A Novel Reordering Model for Statistical Machine Translation

Saeed Farzi,[1] Heshaam Faili,[1] Shahram Khadivi,[2] Jalal Maleki[3]

[1] NLP research Lab,ECE Department University of Tehran, Tehran, Iran
[2] CE/IT Dept, Amirkabir University of Technology, Tehran, Iran
[3] Dept. of Computer and Information Science Linköping University SE-581 83 Linköping, Sweden

{saeedfarzi,hfaili}@ut.ac.ir, jalal.maleki@liu.se, khadivi@aut.ac.ir

Abstract. Word reordering is one of the fundamental problems of machine translation, and an important factor of its quality and efficiency. In this paper, we introduce a novel reordering model based on an innovative structure, named, phrasal dependency tree including syntactical and statistical information in context of a log-linear model. The phrasal dependency tree is a new modern syntactic structure based on dependency relations between contiguous non-syntactic phrases. In comparison with well-known and popular reordering models such as the distortion, lexicalized and hierarchical models, the experimental study demonstrates the superiority of our model regarding to the different evaluation measures. We evaluated the proposed model on a Persian→English SMT system. On average our model retrieved a significant impact on precision with comparable recall value respect to the lexicalized and distortion models, and is found to be effective for medium and long-distance reordering.

Keywords: Reordering, phrase-based SMT, syntactical reordering model, long distance reordering.

1 Introduction

The machine translation task is made of two sub-tasks: collecting the list of words in a translation, which is called the lexical choice, and determining the order of the translated words, which is called reordering [3]. In comparison with the word-based systems, the phrase-based systems can readily address local reorderings whilst the reordering is still a computationally expensive problem at the phrase level. The inability of handling the long-distance reordering problems is known as a pitfall of the Phrase-based SMT, which generally two well-known mechanisms have been introduced so far for it [13, 19]. (1) The distortion penalties, and (2) the lexicalized reordering models. The lexicalized reordering models demonstrate superiority regarding to the distortion models in term of handling the long-distance reorderings because of using phrase content information. The distortion penalty not only forces translation systems not to prefer long-distance reorderings, but also has not considered phrase content information. Thus, it is difficult to obtain a satisfactory translation performance. The lexicalized reordering models suffer from data sparseness problem as well as they are restricted to reorder adjacent phrases, phrases with no gap, whereas the long-distance reorderings especially for syntactically divergent language pairs require much more robust solution in order to predicate the orientations of non-adjacent phrases.

In present research, a new way of integrating the phrase-based and syntactically-informed models is proposed as the form of a new model that supplements the Phrase-based SMT [13, 19]. The crystal clear suggestion is to exploit the syntactically-informed reordering elements (reordering rules) based on novel dependency structure, named, the phrasal dependency tree solely for dealing with the medium- and long-distance reorderings. The phrasal dependency tree is a sort of modern syntactic structure based on dependency relations between contiguous non-syntactic phrases. In addition, rather than standard dependency trees in which words are vertices, our trees have phrases as vertices. In order to handle the short-distance reordering problem, we leverage the achievement of the phrase-based approaches providing a series of target words appropriately ordered as a phrase. In general, the lexicalized reordering model not only learns just the orientation of adjacent phrases but also suffers from the data sparseness problem whereas the proposed model tries to overcome these problems.

Two groups of evaluations have been performed on the proposed reordering model as follows. 1) We follow several sorts of scenarios with different goals to verify the performance and make the experimental work stronger. Two Persian→English translation tasks with different sizes have been employed to imply the accuracy and efficiency of our model. The performance of our model also has been compared with the distortion, lexicalized reordering and hierarchical-based models in term of BLEU [20], TER [24] and LRscore [2] measures. The results illustrate the superiority of our approach. 2) The ability of the proposed model to predicate the medium- and long-distance reorderings has been evaluated in more details. On average our model retrieved a significant impact on precision with comparable recall value respect to the lexicalized and distortion models.

The paper is organized as follows. Related works are reviewed in Section 2. In Section 3 and Section 4, the phrasal dependency tree and the phrase reordering model are explained in more details, respectively. In Section 5, the experimental studies are presented. Finally, we draw some conclusions in Section 6.

2 Related Works

In order to vanquish the long-distance reordering problem; some simple models have been introduced. First one is the distortion model [13, 19], which penalizes translations respect to their jumping distance. Second one is the flat reordering model [14, 29, 31], which is not content dependent either. Last one is the lexicalized reorderings model introduced by several researchers [11, 14, 19, 28]. It is a content dependent approach unlike two previous models. The local orientations of each bilingual phrase pair is learned by the lexicalized reordering model. Performance gains have been observed for systems using the lexicalized reordering model. A hierarchical orientation model, which deals with some global phrase reorderings by a shift reduce algorithm has been proposed by [7, 8]. Due to the heavy use of lexical elements, the last two models tend to suffer from data sparseness problems. Another restriction is that the lexicalized models are limited to reordering phrases with no gaps (adjacent phrases). In comparison to [7, 8], our model uses a systematic approach to fight with the data sparseness problems. Utilizing head words instead of phrases in the phrasal dependency relations reduces side effects of using lexical information. This method benefits from its simplicity, but it suffers from purveying at most a one best guess at syntactic movement. Search-space constraints restrict the decoding search space using syntactic intuitions [1].

There have been many attempts to employ dependency parse in SMT. Quirk et al. [21] integrated a source-side dependency parse with word alignments in order to model dependency relations on biphrases. In contrast to [21], our model employs target-side dependency parser. Shen et al. [23] and Gao et al. [9] introduced an extended version of Hiero [5] in which a dependency parse has been employed in order to inject dependency relations into the non-contiguous phrases. In contrast to the model of [9, 22, 23], our model works on non-syntactic contiguous phrases with left-to-right decoder whereas Shen et al. [23] have to design a string-to-tree machine translation system. Galley and Manning [7] relaxed standard assumptions about dependency parsing because the efficient left-to-right decoding algorithm of phrase based translation could be retained while a dependency parse is included.

3 Phrasal Dependency Tree

Dependency grammar (DG) is a sort of syntactic theories introduced by Lucien Tesniere [27], which is based on a dependency relation between a governor (a word) and its dependents. Because of benefits of DG such as extracting long-distance relations. Wu et al. [30] endeavored to expand the dependency tree node with syntactic phrases. Term "Phrase" usually is used as a syntactic unit in natural language processing tasks. However, a contiguous non-syntactic phrase, which consists of some contiguous words without any syntactic constraints, is another phenomenon which plays important role in NLP applications such as Phrase-based SMT.

3.1 Phrasal Dependency Tree

In order to construct the phrasal dependency tree, we introduce an algorithm which utilizes a lexical word level dependency parser and a segmentation of the sentence. The segmentation provides the non-syntactic contiguous phrases which cover all words of input sentence. A phrasal dependency tree is defined as follows.

Definition 3.1. Let R= $\{r_1,r_2,\ldots, r_m\}$ be a limited set of possible dependency relation type that could hold between any two phrases of a sentence. A phrase dependency graph $G = (V, E)$ is a directed graph consists of nodes, V, and arcs, E, such that for sentence $S=p_0p_1p_2\ldots.p_n$ (p_i is a phrase or segment and p_0 is a dummy phrase as a root). V consists of $p_0p_1p_2\ldots.p_n$ and E is set of triple $<p_i,r,p_j>$. There is no edge between two phrases with the same relation type.

Definition 3.2. *A phrasal dependency tree* T=(V, E) for an input sentence S and dependency relation set R is a spanning tree rooted by node p_0 which is derived from a dependency graph.

The main idea is to replace a word by a contiguous non-syntactic phrase in a sentence. Thus a dependency relation holds between two phrases. One phrase is a governor and other is a dependent. In order to make relationship between phrases regarding to the word dependency relations, we must distinguish a word as head of the phrase. The dependency relations of two head words of phrases play important role in making relations of phrases.

Definition 3.3. The head word of the phrase P_i is the closest word to the root of the word level dependency tree T in comparison with other P_i words. On the other words, the shallowest word of phrase P_i in the word level dependency tree T.

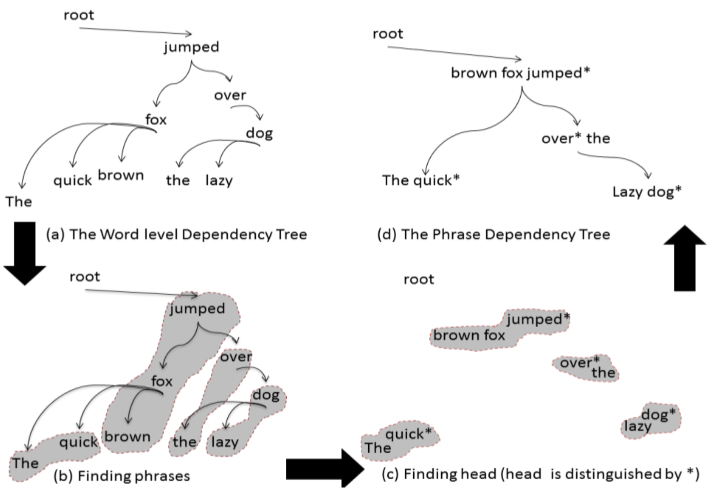

Fig. 1. The procedure of the phrase dependency parsing

The phrasal dependency parsing is conducted in this way: first find a head of each phrase and next travers word level dependency tree in preorder fashion. By visiting a non-head node of a phrase, compact it with the head node of the phrase and remove all its dependency relations. At the end, connect each head node to its nearest ancestor. Consider the following example:

S: "The quick brown fox jumped over the lazy dog"
P[1]: [The quick][brown fox jumped] [over the] [lazy dog]

Fig.1 illustrates the algorithm for the example. Fig.1 (a) demonstrates the word level dependency tree of the sentence S. Fig.1 (b) shows the phrases on the tree. At

[1] P shows a segmentation of S which distinguishes all phrases

the end, each phrase is linked to make a tree. The phrasal dependency tree of sentence S according to the segmentation P is shown in Fig.1 (d).

Note that the output of the algorithm is still a tree because we solely cut edges between neighbor words and generate new edge between a head and its nearest ancestor as well. Additionally, the algorithm guarantees that the output graph's connectivity is maintained and that the graph contains no cycles.

4 Phrasal Reordering Model

As mentioned before, one of the complicated problems in Phrase-based SMT is phrase reordering. We introduce a novel contiguous phrasal reordering model by integrating the phrase dependencies into Phrase-based SMT. The phrase movements are predicated by phrase dependency relations learned from a phrasal dependency corpus. The model depends on calculating the probabilities of the reordering elements, which are estimated via the maximum likelihood estimation from frequencies in a sufficiently-large set of phrasal dependency trees.

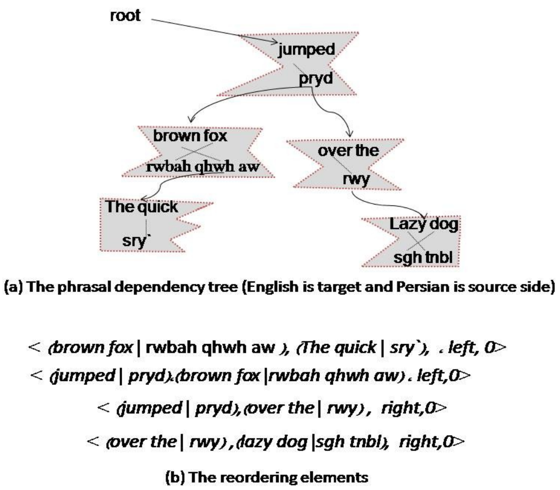

Fig. 2. The phrasal dependency tree and extracted reordering elements (rules)

The reordering element is a branch of the phrasal dependency tree, which depicts the dependency relation between one phrase as a governor (parent) and another phrase as a dependent (child). Fig.2 (a) shows a phrasal dependency tree for two word aligned sentences. The tree has been derived from target order of sentences (English side), and the nodes are constructed by the source and target words plus its word alignments. According to the phrasal dependency tree topology, the dependent node can be settled in the right or left side of the governor node. This direction helps to determine the translation orientation of phrases in decoding phase. Hence, the reordering element has been equipped by the direction of dependent node respect to its governor. All information provided by the reordering elements helps the decoder to score the translation hypothesis more precisely.

The reordering element is shown by:

<governor(target|source), dependent(target|source),direction >

Fig. 2(b) shows the reordering elements, which their probabilities should be computed from a phrasal dependency trees corpus as training data by Eq. 2.

4.1 Training Phase

Given that a reordering element consists of 3elements <g, d, dir >, a total probabilistic model p(<g,d,dir>) is split into 3 partial models as follows.

$$P_{total}(<g,d,dir>) = p(g) * p(d \mid g) * p(dir \mid g,d) \tag{1}$$

P_{total} *(...)* is the probability of the reordering elements called total model. *P(...)* is the probability of features called partial model. All partial models are learned by the maximum likelihood estimation method and smoothed by the modified Kneser-Ney.

4.2 Decoding Phase

During the decoding phase of the left-to-right decoder, the source sentence is segmented into a series of phrases as in a standard phrase-based model. All standard Phrase-based SMT models with the proposed reordering model are incorporated into a log linear fashion to score the partial translations (hypotheses). In order to score the hypothesis, we use Eq.2 to calculate the reordering probability of H.

$$score(H) = -\log\Big(\prod_{p_i, p_j \in phrase(H)} P_{dep}(p_i, p_j)\Big) \tag{2}$$

where *H* is a hypothesis and p_i and p_j are non-syntactic contiguous phrases of *H*. p_i is a governor and p_j is a dependent. *Phrase (H)* returns a list of all phrases of *H*. P_{dep} *(...)* is the probability of the dependency reordering elements calculated by Eq. 3.

$$P_{dep}(p_i, p_j) = P_{total}(<p_i, p_j, dir>) \tag{3}$$

where *dir* is a direction of pi respect to p_j, respectively.

4.3 The data sparseness problem

In order to dominate the data sparseness problem, the head word of a phrase can be utilized by reordering elements instead of incorporating all words of a phrase. Hence the reordering element changes as follows.

<Head (governor), Head (dependent), dir >

Head (governor) and Head (dependent) are head words of governor and dependent phrases, respectively. The big challenge is finding the head words of the governor and dependent during the decoding phase. Due to the lack of word level dependency tree during the decoding phase, determining the head is not straightforward. Here we propose a heuristic approaches to detect head word of phrases during decoding phase. To recognize the head of phrase, POS tags of word has been employed. According to the definition 3.3, the head word is shallowest word respect to the word level dependency tree. In other words, the head plays more luminous syntactically roles than other words of the phrase. For example for given phrase "brown fox jumped" with POS tag sequence "adj[2] noun verb", "verb" is head because of the verb plays more important role than other words.

5 Experimental studies

In order to compare the performance of our reordering model with various reordering models such as the distortion, lexicalized and syntax-based reordering models, some experiments have been carried out by training a Persian→English SMT system. One of the important reasons for choosing Persian and English language pair is lots of differences between English and Persian sentences in the word order.

Persian sentences use SOV (Subject (S), Object (O) and Verb (V)) word order whereas English sentences use SVO structure. Also in Persian language the modifier appears before the modified word whereas English is vice versa. Two validation scenarios have been designed in order to validate the proposed model.

Scenario 1: the aim of the scenario is to validate our model on the small-scale translation task. We intend to understand the impact of our model on translation quality when using the low resource language pairs.

Scenario 2: the performance of our model has been evaluated on large scale training datasets. We intend to show the ability of the model when using several the large-scale translation tasks.

5.1 Data

Table.1 reports translation tasks characteristics. TPC3 [15] with about 400k parallel sentences from novel books has been employed by tsFaEn4-small.

tsFaEn-large utilizes a parallel corpus including about 1 million Persian-English sentences extracted from novel books. PCTS [15] is employed as development and test sets. TMC which is also free Persian monolingual corpus is used to build the target 3-gram language model using the SRILM toolkit with modified Kneser-Ney smoothing [26].

[2] adjective
[3] Tehran Parallel Corpus
[4] Translation task of Persian→English

Table 1. Basic statistics about the novel domain translation tasks (Persian→English)

			Sentence	Token	Unique Token	ASL
tsFaEn-small	Train	source	399K	6M	80K	15.1
		target	399K	6.5M	65K	16.2
	Dev	source	100	1.3K	0.68K	12.4
		target	100	1.4K	0.6K	13.8
	Test	source	300	3.4K	1.4K	11.6
		target	300	3.6K	1.2K	12.1
tsFaEn-large	Train	source	1M	15.6M	250K	15.6
		target	1M	15.7M	210K	15.7
	Dev	source	200	2.1K	1K	10.8
		target	200	2.2K	1K	11.3
	Test	source	200	2.6K	1.1K	13
		target	200	2.2K	0.97K	11.3

5.2 Baseline System Setup

Moses has been employed as a baseline Phrase-based SMT [10, 12] and Hierarchical Phrase-based SMT [4]. Phrase-based SMT utilizes multiple stacks to generate translation hypothesis and SRILM toolkit [25] with interpolated modified Kneyser-Ney smoothing to compute 3-gram language model.

The parameters used for the experiments are: stack size of 100 and the number of target phrases limit of 20. Alignments have been extracted by utilizing the GIZA++ toolkit in words level [17, 18]. Distortion limit equals -1 for the SMT systems equipped by the proposed reordering model.

The hierarchical Phrase-based SMT system utilizes the standard default Moses configuration and relative_thereshold[5]=10 and max_n_item[6]=30.

In order to evaluate the translations, BLEU [20],TER [24] and LRscore [2] measures are used. All model weights have been tuned on development sets via minimum-error rate training (MERT) [16]. The word level dependency tree is generated by the Stanford dependency parser [6].

5.3 Experiments on Small-scale Training Data

In order to illustrate the performance of the reordering models in term of BLEU, TER and LRScore, 4 translation systems with different reordering models have been built on the same conditions. The results on the small-scale translation tasks, tsFaEn-small, have been reported in Table.2. Phrase-based SMT with the distortion, lexicalized and proposed reordering model are denoted by ***pbSMT+d, pbSMT+l*** and ***pbSMT+p***, respectively. ***hpbSMT*** also points to hierarchical Phrase-based SMT.

5 *Relative_threshold* prunes items in a cell which is worse than the best item in that cell

6 The maximum number of items which a cell can maintain

Table 2. BLEU, TER and LRscore scores of all systems with different reordering models

		Metrics			
		DEV	TEST		
		BLEU	BLEU	TER	LRSCORE
tsFaEn-small	*pbSMT+d*	26.71	22.96	62.01	0.32
	pbSMT+l	29.10	25.69	59.75	0.34
	hpbSMT	29.19	26.03	59.23	---
	pbSMT+p	**30.31**	**27.01**	**58.39**	**0.36**

As shown in Table 2, the best BLEU/TER/LRscore scores of translation systems (tsFaEn-small) are about 27/58.3/0.36. The proposed model achieves +1.32/+4.05/+0.98, -1.4/-3.69/-0.84 and +0.02/+0.04/0 point improvements in BLEU, TER and LRscore compared with the lexicalized /distortion/hierarchical models, respectively. Therefore, we can observe that adding our reordering model to Phrase-based SMT brings an illustrious improvement on the small-scale translation tasks with different domains and average sentence length.

In order to investigate the reordering predictive capabilities of models, the minimum number of shifts needed to change a system output so that it exactly matches a given references have been computed.

The shift moves a sequence of words within the translation (for more details [24]), and also shift distance indicates to the number of reordering required to move a word to its right place respect to the given references. Fig. 3 presents the amount of the shifts needed by ***pbSMT+d***, ***pbSMT+l*** and ***pbSMT+p*** on PCTS test set.

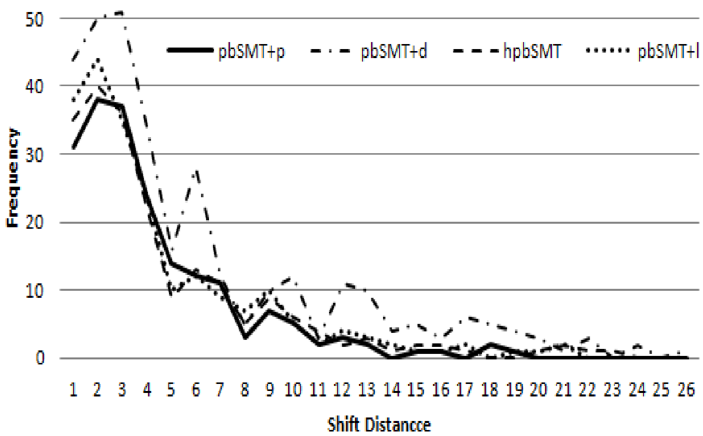

Fig. 3. The number of the reorderings needed by pbSMT+d, pbSMT+l, pbSMT+p to match with the targeted references on PCTS test set

Fig.3 indicates that our model predicates a lot more reordering needed particularly medium- and long- distance reorderings to the translation than the other reordering models. For more analysis, we calculate precision and recall of reorderings [9]. Table

3 reports the total precision and recall which are computed test set and aligned manually.

Table 3. Total Precision and recall

Translation System	Total Precision	Total Recall
pbSMT+d	0.29	0.31
***pbSMT*+l**	0.31	**0.32**
***pbSMT*+p**	**0.33**	**0.32**

From Table 3, we can observe that our model improves precision about +0.02 and +0.04 absolute points respect to the lexicalized and distortion models, respectively. In order to explore the question which word ranges are affected more by the reordering models, Fig.4 shows precision per the reordering distance, respectively. It is figure out that our model has the most positive impact on precision over the most word ranges. The results demonstrate superiority of our model whereas the lexicalized model overtakes in 6, 7, 10 and 11. The error analysis reveals that in the most cases, the proposed model has been predicated the direction of translated phrases correctly. Nevertheless, because of the existence of untranslated phrases, the length of the translations generated by ***pbSMT+p*** is less than others. Consequently, unexpected results observed in 6, 7, 10 and 11. In general, we can observe the momentous improvement on the short-, medium- and even long- distance reorderings. When recall is concerned, our model achieves a comparable recall value respect to the lexicalized reordering model.

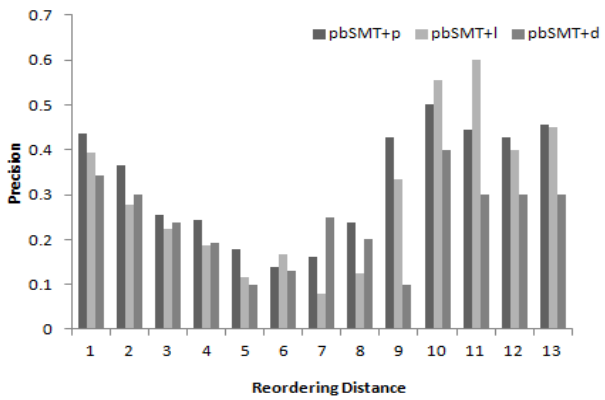

Fig. 4. Precision/reordering distance

5.4 Experiments on Large-scale Training Data

In previous section, we study sparse training data scenarios, in which the reordering and translation models have been learned on two sparse bilingual data sets. In this section we scale the method to a large training set and illustrate that the improvement in terms of translation quality is maintained. Table 4 presents the results of our model in comparison with the other reordering models.

Table 4. BLEU, TER and LRscore scores of all systems with different reordering models

		Metrics			
		DEV	TEST		
		BLEU	BLEU	TER	LRSCORE
tsFaEn-large	*pbSMT+d*	30.11	26.96	59.17	0.35
	pbSMT+l	33.13	29.03	56.57	0.37
	hpbSMT	**33.45**	29.0	56.56	---
	pbSMT+p	33.33	**30.07**	**55.63**	**0.38**

The best BLEU/TER/LRscore scores of translation systems are about 30/55.63/0.38. The proposed model achieves +1.02/+3.11/+1.07 and -0.94/-3.54/-0.94 point improvements in BLEU and TER compared with the lexicalized /distortion/hierarchical models, respectively.

Similar to the small-scale experiments, the reordering predictive capabilities of the models on the large-scale translation tasks have been considered by the number of needed shifts, reordering precision and recall. Fig. 5 depicts the amount of shifts needed by ***pbSMT+d, pbSMT+l*** and ***pbSMT+p***.

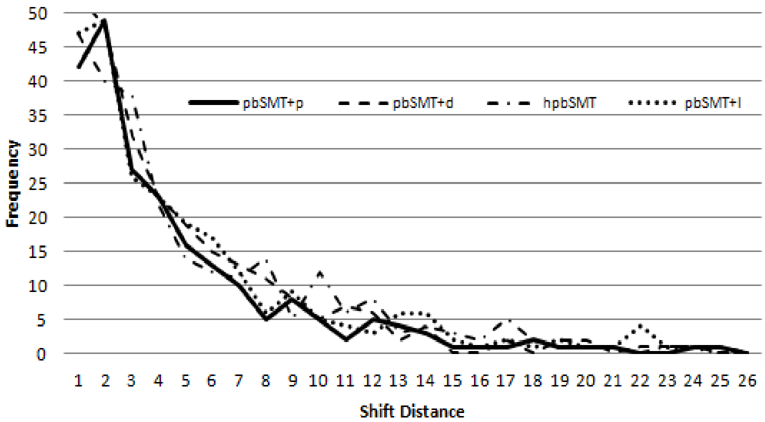

Fig. 5. The number of the reorderings needed by pbSMT+d, pbSMT+l, pbSMT+p

As Fig. 5 shows, pbSMT+p has been more successful than other models in the prediction of the reordering needed. It indicates that our model predicates a lot more reordering needed particularly medium- and long- distance reorderings than the other reordering models. The experiments on the large-scale translation tasks also implies that the proposed model not only obtains better results over the well-known and popular reordering models but also can predicate the medium- and long- distance needed reorderings more than others.

6 Conclusion

In this paper, we introduce a new phrasal reordering model of integrating the phrase dependencies as syntactical structure to the Phrase-base SMT. We exploit the syntactically-informed reordering elements which are included by the translation direction feature in order to deal with the medium- and long- distance reordering problems. The proposed model has been discussed from the theoretical and experimental points of view, and its advantages, disadvantages and constraints in comparison of well-known and popular reordering models have been analyzed. In order to compare the performance of our reordering model with the distortion, lexicalized and hierarchical reordering models, lots of experiments have been carried out by training Persian→English SMT systems. We evaluated the proposed model on two translation tasks in different size. The evaluations illustrate significant improvements in BLEU, TER and LRscore scores comparing to the lexicalized /distortion/hierarchical models. Furthermore, the reordering predictive capabilities of models have been compared by calculating the minimum number of shifts needed to change a system output so that it exactly matches a given references. The results imply that our model predicates a lot more reordering needed particularly medium- and long- distance reorderings than the other reordering models. For a more detailed analysis and answering the question which word ranges are affected more by the reordering models, total precision/recall and precision/recall per distance have been calculated. The proposed model retrieved a significant impact on precision with comparable recall value respect to the lexical reordering model.

References

1. Bach, N., Vogel, S., and Cherry, C.: Cohesive constraints in a beam search phrase-based decoder. In: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers, pp. 1–4 (2009)
2. Birch, A. and Osborne, M.: LRscore for evaluating lexical and reordering quality in MT. In: Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR, pp. 327–332 (2010)
3. Chang, P. C. and Toutanova, K.: A Discriminative SyntacticWord Order Model for Machine Translation. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, pp. 9–16 (2007)
4. Chiang, D.: A hierarchical phrase-based model for statistical machine translation. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pp. 263–270 (2005)
5. Chiang, D.: A hierarchical phrase-based model for statistical machine translation. In: Proc. of ACL (2005)
6. De Marneffe, M.-C., MacCartney, B., Manning, C.D.: Generating typed dependency parses from phrase structure parses. In: Proceedings of LREC, pp. 449–454 (2006)
7. Galley, M., Graehl, J., Knight, J., Marcu, D., DeNeefe, S., Wang, W., and Thayer, I.: Scalable Inference and Training of Context-Rich Syntactic Translation Models. In: Proceedings of the joint conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics, Sydney, Australia (2006)
8. Galley, M. and Manning, C.D.: A Simple and Effective Hierarchical Phrase Reordering Model. In: Proceedings of the EMNLP 2008 (2008)

9. Gao, Y., Koehn, P., and 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 (2011)
10. Koehn, P., Hieu Hoang, Birch, A.: Moses: Open source toolkit for statistical machine translation. In: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (2007)
11. Koehn, P., Axelrod, A., Mayne, A., Callison-Burch, C., Osborne, M., and Talbot, D.: Edinburgh System Description for the 2005 IWSLT Speech Translation Evaluation. In: International Workshop on Spoken Language Translation (2005)
12. Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, R., Bojar, O., Constantin, A., Herbst, A.: Moses: Open source toolkit for statistical machine translation. In: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, 2007.
13. Koehn, P., Och, J., and Marcu, D.: Statistical Phrase-Based Translation. In: Proceedings of HLT/NAACL (2003)
14. Kumar, S. and Byrne, W.: Local phrase reordering models for statistical machine translation. In: Proceedings of HLT-EMNLP (2005)
15. Mansoori, A. and Faili, H.: State-of-the-art English to Persian Statistical Machine Translation System. In: Proceeding of 16th CSI International Symposiums on Artificial Intelligence and Signal Processing (AISP 2012) to appear, Shiraz, Iran (2012)
16. Och, F. J.: Minimum error rate training in statistical machine translation. In Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1, pp. 160–167 (2003)
17. Och, F. J. and Ney, H.: Improved statistical alignment models. In: Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics, pp. 440–447 (2003)
18. Och, F. J., Ney, H.: Improved statistical alignment models, In: Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics, pp. 440–447 (2003)
19. Och, J. and Ney, H.: The alignment template approach to statistical machine translation. Computational Linguistics, vol. 30, pp. 417–449 (2004)
20. Papineni, K., Roukos, S., Ward, T., and Zhu, W.J.: BLEU: a method for automatic evaluation of machine translation. In: Proc. of 40th Annual meeting of the Association for Computational Linguistics, pp. 311–318 (2002)
21. Quirk, C., Menezes, A., and Cherry, C.: Dependency treelet translation: Syntactically informed phrasal SMT. In: proceedings of the 43th Meeting of the Association for Computational Linguistics, pp. 271–279 (2005)
22. Shen, L., Xu, J., and Weischedel, R.: A new string-to-dependency machine translation algorithm with a target dependency language model. In: Proceedings of ACL-08: HLT, pp. 577–585 (2008)
23. Shen, L., Xu, J., and Weischedel, R.: A new string to-dependency machine translation algorithm with a target dependency language model. In: Proc. of ACL (2008)
24. Snover, M., Dorr, B., Schwartz, R., Micciula, L., and Makhoul, J.: A Study of Translation Edit Rate with targeted Human Annotation. In: AMTA 2006, 7th Conference of the Association for Machine Translation in the Americas, Cambridge, pp. 223–231 (2006)
25. Stolcke, A.: SRILM – an extensible language modeling toolkit. In: Proceedings of the international conference on spoken language processing, pp. 901–904 (2002)
26. Stolcke, A.: SRILM – an extensible language modeling toolkit. In: Proceedings of the International Conference on Spoken Language Processing (ICSLP 2002) (2002)
27. Tesniere, L.: El´ements de syntaxe structurale. Editions Klincksieck (1959)
28. Tillmann, C.: A block orientation model for statistical machine translation. In: HLTNAACL, Boston, MA, USA (2004)

29. Wu, D.: Stochastic inversion transduction grammars, with application to segmentation, bracketing, and alignment of parallel corpora. In: Proceeding of IJCAL 1995, Montreal, pp. 1328–1334 (1995)
30. Wu, Y., Zhang, Q., Huang, X., and Wu, L.: Phrase Dependency Parsing for Opinion Mining. In: Proceedings of the 2009 Conference on Empirical Methods in natural Language Processing, pp. 1533–1541 (2009)
31. Zens, R., Ney, H., Watanabe, T., and Sumita, E.: Reordering Constraints for Phrase-Based Statistical Machine Translation. In: Proceedings of CoLing 2004, Geneva, Switzerland, pp. 205–211 (2004)