Chinese Syntactic Reordering for Statistical Machine Translation

Chao Wang
MIT CSAIL
32 Vassar Street, Room 362
Cambridge, MA 02139, USA
wangc@csail.mit.edu

Michael Collins
MIT CSAIL
32 Vassar Street, Room G-484
Cambridge, MA 02139, USA
mcollins@csail.mit.edu

Philipp Koehn
School of Informatics
2 Buccleuch Place, 5BP 2L2
Edinburgh, EH8 9LW, UK
pkoehn@inf.ed.ac.uk

Abstract

Syntactic reordering approaches are an effective method for handling word-order differences between source and target languages in statistical machine translation (SMT) systems. This paper introduces a reordering approach for translation from Chinese to English. We describe a set of syntactic reordering rules that exploit systematic differences between Chinese and English word order. The resulting system is used as a preprocessor for both training and test sentences, transforming Chinese sentences to be much closer to English in terms of their word order. We evaluated the reordering approach within the MOSES phrase-based SMT system (Koehn et al., 2007). The reordering approach improved the BLEU score for the MOSES system from 28.52 to 30.86 on the NIST 2006 evaluation data. We also conducted a series of experiments to analyze the accuracy and impact of different types of reordering rules.

1 Introduction

Syntactic reordering approaches are an effective method for handling systematic differences in word order between source and target languages within the context of statistical machine translation (SMT) systems (Xia and McCord, 2004; Collins et al., 2005). In reordering approaches, sentences in the source language are first parsed, for example using a Treebank-trained parser. A series of transformations is then applied to the resulting parse tree, with the goal of transforming the source language sentence into a word order that is closer to that of the target language. The reordering process is used to preprocess both the training and test data used within an existing SMT system. Reordering approaches have given significant improvements in performance for translation from French to English (Xia and McCord, 2004) and from German to English (Collins et al., 2005).

This paper describes a syntactic reordering approach for translation from Chinese to English. Figure 1 gives an example illustrating some of the differences in word order between the two languages. The example shows a Chinese sentence whose literal translation in English is:

> *this is French delegation at Winter Olympics on achieve DEC best accomplishment*

and where a natural translation would be

> *this is the best accomplishment that the French delegation achieved at the Winter Olympics*

As exemplified by this sentence, Chinese differs from English in several important respects: for example, relative clauses appear *before* the noun being modified; prepositional phrases often appear *before* the head they modify; and so on. It can be seen that some significant reordering of the input is required to produce a good English translation. For this example, application of reordering rules leads to a new Chinese string whose word-by-word English paraphrase is:

Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 737–745, Prague, June 2007.

Before syntactic reordering

```
IP NP PN 这(this)
   VP VC(is)
      NP CP IP NP NR 法国(French)
                  NN 代表团(delegation)
               VP PP P 在(at)
                     LCP NP NN 冬季
                               (Winter)
                            NR 奥运会
                               (Olympics)
                         LC 上(on)
                  VP-A VV 取得(achieve)
            DEC 的(DEC)
         ADJP JJ 最好(best)
         NPB NN 成绩(accomplishment)
```

After syntactic reordering

```
IP NP PN 这(this)
   VP VC(is)
      NP ADJP JJ 最好(best)
         NPB NN 成绩(accomplishment)
         CP DEC 的(DEC)
            IP NP NR 法国(French)
                  NN 代表团(delegation)
               VP VP-A VV 取得(achieve)
                  PP P 在(at)
                     LCP LC 上(on)
                         NP NN 冬季
                               (Winter)
                            NR 奥运会
                               (Olympics)
```

Figure 1: Original (left) and reordered (right) parse trees for the Chinese sentence "这是法国代表团在冬季奥运会上取得的最好成绩," which translates into "*This is the best accomplishment that the French delegation achieved at the Winter Olympics*" in English.

this is best accomplishment DEC French delegation achieve at on Winter Olympics

This reordering is relatively easy to express using syntactic transformations—for example, it is simple to move the entire relative clause "*French delegation at Winter Olympics on achieve DEC*" to a position that is after the noun phrase it modifies, namely "*best accomplishment*." Phrase-based systems are quite limited in their ability to perform transformations of this type. More recently developed hierarchical systems (e.g., (Yamada and Knight, 2001; Chiang, 2005; Marcu et al., 2006)) may be better equipped to deal with reordering of this type; however, in this example they would effectively have to first identify the span of the relative clause, and then move it into the correct position, without any explicit representation of the source language syntax.

In this paper, we describe a set of syntactic reordering rules that exploit systematic differences between Chinese and English word order. The resulting system is used as a preprocessor for both training and test sentences, transforming Chinese sentences to be much closer to English. We report results for the method on the NIST 2006 evaluation data, using the MOSES phrase-based SMT system (Koehn et al., 2007). The reordering rules give an improvement in accuracy from 28.52 to 30.86 BLEU score. A concern for methods that make use of Chinese parsers is that these parsers are typically of relatively low accuracy, particularly given that Chinese requires a word-segmentation step that is not required in languages such as English. Our results show that Chinese parses are useful in SMT in spite of this problem. We report results showing the precision of the reordering rules—essentially testing how often the Chinese sentences are correctly reordered—to give more insight into this issue. We also report experiments which assess the impact of each type of reordering rule on translation accuracy.

2 Related Work

A number of researchers (Brown et al., 1992; Berger et al., 1996; Niessen and Ney, 2004; Xia and McCord, 2004; Collins et al., 2005) have described approaches that preprocess the source language input in SMT systems. We are not, however, aware of work on this topic for translation from Chinese to English. Brown et al. (1992) describe an analysis component for French which moves phrases around (in addition to other transformations) so the source and target sentences are closer to each other in word order. Berger et al. (1996) describe an approach for French that reorders phrases of the form $NOUN_1$ *de* $NOUN_2$. Xia and McCord (2004) describe an approach for French, where reordering rules that operate on context-free rule productions are acquired automatically. Niessen and Ney (2004) describe an approach for translation from German to English that combines verbs with associated particles, and also reorders questions. Collins et al. (2005) also describe an approach for German, concentrating on reordering German clauses, which have quite different word order from clauses in English. Our approach is most similar to that of Collins et al. (2005).

Most SMT systems employ some mechanism that allows reordering of the source language during translation (i.e., non-monotonic decoding). The MOSES phrase-based system that we use has a relatively simple reordering model which has a fixed penalty for reordering moves in the decoder. More sophisticated models include reordering parameters that are sensitive to lexical information (Tillmann, 2004; Kumar and Byrne, 2005; Koehn et al., 2005). The model of Chiang (2005) employs a synchronous context-free grammar to allow hierarchical approaches to reordering. The syntax-based models of Yamada and Knight (2001) and Marcu et al. (2006) build a full parse tree in the target language, again effectively allowing hierarchical reordering based on synchronous grammars. It is worth noting that none of these approaches to reordering make use of explicit syntactic information in the *source* language—for example, none of the methods make use of an existing source-language parser (the systems of Yamada and Knight (2001) and Marcu et al. (2006) make use of a parser in the target language, i.e., English).

Finally, note that a number of statistical MT systems make use of source language syntax in transducer-style approaches; see (Lin, 2004; Ding and Palmer, 2005; Quirk et al., 2005; Liu et al., 2006; Huang et al., 2006). In contrast to the preprocessing approach, they attempt to incorporate syntax directly into the decoding stage.

3 Chinese Syntactic Reordering Rules

We used the Penn Chinese Treebank guidelines (Xue et al., 2005) in searching for a suitable set of reordering rules. We examined all phrase types in the Treebank; potentially phrases of any type could be candidates for reordering rules. Table 1 provides a list of Treebank phrase tags for easy reference. We ruled out several phrase types as not requiring reordering

`ADJP`	adjective phrase
`ADVP`	adverbial phrase headed by `AD` (adverb)
`CLP`	classifier phrase
`CP`	clause headed by `C` (complementizer)
`DNP`	phrase formed by "`XP+DEG`"
`DP`	determiner phrase
`DVP`	phrase formed by "`XP+DEV`"
`FRAG`	fragment
`IP`	simple clause headed by `I` (`INFL`)
`LCP`	phrase formed by "`XP+LC`"
`LST`	list marker
`NP`	noun phrase
`PP`	preposition phrase
`PRN`	parenthetical
`QP`	quantifier phrase
`UCP`	unidentical coordination phrase
`VP`	verb phrase

Table 1: Penn Chinese Treebank phrase tags.

rules. For example, Chinese `ADJP`s, `ADVP`s, `DP`s, `QP`s, and `PP`s all have similar internal word ordering to their English counterparts. Also similar are a group of special structures such as `LST`, `FRAG`, and `PRN`.

We identified three categories that we considered to be the most prominent candidates for reordering. These phrases include `VP`s (verb phrases), `NP`s (noun phrases), and `LCP`s (localizer phrases, which frequently map to prepositional phrases in English). In the following, we discuss each of the three main categories in more detail.

3.1 Verb Phrases

In Chinese, verb phrase modifiers typically occur in pre-verbal position. `VP` modifiers can be `ADVP`s, temporal and spatial `NP`s, `QP`, `PP`s, `CP`s, `IP`s, `DVP`s, and `LCP`s. The `ADVP`s are simple adverbs, which can occur both preverbal and postverbal in an English verb phrase, so we do not attempt to move them. Similarly, the `CP`, `IP`, and `DVP` modifiers are typically adverbial phrases, which do not have a fixed position in English verb phrases. In the following, we only consider cases involving `PP`s, `LCP`s, temporal and spatial `NP`s, and `QP`s.

PPs and LCPs Figure 2 shows an example verb phrase with a `PP` modifier, which translates literally

```
VP PP P 在(at)
      NP-A NPB NN 东部(Eastern)
               NN 联盟(Division)
   VP-A VV 名列(rank)
        QP OD 第十(10th)
```

Figure 2: Example `VP` with `PP` modifier. The phrase translates into "*ranks 10^{th} in the Eastern Division*."

```
VP NP NPB NT 当天(same day)
          NT 上午(morning)
   VP-A VV 发表(issue)
        NP-A NPB NN 声明(statement)
```

Figure 3: Example `VP` with temporal `NP` modifier. The phrase translates into "*issued a statement that morning*."

into "*at Eastern Division rank 10^{th}*." Recognizing that `PPs` in English verb phrases almost always occur after the verb, we use a simple `VP(PP:VP)` reordering rule which states that a `PP` in a parent `VP` needs to be repositioned after the sibling `VP`. `LCPs` are similar to `PPs` and typically map to prepositional phrases in English. Thus they are handled similarly to `PPs`, i.e., `LCPs` in a parent `VP` are repositioned after the sibling `VP`.

NPs Figure 3 gives an example of a verb phrase with a temporal `NP` modifier, which literally translates into "*same day morning issue statement*." In English, temporal phrases such as these almost always occur after the head verb. Conveniently, the Chinese Treebank uses the part of speech (POS) tag `NT` for temporal nouns. Thus, we use a rule which states that a preverbal `NP` will be repositioned after the sibling `VP` if there is at least one `NT` in the `NP` subtree. A similar rule might apply to locative NPS; however, there is no special POS tag in the Treebank marking locations,[1] so we do not have a syntax-based reordering rule to handle locative `NPs`.

QPs `QP` modifiers in verb phrases often correspond to time-related concepts such as duration and frequency. Figure 4 shows an example verb phrase with a `QP` modifier, literally translating into "*many time injured*." Since temporal phrases almost always occur after the verb in English verb phrases, we handle such cases by a simple rule which states that the `QP` in a parent `VP` will be repositioned after the sibling `VP`.

[1]One can argue that `NR` (proper nouns) in that context are likely to be places. However, there also exist many exceptions, and so we decided not to exploit the `NR` tag.

```
VP QP CD 多(many)
      CLP M 次(time)
   VP-A VV 受伤(injured)
```

Figure 4: Example `VP` with `QP` modifier. The phrase translates into "*injured many times*."

```
NP-A DNP PP P 对(to)
            NP-A NPB NR 津巴布韦(Zimbabwe)
         DEG 的(DEG)
     NPB NN 经济(financial)
         NN 援助(aid)
```

Figure 5: An example Chinese `NP` with a `DNP` modifier headed by a `PP`. The phrase translates into "*the financial aid to Zimbabwe*" in English.

3.2 Noun Phrases

Noun phrases in Chinese can take several types of modifiers: for example, phrases of type `QP`, `DP`, `ADJP`, `NP`, `DNP`, and `CP`. The placement of `QP`, `DP`, and `ADJP` modifiers is somewhat similar to English in that these phrases typically occur before the noun they modify. The case of `NP` modifiers in `NPs` is very limited in the Chinese Treebank, since most noun-noun sequences form compounds in a single `NP`. Hence we only developed reordering rules to handle `DNP` and clausal (`CP`) modifiers.

DNPs `DNPs` are formed by "`XP+DEG`," where `XP` can be a phrase of the type `ADJP`, `QP`, `PP`, `LCP`, or `NP`. When the `XP` is an `ADJP` or a `QP`, no reordering is needed because the word order is the same as that of English.

When the `XP` is a `PP` or an `LCP`, the `DNP` essentially corresponds to a prepositional phrase in English, which almost always appears after the noun being modified. Figure 5 shows an example where the `XP` in the `DNP` is a `PP`. The reordering rule to handle these two cases states that, if a parent `NP` has a child `DNP` which in turn has a child `PP` or `LCP`, then the `DNP` is repositioned after the last sibling `NP`.

Figure 6 shows an example noun phrase for which the `XP` in the `DNP` is `NP`. On the surface, the Chinese "$\mathtt{NP}_1$ `DEG` $\mathtt{NP}_2$" sequence is analogous to the English possessive structure of "$\mathtt{NP}_1$'s $\mathtt{NP}_2$" and does

```
NP-A DNP NP DP DT 该(this)
            CLP M 项(measure word)
            NPB NN 技术(technique)
         DEG 的(DEG)
     NPB NN 掌握(mastery)
```

Figure 6: An example Chinese NP phrase with a DNP modifier headed by a NP. The phrase translates into "*the mastery of this technique*" in English.

```
NP CP IP VP VV 参加 (participate)
            NP CP IP VP PP P 在 (at)
                           NP NR 美国(US)
                              NR 盐湖城
                        (Salt Lake City)
                        VP VV 举行 (hold)
                  DEC 的 (DEC)
               QP OD 第八 (8th)
                  CLP M 届 (measure word)
               NPB NN 残疾人
                     (handicap people)
                  NR 冬奥会
                     (Winter Olympics)
      DEC 的 (DEC)
   NPB NR 法国 (French)
   NPB NN 代表队 (delegation)
```

Figure 7: An example with two nested CP modifiers. The phrase translates into "*the French delegation participating in the 8th Special Winter Olympics held in Salt Lake City US*."

```
LCP IP NP-A NPB NN 事故(accident)
       VP VV 发生(happen)
    LC 后(after)
```

Figure 8: An example Chinese localizer phrase. The phrase translates into "*after the accident happened*" in English.

not require reordering, for example, "苏(*Sue*) 的(*'s*) 朋友(*friend*)" in Chinese and "*Sue's friend*" in English. However, the Chinese possessive structure "NP_1 DEG NP_2" can express more sophisticated relationships which are inappropriate for the "NP_1's NP_2" expression. For example, the phrase in Figure 6 can only be translated into "*the mastery of this technique*," but not "*this technique's mastery*." We decide to reorder DNPs of the "NP+DEG" format, because they often can only map to the "NP_2 of NP_1" expression in English. Additionally, the "NP_2 of NP_1" expression is more general and can replace "NP_1's NP_2" in many cases. One exception is when the NP is a pronoun (PN), e.g., "他(*he*) 的(*'s*) 名字(*name*)," in which case the DNP acts simply like a possessive pronoun. Our reordering rule thus states that, if a parent NP has a child DNP which in turn has a child NP that is not a PN, then the DNP is repositioned after the last sibling NP.

CPs Relative clauses correspond to the CP category in the Treebank. Figure 7 shows an example noun phrase with two nested CP modifiers. As illustrated in the figure, relative clauses in Chinese also occur before the noun they modify, which makes the word order of this sentence quite different from that of the English translation. Such distortions in the word reordering will be quite difficult for the word or phrase-based alignment model to capture. However, with the application of a reordering rule to reposition the child CP after its sibling NP under a parent NP, and the PP VP reordering rule for VP introduced previously, the sentence can be easily transformed into "*French delegation participate 8th handicap people Winter Olympics hold at US Salt Lake City*," a sentence whose word order is much closer to that of English.

CP is typically formed by "IP+DEC", in which DEC's only function is to mark the IP as a relative clause, similar to the function of "that" in English. We use a rule to bring DEC to the front of IP under CP, to make it more aligned with the "that + clause" structure of English.

3.3 Localizers

Figure 8 shows an example phrase of the type LCP. Localizers (tagged LC in the Treebank) in Chinese can be thought of as a post-phrasal preposition which is often used with temporal and locative phrases or clauses to mark directional information. They function similarly to prepositions and conjunctions in English such as "before," "on," "when," etc. Constituents of type LCP have a similar function to prepositional phrases. Sometimes they are combined with a pre-phrasal generic preposition "在" (roughly corresponding to "at" in English) to form a PP explicitly. An example is shown in Figure 9.

We developed a simple reordering rule which moves an LC node to immediately before its left sibling under a parent LCP node. This will result in a word order that is more similar to that of the English

```
PP P 在(at)
   LCP IP NP-A NPB NN 事故(accident)
          VP VV 发生(happen)
       LC 后(after)
```

Figure 9: An example Chinese `PP` encompassing an `LCP`. The phrase translates into "*after the accident happened*" in English.

prepositional phrase: the example in Figure 8 has the paraphrase "*after accident happen*" after the reordering rule is applied. In the case where an `LCP` is embedded in a parent `PP` phrase, the `LC` reordering rule will essentially merge the post-phrasal localizer with the pre-phrasal preposition. For example, the phrase in Figure 9 becomes "*at after accident happen*" after reordering. The phrase-based SMT system will have little problem in learning that "at after" translates into "after" in English.

4 Evaluation

Our baseline is a phrase-based MT system trained using the MOSES toolkit (Koehn et al., 2007). The training data consists of nearly 637K pairs of sentences from various parallel news corpora distributed by the Linguistic Data Consortium (LDC).[2] For tuning and testing, we use the official NIST MT evaluation data for Chinese from 2002 to 2006, which have four human generated English reference translations for each Chinese input. The evaluation data from 2002 to 2005 were split into two sets of roughly equal sizes: a tuning set of 2347 sentences is used for optimizing various parameters using minimum error training (also using the MOSES toolkit), and a development set of 2320 sentences is used for various analysis experiments. We report results on the NIST 2006 evaluation data.

A series of processing steps are needed before the reordering rules can be applied, which include segmentation, part-of-speech tagging, and parsing. We trained a Chinese Treebank-style tokenizer and part-of-speech tagger, both using a tagging model based on a perceptron learning algorithm (Collins, 2002). We used the Chinese parser described by Sun and Jurafsky (2004), which was adapted from the parser presented in Collins (1997). We then applied the reordering rules described in the previous section to the parse tree of each input. The reordered sentence is then re-tokenized to be consistent with the baseline system, which uses a different tokenization scheme that is more friendly to the MT system.[3]

	Dev	Nist06
Baseline	31.57	28.52
Reorder	32.86	30.86
Gain	+1.29	+2.34

Table 2: BLEU score of the baseline and reordered systems.

We use BLEU scores as the performance measure in our evaluation (Papineni et al., 2002). Table 2 gives results for the baseline and reordered systems on both the development and test sets. As shown in the table, the reordering method is able to improve the BLEU scores by 1.29 points on the development set, and by 2.34 on the NIST 2006 set.

4.1 Frequency and Accuracy of Reordering Rules

We collected statistics to evaluate how often and accurately the reordering rules are applied in the data. The accuracy is measured in terms of the percentage of rule applications that correctly reorder sentences. The vast majority of reordering errors are due to parsing mistakes.

Table 3 summarizes the count of each rule in the training data, ignoring rules occurring less than 500 times in the training data, and the number of sentences each rule impacts. The most frequent three rules are `NP(CP:NP)`, `VP(PP:VP)`, and `DNP(NP):NP`, which account for over 76% of all the reordering instances and jointly affect 74% of all the training sentences. This shows the prevalence of systematic word order differences between Chinese and English. Only 122,076 (or 19.2%) sentences remain unchanged after the reordering rules are applied.

Each of the processing steps in producing the Chinese parse tree is prone to error and could lead to mistakes in the reordering of the Chinese sentence.

[2] We used 8 corpora for training, including LDC2002E18, LDC2003E07, LDC2003E14, LDC2005E83, LDC2005T06, LDC2006E26, LDC2006E8, and LDC2006G05.

[3] The tokenizer used by the MT system favors smaller word units, and backs off to a character by character scheme for unknown words.

Type	Rule Name	Counts	# Sent.
VP	VP(PP:VP)	331,827	258,214
	VP(NT:VP)	23,353	22,926
	VP(LCP:VP)	8,674	8,661
	VP(QP:VP)	7,834	7,777
NP	NP(CP:NP)	345,165	262,588
	DNP(NP):NP	280,367	218,865
	DNP(PP):NP	38,225	36,295
	DNP(LCP):NP	15,801	15,253
LC	LCP(NP:LC)	146,784	12,8333
	LCP(IP:LC)	36,923	35,749
	LCP(QP:LC)	14,893	14,287
Total		1,249,846	636,686

Table 3: Statistics of various reordering rules in the training data.

To assess the accuracy of reordering rules, we conducted human evaluations on a set of 200 sentences randomly selected from the development set. Within this set, there were in total 155 sentences containing at least one reordering rule, with 339 rules in total. A bilingual speaker was presented with the Chinese parse tree, the sentence before and after the reordering, and the particular reordering rules applied to the sentence. The bilingual rater determined the correctness of each rule by first identifying the scope of the rule and comparing the string before and after reordering, referencing the corresponding parse structure if necessary. Table 4 summarizes the accuracy (precision) for each type of rule. Notice that our human evaluation of the reordering rules does not take into account missed reordering.

Overall, there are a lot of reordering errors caused by incorrect parses. On a sentence level, only 57 out of the 155 reordered sentences (36.8%) are error free. Nevertheless, syntactic reordering seems to be helpful in improving the translation quality, despite noise introduced into the data due to the errors.

4.2 Impact of Individual Reordering Rules

In order to assess the relative effectiveness of the reordering rules, we conducted an experiment in which we trained and tested systems using data that were reordered using different subsets of the reordering rules. Table 5 summarizes the BLEU scores of the reordered system for each rule type.

	Count	Accuracy
VP rules	108	65.7%
NP rules	209	54.6%
LC rules	76	77.6%
All rules	393	62.1%

Table 4: Accuracy of reordering rules on a set of 200 sentences randomly selected from the development set.

	BLEU	Gain
Baseline	31.57	-
VP rules	32.71	+1.14
NP rules	32.23	+0.66
LC rules	31.59	+0.02
All rules	32.86	+1.29

Table 5: Comparison of translation performance with different types of reordering rules. *Gain* is the change in BLEU score when compared to the baseline system. All results are on the development set.

As shown in the table, the VP rules are more effective than the NP rules, even though the NP rules are more frequent than the VP rules in the data. This is perhaps because the reordering of VP modifiers achieves a slightly higher accuracy than that of the NP modifiers. We are a bit surprised by the lack of performance gains with the LC rules only. More analysis is needed to explain this behavior.

4.3 Better Alignment?

There could be two reasons why the syntactic reordering approach improves over the baseline phrase-based SMT system. One obvious benefit is that the word order of the transformed source sentence is much closer to that of the target sentence, which reduces the reliance on the distortion model to perform reordering during decoding. Another potential benefit is that the alignment between the two sides will be of higher quality because of fewer "distortions" between the source and the target, so that the resulting phrase table of the reordered system would be better. However, a counter argument is that the reordering is very error prone, so that the added noise in the reordered data would actually hurt the alignments and hence the phrase table.

Lacking a good way to measure the quality of

	Original Dev	Reordered Dev
Baseline	31.57	32.19
Reorder	30.67	32.86

Table 6: Comparison of BLEU scores in matched and mismatched conditions. The baseline and reordered systems were first tuned on mismatched data before being tested on mismatched data.

the phrase table directly, we conducted an experiment in which we tested the baseline and reordered systems with both the original and reordered development data. The idea is to compare the two systems given the same type of input: if the reordered system learned a better phrase table, then it might outperform the baseline system on un-reordered inputs despite the mismatch; on the other hand, if the baseline system learned a better phrase table, then it might outperform the reordered system on reordered inputs despite the mismatch. However, the results in Table 6 did not settle our question: the reordered system performed worse than the baseline on unreordered data, while the baseline system performed worse than the reordered system on reordered data, both of which can be explained by the mismatched conditions between training and testing. Perhaps more interesting is the performance gap of the baseline system on the reordered data vs. on the original data: it achieved 0.62 BLEU score gain despite the mismatch in training and testing conditions.

5 Discussion and Future Work

In this paper, we described a set of syntactic reordering rules that exploit systematic differences between Chinese and English word order to transform Chinese sentences to be much closer to English in terms of their word order. We evaluated the reordering approach within the MOSES phrase-based SMT system (Koehn et al., 2007). The reordering approach improved the BLEU score for the MOSES system from 28.52 to 30.86 on the NIST 2006 evaluation data. Our manual evaluation of the reordering accuracy indicated that the reordering approach is helpful at improving the translation quality despite relatively frequent reordering errors. The reordering approach even achieved a 0.62 gain in BLEU score when only the test data are reordered.

An important category we examined but did not reorder was clauses of type `IP`, which generally corresponds to declarative sentences in Chinese. Sentences of this form have quite similar top-level constituent ordering to English: both follow SVO (subject-verb-object) order. There are several special cases in which English and Chinese differ, the most notable being the topicalization of objects or temporal and locative noun phrases (which function as adverbial phrases). We did not try to restore them to the canonical order for several reasons. First, topicalization of temporal and locative phrases happens in English as well. For example, "In Israel yesterday, an explosion killed one person and injured twelve" is a perfectly acceptable English sentence. Second, the parser's performance on special constructions is likely to be poor, resulting in frequent reordering errors. Third, special constructions that do not occur often in the data are less likely to have a significant impact on the translation performance. Thus our strategy has been to find reordering rules for syntactic categories that are common in the data and systematically different between the two languages.

In our experiments, the phrase-based MT system uses an un-lexicalized reordering model, which might make the effects of the syntactic reordering method more pronounced. However, in an early experiment[4] submitted to the official NIST 2006 MT evaluation, the reordered system also improved the BLEU score substantially (by 1.34 on NIST 2006 data) over a phrase-based MT system with lexicalized reordering models (Koehn et al., 2005). The same set of reordering rules in the experimental setting in the current paper achieve a 1.82 BLEU improvement on the same data set, which is comparable to the 1.34 gain for the lexicalized system.

We plan to output reordered lattices in the future, so that the approach would be more robust to errors made during parsing/reordering.

Acknowledgements

We would like to thank Brooke Cowan, Stephanie Seneff, and the three anonymous reviewers for their valuable comments. Thanks to Yushi Xu for evaluating the accuracy of the reordering rules. This work

[4]This experiment made use of a subset of the reordering rules we have presented here.

was supported under the GALE program of the Defense Advanced Research Projects Agency, Contract No. HR0011-06-C-0022.

References

Adam L. Berger, Stephen A. Della Pietra, and Vincent J. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–69.

Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, John D. Lafferty, and Robert L. Mercer. 1992. Analysis, statistical transfer, and synthesis in machine translation. In *Proceedings of Conference on Theoretical and Methodological Issues in Machine Translation*.

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

Michael Collins, Philipp Koehn, and Ivona Kučerová. 2005. Clause restructuring for statistical machine translation. In *Poceedings of ACL*, pages 531–540.

Michael Collins. 1997. Three generative, lexicalized models for statistical parsing. In *Proceedings of ACL*.

Michael Collins. 2002. Discriminative training methods for hidden Markov models: Theory and experiments with perceptron algorithms. In *Proceedings of EMNLP*.

Yuan Ding and Martha Palmer. 2005. Machine translation using probablistic synchronous dependency insertion grammars. In *Proceedings of ACL*, pages 541–548, Ann Arbor, Michigan.

Liang Huang, Kevin Knight, and Aravind Joshi. 2006. Statistical syntax-directed translation with extended domain of locality. In *Proceedings of AMTA*.

Philipp Koehn, Amittai Axelrod, Alexandra Birch Mayne, and Chris Callison-Burch. 2005. Edinburgh system description. In *IWSLT Speech Translation Evaluation*.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constrantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of ACL, Demonstration Session*.

Shankar Kumar and William Byrne. 2005. Local phrase reordering models for statistical machine translation. In *Proceedings of HLT-EMNLP*.

Dekang Lin. 2004. A path-based transfer model for machine translation. In *Proceedings of Coling 2004*, pages 625–630, Geneva, Switzerland, Aug 23–Aug 27. COLING.

Yang Liu, Qun Liu, and Shouxun Lin. 2006. Tree-to-string alignment template for statistical machine translation. In *Proceedings of ACL*, pages 609–616.

Daniel Marcu, Wei Wang, Abdessamad Echihabi, and Kevin Knight. 2006. SPMT: Statistical machine translation with syntactified target language phrases. In *Proceedings of EMNLP*, pages 44–52, Sydney, Australia.

Sonja Niessen and Hermann Ney. 2004. Statistical machine translation with scarce resources using morpho-syntactic information. *Computational Linguistics*, 30(2):181–204.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of ACL*.

Chris Quirk, Arul Menezes, and Colin Cherry. 2005. Dependency treelet translation: Syntactically informed phrasal SMT. In *Proceedings of ACL*, pages 271–279, Ann Arbor, Michigan.

Honglin Sun and Daniel Jurafsky. 2004. Shallow semantic parsing of Chinese. In *Proceedings of NAACL-HLT*.

Christoph Tillmann. 2004. A block orientation model for statistical machine translation. In *Proceedings of HLT-NAACL*, Boston, MA, USA.

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

Nianwen Xue, Fei Xia, Fu-Dong Chiou, and Martha Palmer. 2005. The Penn Chinese Treebank: Phrase structure annotation of a large corpus. *Natural Language Engineering*, 11(2):207–238.

Kenji Yamada and Kevin Knight. 2001. A syntax-based statistical translation model. In *Proceedings of ACL*.