A new subtree-transfer approach to syntax-based reordering for statistical machine translation

Maxim Khalilov and **José A.R. Fonollosa**
Universitat Politècnica de Catalunya
Campus Nord UPC, 08034,
Barcelona, Spain
{khalilov,adrian}@gps.tsc.upc.edu

Mark Dras
Macquarie University
North Ryde NSW 2109,
Sydney, Australia
madras@ics.mq.edu.au

Abstract

In this paper we address the problem of translating between languages with word order disparity. The idea of augmenting statistical machine translation (SMT) by using a syntax-based reordering step prior to translation, proposed in recent years, has been quite successful in improving translation quality. We present a new technique for extracting syntax-based reordering rules, which are derived through a syntactically augmented alignment of source and target texts. The parallel corpus with reordered source side is then passed to an N-gram-based machine translation system and the obtained results are contrasted with a monotone system performance. In experiments, we show significant improvement for the Chinese-to-English translation task.

1 Introduction

One of the most challenging problems facing machine translation (MT) is how to place the translated words in the natural order of the target language. A monotone SMT system suffers from weakness in the distortion model, even if it is able to generate correct word-by-word translation. In this study we propose a reordering model that involves both source- and target-side syntax information in the word reordering process.

Our work is inspired by the approach proposed in Imamura et al. (2005), where a complete syntax-driven SMT system based on a two-side subtree transfer is described. In their approach they construct a probabilistic non-isomorphic tree mapping model based on a context-free breakdown of the source and target parse trees; extract alignment templates that incorporate the constraints of the parse trees; and apply syntax-based decoding. We propose to use a similar non-isomorphic subtree mapping to extract reordering rules, but instead of involving them directly in the translation process, we use them to monotonize the source portion of the bilingual corpus.

In the next step, the rules are applied to the source part of the same training corpus changing the source sentence structure such that it more closely matches the word order of the target language. It leads to a simplification of the translation task due to a shorter average length of bilingual units which it is more likely to see when translating an unseen set.

Local and long-range word reorderings are driven by automatically extracted permutation patterns operating with source language constituents and underlaid by non-isomorphic subtree transfer. The target-side parse tree utilization is optional, but it greatly affects system performance: it is considered as a filter constraining the reordering rules to the set of patterns covered by both the source- and target-side subtrees. Apart from the reordering rules representing the order of child nodes, a set of additional rewrite rules based on a deep top-down subtree analysis is considered, which is another novel aspect of the paper.

We used the N-gram-based SMT system of Mariño et al. (2006) to test the proposed syntax-based reordering model, which is an alternative to the phrase-based state-of-the-art Moses[1] system.

2 Related work

In practice, a reordering model operates on a sentence level and is carried out based on word reordering rules derived from the training corpus. Reordering patterns can be purely statistical (see Costa-jussà and Fonollosa (2006), for example), use language-based syntactic information (Collins et al., 2005); the reordering can be driven by a lat-

[1] www.statmt.org/moses/

tice of syntactically motivated alternative translations (Elming, 2008) or be based on automatically extracted patterns driven by syntactical structure of the languages (see Crego and Mariño (2007b) as an example). Another recent implementation of the preprocessing approach to syntax-based reordering though an n-best list generation can be found in Li et al. (2007).

Word class-based reordering patterns were part of Och's Alignment Template system (Och et al., 2004). The modern state-of-the-art phrase-based translation system Moses, along with a distance based distortion model (Koehn et al., 2003), implements the phrase-based reordering (Tillmann and Zhang, 2005).

Reordering algorithms specifically developed for an N-gram system include a constrained distance-based distortion model (Costa-jussà et al., 2006) and a linguistically motivated reordering model employing monotonic search graph extension (Crego and Mariño, 2007a).

An example of a word order monotonization strategy can be found in Costa-jussà and Fonollosa (2006), where a monotone sequence of source words is translated into the reordered sequence using SMT techniques.

In Xia and Mccord (2004) the authors present a hybrid system for French-English translation, based on the principle of automatic rewrite patterns extraction using a parse tree and phrase alignments. This method differs from the one presented in this paper, among other distinctions, by a lexical model underlying the subtree syntax transfer and a different statistical model used for translation.

3 Baseline SMT system

N-gram-based SMT has proved to be competitive with the state-of-the-art systems in recent evaluation campaigns (Khalilov et al., 2008; Lambert et al., 2007).

According to the N-gram-based approach, the translation process is considered as an *arg max* searching for the translation hypothesis $\hat{e}_1^I$ maximizing a log-linear combination of a translation model (TM) and a set of feature models:

$$\hat{e}_1^I = \arg\max_{e_1^I} \left\{ \sum_{m=1}^{M} \lambda_m h_m(e_1^I, f_1^J) \right\} \tag{1}$$

where the feature functions h_m refer to the system models and the set of λ_m refers to the weights corresponding to these models.

A detailed description of the N-gram-based approach can be found in Mariño et al. (2006).

As decoder, we used MARIE[2] (Crego et al., 2005), a beam-search decoder implementing a distance-based constrained distortion model, limited by two parameters: *m* - a maximum distance measured number in words that a phrase can be reordered and *j* - a maximum number of "jumps" within a sentence (Costa-jussà et al., 2006).

4 Syntax-based reordering

Our syntax-based reordering (SBR) system requires access to source and target language parse trees, along with the source-to-target and target-to-source word alignments intersection. In the framework of the study we used the Stanford Parser (Klein and Manning, 2003) for both languages, however the system permits using any other natural language parser allowing for different formal grammars for the source and the target languages.

4.1 Notation

SBR operates with source and target parse trees that represent the syntactic structure of a string in source and target languages in a Context-Free Grammar (CFG) fashion.

This representation is called *"CFG form"*, and is formally defined in the usual way as $G = \langle N, T, R, S \rangle$, where N is a set of nonterminal symbols (corresponding to source-side phrase and part-of-speech tags); T is a set of source-side terminals (the lexicon); R is a set of production rules of the form $\eta \rightarrow \gamma$, with $\eta \in N$ and γ a sequence of terminal and nonterminal symbols; and $S \in N$ is the distinguished symbol.

The reordering rules then have the form

$$\eta_0 @0 \ldots \eta_k @k \rightarrow \\ \eta_{d_0} @d_0 \ldots \eta_{d_k} @d_k | Lexicon | p_1 \tag{2}$$

where $\eta_i \in N$ for all $0 \leq i \leq k$; $(d_o \ldots d_k)$ is a permutation of $(0 \ldots k)$; *Lexicon* includes the source-side set of words for each η_i; and p_1 is a probability associated with the rule. Figure 1 gives two examples of the rule format.

4.2 Rule extraction

Concept. Inspired by the ideas presented in Imamura et al. (2005), where monolingual correspon-

[2] http://gps-tsc.upc.es/veu/soft/soft/marie/

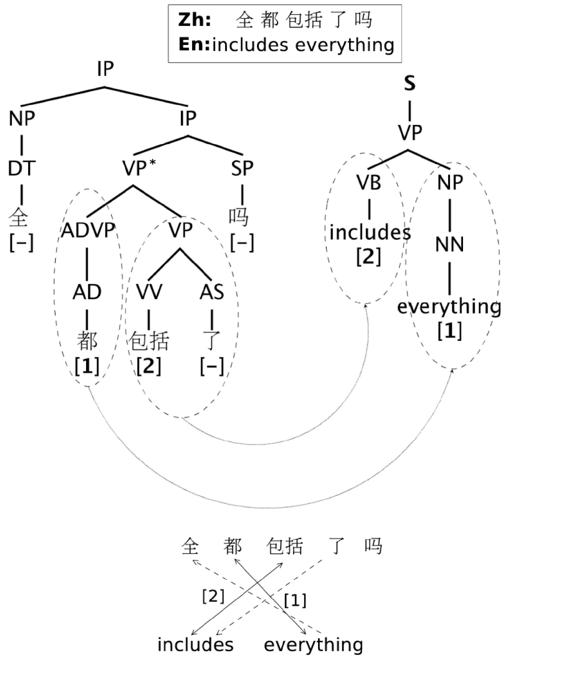
dences of syntactic nodes are used during decoding, we extract a set of bilingual patterns allowing for reordering as described below:

(1) align the monotone bilingual corpus with GIZA++[3] (Och and Ney, 2003) and find the intersection of direct and inverse word alignments, resulting in the construction of the projection matrix P (see below);

(2) parse the source and the target parts of the parallel corpus;

(3) extract reordering patterns from the parallel non-isomorphic CFG-trees based on the word alignment intersection.

Step 2 is straightforward; we explain aspects of Steps 1 and 3 in more detail below. Figure 1 shows an example of the generation of two lexicalized rules; we use this below in our explanations.



Extracted rules:
ADVP@0 VP@1 -> VP@1 ADVP@0 | ADVP@0 << 都 >> VP@1 << 包括 了 >>
AD@0 VP@1 -> VP@1 AD@0 | AD@0 << 都 >> VP@1 << 包括 了 >>

Figure 1: Example of reordering rules extraction.

Projection matrix. Bilingual content can be represented in the form of words or sequences of words depending on the syntactic role of the corresponding grammatical element (constituent or POS).

Given two parse trees and a word alignment intersection, a projection matrix P is defined as an $M \times N$ matrix such that M is the number of words in the target phrase; N is the number of words in the source phrase; and a cell (i, j) has a value based on the alignment intersection — this value is zero if word i and word j do not align, and is a unique non-zero link number if they do.

For the trees in Figure 1,

$$P = \begin{pmatrix} 0 & 0 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Alignment and sub-trees interaction. Each non-terminal from the source and target parse trees is assigned a string carrying information about elements from the alignment intersection which are contained in its child nodes, taking into account the order of their appearance in the tree (AI). For example, the AI string assigned to the source-side internal node VP^* in Figure 1 is "1 2" and to the target-side VP is "2 1". This information is used to indicate the source-side nodes which are to be reordered according to the target language syntactical structure. Reordering patterns are extracted following the source and target-side AIs as shown in Figure 1 ("*main rules*").

If more than one non-zero element of the projection matrix is reachable through the child nodes, the AI has a more complex structure, providing information about elements from alignment intersection belonging to one or another child node. An example can be found in Figure 2.

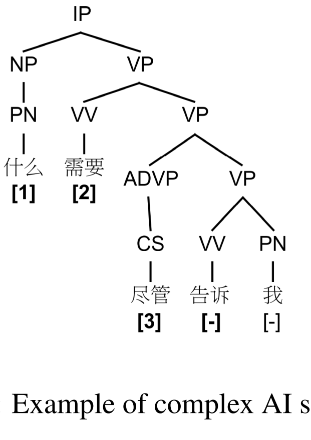


Figure 2: Example of complex AI structure.

Here, the subtree IP is assigned with the AI_{IP} = "1 (2 3)", meaning that it has two child nodes: the first contains the element 1 from the alignment intersection and the second, elements 2 and 3 (we call this subsequence "*closed*"). The reordering system considers nodes assigned with one or more

[3] http://code.google.com/p/giza-pp/

children equally discerning the nodes with different order alignment elements.

Unary chains. Given a unary chain like "$ADVP \rightarrow AD \rightarrow ...$", rules are extracted for each level in this chain. For example in Figure 1, the directly extracted reordering rules are equivalent since the node $ADVP$ leads to the leaf through the node AD and does not have other edges.

The role of target-side parse tree. Conceptually speaking, the use of target-side parse tree is optional. Although reordering is performed on the source side only, the target-side tree is of great importance: the reordering rules can be only extracted if the words covered by the rule are entirely covered by both a node in the source and in the target trees. It allows the more accurate determination of the covering and limits of the rules.

4.3 Secondary rules

There are a lot of nodes for which a comparison of AIs indicates that a subtree transfer can be done, but segmentation of child nodes is not identical.

Figure 3 illustrates this situation. AI strings assigned to the root nodes of the trees contain the same elements, but segmentation and/or order of appearance of elements do not coincide. These subtrees can not be directly used for pattern extraction and more in-depth analysis is required.

We adopt the following six step algorithm for each parent node from the source-side parse tree:

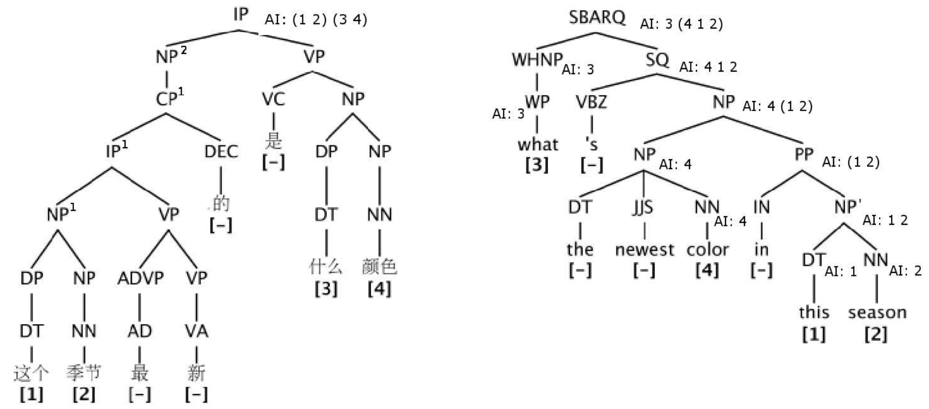
1. Find the AI sequence for the source-side top-level element (considering example, IP node is assigned "(1 2) (3 4)").

2. Go down through the target-side tree, finding AIs for each node.

3. Find all target-side closed subsequences for the source-side AI found on step 1. In example, it is the subsequence "(1 2)".

4. Find all target-side isolated nodes corresponding to the elements which were not covered on step 2. In example, these elements are "3" and "4".

5. Extend the set of source-side nodes found in steps 2 and 3 with equivalent branches. Since the words which are not presented in the alignment intersection do not affect the projection matrix, "equivalence" means here that all the branches spanning the elements from



Example of extracted rules:

```
NP@0 DP@1 NP@2 -> DP@1 NP@2 NP@0 | NP@ << 这个 季节 >> DP@1 << 什么 >> NP@2 << 颜色 >>
NP@0 DP@1 NN@2 -> DP@1 NN@2 NP@0 | NP@ << 这个 季节 >> DP@1 << 什么 >> NN@2 << 颜色 >>
NP@0 DT@1 NP@2 -> DT@1 NP@2 NP@0 | NP@ << 这个 季节 >> DT@1 << 什么 >> NP@2 << 颜色 >>
NP@0 DT@1 NN@2 -> DT@1 NN@2 NP@0 | NP@ << 这个 季节 >> DP@1 << 什么 >> NN@2 << 颜色 >>
CP@0 DP@1 NP@2 -> DP@1 NP@2 CP@0 | CP@0 << 这个 季节 最 新 的 >> DP@1 << 什么 >> NP@2 << 颜色 >>
CP@0 DP@1 NN@2 -> DP@1 NN@2 CP@0 | CP@0 << 这个 季节 最 新 的 >> DP@1 << 什么 >> NN@2 << 颜色>>
CP@0 DT@1 NP@2 -> DT@1 NP@2 CP@0 | CP@0 << 这个 季节 最 新 的 >> DT@1 << 什么 >> NP@2 << 颜色 >>
CP@0 DT@1 NN@2 -> DT@1 NN@2 CP@0 | CP@0 << 这个 季节 最 新 的 >> DT@1 << 什么 >> NN@2 << 颜色>>
NP@0 DP@1 NP@2 -> DP@1 NP@2 NP@0 | NP@0 << 这个 季节 最 新 的 >> DT@1 << 什么 >> NN@2 << 颜色>>
NP@0 DP@1 NN@2 -> DP@1 NN@2 NP@0 | NP@0 << 这个 季节 最 新 的 >> DT@1 << 什么 >> NN@2 << 颜色>>
NP@0 DT@1 NP@2 -> DT@1 NP@2 NP@0 | NP@0 << 这个 季节 最 新 的 >> DT@1 << 什么 >> NN@2 << 颜色>>
NP@0 DT@1 NN@2 -> DT@1 NN@2 NP@0 | NP@0 << 这个 季节 最 新 的 >> DT@1 << 什么 >> NN@2 << 颜色>>
...
NP@0 VP@1 -> VP@1 NP@0 | NP@0 << 这个 季节 最 新 的 >> VP@1 << 是 什么 颜色 >>
```

Figure 3: Example of “secondary“ rules extraction.

the given instance are considered equally (for example, elements NP^1 are equivalent to the nodes IP^1, CP^1 and NP^2).

6. Place them in order corresponding to the target-side AI and construct the final reordering patterns (*"secondary rules"*).

As illustration of the limitations incurred by target-side parse tree, the potential reordering pattern $NP@0\ VP@1 \rightarrow VP@1\ NP@0$ (referring to the top node in the Chinese tree) is not allowed due to distinct source- and target-side tree coverage.

4.4 Rule organization

Once the list of fully lexicalized reordering patterns is extracted, all the rules are progressively processed, reducing amount of lexical information. Initial rules are iteratively expanded such that each element of the pattern is generalized until all the lexical elements of the rule are represented in the form of fully unlexicalized categories. Hence, from each initial pattern with N lexical elements, $2^N - 2$ partially lexicalized rules and 1 general rule are generated. An example of the process of delexicalization can be found in Figure 4.

Thus, finally three types of rules are available: (1) fully lexicalized (initial) rules, (2) partially lexicalized rules and (3) unlexicalized (general) rules.

On the next step, the sets are processed separately: patterns are pruned and ambiguous rules are removed. Fully and partially lexicalized rules are not pruned out, but we set the threshold k_{gener} to 3. All the rules from the corresponding set that appear less than k times are directly discarded. The probability of a pattern is estimated based on frequency in the training corpus, and only the most probable rule is stored.

In this version of the reordering system, only the one-best reordering is used in other stages of the algorithm, so the rule output functioning as an input to the next rule can lead to situations reverting the change of word order that the previously applied rule made. Therefore, the rules that can be ambiguous when applied sequentially are pruned according to the higher probability principle. For example, for the pair of patterns with the same lexicon (which is empty for a general rule leading to a recurring contradiction *NP@0 VP@1 → VP@1 NP@0 p1*, *VP@0 NP@1 → NP@1 VP@0 p2*), the less probable rule is removed.

Finally, there are three resulting parameter tables analogous to the "r-table" as stated in (Yamada and Knight, 2001), consisting of POS- and constituent-based patterns allowing for reordering and monotone distortion.

4.5 Source-side monotonization

Rule application is performed as a bottom-up parse tree traversal following two principles:

(1) the longest possible rule is applied, i.e. among a set of nested rules, the rule with a longest left-side covering is selected. For example, in the case of the appearance of an *NN JJ RB* sequence and presence of the two reordering rules

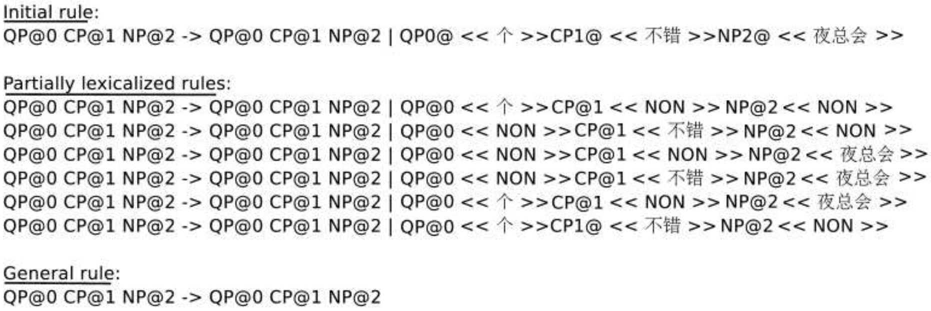
NN@0 JJ@1 → ... and

NN@0 JJ@1 RB@2 → ...

the latter pattern will be applied.

(2) the rule containing the maximum lexical information is applied, i.e. in case there is more than one alternative pattern from different groups, the lexicalized rules have preference over the partially lexicalized, and partially lexicalized over general ones.

Figure 5 shows example of the reordered source-side tree corresponding to the example from Fig-



```
Initial rule:
QP@0 CP@1 NP@2 -> QP@0 CP@1 NP@2 | QP0@ << 个 >>CP1@ << 不错 >>NP2@ << 夜总会 >>

Partially lexicalized rules:
QP@0 CP@1 NP@2 -> QP@0 CP@1 NP@2 | QP@0 << 个 >>CP@1 << NON >> NP@2 << NON >>
QP@0 CP@1 NP@2 -> QP@0 CP@1 NP@2 | QP@0 << NON >>CP@1 << 不错 >> NP@2 << NON >>
QP@0 CP@1 NP@2 -> QP@0 CP@1 NP@2 | QP@0 << NON >>CP@1 << NON >> NP@2 << 夜总会 >>
QP@0 CP@1 NP@2 -> QP@0 CP@1 NP@2 | QP@0 << NON >>CP@1 << 不错 >> NP@2 << 夜总会 >>
QP@0 CP@1 NP@2 -> QP@0 CP@1 NP@2 | QP@0 << 个 >>CP@1 << NON >> NP@2 << 夜总会 >>
QP@0 CP@1 NP@2 -> QP@0 CP@1 NP@2 | QP@0 << 个 >>CP1@ << 不错 >> NP@2 << NON >>

General rule:
QP@0 CP@1 NP@2 -> QP@0 CP@1 NP@2
```

Figure 4: Example of lexical rule expansion.

ure 1 with the applied pattern

$$ADVP@0\ VP@1 \rightarrow VP@1\ ADVP@0$$

and the given lexicon. The resulting reordered Chinese phrase more closely matches the order of the target language and is considered as a result of the subtree transfer.

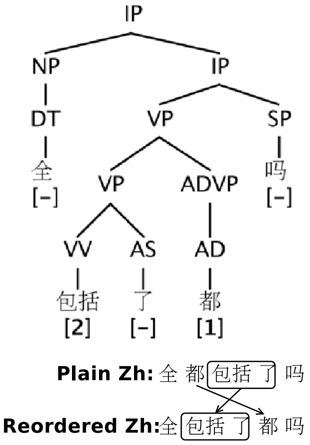


Figure 5: *Reordered source-side parse tree.*

Once the reordering of the training corpus is ready, it is realigned and new more monotonic alignment is passed to the SMT system. In theory, the word links from the original alignment can be used, however, from our experience, running GIZA++ again results in a better word alignment since it is easier to learn on the modified training example.

5 Experiments and results

5.1 Corpus

The experiments were conducted on two Chinese-English corpora: the BTEC corpus consisting of short tourism related sentences and the 50K first-lines extraction from the NIST'06 corpus belonging to the news domain (NIST50K). The main reason why the Chinese-English translation task was chosen for experiments is that European languages are not so crucial for global (long-distance) reordering problem as the translation between Asian languages and English.

We expect that the need for longer distance reorderings would be found in longer sentences, as in the NIST50K corpus, but we also include the BTEC corpus to see whether there is an effect for shorter sentences as well. Basic statistics of the training material can be found in Tables1 and 2.

Both system were optimized and tested on in-domain data. BTEC development and test datasets consist of 489 and 500 sentences, respectively, and are provided with 7 reference translations. NIST50K development and test sets are both 541 sentences long, 4 references are provided.

5.2 Experiment setup

Evaluation conditions were case-insensitive and with punctuation marks considered. We used the Stanford Parser as a NLP parsing engine (Klein and Manning, 2003) trained on the Chinese and English Penn Treebank sets (32 POS/44 constituent categories for Arabic Treebank and 48 POS/14 syntactic tags for English Treebank).

N-gram models were estimated using the SRILM toolkit (Stolcke, 2002). For both tasks TM is represented in a 4-gram model form using modified Kneser-Ney discounting with interpolation, target language model (LM) of words is a 4-gram model with modified Kneser-Ney discounting, while a target-side POS LM is a 4-gram with Good-Turing backing-off.

For all system configurations, apart from monotone experiments, parameters of the distance-based reordering model were set to $m = 5$ and $j = 5$ for a trade-off between efficiency and accuracy.

The optimization criteria was the highest $4NIST + 100BLEU$ score.

5.3 Results and discussion

The following scores are reported in Table 3: final score obtained as a result of model weights tuning for development dataset (*dev*), BLEU and METEOR scores for the test dataset (*test*). We present results for two corpora: BTEC and NIST05K characterized by different domain and sentence length.

We contrast four n-gram-based system configurations comparing the SBR results with the distortion model:

- *Baseline*: the training data is not reordered and allows for *C*onstrained *D*istortion ($m = 5, j = 5$) during decoding, as described in (Costa-jussà et al., 2006);

- *SynBReor*: SBR is applied on the preprocessing step involving *main rules* only, the dev/test sets are monotonically decoded;

- *SynBReor+SecRules*: SBR is applied involving *main* and *secondary rules* and allows for constrained distortion ($m = 5, j = 5$).

	Chinese	English
Sentences	44.9 K	44.9 K
Words	299.0 K	324.4 K
Average sentence length	6.66	7.22
Vocabulary	11.4 K	9 K

Table 1: Basic statistics of the BTEC training corpus.

	Chinese	English
Sentences	50 K	50 K
Words	1.18 M	1.25 M
Average sentence length	23.6	25.03
Vocabulary	27.2 K	30.4 K

Table 2: Basic statistics of the NIST50K training corpus.

Application of the SBR technique demonstrates an improvement in translation quality according to the automatic scores. *SynBReor+mj* is found to be the best system configuration for both sets of experiments, outperforming the baseline configuration by about 0.4 BLEU points (2.9 %) that is not statistically significant for the BTEC task, however, for the NIST50K task the difference is about 0.9 BLEU points (4 %) reaching a statistical significance threshold[4]. The METEOR score also increases with raise of reordering system complexity, supporting the BLEU results. The SBR algorithm is illustrated in Figure 6, where the Chinese block of words is moved to the end of the sentence that better matches the structure of the English counterpart.

As usual, for the tasks with scarce resources the improvements on the *test* and *dev* sets are not coherent. While a clear improvement of test results can be observed in the BTEC results, the development set score degrades when SBR is applied.

It is possible to see from Table 3 that the introduction of secondary rules influences negatively the number of extracted tuples and comparing to the "main rules only" configuration shows a degradation in performance. Generally speaking, secondary rules include more elements than primary ones and are more difficult to be seen in the dataset parsed with the Stanford Parser. However, we speculate that accurate pruning of secondary rules could benefit the system performance significantly.

6 Conclusions and future work

In this paper we introduced a syntax-based reordering technique that monotonizes the word order of

[4] All statistical significance calculations are done for a 95% confidence interval and 1000 resamples (Koehn, 2004).

	dev	test BLEU	test METEOR	# tuples	voc tuples
BTEC experiments					
Baseline	48.17	19.50	47.05	150,378	36,643
SynBReor	47.55	19.91	47.50	157,345	36,936
SynBReor+SecRules	47.83	19.70	47.52	141,430	36,501
NIST50K experiments					
Baseline	19.16	21.28	41.55	240,609	112,947
SynBReor	19.90	22.21	41.77	252,113	114,702
SynBReor+SecRules	19.45	21.80	42.03	251,012	113,985

Table 3: Summary of the experimental results.

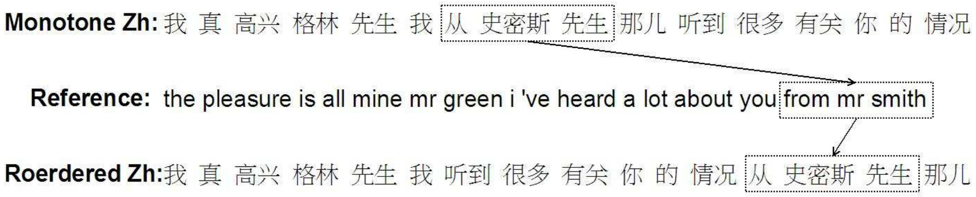

Monotone Zh: 我 真 高兴 格林 先生 我 从 史密斯 先生 那儿 听到 很多 有关 你 的 情况
Reference: the pleasure is all mine mr green i 've heard a lot about you from mr smith
Roerdered Zh:我 真 高兴 格林 先生 我 听到 很多 有关 你 的 情况 从 史密斯 先生 那儿

Figure 6: Example of SBR application.

source and target languages involved in the process of bilingual unit extraction. As can be seen from the results presented, the proposed algorithm shows competitive performance comparing with a fundamental distance-based reordering model.

The comparison is done on two smaller Chinese-English translation tasks with a strong need for word reorderings. The major part of the sentences from the BTEC corpus are short and on the example of the tourism translation task one can observe the SBR capacity to deal with local reordering. The NIST50K task demonstrates potential of the SBR algorithm on the translation task with much longer average sentence length and much need of long-distance reorderings. In this case, the reordered system significantly outperforms the state-of-the-art model.

The proposed approach is flexible and in the next step will be applied to phrase-based systems. Further work also includes the algorithm's application to a different language pair with distinct need for reorderings, analysis of the extracted tuples and development of the algorithm for accurate selection of reordering rules.

7 Acknowledgments

This work has been funded by the Spanish Government under grant TEC2006-13964-C03 (AVIVAVOZ project) and under a FPU grant.

References

Collins, M., Ph. Koehn, and I. Kučerová. 2005. Clause restructuring for statistical machine translation. In *Proc. of the ACL'05*, pages 531–540.

Costa-jussà, M.R. and J.A.R. Fonollosa. 2006. Statistical machine reordering. In *Proc. of EMNLP'06*.

Costa-jussà, M.R., J. M. Crego, A. de Gispert, P. Lambert, M. Khalilov, J. A.R. Fonollosa, J. B. Mariño, and R. E. Banchs. 2006. TALP phrase-based system and TALP system combination for IWSLT 2006. In *Proc. of the IWSLT'06*, pages 123–129.

Crego, J.M. and J.B. Mariño. 2007a. Improving statistical MT by coupling reordering and decoding. *Machine Translation*, 20(3):199–215.

Crego, J.M. and J.B. Mariño. 2007b. Syntax-enhanced N-gram-based SMT. In *Proc. of MT SUMMIT XI*.

Crego, J.M., J.B. Mariño, and A. de Gispert. 2005. An ngram-based statistical machine translation decoder. In *Proc. of INTERSPEECH05*.

Elming, J. 2008. Syntactic reordering integrated with phrase-based SMT. In *Proc. of the ACL'08 Workshop SSST-2*, pages 46–54, June.

Imamura, K., H. Okuma, and E. Sumita. 2005. Practical approach to syntax-based statistical machine translation. In *Proc. of MT Summit X*.

Khalilov, M., A.H. Hernández, M.R. Costa-jussà, J.M. Crego, C.A. Henríquez, P. Lambert, J.A.R. Fonollosa, J.B. Mariño, and R. Banchs. 2008. The TALP-UPC ngram-based statistical machine translation system for ACL-WMT 2008. In *Proc. of WMT'08*, pages 127–131.

Klein, D. and C. Manning. 2003. Accurate unlexicalized parsing. In *Proc. of the ACL'03*.

Koehn, Ph., F.J. Och, and D. Marcu. 2003. Statistical phrase-based machine translation. In *Proc. of the HLT-NAACL'03*, pages 48–54.

Koehn, Ph. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of EMNLP'04*, pages 388–395.

Lambert, P., M.R. Costa-jussà, J. M. Crego, M. Khalilov, J. Mariño, R.E. Banchs, J.A.R. Fonollosa, and H. Shwenk. 2007. The TALP ngram-based SMT system for IWSLT 2007. In *Proc. of IWSLT'07*, pages 169–174.

Li, C., D. Zhang, M. Zhou, M. Li, and Y. Guan. 2007. A probabilistic approach to syntax-based reordering for statistical machine translation. In *Proc. of ACL'07*, pages 720–727.

Mariño, J.B., R.E. Banchs, J.M. Crego, A. de Gispert, P. Lambert, J.A.R. Fonollosa, and M.R. Costa-jussà. 2006. N-gram based machine translation. *Computational Linguistics*, 32(4):527–549.

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

Och, F.J., D. Gildea, S. Khudanpur, A. Sarkar, K. Yamada, A. Fraser, S. Kumar, L. Shen, D. Smith, K. Eng, V. Jain, Z. Jin, and D. Radev. 2004. A Smorgasbord of Features for Statistical Machine Translation. In *Proc. of HLT/NAACL'04*.

Stolcke, A. 2002. SRILM: an extensible language modeling toolkit. In *Proc. of ICSLP'02*.

Tillmann, C. and T. Zhang. 2005. A localized prediction model for statistical machine translation. In *Proc. of the ACL 2005*, pages 557–564.

Xia, F. and M. McCord. 2004. Improving a statistical MT system with automatically learned rewrite patterns. In *Proc. of 2004*.

Yamada, K. and K. Knight. 2001. A syntax-based statistical translation model. In *Proc. of the ACL 2001*, pages 523–530.