A Comparative Study on Discontinuous Phrase Translation

Jiajun Zhang and Chenqqing Zong

NLPR, Institute of Automation Chinese Academy of Sciences, Beijing, China
{jjzhang,cqzong}@nlpr.ia.ac.cn

Abstract. Many research works have reported discontinuous phrase translation can significantly improve translation quality, but experiments are conducted in only one translation direction (e.g. Chinese-to-English) with only one language pair. Thus, two questions remain that whether the discontinuous rules are always much helpful in different language pairs? Furthermore, what kind of discontinuous rules (e.g. source-side discontinuity or target-side discontinuity) contributes most to the performance improvement? To answer these two questions, this paper conducts a comparative study on the contribution of different kinds of discontinuous rules in both translation directions with various language pairs. Then, with this comparative study, this paper proposes a role-based rule filtering strategy to filter the large amount of discontinuous rules that contribute very little to translation quality.

Keywords: statistical machine translation, discontinuous phrases, role-based rule filtering.

1 Introduction

On one hand, many research works[1,2,4,7,10,11] reported discontinuous phrase translation significantly improves the translation quality in statistical machine translation (SMT). However, nearly all the experiments are conducted only in one translation direction (e.g. Chinese-to-English) with only one language pair. Thus, it remains an interesting question whether the discontinuous rules always have a big contribution in different language pairs. This paper plans to answer this question using extensive experiments.

On the other hand, several researchers found that different kinds of discontinuous rules (e.g. source-side discontinuity or target-side discontinuity, Figure 1 for example) contribute to the translation quality quite differently and they reported the finding that source-side discontinuity plays a much more important role than target-side discontinuity in Chinese-to-English translation. This finding is reported at least by two related studies, i.e. [4] and [11]. Specifically, [11] used a syntax-based statistical machine translation system that aims at better translating non-continuous phrases. [4] extended a traditional phrase-based system for discontinuous phrase translation. They made great efforts to translate target-side discontinuity but they disappointedly found that

M. Zhou et al. (Eds.): NLPCC 2012, CCIS 333, pp. 164–175, 2012.

target-side discontinuity is much less useful in Chinese-to-English translation. They also reported the same finding using a hierarchical phrase-based system Joshua[6]. It seems that the finding is independent of translation models. Thus, there is another interesting question whether source-side discontinuity is much more useful than target-side discontinuity in the inverse direction English-to-Chinese translation and in other language pairs? This paper is going to give the answer with comprehensive experiments.

Source-side discontinuity:

在_zài X 中_zhōng → in X

Target-side discontinuity:

考虑_kǎolǜ X → take X into account

Fig. 1. Examples for source-side discontinuity and target-side discontinuity

In statistical machine translation (SMT), translation rules increase dramatically if discontinuity is allowed. Therefore, if we know some kinds of discontinuous rules are useless, we can remove large amount of such redundant rules without performance loss. Correspondingly, we propose a role-based rule filtering strategy.

2 Related Work

Our work described in this paper includes three parts: the first one and also the most important one is to study the relationship between source discontinuity and target discontinuity; the second one is investigating the contribution of discontinuity translation in different language pairs; and the third one is designing a role-based rule filtering strategy.

Relationship between source discontinuity and target discontinuity [11] tried to model discontinuity translation in tree sequence based model and found that source discontinuity is much more effective than target discontinuity in Chinese-to-English translation. [4] also discovered this phenomenon in their extended phrase-based model for non-continuous phrase translation. They meanwhile pointed out the phenomenon exists in hierarchical phrase-based model too. The reason has not been clearly explained.

Contribution of discontinuity translation in different language pairs [8] and [2] suggested that the ability to translate discontinuous phrases is important to model translation. [4,7,11] empirically showed that discontinuity translation indeed leads to substantial improvements. However, they all tested the discontinuity translation in only one translation direction for one language pair. It is worth to study whether it is much useful in other language pairs.

Rule filtering strategy [12] proposed a count-based rule filtering approach which discards rules occurring less than a minimum count. [9] removed those rules whose target parts are not well-formed dependency trees. [5] presented a pattern-based rule filtering method which considers the possible 66 rule patterns.

3 Discontinuous Rule Classification

In order to deeply investigate the discontinuous rules, we first classify them into different kinds. Generally, the discontinuous rules are classified into source-side discontinuity and target-side discontinuity. We can see that these two classes are overlapped. To have a finer classification, we can divide discontinuous rules into three kinds: source-only discontinuity (target is contiguous), target-only discontinuity (source is contiguous), and both-side discontinuity. Table 1 gives the classification details of hierarchical rules[2].

Table 1. Three Categories of Discontinuities. SDR, TDR and BDR denote source-side, target-side, and both-side discontinuous rules respectively

Categories	Rule's source	Rule's target
SDR	uXv $uXvX$ $XuXv$ $uXvXw$	Contiguous ones
TDR	Contiguous ones	$u'Xv'$ $u'Xv'X$ $Xu'Xv'$ $u'Xv'Xw'$ $u'XXv'$
BDR	uXv $uXvX$ $XuXv$ $uXvXw$	$u'Xv'$ $u'Xv'X$ $Xu'Xv'$ $u'Xv'Xw'$ $u'XXv'$

However, we still think this classification is not sufficient because most of both-side discontinuities can be obtained by combining two contiguous rules. For example, the rule $X \rightarrow \langle uX_1 vX_2, u'X_1 v'X_2 \rangle$ [1] is the concatenation of $X \rightarrow \langle uX, u'X \rangle$ and $X \rightarrow \langle vX, v'X \rangle$ if u is aligned to u' and v is aligned to v'. In order to make a strict classification, we define the *strict both-side discontinuity* with word alignments:

Definition: Given a rule $X \rightarrow \langle \gamma, \alpha \rangle$ which is both-side discontinuity with word alignment between γ and α, if a terminal element in one side is aligned to two or

[1] u, v, u', v' are terminal strings, we consider them as terminal element.

more terminal elements in the other side, then the both-side discontinuity is a strict both-side discontinuity.

For example, for the rule $X \rightarrow \langle uX_1 vX_2, u'X_1 v'X_2 \rangle$, if u is aligned to both u' and v', we consider this rule as a strict both-side discontinuous rule.

4 Comparison Study on Discontinuous Phrase Translation

4.1 Experimental Set-Up

The translation system we use in this paper is an open-source hierarchical phrase-based system Joshua[6]. We conduct our experiments on four language pairs: Chinese-English, Spanish-English, French-English, and German-English. For Chinese-English, the training data, extracted from LDC corpora, consists of about 1.92 million parallel pairs. For the direction Chinese-to-English, the tuning set is NIST2006 test data and the test sets include NIST2005 test data (test-1) and NIST2008 test data (test-2). For the direction English-to-Chinese, we split NIST2008 test data into two parts: the first 800 sentences are used as tuning set and the rest 1059 sentences are used as test set (test-1). Another test set (test-2) uses NIST2005 Chinese-to-English first reference as source, and original source as reference. A 5-gram language model is built on the target part in both directions.

For other three language pairs, all data are from the fourth workshop of machine translation[2] (WMT09). The translation model is trained with Europarl corpus: about 1.48M sentence pairs for French-English and German-English, and 1.47M for Spanish-English. We train the 5-gram language models on the large monolingual data: 227M words for German, 218M words for French, 94M for Spanish and 549M for English. The tuning set is *Devset2009-a* consisting of 1025 sentences, the first test set (test-1) is *Devset2009-b* with 1026 sentences, and the second test set (test-2) is the test data of WMT09 including 2525 sentences.

The word alignment is obtained with GIZA++ and case-insensitive BLEU-4 is employed to evaluate the performance of translation results.

To have a detailed comparison, we design six groups of experiments according to the rules that are applied in translation: 1) only with contiguous rules (CR); 2) CR plus source-only discontinuous rules (+SDR); 3) CR plus target-only discontinuous rules (+TDR); 4) CR plus both-side discontinuous rules (+BDR); 5) CR plus both-side discontinuous rules excluding strict both-side discontinuous rules (-SBDR) and 6) with all rules (ALL). The 5th configuration is designed to test the importance of strict both-side discontinuity.

Tables 2-5 report all the translation results in four language pairs. The **bold** figures in tables just show which one, of source discontinuity and target discontinuity, leads to bigger gains.

[2] `http://www.statmt.org/wmt09/`

Table 2. Results of Chinese-English translation

gaps	C→E			E→C		
	tuning	test-1	test-2	tuning	test-1	test-2
CR	28.76	28.07	21.78	30.44	30.21	25.44
+SDR	**29.46**	**28.65**	**22.24**	30.60	30.22	25.54
+TDR	28.86	28.26	21.90	**30.94**	**30.75**	**25.73**
+BDR	29.31	28.78	22.33	30.99	30.85	25.62
-SBDR	29.05	28.37	22.06	30.71	30.64	25.51
ALL	29.69	28.87	22.68	31.47	31.31	25.89

Table 3. Results of French-English translation

gaps	F→E			E→F		
	tuning	test-1	test-2	tuning	test-1	test-2
CR	22.37	22.58	21.94	21.95	22.05	21.36
+SDR	**22.54**	**22.70**	22.01	21.93	21.94	21.10
+TDR	22.51	22.53	**22.05**	**22.06**	**22.00**	**21.35**
+BDR	22.61	22.68	21.98	22.27	22.11	21.53
-SBDR	22.62	22.63	21.91	22.10	22.01	21.46
ALL	22.58	22.72	21.80	22.24	22.08	21.50

Table 4. Results of German-English translation

gaps	G→E			E→G		
	tuning	test-1	test-2	tuning	test-1	test-2
CR	18.25	19.05	15.83	12.78	13.02	10.82
+SDR	**18.50**	**19.44**	**15.85**	**12.97**	13.20	11.00
+TDR	18.41	19.15	15.70	12.91	**13.24**	**11.15**
+BDR	18.51	19.42	15.69	12.82	13.32	11.24
-SBDR	18.43	19.20	15.61	12.72	13.20	11.06
ALL	18.60	19.45	15.92	13.03	13.18	10.98

Table 5. Results of Spanish-English translation

gaps	S→E			E→S		
	tuning	test-1	test-2	tuning	test-1	test-2
CR	23.62	23.50	22.17	23.10	22.95	21.14
+SDR	23.56	23.47	22.07	**23.24**	**23.07**	**21.37**
+TDR	**23.81**	**23.62**	**22.20**	23.17	22.95	21.31
+BDR	23.78	23.55	22.40	23.63	23.15	21.38
-SBDR	23.67	23.38	22.15	23.38	23.12	21.29
ALL	23.59	23.46	22.03	23.50	23.13	21.44

4.2 Contribution of Discontinuous Rules in Different Language Pairs

From Table 2, we can easily see that translation with all discontinuous rules (in ALL lines) substantially outperforms translation with only contiguous rules (in CR lines) and the gains are up to absolute 1.1 BLEU score in English-to-Chinese test-1 set. It verifies that discontinuous rules are very important in Chinese-English translation. It is in line with the conclusions of previous works[4,7,11] in which they only tested on Chinese-to-English translation. However, when coming to European language pairs, the discontinuous rules are much less effective. The best improvement is only 0.4 BLEU point on German-to-English test-1 set. Furthermore, translation with discontinuous rules sometimes even decreases the translation quality (such as that on Spanish-to-English translation). We speculate that the systematic divergency between Chinese and English is much larger than that between these European language pairs; thus the complicated translation rules such as discontinuous rules are very helpful in Chinese-English translation while these rules are not much necessary in European language pairs. In sum, the experimental results show that discontinuous phrase translation does not always contributes much to translation quality in any language pair.

4.3 Source Discontinuity vs. Target Discontinuity

If we focus on the bold figures in Tables 2-5, we can easily see that source discontinuity is not always more helpful than target discontinuity. In Chinese-English translation, the interesting discovery is obvious that the source discontinuity wins in Chinese-to-English translation while the target discontinuity wins in the inverse English-to-Chinese translation. In other three language pairs (Tables 3-5), it also seems if source discontinuity is more helpful in one translation direction, the target discontinuity is more useful in the inverse translation direction (test-2 set in F→E and tuning set in E→G are two exceptions), although the contribution difference between source discontinuity and target discontinuity is not significant. We believe that it is because the discontinuous rules do not contribute much in these three language pairs essentially.

4.4 Contribution of Both-side Discontinuity

Both-side discontinuity does not affect the contribution difference between source and target discontinuity since it has the same effect on both of them. We can see from the tables that both-side discontinuity always improves translation quality (+BDR lines in tables). To figure out how much of the improvement does the strict both-side discontinuity contribute, we re-train and re-test without the strict both-side discontinuities (-SBDR lines in tables). The performance decreases for the lack of SBDR in all cases. This indicates that the strict both-side discontinuity is important though the rules of this kind only account for a small part (less than 1/10 of BDR). By contrast, BDR excluding SBDR occupy a quite large part of all rules but the contribution is relatively small; therefore, lots of such rules of the kind are incline to be filtered.

5 Role-Based Rule Filtering

Knowing the contribution of different kinds of discontinuities can give us a lot of inspiration. For example, since we have known the target discontinuity is more effective in English-to-Chinese translation, we can specially handle the target discontinuity with syntactic or semantic information to further improve the translation quality. On the contrary, there will be little value to deal with target discontinuity with great efforts in Chinese-to-English translation.

The most direct idea we may think of is to filter the rules which have nearly no contribution to translation quality. Just take a look at Table 5 for Spanish-to-English translation, we can use the +TDR configuration as the best setting which obtains nearly the best performance but only contains less than 30% of all the rules. In the rest of this paper, we will introduce a role-based rule filtering strategy.

5.1 Rule Classification by Role

The number of translation rules extracted by the hierarchical systems is usually extremely large. Even though we only keep those which are relevant to the test set, the number of rules is still several millions. For example, the initial rule set for English-to-French test data exceeds 6 million rules. Thus, it is still a challenge to reduce the rule set without decreasing translation quality in order to improve the memory efficiency and speed up the decoder. Each rule filtering method should first classify the translation rules and then throw the rules of some classes away with specific strategy. [12] classified the rules based on their counts and removed the rules which occur less than a minimum count (such as 3). [9] divided the rules into two kinds: one must be well-formed dependency tree in the target part, the other is the rest. They only retain the rules of the first kind. [5] proposed a pattern-based rule classification. A rule pattern is composed by source format (terminal elements, non-terminal elements, and their order) and target format. For example, $\langle uX_1vX_2, u'X_1v'X_2\rangle$ and $\langle uX, u'X\rangle$ are two patterns. In total, there are 66 possible rule patterns. They discarded the rules of patterns which do not reduce translation quality.

In this paper, we propose a role-based rule classification. The role here means the function that the rules show in translation. In principle, the rules in each translation model can be divided into two functions: one is for phrase translation; and the other is for phrase reordering. For hierarchical phrase-based models and syntax-based models, the phrase translation rules can be further classified into continuous phrase rules and discontinuous rules. Additionally, the composed rules[3] are sometimes useful. Based on the analysis and our classification of discontinuous rules, we partition the hierarchical rules into seven categories: (1) Lexical phrase rules (LPR); (2) Reordering rules (RR); (3) Source discontinuous rules (SDR); (4) Target discontinuous rules (TDR); (5) Strict both-side discontinuous rules (SBDR); (6) One non-terminal composed rules (CR1NT); (7) Two non-terminal composed rules (CR2NT)

The lexical phrase rules are just the rules without non-terminals. The rules, such as $X \rightarrow \langle uX, Xu' \rangle$ and $X \rightarrow \langle uX_1vX_2, u'X_2v'X_1 \rangle$, are reordering rules. We have detailed the meaning of the three kinds of discontinuous rules in previous sections. One non-terminal composed rules are the rules with one non-terminal but do not represent reordering and discontinuity translation, i.e. the rule $X \rightarrow \langle uX, u'X \rangle$ which can be concatenated by lexical phrase rule $X \rightarrow \langle u, u' \rangle$ and the general glue rule. Two non-terminal composed rules are similar to CR1NT. It is worth noting that RR and SBDR maybe overlap, i.e. if source terminals and target terminals in $X \rightarrow \langle uX_1vX_2, u'X_2v'X_1 \rangle$ cross linked, the rule also belongs to SBDR. To enable the role-based rule classification to form a partition of the rules, we consider the rules of this kind as SBDR.

With the role-based rule classification, we have investigated which kind rules are essential and which ones can be removed safely. The premise of our filtering strategy is that some kinds of rules are useless to translation quality in certain language pairs.

5.2 Filtering Strategy

Because we have known the contribution of different kinds of discontinuous rules, we can design different rule filtering strategies according to the different language pairs. For example, in Spanish-to-English translation, the +TDR configuration can be used as our basis for filtering since it has nearly the same performance with that using all rules and meanwhile excludes more than 70% rules. However, in Chinese-to-English translation, the ALL configuration cannot be substituted. Thus, in this case, we adopt the leave-one-out (LOO) method to test the effectiveness of each kind rule. In this subsection, the Spanish-to-English and Chinese-to-English translations (test-1 data set used as the test set) are used as instances to introduce the filtering strategies.

Spanish-to-English: Since the CR configuration in Table 2-5 includes LPR, and part of RR, CR1NT and CR2NT[3], we first test the case of +TDR configuration with CR1NT, CR2NT, and RR, and then study whether they are necessary[4]. To have a better comparison, we also try the pattern-based rule filtering strategy[5]. Table 6 gives the patterns which occupy the most part but are proved to be useless in Arabic-to-English translation[5].

The experimental results are shown in Table 7 and Table 8. An interesting thing we can see from Table 7 that the role-based rule filter can further discard a large number of rules (Line 3 in Table 7) and keep approximately the same performance with that using all rules. The rules of this configuration only account for about 16% of all rules. We are much surprised that the reordering rules are not important and can be removed in

---

[3] The remaining RR are $\langle Xu, u'X \rangle$ and $\langle uX, Xu' \rangle$. The remaining CR1NT include $\langle uX, u'X \rangle$ and $\langle Xu, Xu' \rangle$. The rest CR2NT are $\langle X_1uX_2, X_1u'X_2 \rangle$, $\langle X_1uX_2, u'X_1X_2 \rangle$ and $\langle X_1uX_2, X_1X_2u' \rangle$.

[4] Lexical phrase rules are basis for every translation model, and they should be included always. However, they can partly filtered by other strategies such as count-based method.

Spanish-to-English translation. The reason will be further investigated in our future work. Line 4 of Table 7 shows the configuration that only keeps lexical phrase rules (LPR) and target discontinuous rules (TDR) leads to degradation in translation quality.

Although Table 8 tells us that the pattern-based rule filter decreases the rule set without harming the translation quality, the reduced size of rule set is not as big as that done by our strategy.

Table 6. The useless patterns in [5]

a	$\langle uX, u'X \rangle$, $\langle Xu, Xu' \rangle$
b	$\langle X_1 u X_2, * \rangle$
c	$\langle X_1 u X_2 v, X_1 u' X_2 v' \rangle$ $\langle u X_1 v X_2, u' X_1 v' X_2 \rangle$
d	$\langle u X_1 v X_2 w, * \rangle$

Table 7. Effect of role-based rule filters and number of rules (in millions) for Spanish-to-English translation

role	tuning		test-1	
	rules	BLEU	rules	BLEU
+TDR	1.27	23.81	1.28	23.62
-RR1NT	**0.70**	**23.70**	**0.71**	**23.53**
-RR1NT -RR	**0.66**	**23.86**	**0.67**	**23.61**
-RR1NT -RR2NT -RR	0.40	23.35	0.41	23.29
ALL	4.13	23.59	4.23	23.46

Table 8. Impact of pattern-based rule filters and number of rules for Spanish-to-English translation

pattern	tune		test-1	
	rules	BLEU	rules	BLEU
-(a~d)	1.15	23.84	1.31	23.60

Chinese-to-English: As we have no idea which kind rules could be excluded in Chinese-to-English translation (see Table 2), we just apply the LOO strategy to filter rules from the start. Similarly, we also apply the pattern-based rule filtering strategy for comparison. The statistics are reported in Table 9-10. Different from Spanish-to-English translation, the rules of most roles are indispensable. Even for the composed rules with one non-terminal (CR1NT), [5] argued these rules are redundant for translation from Arabic to English and it is also verified by us in Spanish-to-English translation. However, the rules of this role are much useful in Chinese-to-English translation. Fortunately, we can still discard a lot of useless rules (CR2NT) and reduce the rule set by 20%.

Using the pattern-based strategy, the rules belong to pattern (b) and (c) can be thrown away. But, less than 20% rules can be removed with safe. We find that the reason why the rules of pattern (d) cannot be taken away is that about half of the discarded rules are strict both-side discontinuous rules that we have proved to be useful.

It is worth noting that our remaining rules after filtering are possible to be further filtered by other strategies (such as count-based). To sum up, our role-based strategy can give a preview of the contribution of the rules belonging to different roles. Of course, our strategy is effective since a large number of rules can be discarded if the translation quality is not degraded without these rules.

Another important thing we should notice is that, compared to 66 possible patterns in pattern-based rule filtering strategy, our strategy only classifies the rules into 7 roles and it is much simpler for experimentation. Furthermore, this classification method is also valid for other translation models, such as string-to-tree model[3] in which rules are made up of minimal rules (like the first 5 roles we discussed in this paper) and composed rules (similar to our CR1NT and CR2NT). Therefore, our rule filtering strategy is more robust.

Through using our role-based strategy in different language pairs, we observe that translation between Chinese and English requires complicated translation rules, but simpler rules are demanded in Spanish-English translation. Therefore, we conclude that the larger divergent between two languages, i.e. Chinese and English, the more complicated rules needed for translation between them.

Table 9. Effect of role-based rule filters for Chinese-to-English, rules in millions

role	tune		test-1	
	rules	BLEU	rules	BLEU
ALL	0.6	29.69	0.69	**28.87**
-CR1NT	0.37	29.20	0.43	28.27
-CR2NT	0.48	**29.85**	0.56	28.86
-RR	0.58	29.24	0.67	28.34
-SDR	0.58	29.35	0.67	28.50
-TDR	0.59	29.64	0.68	28.72
-SBDR	0.58	29.40	0.69	28.58

Table 10. Impact of pattern-based rule filters for Chinese-to-English translation

pattern	tune		test-1	
	rules	BLEU	rules	BLEU
a	0.46	29.22	0.54	28.37
b	0.56	**29.84**	0.65	**28.88**
c	0.53	29.76	0.62	28.84
d	0.59	29.49	0.68	28.69

6 Conclusions

In this paper, we present an empirical study on the contribution of different kinds of discontinuities in bidirectional translations in various language pairs. First, a method of discontinuity classification is proposed. Then, we analyze the symmetry of linguistic phenomena and conduct large scale experiments in different language pairs to answer the question why source discontinuity performs much better than target discontinuity in Chinese-to-English translation. Our interesting finding is that once the source discontinuity is more effective in one translation direction, the target discontinuity usually be more effective in the inverse direction. This gives us clues to determine source discontinuity or target discontinuity should be specially dealt with in certain translation direction for specific language pair. Furthermore, we have shown that discontinuity translation in some main European language pairs is not as useful as in Chinese-English language pair.

Having known the contribution of discontinuity translation, we finally propose a role-based rule filtering approach, which shows efficiency and more robust compared with the pattern-based strategy. Through our rule filtering strategy, we have found that the more divergent between two languages, the more complicated rules needed for translation between them.

Although the experiments are conducted using a hierarchical phrase-based translation model, we believe our findings are also valid in other translation models.

Acknowledgements. We would like to thank the anonymous reviewers for their great comments. The research work has been partially funded by the Natural Science Foundation of China under Grant No. 60975053 and 61003160 and also supported by Hi-Tech Research and Development Program ("863" Program) of China under Grant No. 2011AA01A207.

References

1. Bod, R.: Unsupervised syntax-based machine translation: The contribution of discontiguous phrases. In: Proc. of MT Summit 2007, pp. 51–57 (2007)
2. Chiang, D.: Hierarchical phrase-based translation. Computational Linguistics 33(2), 201–228 (2007)
3. Gally, M., Graehl, J., Knight, K., Marcu, D., DeNeefe, S., Wang, W., Thayer, I.: Scalable inference and training of context-rich syntactic translation models. In: Proc. of ACL-COLING 2006 (2006)
4. Galley, M., Manning, C.D.: Accurate Non-Hierarchical Phrase-Based Translation. In: Proc. of NAACL 2010 (2010)
5. Iglesias, G., Gispert, A., Banga, E.R., Byrne, W.: Rule filtering by pattern for efficient hierarchical translation. In: Proc. of EACL 2009, pp. 380–388 (2009)
6. Li, Z., Callison-Burch, C., Dyer, C., Ganitkevitch, J., Khudanpur, S., Schwartz, L., Thornton, W.N.G., Weese, J., Zaidan, O.F.: Joshua: An open source toolkit for parsing-based machine translation. In: Proc. of ACL 2009, pp. 135–139 (2009)

7. Lopez, A.: Machine translation by pattern matching. Ph.D. Thesis, University of Maryland at College Park (2007)
8. Quirk, C., Menezes, A.: Do we need phrases? Challenging the conventional wisdom in statistical machine translation. In: Proc. of HLT-NAACL 2006, pp. 9–16 (2006)
9. Shen, L., Xu, J., Weischedel, R.: A new string-to-dependency machine translation algorithm with a target dependency language model. In: Proc. of ACL 2008: HLT, pp. 577–585 (2008)
10. Simard, M., Cancedda, N., Cavestro, B., Dymetman, M., Gaussier, E., Goutte, C., Yamada, K., Langlais, P., Mauser, A.: Translating with non-contiguous phrases. In: Proc. of HLT-EMNLP 2005 (2005)
11. Sun, J., Zhang, M., Tan, C.L.: A non-contiguous tree sequence alignment-based model for statistical machine translation. In: Proc. of ACL-IJCNLP 2009, pp. 914–922 (2009)
12. Zollmann, A., Venugopal, A., Och, F., Ponte, J.: A systematic comparison of phrase-based, hierarchical and syntax-augmented statistical MT. In: Proc. of COLING 2008, pp. 1145–1152 (2008)