Case Frame Constraints for Hierarchical Phrase-Based Translation: Japanese-Chinese as an Example

Jiangming Liu[1], JinAn Xu[1,⋆], Jun Xie[2], and Yujie Zhang[1]

[1] School of Computer and Information Technology, Beijing Jiaotong University
[2] Institute of Computing Technology Chinese Academy of Science
{jaxu,yjzhang}@bjtu.edu.cn, {jmliunlp,stiffxj}@gmail.com

Abstract. Hierarchical phrase-based model has two main problems. Firstly, without any semantic guidance, large numbers of redundant rules are extracted. Secondly, it cannot efficiently capture long reordering. This paper proposes a novel approach to exploiting case frame in hierarchical phrase-based model in both rule extraction and decoding. Case frame is developed by case grammar theory, and it captures sentence structure and assigns components with different case information. Our case frame constraints system holds the properties of long distance reordering and phrase in case chunk-based dependency tree. At the same time, the number of HPB rules decrease with the case frame constraints. The results of experiments carried out on Japanese-Chinese test sets shows that our approach yields improvements over the HPB model (+1.48 BLEU on average).

1 Introduction

The hierarchical phrase-based (HPB) model (Chiang, 2007) is widely used in statistical machine translation. Extended from phrase-based (PB) rules (Koehn et al., 2003), HPB rules are capable of capturing phrase-level reordering by exploiting the underlying hierarchical structures in natural language. HPB model is formally synchronous context-free grammar but this is learned from a bitext without any syntactic information, so that HPB suffers from limited phrase reordering in the case of combining translated phrases with monotonic glue rules. As a result, it performs not so well in long distance reordering. Furthermore, without phrase boundary determination, the number of HPB rules increases explosively with the increase in training data. To address the HPB model limitation, a number of work is motivated in two aspects.

In the process of HPB decoding, many recent work are motivated to preserve linguistic information in HPB model derivation. Syntactic features are derived from the source dependency parsing to directly guide derivation in HPB model (Marton and Resnik, 2008; Huang et al., 2010; Gao et al., 2011; Marton et al., 2012). However, these systems perform not so well in agglutinative language

⋆ Corresponding author.

C. Zong et al. (Eds.): NLPCC 2014, CCIS 496, pp. 123–137, 2014.

translation due to the agglutinative properties of complex and varied morphology. The agglutinative languages are usually provided with relatively accurate chunk-based dependency analysis. The structure impedes the utilization of word-based dependency to string translation model (Flannery, et al. 2011).

In terms of extraction and presentation of HPB rules, many significant works focus on assigning HPB rules with extra constraints to explore search space (Li et al., 2012; He et al., 2010), and the suited HPB rules can be selected from the rule selection model (Liu et al., 2008; He et al., 2008). However, the total number of HPB rules remains the same, and a large number of redundant rules are extracted.

To solve the above two problems, we exploit case frame constraints (CFCs) in this paper. The description of case frame will be introduced in section 3. At the same time, this paper presents case chunk-based dependency, and the purposes of our works include alleviating reordering problem and restricting HPB rule extraction in case frame, and finally case frame HPB rules (CF-HPBs) are extracted. In terms of the reorder- ing process, case frame reordering rules (CF-Rs) are automatically extracted from the source side parsing and aligning parallel corpus, and this aims to alleviate the reordering problems under the condition of preserving all the components in the sentence.

This paper proposes a novel approach to use case frame constraints in Japanese-Chinese statistical machine translation as an example and achieve better performance than HPB model and word-based dependency model as shown in our experimental results.

According to our knowledge, case frame is rarely used in statistical machine translation. Our work is the first to try case frame in statistic ma- chine translation. The main contributions of our works are using case frame constraints in HPB rule extraction and decoding.

The remainder of this paper is organized as follows. Section 2 introduces some related work and mainly contributes to this paper. We present case frame constraints HPB rules (CF-HPBs) extraction in section 3, then we define the case chunk-based dependency tree and describe case frame reordering rules in section 4. Section 5 presents our model. Section 6 reports our experiments. Section 7 presents the analysis on the experimental. Section 8 concludes this paper with prospects for future work.

2 Related Work

In recent years, word-based dependency structure is widely used to incorporate linguistic information into machine translation (Lin, 2004; Quirk et al., 2005; Ding and Palmer, 2005; Xiong et al., 2007). The reordering problem can be alleviated, especially in long distance reordering problem (Xie et al., 2011). Dependency-to-string model employs rules whose source-side is a word-based dependency structure with POS and target as string. Reordering problem can be alleviated by simple nodes exchange.

Many novel approaches are presented for restricting HPB rules extraction (He et al., 2010; Xiong et al., 2010). These methods employ supervised learning technology, and suitable features are selected to train a boundary classifier as soft constraint for decoding. However, a total number of HPB rules is not decreased and a large-scale corpus is needed for training classifier.

Our proposed approach focuses on case frame constraints (CFCs) to improve the quality of extracted rule and decoding. Moreover, this paper defines a new structure dependency tree, which is more suitable for agglutinative language than word-based dependency tree. The derivation based on new structure tree holds two merits. Firstly, the number of HPB rules is decreased. Secondly, the decoding efficiency and translation quality are improved.

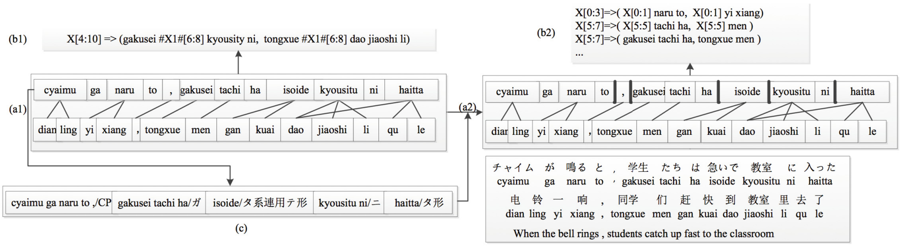

Fig. 1. An example of phrase boundary determination for CF-HPBs extraction

where (a1) is sentence pair with word alignments; (b1) is an example of HPB rules without case frame constraints; (c) is source side sentence with case boundary generated by KNP tools; (a2) is a sentence pair with word alignment and case boundary (marked in bold); (b2) is a set of CF-HPBs examples.

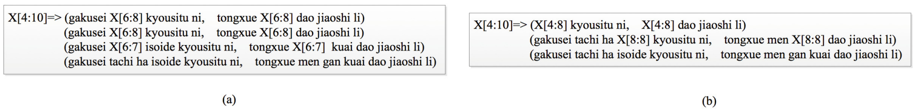

Fig. 2. Examples of unreasonable derivation (a) and reasonable derivation (b)

3 Case Frame Constraints

3.1 Case Frame

Case grammar created by Fillmore (1968) and developed by Cook (1989) in English case grammar, is used to linguistically analyze the surface syntactic structure of sentences by investigating the combination of cases. Case frame is analyze by case grammar. Case grammar has been developed in different languages. In Japanese, a case frame corpus is extended and built from web resources (Kawahara and Kurohashi, 2006). Under the case frame corpus, the system of Japanese syntactic and case structure analysis turns to be a state-of-the-art (Buchholz and Marsi, 2006).

3.2 CFCs on HPB Rules

HPB rules replaces common phrase with non-terminal variables, which confuses primary with secondary linguistic. A large number of long HPB rules are slightly linguistic, especially in the case of the Japanese language.

Let $S = (sw_0, sw_1, ..., sw_l)$be a source word sequence and $T = (tw_0, tw_1, ..., tw_m)$ be a target sequence, where sw_i and tw_j are source word and target word respectively. With word alignment, HPB rules can be extracted as:

$$(sw_0\ X_{(1)\to(0)}\ sw_2, X_{(1)\to(0)}\ tw_1),$$
$$(sw_0\ X_{(1...2)\to(0...1)}\ sw_3\ sw_4, X_{(1...2)\to(0...1)}\ tw_2\ tw_3),$$
$$(sw_0\ X_{(1...3)\to(0...2)}\ sw_4, X_{(1...3)\to(0..2)}\ tw_3),$$

where X is non-terminal variable and its indices denotes the word alignment. The non-terminal variables in HPB rules can be generated by replacing common phrase without distinguishing what the component means. To make derivation more reasonable, we use syntax to assign components with specific semantic information that makes sense.

According to case frame theory, a sentence is divided into many components. CFCs are used in determining component boundaries during HPB rule extraction, at the same time, each component can be labeled with specific case information. Each case boundary is regarded as the phrase boundary in the process of HPB rule extraction. Suppose a case frame in source language given like $CF = \{(sw_0)_{subject}, (sw_1\ sw_2)_{verb-head}\ (sw_3\ sw_4)_{object} ...\}$where the phrase sw_0 is marked as subjective case, phrase $sw_{1...2}$ as head, and $sw_{3...4}$ as objective case. The traditional HPB rule with the non-terminal variable $X_{(1...3)\to(0...2)}$ is filtered due to the fact that $X_{(1...3)}$ is over verb-head case boundary (phrase boundary). As a result, phrases without over the case boundary (phrase boundary) can be generalized as non-terminal variable. It means that many rules without suitable to case frame are filtered, and finally CF-HPBs will be achieved semantically. An example is shown in Figure 1, where “tongxue men” means students, and “gan kuai dao” means “catch up fast”. “men gan kuai” is unreasonably generalized for a non-terminal variable in Figure 1(b1). Also, in case frame constraints, “men gan kuai” is aligned to source side sequence “tachi ha isoide” over the case boundary that is forbidden in CF-HPBs, and then it will be filtered. In this way, each component in CF-HPBs can be assigned with a semantic label, namely case.

The extra properties of CF-HPBs are maintained bellow:

Property 1. *An acceptable non-terminal variable is only generalized by the phrase without over the component boundary (case boundary).*

Property 2. *An acceptable non-terminal variable can be generalized by any common phrase inside one component.*

Due to these properties, a reasonable derivation can be obtained as shown in Figure 2, which is part of derivation in the case of Figure 1.

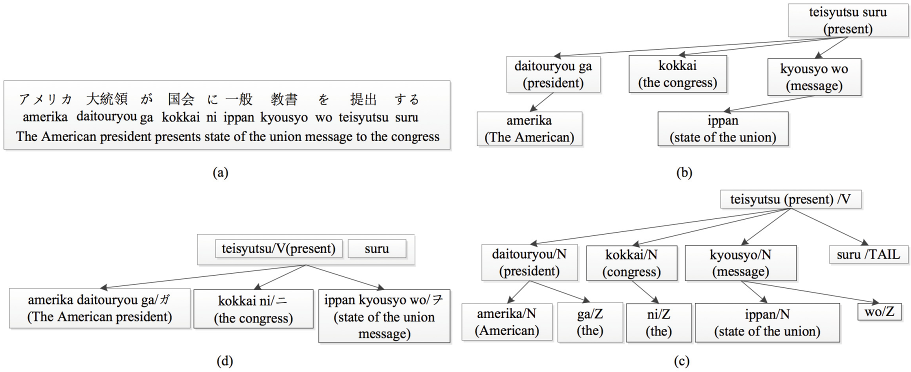

Fig. 3. Example of chunk-based dependency structure and word-based dependency structure

where (a) is the original source side sentence; (b) is a chunk-based dependency structure given by KNP; (c) is a word-based dependency structure with a simple change from chunk-based structure and (d) is a case chunk-based dependency structure, where "The American president" is subjective denoted by specific case tag, "the congress" is objective and "state of the union message" is direction

4 Case Frame Reordering

4.1 Case Chunk-Based Dependency

Before extraction of case frame reordering information, a specific structure, namely case chunk-based dependency structure, is firstly defined.

In case frame, a sentence can be divided into many components with different cases, where each component is a word or a phrase, which is defined as a chunk in this paper. Case chunk-based dependency tree can be defined as a tuple $CT = (\Sigma, \Lambda, C, D)$, where Σ is a set of words, Λ is a set of chunks corresponding to case boundary, C is a set of possible case tag for chunks, and D is dependency relation2 between chunks. It is distinguishable for case chunk-based dependency tree with word-based dependency tree shown in Figure 3. In case chunk-based dependency tree, each node consists of chunk and related case tag.

Due to parallel sentence pairs given for statistical machine translation, in our model, source side sentence is represented by case chunk-based dependency structure and target side sentence is represented by word sequence. With word alignments, our initial model is defined by a tuple (CT, Δ, A) where CT is source side case chunk-based dependency tree, Δ is a set of target side words and A is word-to-word alignment. An example is shown in Figure 4(a). Since source side item (node) is chunk-level and target side item is word-level, the change is carried out from word-to-word alignment A to chunk-to-word alignment A'.

$$A' = \{(c, tw) | \exists sw \in c, (sw, tw) \in A, c \in \Lambda, tw \in \Delta\}$$

So our final model is defined by $M = (CT, \Delta, A')$. Based on the model, case frame reordering rules can be extracted.

4.2 Case Frame Reordering Rule

Case frame reordering rules (CF-Rs) are represented by a tuple $(t, s, \sim)$ where:

- t is a dependent relation of the source dependency structure, with each node labeled with a variable from a set $X = \{x_1, x_2, ...\}$ constrained by a case from C. Specially, head node can be also labeled by a chunk constrained by a case from C.
- $s \in X$ are the target side chunk slots corresponding to source side non-terminal variables.
- $\sim$ is a one-to-one mapping from slot in s to variables in t

One example is shown in Figure 4(c).

According to our rule definition, CF-Rs have two properties bellow:

Property 3. *Each node, except head node in source side of rules, is unlexicalized, and each item in target side is slot with variable corre- sponding to source side variable.*

Property 4. *Head node in source side can be lexicalized or unlexicalized. And thus CF-Rs can be classified into CF-LRs (lexicalized) and CF-URs (unlexicalized).*

Prior work on rule extraction, reordering and lexical translation are both considered at the same time. Also, alignment error propagation impacts reordering rule and lexical translation. Instead, during the extraction process of CF-Rs, we only consider the variables reordering on the target side. In the following section, we will present how to extract rules using our model.

4.3 CF-Rs Acquisition

Now, it focuses on reordering rule extraction. Before extraction, anchor is defined to assist reordering formation extraction among each item on the target side. Anchor can be represented as a function $Ach(sp)$, where $sp \in SP$ denotes possible span on the target side. Here, span is a set of word index on the target side corresponding to certain node on the source side tree (same index may occur twice or more), where spans of all child nodes are at chunk-level and spans of head nodes are at word-level. So the amount of spans is larger than the number of nodes on the source side tree. In Figure 4(b), the head node has two spans and each child node has only one span. The value of $Ach(sp)$ is a real number which is computed using following formula

$$Ach(sp_i) = sum(Sign(sp_i, SP_i))$$

Where sp_i denotes the ith span in SP, SP_i denotes set of spans except sp_i, $Sign(sp_i, SP_i)$ returns a $|SP_i|$ size vector (V_i) where each item is 1 or 0 and

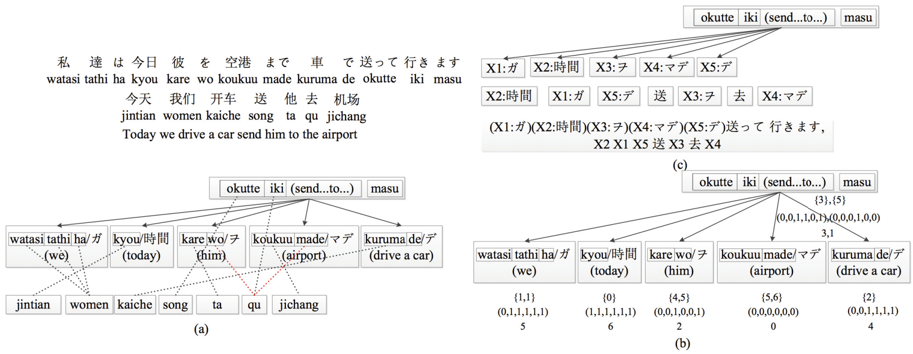

Fig. 4. An example of CF-Rs extraction

where (a) is a case chunk-based dependency-to-string with word alignment, where the red dotted line is noise alignment; (b) is with chunk alignment, where each node has three extra items in extraction, first is spans, second is signal vector, and third is *anchor*; (c) is an extracted rule.

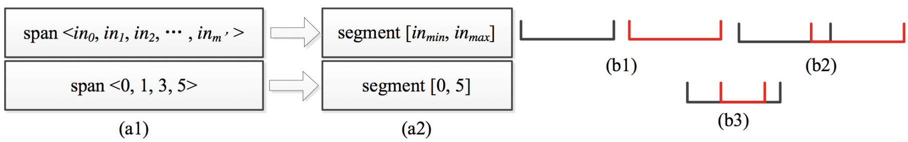

Fig. 5. Span, segment and three spans relation

where (a1) is a span. (a2) is a segment from the span. (b) is a separation span relation, (c) is an interaction span relation and (d) is an inclusion span relation, where each of them is a segment.

sum is a function to sum up all item in that vector. $V_i = (v_0, v_1, ..., v_j, ..., v_{|SP_i|})$ where according to chunk-word alignment, v_j is 1 if and only if the j^{th} span is relatively left to the i^{th} span, otherwise it is 0. To formulate the span relation, function $F(i,j)$ is defined to capture relation between i^{th} spans and j^{th} span. $F(i,j)$ follows three strategies, which respectively deal with three situations.

– **Separation** segment i is separated from segment j, where segment i is generated by minimum index and maximum index in sp_i, and segment j is similar. Under this condition as shown in Figure 5(b), $F(i,j)$ is 1 if and only if largest index in sp_i is smaller than or equal to smallest index in sp_j, otherwise it is 0.
– **Intersection** segment i and segment j are interacted. Under this condition as shown in Figure 5(c), $F(i,j)$ is 1 if and only if sum of all index in sp_i is smaller than in sp_j, otherwise 0.
– **Inclusion** segment i cover segment j, or segment j covers segment i. Under this condition as shown in Figure 5(d), $F(i,j)$ is 0, which means the default value remains the same.

Generally speaking, each node on the source side will place its own non-terminal variable on the target side with left-right order according to its anchor (may also be called a rank). Briefly speaking, Anchor ensures the order of target side items corresponds to the source side items. In this way, CF-Rs are extracted as shown in Figure 4(c).

CF-Rs are similar with dependency-string rules as mentioned in (Xie, et al. 2011). However, CF-Rs are guided by a case frame, and their semantic labels considers case frame structure in the whole sentence, conversely, POS only consider one word and components are neglected in the whole sentence. Moreover, some alignment errors will be alleviated in obvious anchor function. Under case constraints, fuzzy reordering information extraction is useful in agglutinative language due to its complex morphemes.

Following the case grammar, each component in a sentence will have a complete semantic representation. CF-Rs only achieve reordering information among components. Translation inside components is done using CF-HPBs as described in previous section.

Generally, in case frame, outside reordering of each component in sentence is done using case frame. And then, inside translation each component it is done using phrase-based rules, which is superior in HPB model in terms of short-distance reordering and lexical translation.

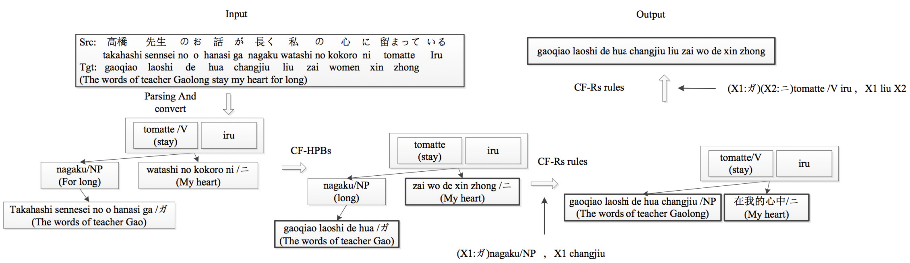

Fig. 6. An example of derivation

where in each step, the bolded box represents current translation focus.

5 Translation

5.1 Derivation

In this paper, CF-HPBs and CF-Rs are defined, and are integrated in derivation, where both CF-HPBs and CF-Rs are named case frame rules. This subsection describes one possible derivation in details as shown in Figure 6.

Under the case frame, a sentence is decomposed into a number of components, each of which has completed semantic content. Furthermore, a sentence is expressed in a form of a case chunk-based dependency tree. For one derivation, CF-HPBs are used in inside components derivation, and CF-Rs are used in outside components derivation, i.e. components reordering.

5.2 Log-Linear Model

Following (Och and Ney, 2002), we adopt a general log-linear model. Let d be a derivation that translate source chunk-based dependency tree T into a target string e. The probability of d is defined as:

$$P(d) \propto \Pi_i \phi_i(d)^{\lambda_i}$$

Where ϕ_i are features defined on derivations and λ_i are feature weights. Due to the fact that our translation rules have two sets, namely CF-HPBs and CF-Rs, during derivation, two kinds of rules are integrated. In our experiments of this paper, we used nine features which are similar with (Xie et al., 2011) as follow:

- CR-HPBs translation probabilities $P_{HPB}(t|s)$ and $P_{HPB}(s|t)$;
- CR-HPBs lexical translation probabilities $P(_{HPBlex}t|s)$ and $P_{HPBlex}(s|t)$;
- CF-Rs translation probabilities $P_R(t|s)$ and $P_R(s|t)$;
- CF-Rs lexical translation probabilities $P_{Rlex}(t|s)$ and $P_{Rlex}(s|t)$;
- Rule penalty $exp(-1)$;
- Language model $P_{lm}(e)$;
- Word penalty $exp(|e|)$.

During feature tuning process, different features are added into log-linear model and each weight of features can be discriminatively trained by MERT (Och, 2003), which is similar to (Li et al., 2012; Xie et al., 2011). Features include translation probabilities, lexical translation probabilities, language model, rule penalty, and word penalty.

5.3 Decoding

Our decoder is based on bottom up chart parsing. It determines the best derivation d^* that translates the input case chunk-based dependency structure into a target string among all possible derivations D: $d^* = argmax_{d \in D} P(D)$

Given a source case chunk-based dependency structure T. For each accessed internal node n, it gets a case frame corresponding to the node n, and checks if the CF-Rs are set for the matched reordering rules, and then checks if the CF-HPBs rule is set for matched translation rules. If there is no matched rule, we construct a *pseudo translation rule* according to the case frame, which has no reordering information like glue rules. Due to a large search space, a large number of translation is generated by substituting the variables in the target side of a translation rule with the translations of the corresponding slots in the source case frame. Similar to (Xie, et al. 2011), we make use of cube pruning (Chiang, 2007; Huang and Chiang, 2007) to find candidates with integrated language model for each node.

6 Experiments

We evaluate the case frame constraints in the replications of hierarchical phrase-based model in Japanese-Chinese translation. In these experiments, a replication of hierarchical phrase-based model is taken as a baseline model with beam size is 200 and the beam threshold of 0. The maximum initial phrase length is 10. In order to compare chunk-based dependency and word-based dependency, we also take dependency to string (*dep2str*) system by simply changing from chunk dependency to word dependency in word-POS process as shown in Figure 3. Under the same condition, this paper utilizes our model to constrain rule extraction and decoding.

6.1 Data

Due to that Japanese-Chinese parallel corpus is rare, our corpus consists of 280k sentence pairs for training which come from CWMT 2011 (Zhao et al., 2011) Japanese-Chinese evaluation task data in news domain. 500 sentence pairs are for parameters optimization. For testing, we use 900 sentence pairs provided by the task. In addition, we mix all the sentence pairs (including training, developing and testing data), and randomly select 500 sentence pairs for developing, 900 sentence pairs for testing and the rest of the sentences for training.

The source side sentences are parsed by KNP (Kurohashi and Nagao, 1994) into chunk dependency structures whose nodes are at chunk-level. Also we achieve corresponding case frame analysis from byproduct of KNP. The word alignment is obtained by running GIZA++ (Och and Ney, 2003) on the corpus in both direction and applying "grow-diag-and" refinement (Koehn et al., 2003). We apply SRI Language Modeling Toolkit (Stolcke, 2002) to train a 5-gram language model for target side sentences.

6.2 Baseline Model

In order to evaluate our system performance, we take a replication of Hiero (Chiang, 2007) as the hierarchical phrase-based model baseline (*hiero-re* for short), where we set the beam size $b = 200$ and the beam threshold $\beta = 0$. The maximum initial phrase length is 10.

Also, we use *dep2str* as the dependency-to-string model baseline, which consider word based dependency as provided by (Xie et al., 2011), where the same parameters are used for the experiment.

6.3 Result

Table 1 illustrates the translation experimental results. It shows that our system has achieved the best results on test sets, with +2.83 BLEU points on average higher than that of *dep2str*, and +1.23 BLEU points on average higher than that of *hiero-re*. It demonstrates that case frame constraints are useful to improving translation quality for HPB model. Compared with *dep2str*, chunk-based dependency tree performs better than word-based dependency does. In terms of the rule amount, the number of CF-Rs and CF- HPBs is decreased by more than half in the corpus of 280k sentence pairs. We believe case frame constraints superiority can be more obvious in larger corpus.

Table 1. The BLEU-4 score (%) on test sets of different system

System	Rule#	CWMT	Mix	Avg
hiero-re	24.0M	22.26	18.46	20.33
dep2str	2.8M	19.34	18.12	18.73
ours	1.4M+10.0M	22.62*	20.50*	21.56*

where the "+" denotes the 1.4 million CF-Rs and 10 million CF-HPBs on case frame constraints. The "*" denotes that the results show significant improvements over all of the other systems (p¡0.01)

7 Analysis

In HPB model, glue rule is frequently used for combining long sub-sentence without considering possible reordering. The agglutinative language, Japanese for example, has complex and varied morphology. Although the utilization of POS is general for the dependency rule variables in *dep2str*, it has local lexicalization, and some translation words are omitted. CF-HPBs maintain phrase translation with semantic label and CF-Rs alleviate long distance reordering problem. To further our analysis, we compare some actual translations generated by *hiero-re*, *dep2str* and our system. Figure 7 give one translation of our test set, which is helpful to elucidate these *problems* in terms of reordering and lexical *translation.*

7.1 Better Reordering

Main structure in Japanese structure is SOV-style, which is different from Chinese SVO-style. Reordering problem is significant in Japanese-Chinese translation, especially with long phrase for S and/or V. Compared with hierarchical phrase-based rules, CF rules have better phrase reordering. In the first example as shown in Figure 7, the source sentence main centered verb chunk is "tuujite (rely on)", and however, the objective is a long phrase (15 words) depending on the left of that verb chunk, which is a typical SOV-style. *Hiero-re* mistakenly treats that long phrase as subjective, thus results in translation with different meaning from source sentence. Conversely, our system captures this component relations in case frame and translates it into "tuujite (rely on)...". Although adverb "sarani (further)" is translated with incorrect ordering, the lexical translation is correct, and it makes sense that it cannot influence the understanding of source sentence.

7.2 Better Lexical Translation

Although word-based dependency tree-to-string model can also capture distance reordering problem (Xie, et al., 2011), depending strictly on word alignment in

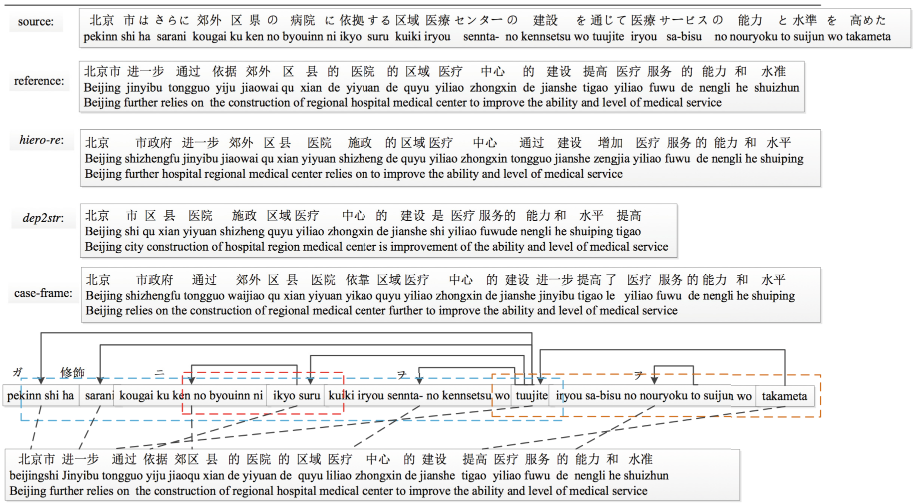

Fig. 7. Actual translations produced by the baselines and our system

For our system, we also display the long distance case chunk-based dependencies correspondence in Japanese and Chinese. In source side dotted box is a case frame.

dep2str, this does not lead to good performance on phrase translation as indicated in Figure 7. The adverb “sarani (further)” and “no (of)” has no corresponding translation or incorrect translation in *dep2str* because they are aligned into a common or NULL. Moreover, complex morphology expressed by long suffixes caused many words to be aligned to incorrect word. Complex alignment brings about some rules that cannot be extracted. Conversely, chunk-based dependency with fuzzy alignment can maintain the phrase-based rule (done with CF-HPBs) extraction of inside components without reordering deficiency (done with CF-Rs).

7.3 Summary

All these results prove the effectiveness of case frame constraints in both long reordering and translation. We believe that case chunk-based dependency tree-string model has an advantage of tending to assign semantic information on variables in rules with case grammar, and not the POS of a word in dependency-to-string model, and also it has an advantage of maintaining phrase structure inside of components with semantic boundary.

The incapability of *hiero-re* in handling long distance reordering is not caused by the limitation of rule representation but by the compromise in rule extraction and decoding for balance be- tween the decoding speed and performance. The hierarchical phrase-based model prohibits any nonterminal X from spanning a substring longer than 10 on the source side that makes the decoding algorithm asymptotically linear-time (Chiang, 2005).

The *dep2str* has a good performance in long distance reordering. However, local lexicalization is restricted by word alignment. Therefore, compatibility with phrases is necessary (Meng, et al. 2013).

8 Conclusion and Future Work

This paper presents case frame constraints for rule extraction and decoding in hierarchical phrase-based model. Compared with HPB rules, the amount of CF-HPBs is decreased. The CF-Rs take the source side as case frame and the target side as string. Our system has an advantage of both long distance reordering and phrase constituency. Moreover, CF-Rs distinguish variables with cases. According to the case frame theory, we interestedly discovered that it can disambiguate some translations. For example, NP with object case or subjective case has different translation.

Case frame constraints are linguistic constraints according to the case grammar theory. It is available for many languages. Case frame can also be used in many aspect of natural language processing, such as summarization, semantic role labeling and bilingual alignment. Meanwhile, more deep semantic case information is expected to further improve the translation quality. Furthermore, It is meaningful to transmit the case information from Japanese to more other languages, and it can be useful to improve the translation quality between more languages.

References

1. Chiang, D.: Hierarchical phrase-based translation. Computational Linguistics 33(2) (2007)
2. Koehn, P., Och, F.J., Marcu, D.: Statistical phrase-based translation. In: Proceedings of HLT-NAACL, pp. 127–133 (2003)
3. Marton, Y., Resnik, P.: Soft Syntactic Constraints for Hierarchical Phrased-Based Translation. In: Proceedings of the Association for Computational Linguistics, pp. 1003–1011 (2008)
4. Huang, Z., Cmejrek, M., Zhou, B.: Soft Syntactic Constraints for Hierarchical Phrase-based Translation Using Latent Syntactic Distributions. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp. 138–147. Association for Computational Linguistics (2010)
5. Gao, Y., Koehn, P., Birch, A.: Soft Dependency Constraints for Reordering in Hierarchical Phrase-Based Translation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 857–868. Association for Computational Linguistics (2011)
6. Marton, Y., Chiang, D., Resnik, P.: Soft syntactic constraints for Arabic-English hierarchical phrase-based translation. Machine Translation 26(1-2), 137–157 (2012)
7. Flannery, D., Miyao, Y., Neubig, G., Mori, S.: Training Dependency Parsers from Partially Annotated Corpora. IJCNLP 2011, 776–784 (2011)
8. Li, J., Tu, Z., Zhou, G., van Genabith, J.: Head-Driven Hierarchical Phrase-based Translation. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pp. 33–37 (2012)

9. He, Z., Meng, Y., Yu, H.: Maximum entropy based phrase reordering for hierarchical phrase-based translation. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp. 555–563. Association for Computational Linguistics (2010)
10. Liu, Q., He, Z., Liu, Y., Liu, S.: Maximum entropy based rule selection model for syntax-based statistical machine translation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 89–97. Association for Computational Linguistics (2008)
11. He, Z., Liu, Q., Lin, S.: Improving statistical machine translation using lexicalized rule selection. In: Proceedings of the 22nd International Conference on Computational Linguistics, vol. 2008, pp. 321–328. Association for Computational Linguistics (2008)
12. Lin, D.: A path-based transfer model for machinetranslation. In: Proceedings of Coling 2004, Geneva, Switzerland, August 23-27, pp. 625–630 (2004)
13. Quirk, C., Menezes, A., Cherry, C.: Dependency treelet translation: Syntactically informed phrasal smt. In: Proceedings of ACL 2005, pp. 271–279 (2005)
14. Ding, Y., Palmer, M.: Machine translation using probabilistic synchronous dependency insertion grammars. In: Proceedings of ACL (2005)
15. Xiong, D., Liu, Q., Lin, S.: A dependency treelet string correspondence model for statistical machine translation. In: Proceedings of the Second Workshop on Statistical Machine Translation, Prague, Czech Republic, pp. 40–47 (June 2007)
16. Xie, J., Mi, H., Liu, Q.: A Novel Dependency-to-String Model for Statistical Machine Translation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 216–226. Association for Computational Linguistics (2011)
17. He, Z., Yao, M., Yu, H.: Learning phrase boundaries for hierarchical phrase-based translation. In: Proceedings of the 23rd International Conference on Computational Linguistics, pp. 383–390. Association for Computational Linguistics (2010)
18. Xiong, D., Zhang, M., Li, H.: Learning translation boundaries for phrase-based decoding. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 136–144. Association for Computational Linguistics (2010)
19. Fillmore, C.J.: The Case for Case. In: Bach, Harms (eds.) Universals in Linguistic Theory, pp. 1–88. Holt, Rinehart, and Winston, New York (1968)
20. Cook, W.A.: SJ: Case Grammar Theory. Georgetown University Press, Washington, DC (1989)
21. Kawahara, D., Kurohashi, S.: Case frame compilation from the web using high-performance computing. In: Proceedings of the 5th International Conference on Language Resources and Evaluation, pp. 1344–1347 (2006)
22. Buchholz, S., Marsi, E.: CoNLL-X shared task on Multilingual Dependency Parsing. In: Proceedings of the Tenth Conference on Computational Natural Language Learning, pp. 149–164. Association for Computational Linguistics (2006)
23. Och, F.J.: Minimum error rate training in statistical machine translation. In: Proceedings of ACL-2003, Sapporo, Japan, pp. 160–167 (July 2003)
24. Zhao, H., Lv, Y., Ben, G., Liu, Q.: Evaluation report for the 7th china workshop on machine translation. In: The 7th China Workshop on Machine Translation, CWMT 2011 (2011)
25. Kurohashi, S., Nagao, M.: Knparser: Japanese dependency/case structure analyzer. In: Proceedings of the Workshop on Sharable Natural Language Resources, pp. 48–55 (1994)

26. Och, F.J., Ney, H.: A systematic comparison of various statistical alignment models. Computational Linguistics 29(1), 19–51 (2003)
27. Koehn, P., Och, F.J., Marcu, D.: Statistical phrase-based translation. In: Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, Edmonton, Canada (July 2003)
28. Stolcke, A.: Srilm - an extensible language modeling toolkit. In: Proceedings of ICSLP, vol. 30, pp. 901–904 (2002)
29. Chiang, D.: A hierarchical phrase-based model for statistical machine translation. Proceedings of ACL 2005, 263–270 (2005)
30. Meng, F., Xie, J., Song, L., Lv, Y., Liu, Q.: Translation with Source Constituency and Dependency Trees. In: Proceedings of EMNLP 2013, Seattle, Washington, USA (October 2013)