [TARGET buy] [ARG1 a SLR].

Fig. 1 A sample SRL result for an English sentence

2 Methods

2.1 Constructing SRT

To construct SRT, we first apply syntax analysis and semantic role labeling on the target-side training corpus to obtain syntax trees and SRL results. Then we add the semantic role information to the syntax tree.

For example, the syntactic parsing result on the English sentence 'When you get it, run at top speed to buy an SLR.' is shown in Fig. 2. The result of SRL (Fig. 1) has multiple lines. Each line corresponds to one predicate and its arguments.

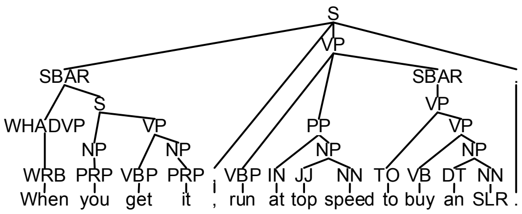

Fig. 2 A syntax tree for an English sentence

We add semantic role labels to the syntax tree. For each label, we obtain all the words it governs and find the largest syntactic structure in the syntax tree that contains exactly those words. We insert the label as a new node, whose child node would be the largest structure and whose parent node would be the structure's old parent. We obtain a new tree, called the 'syntax-role tree' (Fig. 3). For example, the argument [ARGM-MNR] governing leaf nodes 'at', 'top', 'speed', and 'PP' is the largest structure that exactly contains the three words. Then we insert a new node labeled ARGM-MNR between the node PP and its parent node VP. In Fig. 3, the shaded nodes are the added semantic role labels.

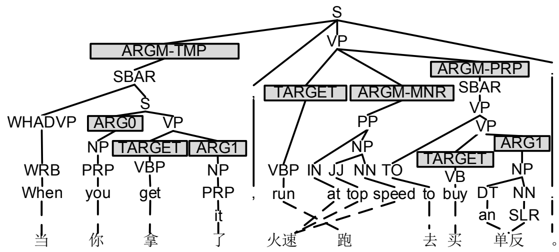

Fig. 3 A syntax-role tree for an English sentence with word alignment for an English-Chinese sentence pair

2.2 Constructing RST

To construct RST, we first transform the results of SRL into a tree, and then incorporate the syntactic information.

We find that the multiple lines in SRL results can be transformed to one tree, called the 'semantic role tree' as shown in Fig. 4. We also drop some overlapped role labels. We add suffixes (a, b, c) to differentiate the arguments corresponding to different predicates. If there are more than three predicates in a sentence, we use more suffixes (d, e, etc.). For example, the labels ending with suffixes 'a', 'b', and 'c' correspond to the predicate 'get', 'run', and 'buy', respectively.

The tree in Fig. 4 is too flat and lacks syntactic and lexical information. This would lead to a poor translation performance. To overcome it, we incorporate the results of syntax analysis and obtain a role-syntax tree (Fig. 5). In Fig. 5, the shaded nodes are the added syntax information.

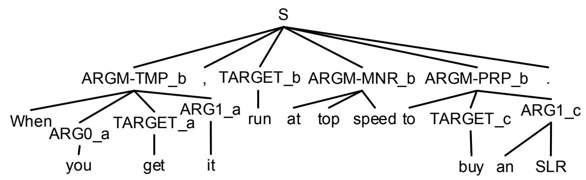

Fig. 4 A constructed semantic role tree for an English sentence (The tree is transformed from the SRL results)

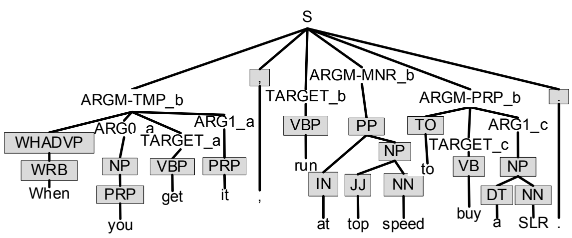

Fig. 5 An RST for an English sentence

2.3 Learning translation rules

In the training phase of SMT, we replace the syntax tree in the string-to-tree model with our SRT or RST, and train a translation system respectively. We obtain some improvement in terms of both BLEU and Meteor scores (see Section 5).

Like the learning of all other synchronous grammars, the string-to-tree model is learned from the word-aligned corpus. The grammar uses non-terminals like ‘X’ on the source side and tree labels like ‘NP’ on the target side. For example, the Chinese token ‘火速’ is word-aligned with the English words ‘at top speed’. When including the part-of-speech (POS), we obtain the correspondence as shown in Fig. 6a. To be a grammar rule, it needs a single node governing the three words on the target side. Fortunately, the three words are all child nodes of the semantic role ARGM-MNR. Using this role node, we obtain the complete sub-tree on the target side (Fig. 6b). Then we obtain the rule as shown in Fig. 6c.

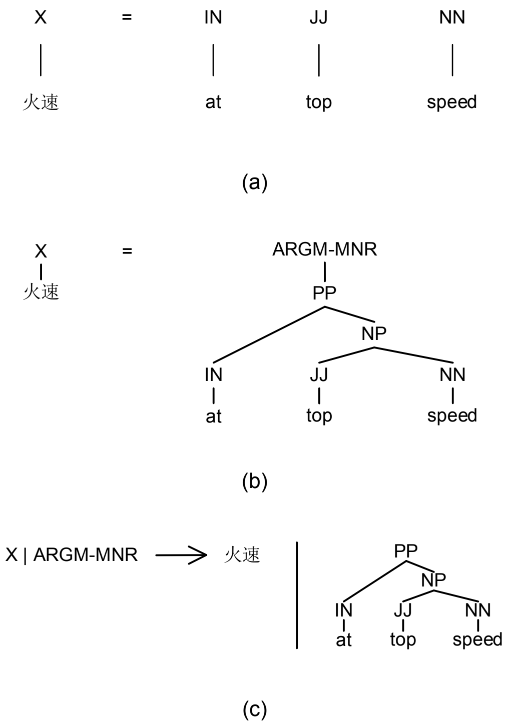

Fig. 6 An example of the learning translation rule from the example in Fig. 3: (a) a correspondence between a Chinese word and three English words with POS information; (b) a complete sub-tree on the target side; (c) a translation rule learned from this example

3 Experiments

3.1 Setup

We applied our methods on both spoken language data and news data. All three Chinese-English training corpora, BOLT (LDC2013E81, LDC2013E85, LDC2013E118, LDC2013E125, LDC2013E132, LDC2013E80, LDC2013E83, LDC2014E08, LDC2014E50, LDC2014E69, LDC2014E99, LDC2013E119, LDC2014E110, LDC2014E111), FBIS (LDC2003E14), and large-scale data (LDC2003E14, LDC2000T46, LDC2007T09, LDC2005T10, LDC2008T06, LDC2009T15, LDC2010T03, LDC2009T02, LDC2009T06, LDC2013T11, LDC2013T16, LDC2007T23, LDC2008T08, LDC2008T18, LDC2014T04, LDC2014T11, LDC2005T06, LDC2007E101, LDC2002E18) come from the Linguistic Data Consortium (LDC). The BOLT corpus contains mostly spoken data, including conversation, short message service (SMS), and chat. The FBIS and large-scale corpora are mainly news data. The FBIS corpus is sentence-aligned using the Champollion Toolkit (CTK) (https://www.ldc.upenn.edu/language-resources/tools). The large-scale data is used mainly to verify the effectiveness of our methods on a larger data set.

We used the string-to-tree translation model (Str2tr) provided by the Moses translation model (Koehn *et al.*, 2007) as our baseline system. We also used a strong phrase-based translation system (Phb) for a comparison. The English corpora were parsed by the Berkeley parser (Petrov *et al.*, 2006) and semantic role labeled by ASSERT (Pradhan *et al.*, 2004).

We word-aligned the parallel corpora with the MGiza tool (http://www.kyloo.net/software/). In the experiments on the spoken corpus, we used only the 5-gram language model trained on the English training data, whereas in the experiments on the news corpora, we used the 5-gram language models trained on both English training data and the English Gigaword Fifth Edition corpora. NIST MT02 was used as the tuning set. NIST MT03, MT04, MT05, MT06 (NIST part), MT08, MT12 (general), and MT08-12 (progress) were used as the test data. Both the case-insensitive BLEU score (Papineni *et al.*, 2002) and Meteor score (Denkowski and Lavie, 2014) were employed as the evaluation metrics. To test whether a performance difference was statistically significant, we conducted significance tests following Koehn (2004) for BLEU and Clark *et al.* (2011) for Meteor. The English Gigaword corpora contain 9 876 086 documents, about 4 032 686 000 tokens in total. The statistics of other data are shown in Table 1.

Table 1 Data statistics of the SMT experiment

Data	Number of sent pairs	Number of tokens (C/E)
BOLT (train)	121 078	4 300 712/5 221 584
BOLT (dev)	4935	189 217/226 015
BOLT (test)	4977	160 786/207 641
FBIS	252 384	8 161 546/10 233 829
Large-scale data	2 032 497	55 636 231/61 062 511
MT02	878	22 350/25 339
MT03	919	23 992/25 999
MT04	1597	43 128/46 952
MT05	1082	29 475/30 882
MT06 (NIST part)	1664	37 822/41 014
MT08	1357	32 042/37 307
MT12 (general)	820	21 321/25 316
MT08-12 (progress)	1370	30 935/36 043

3.2 Results

Table 2 illustrates the final translation results on the spoken corpus (BOLT). As can be seen, the two methods using SRT and RST outperform the baseline by 0.56 and 0.22 BLEU points, respectively. This comparison means that SRT performs better by combination with role labels. Note that none of the tree-based methods (Str2tr, SRT, and RST) can beat the phrase-based system on this spoken corpus.

Table 3 illustrates the results on the FBIS news corpus. As for the results on the spoken corpus, SRT performs better than RST and exhibits an improvement of 1.33 BLEU points on average. However, both SRT and RST show much greater improvement than those on the spoken corpus, and SRT beats the Phb system. This is attributable to the better grammaticality of the news corpus compared to the spoken one.

Table 4 illustrates the results on the large-scale news corpus. Both SRT and RST beat the baseline and Phb system. RST shows the best result with an improvement of 1.13 BLEU points.

3.3 Analysis

From Table 2, we can see that both methods obtain better results than the baseline system, but all results are worse than those of the phrase-based system. We think the reason is that the performance of the three tree-based methods depends on the accuracy of the syntactic parser or SRL, which is fairly poor on the spoken corpus. However, there is quite a large difference in Tables 3 and 4. The accuracy of the parser or SRL gets much better on the news corpora. The performance of the string-to-tree model gets closer to that of the phrase-based model, and the two proposed methods surpass them both. Therefore, we believe that the three tree-based models are much more applicable to the news corpus than the spoken corpus, due to the better parsing accuracy.

Table 2 Experimental results on the BOLT corpus

Method	BLEU score	Meteor score
Str2tr	13.29	22.52
SRT	13.85* (+0.56)	23.14* (+0.62)
RST	13.51 (+0.22)	22.73 (+0.21)
Phb	**14.17**	**23.25**

In Tables 2 and 3, the performance of SRT is better than that of RST, but it is just the opposite in Table 4. We believe the reason is that RST faces a more severe data sparsity condition. The difference between the two methods relates to the fact that the majority of SRT is syntactic structure, but the higher layers in RST are role labels, which dominate the whole tree, and the semantic role structure is flatter. When trained on larger data, data sparsity is alleviated with larger data and the effectiveness of semantic information is established.

We give three translation examples in Table 5 to show specifically the effectiveness of our methods. In the first example, Phb and Str2tr position the noun phrase ‘the refugees’ incorrectly and produce bad translations. This is because they cannot identify the semantic relationship between the word ‘affected’ and the phrase ‘the refugees’. Conversely, our SRT system reorders the phrases based on its learning of the orders of the semantic roles in English. In the second example, our SRT method successfully recognizes the prepositional phrase “on the company’s business or contract” as an [ARG3] argument and moves it to the end of the sentence. Furthermore, in the third example, our SRT method recognizes the phrase “to promote the reunification of the motherland” as the [ARGM-PNC] argument and moves it to the end of the sentence. The Phb and Str2tr systems perform translation without reordering.

Table 3 Experimental results on the FBIS corpus

Test set	BLEU score (%)				Meteor score (%)			
	Str2tr	SRT	RST	Phb	Str2tr	SRT	RST	Phb
MT03	28.83	**30.35*** (+1.52)	29.29 (+0.46)	29.57	28.20	28.71* (+0.51)	28.01 (–0.19)	**28.84**
MT04	31.33	**33.15**$^{*\#}$ (+1.82)	31.87* (+0.54)	31.46	29.31	**30.11**$^{*\#}$ (+0.80)	29.53* (+0.22)	29.97
MT05	28.27	**29.92**$^{*\#}$ (+1.65)	28.69 (+0.42)	28.33	28.66	29.26* (+0.60)	28.62 (–0.04)	**29.47**
MT06	27.76	**29.27**$^{*\#}$ (+1.51)	28.32* (+0.56)	28.72	27.03	27.97* (+0.94)	27.20 (+0.17)	**28.10**
MT08	22.27	**22.71**$^{*\#}$ (+0.44)	21.76 (–0.51)	22.11	24.09	**24.40*** (+0.31)	23.59 (–0.50)	24.37
MT08-12	21.12	**22.39**$^{*\#}$ (+1.27)	21.00 (–0.12)	21.84	23.82	24.35* (+0.53)	23.26 (–0.56)	**24.57**
MT12	20.57	**21.68**$^{*\#}$ (+1.11)	20.79 (+0.22)	21.00	22.90	23.34 (+0.44)	22.81 (–0.09)	**23.45**
Average	25.74	**27.07** (+1.33)	25.96 (+0.22)	26.14	26.29	26.88 (+0.59)	26.15 (–0.14)	**26.97**

* and # denote that the result is significantly better than those of Str2tr and Phb, respectively (at significance level $p<0.01$)

Table 4 Experimental results on the large-scale corpus

Test set	BLEU score (%)				Meteor score (%)			
	Str2tr	SRT	RST	Phb	Str2tr	SRT	RST	Phb
MT03	31.80	33.74* (+1.94)	**34.26**$^{*\#}$ (+2.46)	33.50	30.55	30.92$^{\#}$ (+0.37)	**31.50**$^{*\#}$ (+0.95)	30.64
MT04	34.29	35.60$^{*\#}$ (+1.31)	**36.28**$^{*\#}$ (+1.99)	34.79	31.20	31.57$^{\#}$ (+0.37)	**32.13**$^{*\#}$ (+0.93)	31.09
MT05	33.19	34.24* (+1.05)	**34.73**$^{*\#}$ (+1.54)	33.92	31.55	31.51 (–0.04)	**32.41**$^{*\#}$ (+0.86)	31.56
MT06	32.85	32.89$^{\#}$ (+0.04)	**33.28**$^{\#}$ (+0.43)	31.93	29.76	29.75$^{\#}$ (–0.01)	**30.28**$^{*\#}$ (+0.52)	29.15
MT08	26.11	**26.32**$^{\#}$ (+0.21)	26.29$^{\#}$ (+0.18)	23.99	26.48	26.61$^{\#}$ (+0.13)	**26.84**$^{\#}$ (+0.36)	25.39
MT08-12	25.34	25.13$^{\#}$ (–0.21)	**25.39**$^{\#}$ (+0.05)	24.06	26.32	26.20$^{\#}$ (–0.12)	**26.60**$^{\#}$ (+0.28)	25.70
MT12	22.88	23.72* (+0.84)	**24.08**$^{*\#}$ (+1.20)	23.48	24.89	25.15$^{\#}$ (+0.26)	**25.44**$^{*\#}$ (+0.55)	24.94
Average	29.49	30.23 (+0.74)	**30.62** (+1.13)	29.38	28.68	28.82 (+0.14)	**29.31** (+0.63)	28.35

* and # denote that the result is significantly better than those of Str2tr and Phb, respectively (at significance level $p<0.01$)

4 Related work

Recent work leveraging semantic roles in SMT can be categorized into three directions (Zhai *et al.*, 2012): pre- and post-processing, designing semantic role based features and using them to decode, and refining the non-terminals of the syntax-based translation model.

Semantic roles are first used to improve the SMT system by Komachi *et al.* (2006). They reordered the source-side language chunks in the pre-processing stage based on semantic roles. Wu and Fung (2009) used semantic role labeling to improve a phrase-based SMT system in the post-processing stage, through reproducing translation hypotheses by moving phrases whose predicate or semantic role was mismatched to

Table 5 Examples of the MT outputs

System	Translation
Phb	[受到 这 一 变更 影响] 的 [难民] 主要 分散 在 约旦 、 叙利亚 和 土耳其 。 [Affected by this change] [refugees] scattered mainly in Jordan , Syria and Turkey .
Str2tr	[受到 这 一 变更 影响] 的 [难民] 主要 分散 在 约旦 、 叙利亚 和 土耳其 。 [Affected by this change] [the refugees] scattered in Jordan , Syria and Turkey .
SRT	[受到 这 一 变更 影响] 的 [难民] 主要 分散 在 约旦 、 叙利亚 和 土耳其 。 [ARG1] [TARGET][ARG0] [ARG0] [TARGET] [ARG2] [The refugees] [affected by this change] mainly scattered in Jordan , Syria and Turkey .
Ref.	[受到 这 一 变更 影响] 的 [难民] 主要 分散 在 约旦 、 叙利亚 和 土耳其 。 [Refugees] [affected by this change] mainly spread around Jordan , Syria and Turkey .
Phb	MobilCom 在 今天 稍后 发表 的 声明 中 说 ， [施密德 个人 破产] [对 公司 的 业务 或 合约] [没有 影响] 。 MobilCom in a statement later today , said that [the company 's personal bankruptcy 施密德] [business or contract] [has not affected] .
Str2tr	MobilCom 在 今天 稍后 发表 的 声明 中 说 ， [施密德 个人 破产] [对 公司 的 业务 或 合约] [没有 影响] 。 MobilCom said in a statement issued later today , [施密德 personal bankruptcies] [in the business or contract] [has not been affected] .
SRT	MobilCom 在 今天 稍后 发表 的 声明 中 说 ， [施密德 个人 破产] [对 公司 的 业务 或 合约] [没有] [影响] 。 [ARG0] [TARGET] [ARG1] [ARG3] MobilCom said in a statement issued later today , [施密德 personal bankruptcy] [has no] [impact] [on the company 's business or contract] .
Ref.	MobilCom 在 今天 稍后 发表 的 声明 中 说 ， [施密德 个人 破产] [对 公司 的 业务 或 合约] [没有 影响] 。 MobilCom said in a statement later today that [Schmid 's personal bankruptcy] [has no effect] [on the company 's operation or contract] .
Phb	声明 呼吁 [海外 同胞] [为 推动 祖国 统一] [而 不断 作出 新 的 努力] 。 The statement urged [overseas compatriots] [to promote the reunification of the motherland] [and continue to make new efforts] .
Str2tr	声明 呼吁 [海外 同胞] [为 推动 祖国 统一] [而 不断 作出 新 的 努力] 。 The statement urged [the overseas compatriots] [to promote the reunification of the motherland] [and continue to make new efforts] .
SRT	声明 呼吁 [海外 同胞] [为 推动 祖国 统一] 而 不断 [作出] [新 的 努力] 。 [ARG0] [TARGET] [ARG1] [ARGM-PNC] The statement urged [overseas compatriots] [to make] [new efforts] [to promote the reunification of the motherland] .
Ref.	声明 呼吁 [海外 同胞] [为 推动 祖国 统一] 而 不断 [作出] [新 的 努力] 。 The statement urged [overseas compatriots] to [make] [unceasing and new efforts] [to promote the reunification of the motherland] .

the input.

Liu and Gildea (2010) modeled two types of features, reordering and deleting source-side semantic roles, into the decoding phase of tree-to-string translation. Xiong *et al.* (2012) proposed two models, a predicate translation model and an argument reordering model. The former model takes lexical and semantic contexts into consideration when translating predicates, whereas the latter predicts the moving direction of arguments from the source to the target language sentence. Zhai *et al.* (2012) proposed a framework which is divided into three steps, including source-side predicate-argument structure (PAS) acquisition, transforming source-side PASs to their target counterparts, and translation.

Liu and Gildea (2008) replaced syntactic labels with semantic roles, and combined them to generate more refined tree labels. Aziz *et al.* (2011) created shallow semantic trees to improve the tree-to-string translation model. Similarly, Bazrafshan and Gildea (2013) attached only core arguments of each predicate to the syntax tree for rule extraction. They all combined the syntactic label and semantic label into one, which leads to data sparsity.

Our work is different from the above in the following aspects: (1) We use semantic roles by inserting them into a syntax tree, or creating a 'semantic role tree' and then incorporating syntactic information into it; (2) We use multiple lines in the SRL results and the hierarchy structure of semantic roles; (3) We use all the arguments of the predicates, not only the core arguments; (4) We do not differentiate among the predicates in SRT labels, which will lead to data sparsity and reduce the translation performance.

5 Conclusions and perspectives

We have presented an effort devoted to the use of semantic information for SMT. We first incorporate semantic roles into target-side parse trees, and then perform string-to-tree machine translation using the newly generated trees. Experimental results demonstrate that our methods improve the translation performance significantly.

Our methods improve the translation performance in the following aspects: (1) They use the semantic roles in the target language to reorder the semantic chunks in output; (2) They use the hierarchy of the predicate-argument structure to generate semantic chunks and gather them.

In the future, we will explore the situations in which the semantic roles help improve or reduce the translation performance in detail. We will also use the overlapping semantic roles which are dropped in this study.

References

Aziz, W., Rios, M., Specia, L., 2011. Shallow semantic trees for SMT. Proc. 6th Workshop on Statistical Machine Translation, p.316-322.

Baker, C.F., Fillmore, C.J., Lowe, J.B., 1998. The Berkeley Framenet Project. Proc. 17th Int. Conf. on Computational Linguistics, p.86-90. https://doi.org/10.3115/980451.980860

Bazrafshan, M., Gildea, D., 2013. Semantic roles for string to tree machine translation. Proc. 51st Annual Meeting of the Association for Computational Linguistics, p.419-423.

Brown, P.F., Cocke, J., Pietra, S.A.D., *et al.*, 1990. A statistical approach to machine translation. *Comput. Ling.*, **16**(2): 79-85.

Brown, P.F., Pietra, V.J.D., Pietra, S.A.D., *et al.*, 1993. The mathematics of statistical machine translation: parameter estimation. *Comput. Ling.*, **19**(2):263-311.

Chiang, D., 2005. A hierarchical phrase-based model for statistical machine translation. Proc. 43rd Annual Meeting on Association for Computational Linguistics, p.263-270. https://doi.org/10.3115/1219840.1219873

Clark, H.J., Dyer, C., Lavie, A., *et al.*, 2011. Better hypothesis testing for statistical machine translation: controlling for optimizer instability. Proc. 49th Annual Meeting of the Association for Computational Linguistics, p.176-181.

Denkowski, M., Lavie, A., 2014. Meteor universal: language specific translation evaluation for any target language. Proc. 9th Workshop on Statistical Machine Translation, p.376-380. https://doi.org/10.3115/v1/W14-3348

Galley, M., Hopkins, M., Knight, K., *et al.*, 2004. What's in a translation rule. Proc. Human Language Technology Conf. of the North American Chapter of the Association for Computational Linguistics. https://doi.org/10.21236/ada460212

Gildea, D., Jurafsky, D., 2002. Automatic labeling of semantic roles. *Comput. Ling.*, **28**(3):245-288. https://doi.org/10.1162/089120102760275983

Huang, L., Chiang, D., 2005. Better *k*-best parsing. Proc. 9th Int. Workshop on Parsing Technology, p.53-64. https://doi.org/10.3115/1654494.1654500

Koehn, P., 2004. Statistical significance tests for machine translation evaluation. Proc. Conf. on Empirical Methods in Natural Language Processing, p.388-395.

Koehn, P., Och, F.J., Marcu, D., 2003. Statistical phrase-based translation. Proc. Conf. of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, p.48-54. https://doi.org/10.3115/1073445.1073462

Koehn, P., Hoang, H., Birch, A., *et al.*, 2007. Moses: open source toolkit for statistical machine translation. Proc. 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, p.177-180. https://doi.org/10.3115/1557769.1557821

Komachi, M., Matsumoto, Y., Nagata, M., 2006. Phrase reordering for statistical machine translation based on predicate-argument structure. Int. Workshop on Spoken Language Translation, p.77-82.

Liu, D., Gildea, D., 2008. Improved tree-to-string transducer for machine translation. Proc. 3rd Workshop on Statistical Machine Translation, p.62-69. https://doi.org/10.3115/1626394.1626402

Liu, D., Gildea, D., 2010. Semantic role features for machine translation. Proc. 23rd Int. Conf. on Computational Linguistics, p.716-724.

Liu, Y., Liu, Q., 2010. Joint parsing and translation. Proc. 23rd Int. Conf. on Computational Linguistics, p.707-715.

Liu, Y., Liu, Q., Lin, S., 2006. Tree-to-string alignment template for statistical machine translation. Proc. 21st Int. Conf. on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, p.609-616. https://doi.org/10.3115/1220175.1220252

Marcu, D., Wang, W., Echihabi, A., *et al.*, 2006. SPMT: Statistical machine translation with syntactified target language phrases. Proc. Conf. on Empirical Methods in Natural Language Processing, p.44-52. https://doi.org/10.3115/1610075.1610083

Meyers, A., Reeves, R., Macleod, C., *et al.*, 2004. The nombank project: an interim report. HLT-NAACL Workshop: Frontiers in Corpus Annotation, p.24-31.

Mi, H., Huang, L., Liu, Q., 2008. Forest-based translation. Proc. ACL-08: HLT, p.192-199.

Och, F.J., Ney, H., 2004. The alignment template approach to statistical machine translation. *Comp. Ling.*, **30**(4):417-449. https://doi.org/10.1162/0891201042544884

Palmer, M., Gildea, D., Kingsbury, P., 2005. The proposition bank: an annotated corpus of semantic roles. *Comp. Ling.*, **31**(1):71-106. https://doi.org/10.1162/0891201053630264

Papineni, K., Roukos, S., Ward, T., *et al.*, 2002. BLEU: a method for automatic evaluation of machine translation. Proc. 40th Annual Meeting on Association for Computational Linguistics, p.311-318. https://doi.org/10.3115/1073083.1073135

Petrov, S., Barrett, L., Thibaux, R., *et al.*, 2006. Learning accurate, compact, and interpretable tree annotation. Proc. 21st Int. Conf. on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, p.433-440. https://doi.org/10.3115/1220175.1220230

Pradhan, S.S., Ward, W., Hacioglu, K., *et al.*, 2004. Shallow semantic parsing using support vector machines. Human Language Technologies: the Annual Conf. of the North American Chapter of the Association for Computational Linguistics, p.233-240.

Wu, D., 1995. Grammarless extraction of phrasal translation examples from parallel texts. Proc. 6th Int. Conf. on Theoretical and Methodological Issues in Machine Translation, p.354-372.

Wu, D., 1996. A polynomial-time algorithm for statistical machine translation. Proc. 34th Annual Meeting on Association for Computational Linguistics, p.152-158. https://doi.org/10.3115/981863.981884

Wu, D., Fung, P., 2009. Semantic roles for SMT: a hybrid two-pass model. Proc. Human Language Technologies: the Annual Conf. North American Chapter of the Association for Computational Linguistics, p.13-16. https://doi.org/10.3115/1620853.1620858

Xiong, D., Zhang, M., Li, H., 2012. Modeling the translation of predicate-argument structure for SMT. Proc. 50th Annual Meeting of the Association for Computational Linguistics, p.902-911.

Yamada, K., Knight, K., 2001. A syntax-based statistical translation model. Proc. 39th Annual Meeting on Association for Computational Linguistics, p.523-530. https://doi.org/10.3115/1073012.1073079

Zhai, F., Zhang, J., Zhou, Y., *et al.*, 2012. Machine translation by modeling predicate-argument structure transformation. Proc. Int. Conf. on Computational Linguistics, p.3019-3036.