The Impact of Source–Side Syntactic Reordering on Hierarchical Phrase-based SMT

Jinhua Du, Andy Way
CNGL, School of Computing,
Dublin City University,
Dublin, Ireland
jdu,away@computing.dcu.ie

Abstract

Syntactic reordering has been demonstrated to be helpful and effective for handling different word orders between source and target languages in SMT. However, in terms of hierarchial PB-SMT (HPB), does the syntactic reordering still has a significant impact on its performance? This paper introduces a reordering approach which explores the 的 (DE) grammatical structure in Chinese. We employ the Stanford DE classifier to recognise the DE structures in both training and test sentences of Chinese, and then perform word reordering to make the Chinese sentences better match the word order of English. The annotated and reordered training data and test data are applied to a re-implemented HPB system and the impact of the DE construction is examined. The experiments are conducted on the NIST 2008 evaluation data and experimental results show that the BLEU and METEOR scores are significantly improved by 1.83/8.91 and 1.17/2.73 absolute/relative points respectively.

1 Introduction

Syntactic structure-based reordering has been shown to be significantly helpful for handling word order issues in phrase-based machine translation (PB-SMT) (Xia and McCord, 2004; Collins et al., 2005; Li et al., 2007; Wang et al., 2007; Elming, 2008; Chang et al., 2009). Generally, PB-SMT has an independent reordering model because the phrases themselves in PB-SMT do not have an ability to perform word reordering. However, as regards the hierarchical PB-SMT (HPB) (Chiang, 2005), it possesses an inherent word reordering capability because each rule is a hierarchical structure which contains the sub-phrases. The order of a sub-phrase pair in a source–target hierarchial phrase pair is decided by the positions of non-terminals when generating the hierarchical phrase based on the word alignment links between the source and target language sentence. Generally, there is no specific reordering model inside the HPB system. When the hierarchical phrase is expanded during the decoding process, the word order is carried out automatically according to the positions of the corresponding non-terminal symbols and the synchronous context-free grammars (SCFG) between the source and target phrases.

The phrases in HPB are generally variable-based phrases, i.e., short phrases are substituted in a long phrase with variables or non-terminals based on SCFG rules. The method of constructing variable-based phrases facilitates learning of the different orderings between the source and target language to some extent. Therefore, in some sense, it can be regarded that the hierarchical phrases contain a hidden reordering model. However, this hidden model cannot handle some flexible syntactic structures well such as the 的 (DE) in Chinese. In this paper, we utilise the Stanford DE-annotated approach (Chang et al., 2009) in both training and test sentences of Chinese, and then carry out word reordering on the Chinese side so as to adjust the Chinese sentences to have a closer word order with English. The purpose of this work is to examine how and why the syntactic DE reordering approach affects the HPB system despite the fact that the hierarchical phrases already have a reordering capability. Consequently, the main contributions of this paper are,

- applying the DE syntactic reordering approach into the HPB system;
- exploring how hierarchical phrases perform

word reordering, together with any deficiencies therein;

- examining how the DE structural phrases influence the HPB from the aspects of word alignment, phrase extraction and hierarchical phrase generalisation etc.

We employ the Stanford DE classifier to preprocess both the training and test data by explicitly labeling 的 (DE) constructions, as well as reordering phrases. Then we re-train the word alignment using GIZA++ (Och and Ney, 2003) and build a reordered initial phrase table and a hierarchical phrase table. The experimental results within a re-implemented HPB system show significant improvements on NIST 2008 evaluation data in terms of BLEU (Papineni et al., 2002), METEOR (Banerjee and Lavie, 2005) and TER (Snover et al., 2006) scores.

The remainder of this paper is organised as follows. In section 2, we introduce the phenomenon of word order errors caused by the DE construction. Section 3 studies the reordering mechanism of hierarchical phrases in HPB. In Section 4, we describe 5 different classes of DE construction in Chinese and the algorithm of the Stanford DE classifier. In Section 5, the experiments conducted on NIST 2008 Chinese-to-English evaluation data are reported. Meanwhile, in Section 6, an in-depth analysis on how the syntactic DE reordering affects HPB is carried out. Section 7 concludes and gives avenues for future work.

2 The Problem of Chinese DE Construction Translation

It is well-known that in MT, it is difficult to translate from Chinese to English because of the different word orders (cf. the different orderings of head nouns and relative clauses). (Chang et al., 2009) pointed out that many of these structural differences are related to the ubiquitous Chinese structural particle phrase 的 (DE) construction, used for a wide range of noun modification constructions (both single word and clausal) and other uses. The examples shown in Figure 1 illustrate the errors of three translation results from different MT systems, and many errors relate to incorrect reordering for the 的 (DE) structure.

These three translations are from different Hiero systems. Although HPB has an inherent reordering capability, none of them reordered "bad reputation" and "middle school" around the DE. (Chang et al., 2009) analysed that this is because it is not sufficient to have a formalism which supports phrasal reordering. They claimed it is necessary to have sufficient linguistic modeling, so that the system knows when and how much to rearrange.

Source: 当地(local) 一所(a) 名声不佳(bad reputation) 的(with) 中学(middle school)
Reference: 'a local middle school with a bad reputation'
Team 1: 'a bad reputation of the local secondary school'
Team 2: 'the local a bad reputation secondary school'
Team 3: 'a local stigma secondary schools'

Figure 1: Examples of DE Construction Translation Errors from (Chang et al.,2009)

To solve this problem, (Wang et al., 2007) proposed a syntactic reordering approach to deal with structural differences and to reorder source language sentences to be much closer to the order of target language sentences. They presented a set of syntactic rules to determine whether a 的 (DE) construction should be reordered or not before translation. The deficiency of their algorithm is that they did not fully consider the flexibility of the DE construction, such that it can could be translated in many different ways. (Chang et al., 2009) extended the work of (Wang et al., 2007) and characterised the DE structures into 5 classes based on their behaviour. We give a detailed description of Chang's DE classifier in Section 4.

3 HPB-style Reordering

The hierarchical phrases not only have a powerful generalisation ability, but also a strong reordering capability. The idea of presenting hierarchical phrases is to learn reordering of phrases in the same way that the phrases are good for learning reordering of words, cf. Figure 2 as an illustration (Chiang, 2005). In Figure 2, there are three hierarchical phrase pairs (lexicalized synchronous grammar rule) which are related to syntactic reordering (Chiang, 2005), namely,

- <yu X_1 you X_2, have X_2 with X_1>

  X_1 and X_2 are placeholders for sub-phrases (See Figure 2). This rule shows that the Chinese PPs almost always modify VPs on the left, whereas English PPs usually modify VPs on the right. It can be found that this rule is a phrase-reordering rule which generalises the

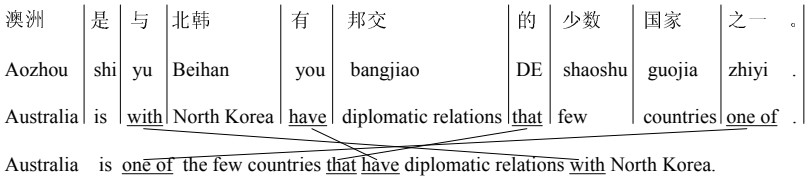

Figure 2: Examples of reordering constructions from (Chiang, 2005)

reordering of verb phrases and prepositional constituents.

- $<X_1$ DE X_2, the X_2 that $X_1>$

  This rule indicates that Chinese relative clauses modify NPs on the left while English relative clauses modify on the right. Therefore, this reordering rule generalises the reordering of relative clauses and the DE structure.

- $<X_1$ zhiyi, one of $X_1>$

  This rule captures the construction *zhiyi* in English word order which is different from Chinese word order.

Based on the reordering rules above, the translation goal for HPB is to rotate the noun head and the preceding relative clause around 的 (DE), so that "[one of few countries] 的 [have diplomatic relations with North Korea]" can be correctly translated. However, although the HPB system has a strong reordering capability in its generalised phrases, it still cannot process some complicated and flexible cases of DE construction like those in Figure 1. Therefore, in the next section, we introduce the Stanford DE classification and use it as a necessary complementary component to improve the HPB system.

4 Stanford DE Classifier

(Chang et al., 2009) argued that one possible reason why the 的(DE) construction remains unsolved is that previous work has paid insufficient attention to the many ways in which the 的 (DE) construction can be translated, as well as the rich structural cues which exist for these translations. In reality, there are many strategies for translating Chinese [A 的 B] phrases into English. The Stanford DE classifier uses a statistical classifier trained on various features to predict for a given Chinese 的 (DE) construction both whether it will reorder in English and which construction it will translate to in English.

4.1 Five classes of DE Construction

For a Chinese noun phrase [A 的 B], it can be categorized into one of the following five classes:

- A B

  In this category, A in the Chinese side is translated as a pre-modifier of B. In most cases A is an adjectival form.

- B preposition A

  There are several cases that are translated into the form B preposition A.

- A's B

  In this class, the English translation is an explicit s-genitive case. This class occurs much less often but is still interesting because of the difference from the of-genitive.

- relative clause

  In this class, the relative clause would be introduced by a relative pronoun or be a reduced relative clause.

- A preposition B

  This class is another small one. The English translations that fall into this class usually have some number, percentage or level word in the Chinese A.

4.2 Log-linear DE Classifier

A log-linear classifier is trained to classify each DE based on features extracted from the parsed data. The features used in the classification model and the accuracies are shown in Table 1.

In Table 1, DEPOS is the part-of-speech tag of DE; A-pattern indicates the Chinese syntactic patterns appearing before 的 (DE); SemClass represents the semantic class of words; and Topicality denotes the re-occurrence of nouns. The 2-class is the categorised classes of DE in (Wang et al., 2007).

	5-class Acc.(%)	2-class Acc. (%)
baseline	-	76.0
DEPOS	54.8	71.0
+A-pattern	67.9	83.7
+POS-ngram	72.1	84.9
+Lexical	74.9	86.5
+SemClass	75.1	86.7
+Topicality	75.4	86.9

Table 1: Features and accuracies of DE classification in Stanford DE Classifier from (Chang et al., 2009)

5 Syntactic Reordering for HPB System

In this section, we firstly perform experiments using the DE annotated and reordered data into the HPB system to verify whether it works or not; secondly we give an in-depth analysis as to how the DE construction affects the HPB system. Meanwhile, we also test the DE classified and reordered data into a phrase-based system to verify whether this approach would have a consistent improvement if it were applied to different types of MT systems.

5.1 Experimental Settings

For our MT experiments, we used Moses (Koehn et al., 2007) as the phrase-based system and employed a re-implementation of HPB (Chiang, 2005) as our hierarchical phrase-based system[1]. The alignment is carried out by GIZA++ (Och and Ney, 2003) and then we symmetrized the word alignment using the grow-diag-final heuristic. Parameter tuning is performed using Minimum Error Rate Training (MERT) (Och, 2003).

The training data contains 2,159,232 sentence pairs, including the HK parallel corpus, ISI parallel data, UN data and other news data. The 5-gram language model is trained on the English part of the parallel training data. The development set (devset) is the NIST MT2006 test set which contains 1,664 sentences. The test set is the NIST MT2008 "current" test set which has 1,357 sentences from two different domains, namely newswire and web-data translation genres. All the dev and test sets have 4 references per source sentence.

To run the DE classifier, we firstly use the Stanford Chinese parser (Levy and Manning, 2003) to parse the Chinese side of the MT training data and the devset and test set.

[1]Our re-implemented HPB doesn't work better than Moses. In future, we will use Moses Chart decoder that is a HPB as well to re-do the experiments.

5.2 Statistics of 5-class DE Annotation

For the DE-annotated MT experiments, after we parse the training data and the devset and the test set, we use the DE classifier to annotate the DE constructions in NPs in all of the parsed data. The 5 classes in section 4.1 are represented by 的$_{AB}$, 的$_{AsB}$, 的$_{BprepA}$, 的$_{relc}$ and 的$_{AprepB}$ to replace the original 的 (DE) character. Once the DE data are labeled, we pre-process the Chinese data by reordering the sentences only with 的$_{BprepA}$ and 的$_{relc}$ annotations. Table 2 lists the statistics of the DE classes in the MT training data, devset and test set.

	training		devset		testset	
DE-class	count	%	count	%	count	%
的$_{AB}$	312,590	23.07	544	26.84	464	29.50
的$_{AprepB}$	6,953	0.51	9	0.44	7	0.44
的$_{AsB}$	13,105	0.97	21	1.04	11	0.70
的$_{BprepA}$	658,692	48.62	974	48.05	663	42.12
的$_{relc}$	316,675	23.37	392	19.34	326	20.71
的$_{non}$	46,752	3.45	87	4.29	103	6.54
Total 的	1,354,767	100	2027	100	1574	100

Table 2: The number of different DE classes labeled for training data, devset and testset

"的$_{non}$" in Table 2 denotes the unlabeled 的 instances which do not belong to any of the 5 classes. We can see that the reordering DE constructions of 的$_{BprepA}$ and 的$_{relc}$ account for 71.99%, 67.39% and 62.83% of the total DE constructions in the training data, devset and test set respectively.

After this preprocessing, we restart the whole MT pipeline: we run GIZA++ (Och and Ney, 2003) to align the reordered data, build phrase tables, tune the MT system and finally evaluate the translation output.

5.3 Experimental Results

The experiments are conducted on a phrase-based SMT system and a re-implemented HPB system for the purpose of verifying whether the DE classification approach has a consistent impact on different types of systems. The experimental results are shown in Table 3.

"+DE Cls." in Table 3 denotes that the data for MT systems are preprocessed using the DE classifier. The baseline systems indicate that the data is neither categorised into DE classes nor reordered on the Chinese side. In the "Imp." row, the number before the slash is the absolute gain in BLEU score while the number after the slash is the absolute im-

Metric	PB-SMT		HPB	
	Baseline	+DE Cls.	Baseline	+DE Cls.
BLEU4	22.42	23.47	20.53	**22.36**
MTR	43.61	44.39	42.91	**44.08**
TER	62.25	61.34	62.81	**61.90**
Imp.	1.05/0.78		**1.83/1.17**	

Table 3: Experimental results on PB-SMT and HPB by applying DE annotation

provement in terms of METEOR score. We can see that the results with DE application achieved consistent and big improvements (the significance test will be done in later version) compared to the Baseline systems for PB-SMT and HPB in terms of BLEU, METEOR and TER scores.

6 Analysis: How DE Annotation Affects HPB

The Stanford DE classifier focuses on the annotation on DEs and on how this can improve translation quality. It was shown in (Chang et al., 2009) that handling the different DE constructions can significantly improve the performance of phrase-based MT. Our goal is to verify whether it is helpful to the HPB system. In the experiments in section 5.3, the results show that HPB with DE annotation gained a significant improvement over the baseline. In this section, we take examples of DE reordering annotations of DE$_{BprepA}$ and DE$_{relc}$ to analyse precisely how DE annotation influences the HPB system.

6.1 Example 1: DE$_{BprepA}$ Annotation

DE$_{BprepA}$ is one of two annotations which would have to be reordered in Chinese sentences. This annotation indicates that the Chinese "A" is translated into a prepositional phrase. Figure 3 is an example of DE$_{BprepA}$ which is a translation from the NIST2008 test set.

In Figure 3, the position of Chinese A "临汾 市 中院 刑 庭" and the position of Chinese B "工作 人员" are swapped in the 4 English references. The Chinese A is translated into a prepositional phrase in English side. In the translation of the HPB baseline system, the DE constructions were not translated well, with the 的 (DE) being translated into "of" independently, as shown in Figure 4. Figure 4(a) illustrates the SCFG derivation in HPB without any DE annotation, while Figure 4(b) describes the SCFG derivation in HPB with both DE annotation and DE reordering.

We can see that in Figure 4(a), because there is not a long SCFG rule to cover the DE construction in this example, the hierarchical phrase does not have such a good generalisation capability, so that the "的" is independently translated as "of", which causes a poor reordering. However, the 的 (DE) in Figure 4(b) is annotated $BprepA$ and the sentence is correctly reordered before decoding by exchanging the positions of contexts A and B of 的 (DE). Therefore, by using the DE-annotated approach, even if the Chinese sentence cannot be fully generalised or covered by a hierarchical rule, it still can be correctly translated because of the reordering pre-processing.

6.2 Example 2: DE$_{relc}$ Annotation

DE$_{relc}$ is another annotation which requires reordering of the Chinese sentences. This annotation indicates that the Chinese "A" is translated into a relative clause. Figure 5 is an example of DE$_{relc}$ which is also a translation from the NIST2008 test set.

In Figure 5, the position of Chinese A "现在 无法 做" and the position of Chinese B "事情" are swapped in the 4 English references. The Chinese A is translated into a relative clause on the English. For the HPB baseline system, the translation is bad because it cannot correctly translate the context A of DE construction, and so the 的 (DE) is translated into "in the" as shown in Figure 6. Figure 6(a) illustrates the SCFG derivation of Example 2 in HPB without DE annotation, while Figure 6(b) describes the SCFG derivation in the HPB with DE annotation and DE reordering.

We can see that in Figure 6(a), although the DE construction is generalised by a SCFG rule, it still cannot capture the reordering between the contexts A and B of DE. However, the 的 (DE) in Figure 6(b) is annotated $relc$ and the sentence is correctly reordered before decoding. Furthermore, in this example the most important point is that there is a DE rule which has two non-terminals to generalise the contexts A and B of the DE constructions.

6.3 The Influence of DE Annotation and Reordering on Word Alignment

Word alignment has a significant impact on the phrase extraction and probability calculation. Consequently, as for the HPB system, the initial phrase table and the word alignment links in the phrase pair will impact on the hierarchical phrase generation. Therefore, in this section, we examine how

Chinese	据 [临汾市 中院 刑 庭]$_A$ 的 [工作 人员]$_B$ 透露
Ref 1	[the staff]$_B$ [at the linfen intermediate criminal court]$_A$ revealed
Ref 2	[a staff member]$_B$ [at the criminal court of linfen intermediate court]$_A$ disclosed
Ref 3	[a staff member]$_B$ [at the criminal court of the linfen intermediate people 's court]$_A$ disclosed
Ref 4	[the staff]$_B$ [of linfen city intermediate people 's criminal court]$_A$ revealed
BAS_HPB	[linfen intermediate court of the criminal court]$_{\mathbf{A}}$ of [staff]$_{\mathbf{B}}$ revealed
DE_HPB	[according to the staff]$_{\mathbf{B}}$ [of the lifen intermediate criminal court]$_{\mathbf{A}}$ revealed

Figure 3: Example of DE$_{BprepA}$ annotation

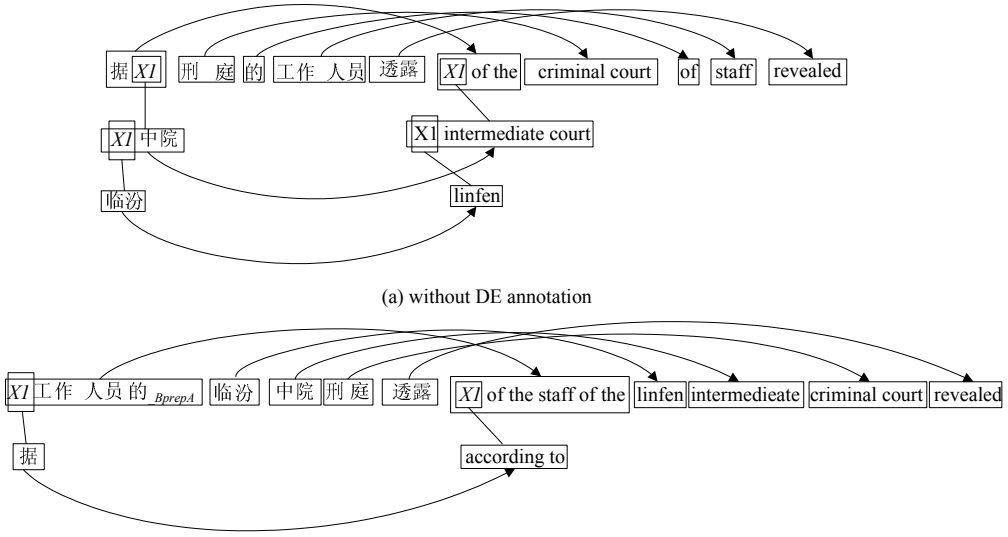

Figure 4: The process of SCFG derivations of the example in Figure 3

Chinese	我 有 太 多 [现在 无法 做]$_A$ 的 [事情]$_B$ 等待 着 我 去 做 。
Ref 1	i have too many [things]$_B$ waiting for me to do [that i can 't do at the moment]$_A$.
Ref 2	i have too many [things]$_B$ [which i cannot do now]$_A$ waiting for me to do .
Ref 3	i have too many [things]$_B$ waiting for me [that i cannot do now]$_A$.
Ref 4	i have too many [things]$_B$ [that i cannot do now]$_A$ waiting for me to do .
BAS_HPB	i have too much to do things in the waiting for me to do .
DE_HPB	i have too many [things]$_{\mathbf{B}}$ [that couldn 't make]$_{\mathbf{A}}$ waiting for me to do .

Figure 5: Example of DE$_{relc}$ annotation

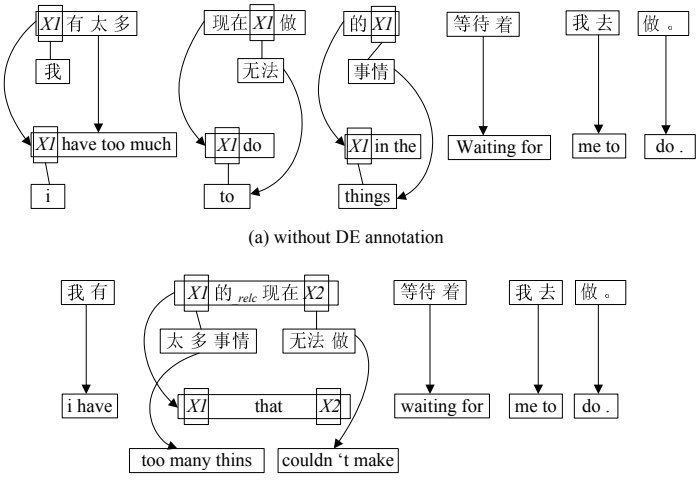

Figure 6: The process of SCFG derivations of the example in Figure 5

the DE annotation and reordering have an influence on the word alignment. See Figure 7(a) and (b) as an illustration.

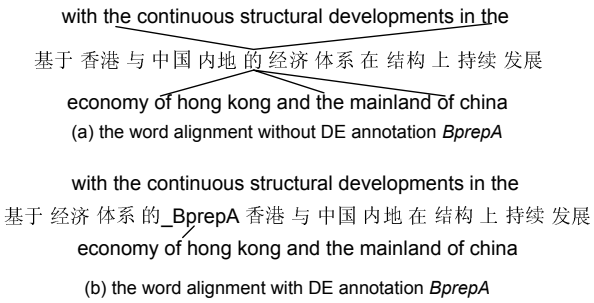

Figure 7: Examples of word alignment between DE-annotated and non-annotated

In Figure 7(b), the DE is annotated with $BprepA$ and the Chinese sentence is reordered. Then all word alignment links of the DE annotated and non-annotated pairs between the Chinese and English sentences in Figure 7 after the grow-diag-final symmetrization are as follows,

(a) 1:1 2:10 2:11 3:12 4:16 5:14 6:2 6:7 6:9 6:13 6:15 7:8 8:8 9:6 10:4 12:3 13:5

(b) 1:1 1:6 2:8 3:8 4:9 5:10 5:11 6:12 7:16 8:13 8:14 9:2 10:4 12:3 13:5

We can see that the 的(DE) in (a) is aligned to 5 words on English side (6:2 6:7 6:9 6:13 6:15), whereas in (b) it is only aligned to 1 word of "of" on English side (6:12). The multiple alignment links in (a) decrease the word alignment accuracy and limit the phrase extraction because of the consistency constrain. Table 8 lists parts of phrases extracted from the (a) and (b) alignment.

without DE	with DE
–	经济 体系 的$_{BprepA}$ 香港 || economy of hong kong
经济 体系 || economy	经济 体系 || economy
香港 || hong kong	香港 || hong kong
–	*X1* 的$_{BprepA}$ *X2* || *X1* of *X2*

Figure 8: Initial phrases and SCFG rules extracted from both "non-annotated DE " and "annotated DE" alignment

In Figure 8, we find that the multiple alignment of "的" in the non-annotated alignment pair cannot extract a phrase with "的". However, in the DE-annotated alignment pair, we can extract the long phrase "经济 体系 的$_{BprepA}$ 香港" which includes the sub-phrases. Therefore, an SCFG rule containing two non-terminals can be generated as shown in Figure 8. In fact, in Figure 7, there are only 16 phrases to be extracted from the non-annotated alignment pair, while there are 34 phrases extracted from the DE-annotated alignment pair. Therefore, DE annotation and reordering have a significant influence on both word alignment and phrase extraction, as well as the hierarchical phrase generalisation, which will indirectly affect the performance of the HPB system.

6.4 Analysis on Altered Sentences by Reordering DE Constructions

Since the DE-annotated approach is to label the Chinese sentences and then reorder the DE constructions with labels $BprepA$ and $relc$, knowing what percentage of sentences are altered and comparing the results of altered sentences will be useful as to how much the DE reordering has an impact on system performance. In our NIST2008 testset, there are 839 out of 1357 sentences (61.8%) that have DEs under NPs, and there are 664 out of 839 sentences (79.1%) that have $BprepA$ or $relc$ labels and are reordered. These show that 1) the preprocessing affects the majority of the sentences; 2) there is a significant reordering in the majority of the affected sentences.

The experimental results of altered sentences (664 reordered sentences) are shown in Table 4.

Metric	HPB	
	Baseline	+DE Cls.
BLEU4	19.77	**22.47**
MTR	42.99	**44.48**
TER	64.44	**62.86**
Improvement	**2.70/1.49**	

Table 4: Experimental results of altered sentences on HPB systems

We can see that for the HPB system, the reordered sentences with DE annotation and reordering obtained improvements of 2.70/13.66 absolute/relative BLEU and 1.49/3.47 absolute/relative METEOR, which demonstrate significantly better than the baseline system. Therefore, as for the altered sentences, the DE-annotated approach is helpful in choosing better English translations.

6.5 Summary of the Analysis

Giving the analysis in the above sections, we summarise the reasons how the DE-annotated approach affects the HPB system, namely,

- When a normal hierarchical phrase has a poor generalisation ability that cannot cover the

contexts of DE constructions, it will cause reordering problems. Thus, the DE reordering approach can reduce the influence of the poor generalisation capability.

- The reordered hierarchical rules have a better generalisation capability.
- The reordered data has an influence on the word alignment and phrase extraction as well as the hierarchical phrase generalisation.

7 Conclusions and Future Work

In this paper, we applied the Stanford DE-annotated approach into the hierarchical phrase-based system. The Stanford DE classifier firstly categorised the DE constructions in Chinese sentences into 5 classes according to its different constituents and then reordered the contexts of the DE constructions with labels of $BprepA$ and $relc$ so as to better match the English word order. Our experiments on the Chinese-to-English NIST2008 test set showed that the DE-annotated approach achieved significant gains over the baseline system in terms of BLEU, METEOR and TER scores. Based on two typical examples, we performed an in-depth analysis of how the DE-annotated approach affects the HPB system and summarised the reasons from three viewpoints.

As for future work, firstly we plan to carry out a larger scale experiments on the HPB system and verify the consistency of the improvements. Secondly, we plan to apply the DE-annotated approach into a syntax-based MT system and examine the effects. We also intend to improve the classification accuracy of the DE classifier to further improve the translation quality.

Acknowledgments

Many thanks to Pi-Chuan Chang for providing the DE classifier and giving valuable instructions. This work is supported by Science Foundation Ireland (Grant No. 07/CE/I1142) as part of the Centre for Next Generation Localisation (www.cngl.ie) at Dublin City University. Thanks also to the reviewers for their insightful comments.

References

Satanjeev Banerjee and Alon Lavie. 2005. METEOR: an automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, ACL-2005*, pages 65–72.

Pi-Chuan Chang, Dan Jurafsky and Christopher D. Manning. 2009 Disambiguating "DE" for Chinese-English machine translation. In *Proceedings of the Fourth Workshop on SMT*, pages 215–223.

David Chiang. 2005. A hierarchical phrase-based model for statistical machine translation. In *Proceedings of ACL'05*, pages 263–270.

Michael Collins, Philipp Koehn, and Ivona Kucerova. 2005. Clause restructuring for statistical machine translation. In *Proceedings of ACL05*, pages 531–540.

Jakob Elming. 2008. Syntactic reordering integrated with phrase-based SMT. In *Proceedings of ACL-08 SSST-2*, pages 46–54.

Philipp Koehn, Hieu Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, Wade Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of ACL'07*, pages 177–180.

Roger Levy and Christopher Manning. 2003. Is it harder to parse Chinese, or the Chinese treebank? In *Proceedings of ACL'03*, pages 439–446.

Chi-Ho Li, Dongdong Zhang, Mu Li, Ming Zhou, Minghui Li and Yi Guan. 2007. A probabilistic approach to syntax-based reordering for statistical machine translation. In *proceedings of the ACL'07*, pages 720–727.

Franz Och. 2003. Minimum Error Rate Training in Statistical Machine Translation. In *Proceedings of ACL'03*, pages 160–167.

Franz Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.

Kishore Papineni, Salim Roukos, Todd Ward and Wei-Jing Zhu. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the ACL-02*, pages 311–318.

Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla and John Makhoul. 2006. A study of translation edit rate with targeted human annotation. In *Proceedings of the AMTA'06*, pages 223–231.

Chao Wang, Michael Collins, and Philipp Koehn. 2007. Chinese syntactic reordering for statistical machine translation. In *Proceedings of EMNLP-CoNLL*, pages 737–745.

Fei Xia and Michael McCord. 2004. Improving a statistical MT system with automatically learned rewrite patterns. In *Proceedings of COLING 2004*, pages 508–514.