

Integrating Case Frame into Japanese to Chinese Hierarchical Phrase-based Translation Model

JinAn Xu, JiangMing Liu, YuFeng Chen, YuJie Zhang, Fang Ming and Shaotong Li

Beijing Key Lab of Traffic Data Analysis and Mining

School of Computer and Information Technology, Beijing Jiaotong University

Haidian District, Beijing, China

{jaxu, 12120432, chenyf, yjzhang, 14120416, 11281094}@bjtu.edu.cn

Abstract

This paper presents a novel approach to enhance hierarchical phrase-based (HPB) machine translation systems with case frame (CF). We integrate the Japanese shallow CF into both rule extraction and decoding. All of these rules are then employed to decode new sentences in Japanese with source language case frame. The results of experiments carried out on Japanese-Chinese test sets. It shows that our approach maintains the advantages of HPB translation systems while at the same time naturally incorporates CF constraints. The case frame rules can complement Hiero-style rules. Our approach is especially effective for language pairs with large word order differences, such as Japanese-to-Chinese.

1 Introduction

In the Japanese-Chinese machine translation task, reordering is the main problem due to substantial differences in sentence structures between these two languages. For example, Japanese has a subject-object-verb (SOV) structure, while, Chinese has a subject-verb-object (SVO) structure.

The pre-ordering technology is one way to handle this problem (Wu, et al., 2011), but it needs to train a pre-ordering model. An hierarchical phrase-based (HPB) model (Chiang, 2005; Chiang, 2007) is a powerful method to cover any format of translation pairs by using synchronous context free grammar. Hiero grammars can capture complex nested translation relationships to handle reordering. However, due to its compromise on the efficiency of rule extraction and decoding, (a) a source language span limit is applied with 10, (b) the number of non-terminals in one rule is set to 2, (c) there is a prohibition of consecutive non-terminals on the source language side of a rule and

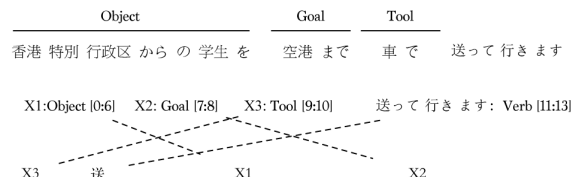


Figure 1: The reordering problem in Japanese-Chinese translation

(d) coarse-grained rules. The HPB model does not perform well when reordering in a Japanese-Chinese machine translation task as shown in Figure 1, which shows an example of long distance reordering covering 13 words.

With a traditional approach, the typical HPB model fails to capture complex reordering information as shown in Figure 1. By contrast, Fillmore (1968) has proposed case grammar, which is effectively proved and originally used in rule-based machine translation (RBMT) system (Yamabana, 1997). Furthermore, Kawahara (1994, 2002) defines the Japanese shallow CF that is widely and successfully used in Japanese dependency tasks provided by CoNLL-09 (Hajič, 2011). Figure 2 shows the CF's ability to capture reordering information.

In this paper, we describe effective approaches to introducing source language Japanese CF in the Japanese-Chinese translation task. Unlike previous work, we are the first to use Japanese CF information on the HPB model, and to transform CF information into SCFG style rules, which is suitable and useful in the original HPB decoder. By importing CF into the HPB model, we expand search space and introduce fine-grained rules.

The remainder of this paper is organized as follows. After introducing Japanese CF, the proposed approach is introduced in Section 3; the experimental results and associate analysis are given

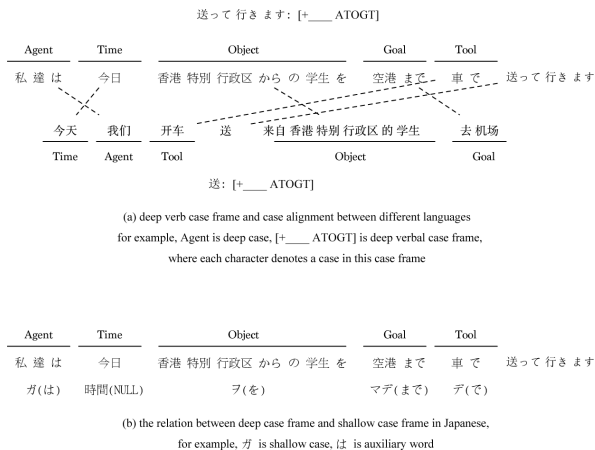


Figure 2: Deep verbal CF and shallow verbal case between Japanese and Chinese

in Section 4. Section 5 briefly discusses related work; Finally, conclusions are drawn in Section 6.

2 Case Frame

Unlike HPB model’s format grammar, case grammar is linguistically sensible and is applied to semantically analyze sentence. Based on case grammar, a sentence will be analyzed using different deep case components (agent, instrument, experiencer, object, location, benefactive, factitive, goal, source and time). This way, Fillmore has defined the deep verb CF, where one example is shown in Figure 2(a).

Deep case is language independent. If two sentences from different languages have the exactly same meaning and description, they will have the same deep case grammar analysis. Figure 2(a) shows the sentence “today we will get him to the airport by car” described respectively in Japanese and Chinese. Meanwhile, Figure 2(a) shows deep case alignment between these two different languages. Deep case alignment in two different languages is one to one mapping. For example, in Figure 2(a), “私達は” (we) is the agent in Japanese, mapping “我们” (we) (agent) in Chinese.

The deep CF is well known, but it is rarely used in statistical machine translation due to the difficulty of the auto-analysis for all languages including Japanese. However, due to the explicit case in Japanese, Kawahara (2002) redefines the shallow verbal CF in Japanese shown in Figure 2(b), where an auxiliary word contributes to the shallow CF

Type	Examples of Japanese shallow case
Agent	ガ (ga)
Object	ニ (ni), ヘ (he), ヲ (wo)
Instrument	ニ (ni), デ (de)
Experiencer	ニ (ni), ト (to), ヘ (he)
Source	ニ (ni), ヘ (he), カラ (kara), ヨリ (uori) マデ (made), ノ (no)
Goal	ニ (ni), ヘ (he)
Location	ニ (ni), デ (de), カラ (kara) ヨリ (yori), マデ (made)
Time	ニ (ni), デ (de), カラ (kara) ヨリ (yori), マデ (made)
Others	無格 (none), 修飾 (modification)

Table 1: The mapping between a deep case and a Japanese shallow case.

analysis. As a result, recent research has achieved high accuracy (more than 90%) on Japanese shallow CF analysis (Kawahara and Kurohashi, 2006). Between the deep case and the Japanese shallow case, there is a many-to-many relation shown in Table 1. In this paper, we will only use “case frame” to represent Japanese verbal shallow CF for short.

3 The proposed approach

A case frame is the linguistic concept, which provides linguistic guidance for derivation. Here, we present a method to alleviate complex reordering problems in the Japanese-Chinese machine translation task with case frame.

Generally, we obtain both the case frame and the hiero-style SCFG from the training data, and then transfer the case frame rule (CFR) to SCFG style and use both of them in decoding with the SCFG. The method benefits from both hiero-style translation and linguistic information. In the rule extraction of our approach, we acquire case frame rules using fuzzy strategy and hiero-style rules using traditional HPB rule extraction method. In decoding, we use the traditional HPB decoder with CYK and cube pruning.

Figure 3 shows an example of CFRs extraction processing from a pair of word-aligned Japanese-Chinese sentences with a source language CF, and their SCFG style.

3.1 Case Frame Rules Extraction

As described in section 2, the Japanese shallow case frame can be obtained through surface analysis. This way, we can extract case frame reordering rules from sentence pairs with alignment information as shown in Figure 4, where original case

Agent	Time	Object	Goal	Tool
私達は	今日	香港特別行政区からの学生を	空港まで	車で送って行きます
ガ(は)	時間(NULL)	ヲ(を)	マデ(まで)	デ(で)
今天	我們	开车送	来自香港特别行政区的学生	去机场

Original Case Frame Rules

- o1: (ガ)私達は、我們
o2: (時間)今日、今天
o3: (ヲ)香港特別行政区からの学生を、来自香港特别行政区的学生
o4: (マデ)空港まで、去机场
o5: (デ)車で、开车
o6: (x1:ガ) (x2:時間) (x3:ヲ) (x4:マデ) (x5:デ)送って行きます、
x2 x1 x5 送 x3 x4)

Case Frame Rules with SCFG style converting from o1...o6

- r1: X_ガ -> (私達は、我們)
r2: X_時間 -> (今日、今天)
r3: X_ヲ -> (香港特別行政区からの学生を、
来自香港特别行政区的学生)
r4: X_マデ -> (空港まで、去机场)
r5: X_デ -> (車で、开车)
r6: X -> (X_ガ X_時間 X_ヲ X_マデ X_デ 送って行きます、
X_時間 X_ガ X_デ 送 X_ヲ X_マデ)

- o1: (ガ)私達は、我們
r1: X_ガ -> (私達は、我們)

(a) the example of phrase rule transformation

- o6: (x1:ガ) (x2:時間) (x3:ヲ) (x4:マデ) (x5:デ) 送って行きます (NULL),
x2 x1 x5 送 x3 x4)
r6: X -> (X_ガ X_時間 X_ヲ X_マデ X_デ 送って行きます、
X_時間 X_ガ X_デ 送 X_ヲ X_マデ)

(b) the example of reordering rule transformation

Figure 3: Example of CFRs extraction processing.

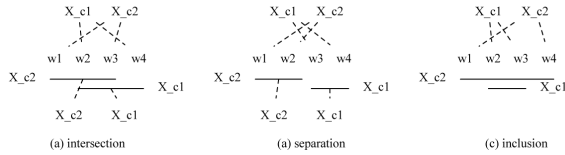


Figure 4: The fuzzy extraction strategy for CF reordering rule.

frame rules are from $o1$ to $o6$.

Given a source language case frame and related word alignment, one case frame is mapped to the case frame reordering rule set, where there are two kinds of rules: reordering rule and phrase rule.

- Phrase rule: Each component in a case frame generates one phrase rule. We extract the phrase rule by following the traditional phrase-based model's strategy (Och and Ney, 2004). Each phrase rule has a case distinction associated with a shallow case in a case frame like $r1$ to $r5$ in Figure 3.

- Reordering rule: One case frame generates one reordering rule. For reordering rule extraction, we need to compute the relatively order of target language span associated with each case slot. The order is relatively soft to the word alignment. For example, if a source language phrase A covers target span $[2, 4]$ and the other source language phrase B covers target span $[1, 3]$, then the phrase A is relatively right to the phrase B in target side; if a span is covered by the other one, the rule is forbidden during extraction. All of the

Figure 5: The example of transformation for phrase rule and reordering rule.

possible case frames with word alignment can be seen in Figure 4, where only (c) rule is forbidden. The reordering rule is like $r6$ in Figure 3.

3.2 Transforming Case Frame Rule into SCFG style

To make case frame rules directly accessible to the Hiero-style decoder with performs decoding with SCFG rules, we convert original case frame rules into SCFG style. And then, case frame rule is defined as SCFG-style, which is a little different from hiero rules.

- Phrase rule transformation: We take $o1$ as an example transforming into $r1$ shown in Figure 5(a). We use $o1$'s case distinction as case distinction of $r1$'s left. The source side of the $r1$'s right is source phrase in $o1$ and the target side is target phrase in $o1$.

- Reordering rules transformation: We take $o6$ as an example transforming into $r6$ shown in Figure 5(b). We also use $o6$'s verb case distinction as case distinction of $r6$'s left. (default X if there is no case distinction in this example). Each slot of $o6$ is transformed into related X with respective case distinction in $r6$. The target side of the rule's right is target language's reordering. It is clearly seen that if there is no non-terminals in the right of reordering rule, reordering rule is the same with phrase rule.

In this way, each case frame rule is associated with exactly one SCFG rule. Therefore, we can obtain a fine-grained SCFG from case frames due to case distinction. On one hand, non-terminals associated with case are linguistically sensible. For

example of r_4 , “空港 まで” with “マデ” case is translated to “去 机场” that means “to airport”. On the other hand, it can capture complex reordering information. For example of r_6 , the source side of the rule’s right means that “ガ” (who) “時間” (when) “ヲ” (whom) “マデ” (where) “デ” (how) “送って 行きます” (send) in Japanese order, and the target side of the rule’s right means that “時間” (when) “ガ” (who) “デ” (how) “送” (send) “ヲ” (whom) “マデ” (where) in Chinese order.

For reordering the rule extraction, we need to compute the relatively order of target language span associated with each case slot. The order is relatively soft to the word alignment. For example, if a source language phrase A covers target span [2, 4] and the other source language phrase B covers target span [1, 3], then the phrase A is relatively right to the phrase B on the target side; if a span is covered by the other one, the rule is forbidden during extraction. All the possible CFs with word alignment can be seen in Figure 4, where only (c) rule is forbidden.

Generally, we define the transformed case frame rules as SCFG style:

$$X \rightarrow \langle \gamma, \alpha, \sim \rangle \quad (1)$$

Where X is non-terminal, γ and α are both strings of terminals and non-terminals as the same with SCFG in the HPB model. Compared with SCFG in the HPB model, the only difference is that non-terminals are distinguished by case as shown in Figure 3 from r_1 to r_6 .

3.3 Decoding

Both transformed case frame rules and HPB rules can be applied using traditional Hiero decoders with a slight modification. Here we follow the description of Hiero decoding by Chiang (2007). The source sentence is parsed under the Hiero grammar using the CYK algorithm. Each cell in the CYK grid is associated with a list of rules that applies to its span from the bottom up. For each derivation, we apply cube pruning (Chiang, 2007) and beam search technology.

This procedure accommodates traditional HPB rules directly. We use traditional HPB rules for translation as shown in Figure 6(a). For example, the traditional rule can be applied in the span (14,16). Since the span (4, 18) is longer than 10 words, the traditional rule cannot be applied in the span.

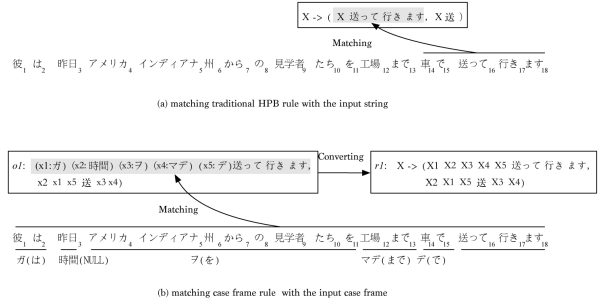


Figure 6: Decoding with both traditional hiero grammar and case frame.

We move our focus towards case frame reordering rules, and analyze sentences and obtain all the case frames, and then for each CF, we match rules to the span related to the CF. If a match is found, the CYK cell for the span is selected, and that rule is added to the list of rules in the selected CYK cell as shown in Figure 6(b). For example, the span (1,18) can be matched with r_6 . The complex reordering can be captured by r_6 .

It is clear that the HPB rules have non-terminals without any distinction and the case frame rules have non-terminals with case distinction. Generally, there are two kinds of non-terminals: X and X with case. During decoding, we respectively use three kinds of constraints on case frame rule matching:

Without constraints ignore all the case distinction in case frame rules, so case frame rule format is the same with HPB rules. In this way, we just expand SCFG.

Soft constraints admit the match between different case distinctions by adding extra dynamic feature – soft count. For example, X with “ヲ” is allowed to match X with “マデ” by adding 1 to soft count.

Hard constraints only admit the completed and exact match. On one hand, we admit X to match all of the X with or without distinction, on the other hand, we only allow X with distinction to match X with the same distinction.

3.4 Features

The baseline feature set used in this work consists of 7 features, including a strong 5-gram language model, bidirectional translation probabilities, bidirectional lexical probabilities, and a word count, a glue rule count. In the CF reordering rule, bidirectional translation probabilities and bidirection-

al lexical probabilities are also used during decoding. In addition, we introduce several features for applying case frame rules, and we adopt these features to log-linear model during decoding.

- Rule type indicators. For soft or hard constraint, we consider two indicator features, indicating case frame rules, case frame reordering rules. Case frame rules indicator feature is used to distinguish case frame rules and original HPB rules. Case frame reordering rules indicator feature is used to distinguish phrase rules and reordering rules in case frame rule set.

- Dynamic soft constraints. For soft constraints, we consider the soft constraints. Note that when X with case mismatches X with other different case, we add dynamic soft constraints count for this mismatching instead of prohibition.

4 Evaluation

4.1 Experimental Setup

We report results for this Japanese-Chinese task. We use two data sets, where one uses news from the 7th China Workshop on Machine Translation (CWMT) including 280 thousand sentence pairs for training, 500 sentence pairs for parameter optimization and 900 sentence pairs for testing, the other, from Asian Scientific Paper Excerpt Corpus-Japanese to Chinese (ASPEC-JC) includes 680 thousand pairs for training, 2090 sentence pairs for parameter optimization and 1800 sentence pairs for testing.

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.

4.2 Results

For comparison, we also manually modify the extracted case frame rules of development and test data with case frame information according to the Japanese and Chinese grammar. We report machine translation performance in Table 2 using case insensitive BLEU-4 metric (Papineni et al., 2002), considering the balance of the performance

entry	system	CWMT	
		Rule #	BLEU-4
exp1	baseline	14.70M	0.2805
exp2*	exp1 + cfr	14.70M+0.71M	0.2836
manually	exp1 + cfr	/	0.2865
entry	system	ASPEC	
		Rule #	BLEU-4
exp1	baseline	215.67M	0.2717
exp2*	exp1 + cfr	215.67+7.21M	0.2748
manually	exp1 + cfr	/	0.2763

Table 2: BLEU[%] scores of various systems. * means that a system is significantly different from the baseline at $p < 0.01$. M means million and + means hierarchical rules with CFRs.

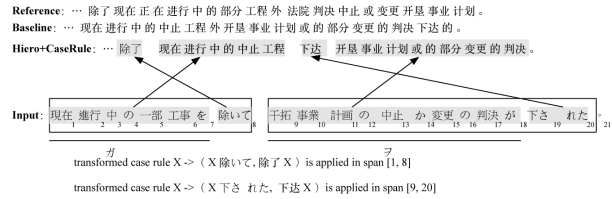


Figure 7: Comparison of translations generated by the baseline and improved systems.

of lexical and phrase. The experiments are organized as follows:

- exp1: we use the NiuTrans (Xiao, 2012) hierarchical phrase-based model as strong baseline system.
- exp2: we transform CFRs into SCFG-style rules without any case distinction, and add these rule into exp1 system.

4.3 Analysis

Finally, we discuss an example of real translation from our test set. See Figure 7 for translations generated from different systems. The Japanese input sentence contains “...下された” which is usually translated into “下达...” (i.e. a transformed CF reordering rule “ $X \rightarrow X$ (下された, 下达 X)”). However, because the “...下された” pattern spans 12 words and that is beyond the span limit, our baseline is unable to apply this desired rule and so it chooses the wrong reordering translation. When importing CF reordering rule which captures the CF “(を) 下された”, we can transform the CF reordering rule into one that is SCFG-style and achieve right reordering information.

- Better reordering Main structure in Japanese structure is SOV-style, which is different from

Source	集合時間をご確認の上, 時間厳守にてご集合ください。 We need to be on time, please be punctual!
Ref	确认好集合时间后, 请严格按时间集合。
Hiero	集合时间确认, 时间严格集合。
CFs	确认好集合时间, 请准时集合。

Figure 8: Actual translations produced by the baselines and our system.

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, case frame rules have better phrase reordering. In the example as shown in Figure 8, the source sentence main centered verbs contain the word “確認(confirm)” and the word “集合(gather)”. The Hiero result mistakenly treats that objective phrase as subjective (SOV), thus results in translation with different structure from source sentence. Conversely, our system captures this component relations in case frame and translates it into the SVO structure.

- Better exical translation results Moreover, we also find that our system can get better lexical translation results, for instance, the result of the word “時間厳守(punctuality)”, as indicated in Figure 8.

5 Related Work

Recently linguistically-motivated models have been intensively investigated in MT. In particular, source tree-based models (Liu et al., 2006; Huang et al., 2006; Eisner, 2003; Zhang et al., 2008; Liu et al., 2009a; Xie et al., 2011) have received growing interest due to their excellent ability to model source language syntax for better lexical selection and reordering. Alternatively, the hierarchical phrase-based approach (Chiang, 2005) considers the underlying hierarchical structures of sentences but does not require linguistically syntactic trees on either language’s side.

There are several lines of work for augmenting hierarchical phrase-based systems with the use of source language linguistic information. Xiao (2014) incorporates source syntax into the hierarchical phrase-based model. They develop procedures for joint decoding and optimization within a single system by transforming tree-to-string rules

into SCFG rules. By enlarging SCFG grammar, they perform well on Chinese-English tasks. Our approach is motivated by high-precision Japanese case analysis, and aims to augment the search space of Hiero with linguistically-motivated hypotheses. Moreover, we consider hiero as the backbone model and only introduce and transform Japanese CF into SCFG rules where they can contribute.

Another related line of work is to introduce pre-ordering approach for Japanese main structure. Wu (2011) and Sudoh (2013) propose several methods to train pre-ordering model for pre-ordering. We note that, we have no need to train extra pre-ordering models for the Japanese main structure, and we only use the high-precision Japanese explicit case analysis to improve Japanese-Chinese translation performance described in this paper.

6 Conclusion and Future Work

We have presented an approach to improving Hiero-style systems by augmenting the SCFG with Japanese case frame rules. The input case frame are used to introduce new linguistically-sensible hypotheses into the translation search space while maintaining the Hiero robustness qualities and avoiding computational explosions. We obtain significant improvements over a strong Hiero baseline in the Japanese-to-Chinese task.

This paper presented an approach to improve HPB model systems by augmenting the SCFG with Japanese CFRs. The CF are used to introduce new linguistically-sensible hypotheses into the translation search space while maintaining the Hiero robustness qualities and avoiding computational explosions. We obtain significant improvements over a strong HPB baseline in the Japanese-to-Chinese task. We will try to improve the performance of our system with soft constraint or hard constraint using case frame rules, and we will challenge to resolve the problem of tense, aspect and some special grammatical sentences of Japanese to Chinese translation.

Acknowledgments

The authors were supported by National Nature Science Foundation of China (Contract 61370130 and 61473294), the Fundamental Research Funds for the Central Universities (2014RC040), and Project 20141004065 supported by National

Training Program of Innovation and Entrepreneurship for Undergraduates.

References

- Wu, X., Sudoh, K., Duh, K., Tsukada, H., and Nagata 2011. *Extracting preordering rules from chunk-based dependency trees for Japanese-to-English translation.*, Proceedings of the 13th Machine Translation Summit. pages 300–307.
- Chiang, David. 2005. *A hierarchical phrase-based model for statistical machine translation.*, Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2005. pages 263–270.
- Chiang, David. 2007. *Hierarchical phrase-based translation.*(J) computational linguistics, 33(2), pages 201–228.
- Fillmore, C. J. 1967. *The case for case.*
- Kiyoshi Y., KAMEI S. and MURAKI K. A. 1997. *A Hybrid Approach to Interactive Machine Translation.*, In Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, Nagoya, Japan, August 23-29, 1997. Morgan Kaufmann, 1997, pages 977–982.
- Kawahara, D., and Kurohashi, S. 2002. *Fertilization of case frame dictionary for robust Japanese case analysis.* In Proceedings of the 19th international conference on Computational linguistics-Volume 1, 2002, pages 1–7, Association for Computational Linguistics.
- Kawahara D, Kurohashi S. 2006. *Case frame compilation from the web using high-performance computing.* In Proceedings of the 5th International Conference on Language Resources and Evaluation. 2006, pages 1344–1347.
- Och F. J. and Ney H. 2004. *TextRank: The alignment template approach to statistical machine translation[J]*, Computational linguistics, 2004, 30(4): 417–449.
- Hajič J., Ciaramita M., Johansson R., et al. 2009. *shared task: Syntactic and semantic dependencies in multiple languages.* In Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task. Association for Computational Linguistics, 2009, pages 1–18.
- Liu Y, Liu Q, Lin S. 2006. *Tree-to-string alignment template for statistical machine translation.* In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2006, pages 609–616.
- Huang, L., Knight, K. and Joshi, A. 2006. *Statistical syntax-directed translation with extended domain of locality.* In Proceedings of AMTA pages 66–73.
- Eisner J. 2003. *Learning non-isomorphic tree mappings for machine translation.* In Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 2. Association for Computational Linguistics, 2003, pages, 205–208.
- Min Zhang, Hongfei Jiang, Haizhou Li, Chew Lim Tan and Sheng Li. 2008. *A tree sequence alignment-based tree-to-tree translation model.* In Proceedings of ACL-08: HLT (2008): 559.
- Liu Y, Lü Y, Liu Q. 2009. *Improving tree-to-tree translation with packed forests.* In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2. Association for Computational Linguistics, 2009, pages 558–566.
- Xie J., Mi H. and Liu Q. 2011. *A novel dependency-to-string model for statistical machine translation.* In Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2011, pages 216–226.
- Xiao, T., Zhu, J., Zhang, H. and Li, Q. 2012. *Niu-Trans: an open source toolkit for phrase-based and syntax-based machine translation.*, In Proceedings of the ACL 2012 System Demonstrations. Association for Computational Linguistics, pages 19–24.
- Tong Xiao, Adrià de Gispert, Jingbo Zhu and Bill Byrne. 2014. *Effective Incorporation of Source Syntax into Hierarchical Phrase-based Translation.* In Proceedings of Coling 2014, pages 2064–2074.
- Sudoh K, Wu X, Duh K, et al. 2013. *Syntax-Based Post-Ordering for Efficient Japanese-to-English Translation.* ACM Transactions on Asian Language Information Processing (TALIP), 2013, 12(3): 12.