

Syntactic Reordering Integrated with Phrase-based SMT

Jakob Elming

Computational Linguistics
Copenhagen Business School
jel.isv@cbs.dk

Abstract

We present a novel approach to word reordering which successfully integrates syntactic structural knowledge with phrase-based SMT. This is done by constructing a lattice of alternatives based on automatically learned probabilistic syntactic rules. In decoding, the alternatives are scored based on the output word order, not the order of the input. Unlike previous approaches, this makes it possible to successfully integrate syntactic reordering with phrase-based SMT. On an English-Danish task, we achieve an absolute improvement in translation quality of 1.1 % BLEU. Manual evaluation supports the claim that the present approach is significantly superior to previous approaches.

1 Introduction

The emergence of phrase-based statistical machine translation (PSMT) (Koehn et al., 2003) has been one of the major developments in statistical approaches to translation. Allowing translation of word sequences (phrases) instead of single words provides SMT with a robustness in word selection and local word reordering.

PSMT has two means of reordering the words. Either a phrase pair has been learned where the target word order differs from the source (phrase internal reordering), or distance penalized orderings of target phrases are attempted in decoding (phrase

external reordering). The first solution is strong, the second is weak.

The second solution is necessary for reorderings within a previously unseen sequence or over distances greater than the maximal phrase length. In this case, the system in essence relies on the target side language model to get the correct word order. The choice is made without knowing what the source is. Basically, it is a bias against phrase external reordering.

It seems clear that reordering often depends on higher level linguistic information, which is absent from PSMT. In recent work, there has been some progress towards integrating syntactic information with the statistical approach to reordering. In works such as (Xia and McCord, 2004; Collins et al., 2005; Wang et al., 2007; Habash, 2007), reordering decisions are done “deterministically”, thus placing these decisions outside the actual PSMT system by learning to translate from a reordered source language. (Crego and Mariño, 2007; Zhang et al., 2007; Li et al., 2007) are more in the spirit of PSMT, in that multiple reorderings are presented to the PSMT system as (possibly weighted) options.

Still, there remains a basic conflict between the syntactic reordering rules and the PSMT system: one that is most likely due to the discrepancy between the translation units (phrases) and units of the linguistic rules, as (Zhang et al., 2007) point out.

In this paper, we proceed in the spirit of the non-deterministic approaches by providing the decoder with multiple source reorderings. But instead of scoring the input word order, we score the order of the output. By doing this, we avoid the integration problems of previous approaches.

It should be noted that even though the experi-

© 2008. Licensed under the *Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported* license (<http://creativecommons.org/licenses/by-nc-sa/3.0/>). Some rights reserved.

ments are conducted within a source reordering approach, this scoring is also compatible with other approach. We will, however, not look further into this possibility in the present paper.

In addition, we automatically learn reordering rules based on several levels of linguistic information from word form to subordination and syntactic structure to produce reordering rules that are not restricted to operations on syntactic tree structure nodes.

In the next section, we discuss and contrast related work. Section 3 describes aspects of English and Danish structure that are relevant to reordering. Section 4 describes the automatic induction of reordering rules and its integration in PSMT. In section 5, we describe the SMT system used in the experiments. Section 6 evaluates and discusses the present approach.

2 Related Work

While several recent authors have achieved positive results, it has been difficult to integrate syntactic information while retaining the strengths of the statistical approach.

Several approaches do deterministic reordering. These do not integrate the reordering in the PSMT system; instead they place it outside the system by first reordering the source language, and then having a PSMT system translate from reordered source language to target language. (Collins et al., 2005; Wang et al., 2007) do this using manually created rules, and (Xia and McCord, 2004) and (Habash, 2007) use automatically extracted rules. All use rules extracted from syntactic parses.

As mentioned by (Al-Onaizan and Papineni, 2006), it can be problematic that these deterministic choices are beyond the scope of optimization and cannot be undone by the decoder. That is, there is no way to make up for bad information in later translation steps.

Another approach is non-deterministic. This provides the decoder with both the original and the reordered source sentence. (Crego and Mariño, 2007) operate within Ngram-based SMT. They make use of syntactic structure to reorder the input into a word lattice. Since the paths are not weighted, the lattice merely narrows down the size of the search space. The decoder is not given reason to trust one path (reordering) over another.

(Zhang et al., 2007) assign weights to the paths of their input word lattice. Instead of hierarchical

linguistic structure, they use reordering rules based on POS and syntactic chunks, and train the system with both original and reordered source word order on a restricted data set (<500K words). Their system does not out-perform a standard PSMT system. As they themselves point out, a reason for this might be that their reordering approach is not fully integrated with PSMT. This is one of the main problems addressed in the present work.

(Li et al., 2007) use weighted n-best lists as input for the decoder. They use rules based on a syntactic parse, allowing children of a tree node to swap place. This is excessively restrictive. For example, a common reordering in English-Danish translation has the subject change place with the finite verb. Since the verb is often embedded in a VP containing additional words that should not be moved, such rules cannot be captured by local reordering on tree nodes.

In many cases, the exact same word order that is obtained through a source sentence reordering, is also accessible through a phrase internal reordering. A negative consequence of source order (SO) scoring as done by (Zhang et al., 2007) and (Li et al., 2007) is that they bias against the valuable phrase internal reorderings by only promoting the source sentence reordering. As described in section 4.3, we solve this problem by reordering the input string, but scoring the output string, thus allowing the strengths of PSMT to co-exist with rule-based reordering.

3 Language comparison

The two languages examined in this investigation, English and Danish, are very similar from a structural point of view. A word alignment will most often display an almost one-to-one correlation. In the hand-aligned data, only 39% of the sentences contain reorderings (following the notion of reordering as defined in 4.1). On average, a sentence contains 0.66 reorderings.

One of the main differences between English and Danish word order is that Danish is a verb-second language: the finite verb of a declarative main clause must always be the second constituent. Since this is not the case for English, a reordering rule should move the subject of an English sentence to the right of the finite verb, if the first position is filled by something other than the subject. This is exemplified by (1) (examples are annotated with English gloss and translation), where 'they'

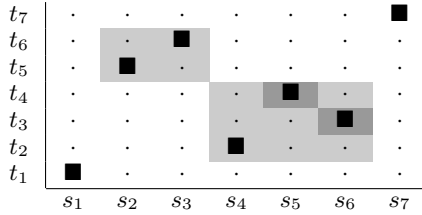


Table 1: Reordering example

should move to the right of 'come' to get the Danish word order as seen in the gloss.

- (1) nu kommer de
 [now come **they**]
 'here **they** come'

Another difference is that Danish sentence adverbials in a subordinate clause move to the left of the finite verb. This is illustrated in example (2). This example also shows the difficulty for a PSMT system. Since the trigram 'han kan ikke' is frequent in Danish main clauses, and 'han ikke kan' is frequent in subordinate clauses, we need information on subordination to get the correct word order. This information can be obtained from the conjunction 'that'. A trigram PSMT system would not be able to handle the reordering in (2), since 'that' is beyond the scope of 'not'.

- (2) han siger at han ikke kan se
 [he says that he **not** can see]
 'he says that he can **not** see'

In the main clause, on the other hand, Danish prefers the sentence adverbial to appear to the right of the finite verb. Therefore, if the English adverbial appears to the left of the finite verb in a main clause, it should move right as in example (3).

- (3) hun så aldrig skibet
 [she saw **never** the ship]
 'she **never** saw the ship'

4 Reordering rules

4.1 Definition of reordering

In this experiment, reordering is defined as two word sequences exchanging positions. These two sequences are restricted by the following conditions:

- **Parallel consecutive:** They have to make up consecutive sequences of words, and each has to align to a consecutive sequence of words.
- **Maximal:** They have to be the longest possible consecutive sequences changing place.
- **Adjacent:** They have to appear next to each other on both source and target side.

The sequences are not restricted in length, making both short and long distance reordering possible. Furthermore, they need not be phrases in the sense that they appear as an entry in the phrase table.

Table 1 illustrates reordering in a word alignment matrix. The table contains reorderings between the light grey sequences (s_2^3 and s_4^6)¹ and the dark grey sequences (s_5^5 and s_6^6). On the other hand, the sequences s_3^3 and s_4^5 are e.g. not considered reordered, since neither are maximal, and s_4^5 is not consecutive on the target side.

4.2 Rule induction

In section 3, we pointed out that subordination is very important for word order differences between English and Danish. In addition, the sentence position of constituents plays a role. All this information is present in a syntactic sentence parse. A subordinate clause is defined as inside an SBAR constituent; otherwise it is a main clause. The constituent position can be extracted from the sentence start tag and the following syntactic phrases. POS and word form are also included to allow for more specific/lexicalized rules.

Besides including this information for the candidate reordering sequences (left sequence (LS) and right sequence (RS)), we also include it for the set of possible left (LC) and right (RC) contexts of these. The span of the contexts varies from a single word to all the way to the sentence border. Table 2 contains an example of the information available to the learning algorithm. In the example, LS and RS should change place, since the first position is occupied by something other than the subject in a main clause.

In order to minimize the training data, word and POS sequences are limited to 4 words, and phrase structure (PS) sequences are limited to 3 constituents. In addition, an entry is only used if at least one of these three levels is not too long for

¹Notation: s_x^y means the consecutive source sequence covering words x to y .

Level	LC	LS	RS	RC
WORD	<s> today , today , ,	he	was driving	home home . home . < /s>
POS	<S> NN , NN , ,	PRP	AUX VBG	NN NN . NN . < /S>
PS	<S> NP , NP , ,	NP	AUX VBG	ADVP ADVP . ADVP . < /S>
SUBORD	main	main	main	main

Table 2: Example of experience for learning. Possible contexts separated by ||.

both LS and RS, and too long contexts are not included in the set. This does not constrain the possible length of a reordering, since a PS sequence of length 1 can cover an entire sentence.

In order to extract rules from the annotated data, we use a rule-based classifier, Ripper (Cohen, 1996). The motivation for using Ripper is that it allows features to be sets of strings, which fits well with our representation of the context, and it produces easily readable rules that allow better understanding of the decisions being made. In section 6.2, extracted rules are exemplified and analyzed.

The probabilities of the rules are estimated using Maximum Likelihood Estimation based on the information supplied by Ripper on the performance of the individual rules on the training data. These logarithmic probabilities are easily integratable in the log-linear PSMT model as an additional parameter by simple addition.

The rules are extracted from the hand-aligned, Copenhagen Danish-English Dependency Treebank (Buch-Kromann et al., 2007). 5478 sentences from the news paper domain containing 111,805 English words and 100,185 Danish words. The English side is parsed using a state-of-the-art statistical English parser (Charniak, 2000).

4.3 Integrating rule-based reordering in PSMT

The integration of the rule-based reordering in our PSMT system is carried out in two separate stages:

1. Reorder the source sentence to assimilate the word order of the target language.
2. Score the target word order according to the relevant rules.

Stage 1) is done in a non-deterministic fashion by generating a word lattice as input in the spirit of e.g. (Zens et al., 2002; Crego and Mariño, 2007; Zhang et al., 2007). This way, the system has both the original word order, and the reorderings predicted by the rule set. The different paths of the

word lattice are merely given as equal suggestions to the decoder. They are in no way individually weighted.

Separating stage 2) from stage 1) is motivated by the fact that reordering can have two distinct origins. They can occur because of stage 1), i.e. the lattice reordering of the original English word order (phrase *external* reordering), and they can occur inside a single phrase (phrase *internal* reordering). We are, however, interested in doing *phrase-independent*, word reordering. We want to promote rule-predicted reorderings, regardless of whether they owe their existence to a syntactic rule or a phrase table entry.

This is accomplished by letting the actual scoring of the reordering focus on the target string. The decoder is informed of where a rule has predicted a reordering, how much it costs to do the reordering, and how much it costs to avoid it. This is then checked for each hypothesized target string by keeping track of what *source position target order* (SPTO) it corresponds to.

The SPTO is a representation of which source position the word in each target position originates from. Putting it differently, the hypotheses contain two parallel strings; a target word string and its SPTO string. In order to access this information, each phrase table entry is annotated with its internal word alignment, which is available as an intermediate product from phrase table creation. If a phrase pair has multiple word alignments, the most frequent is chosen.

Table 3 exemplifies the SPTO scoring. The source sentence is *'today he was late'*, and a rule has predicted that word 3 and 4 should change place. When the decoder has covered the first four input words, two of the hypotheses might be H1 and H2. At this point, it becomes apparent that H2 contains the desired SPTO (namely '4 3'), and it get assigned the reordering cost. H1 does not contain the rule-suggested SPTO (in stead, the words are in the order '3 4'), and it gets the violation cost. Both these scorings are performed in a phrase-

Source sentence:	today ₁ , ₂ he ₃ was ₄ late ₅	
Rule:	3 4 → 4 3	
Hypothesis	Target string	SPTO
H1	idag han var	1 3 4
H2	idag var han	1 4 3

Table 3: Example of SPTO scoring during decoding at source word 4.

independent manner. The decoder assigns the reordering cost to H2 without knowing whether the reordering is internal (due to a phrase table entry) or external (due to a syntactic rule).

Phrase internal reorderings at other points of the sentence, i.e. points that are not covered by a rule, are not judged by the reordering model. Our rule extraction does not learn every possible reordering between the two languages, but only the most general ones. If no rule has an opinion at a certain point in a sentence, the decoder is free to choose the translation it prefers without reordering cost.

Separating the scoring from the source language reordering also has the advantage that the SPTO scoring in essence is compatible with other approaches such as a traditional PSMT system. We will, however, not examine this possibility further in the present paper.

5 The PSMT system

The baseline is the PSMT system used for the 2006 NAACL SMT workshop (Koehn and Monz, 2006) with phrase length 3 and a trigram language model (Stolcke, 2002). The system was trained on the English and Danish part of the Europarl corpus version 3 (Koehn, 2005). Fourth quarter of 2000 was removed in order to use *the common test set* of 11369 sentences (330,082 English words and 309,942 Danish words with one reference) for testing. In addition, fourth quarter of 2001 was removed for development purposes. Of these, 10194 were used for various analysis purposes, thereby keeping the test data perfectly unseen. 500 sentences were taken from the development set for tuning the decoder parameters. This was done using the Downhill Simplex algorithm. In total, 1,137,088 sentences containing 31,376,034 English words and 29,571,518 Danish words were left for training the phrase table and language model.

The decoder used for the baseline system is Pharaoh (Koehn, 2004) with its distance-

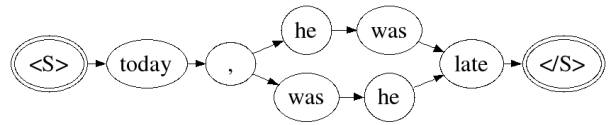


Figure 1: Example word lattice.

System	Dev	Test	Swap Subset
Baseline	0.262	0.252	0.234
no scoring	0.267	0.256	0.241
SO scoring	0.268	0.258	0.244
SPTO scoring	0.268	0.258	0.245

Table 4: BLEU scores for different scoring methods.

penalizing reordering model. For the experiments, we use our own decoder which — except for the reordering model — uses the same knowledge sources as Pharaoh, i.e. bidirectional phrase translation model and lexical weighting model, phrase and word penalty, and target language model. Its behavior is comparable to Pharaoh when doing monotone decoding.

The search algorithm of our decoder is similar to the RG graph decoder of (Zens et al., 2002). It expects a word lattice as input. Figure 1 shows the word lattice for the example in table 3.

Since the input format defines all possible word orders, a simple monotone search is sufficient. Using a language model of order n , for each hypothesized target string ending in the same $n-1$ -gram, we only have to extend the highest scoring hypothesis. None of the others can possibly outperform this one later on. This is because the maximal context evaluating a phrase extending this hypothesis, is the history ($n-1$ -gram) of the first word of that phrase. The decoder is not able to look any further back at the preceding string.

6 Evaluation

6.1 Results and discussion

The SPTO reordering approach is evaluated on the 11369 sentences of *the common test set*. Results are listed in table 4 along with results on the development set. We also report on the *swap subset*. These are the 3853 sentences where the approach actually motivated reorderings in the test set, internal or external. The remaining 7516 sentences were not influenced by the SPTO reordering approach.

System	BLEU	Avr. Human rating
Baseline	0.234	3.00 (2.56)
no scoring	0.240	3.00 (2.74)
SO scoring	0.239	3.00 (2.62)
SPTO scoring	0.244	2.00 (2.08)

Table 5: Evaluation on the set where SO and SPTO produce different translations. Average human ratings are medians with means in parenthesis, lower scores are better, 1 is the best score.

We report on 1) the baseline PSMT system, 2) a system provided with a rule reordered word lattice but no scoring, 3) the same system but with an SO scoring in the spirit of (Zhang et al., 2007; Li et al., 2007), and finally 4) the same system but with the SPTO scoring.

The SPTO approach gets an increase over the baseline PSMT system of 0.6 % BLEU. The swap subset, however, shows that the extracted rules are somewhat restricted, only resulting in swap in $\frac{1}{3}$ of the sentences. The relevant set, i.e. the set where the present approach actually differs from the baseline, is therefore the swap subset. This way, we concentrate on the actual focus of the paper, namely the syntactically motivated SPTO reordering. Here we achieve an increase in performance of 1.1 % BLEU.

Comparing to the other scoring approaches does not show much improvement. A possible explanation is that the rules do not apply very often, in combination with the fact that the SO and SPTO scoring mechanisms most often behave alike. The difference in SO and SPTO scoring only leads to a difference in translation in 10% of the sentences where reordering is done. This set is interesting, since it provides a focus on the difference between the SO and the SPTO approaches. In table 5, we evaluate on this set.

The BLEU scores on the entire set indicate that SPTO is a superior scoring method. To back this observation, the 100 first sentences are manually evaluated by two native speakers of Danish. (Callison-Burch et al., 2007) show that ranking sentences gives higher inter-annotator agreement than scoring adequacy and fluency. We therefore employ this evaluation method, asking the evaluators to rank sentences from the four systems given the input sentence. Ties are allowed. The annotators had reasonable inter-annotator agreement ($\kappa = 0.523$, $P(A) = 0.69$, $P(E) = 0.35$). Table 5

Decoder choice	SO	SPTO
Phrase internal reordering	401	1538
Phrase external reordering	3846	2849
Reject reordering	1468	1328

Table 6: The choices made based on the SO and SPTO scoring for the 5715 reorderings proposed by the rules for the test data.

shows the average ratings of the systems. This clearly shows the SPTO scoring to be significantly superior to the other methods ($p < 0.05$).

Most of the cases (55) where SPTO outperforms SO are cases where SPTO knows that a phrase pair contains the desired reordering, but SO does not. Therefore, SO has to use an external reordering which brings poorer translation than the internal reordering, because the words are translated individually rather than by a single phrase (37 cases), or it has to reject the desired reordering (18 cases), which also hurts translation, since it does not get the correct word order.

Table 6 shows the effect of SO and SPTO scoring in decoding. Most noticeable is that the SO scoring is strongly biased against phrase internal reorderings; SPTO uses nearly four times as many phrase internal reorderings as SO. In addition, SPTO is a little less likely to reject a rule proposed reordering.

6.2 Rule analysis

The rule induction resulted in a rule set containing 27 rules. Of these, 22 concerned different ways of identifying contexts where a reordering should occur due to the verb second nature of Danish. 4 rules had to do with adverbials in main and in subordinate clauses, and the remaining rule expressed that currency is written after the amount in Danish, while it is the other way around in English. Since the training data however only includes Danish Crowns, the rule was lexicalized to 'DKK'.

Table 7 shows a few of the most frequently used rules. The first three rules deal with the verb second phenomenon. The only difference among these is the left context. Either it is a prepositional phrase, a subordinate clause or an adverbial. These are three ways that the algorithm has learned to identify the verb second phenomenon conditions. Rule 3 is interesting in that it is lexicalized. In the learning data, the Danish correspondent to 'however' is most often not topicalized, and the subject

No	LC	LS	RS	RC
1	PS: <S> PP ,	PS: NP	POS: FV	
2	PS: SBAR ,	PS: NP	POS: FV	
3	PS: ADVP , ! WORD: however ,	PS: NP	POS: FV	
4		PS: FV	POS: RB	PS: VP SUB: sub
5	PS: <S> NP SUB: main	PS: ADVP	POS: FV	

Table 7: Example rules and their application statistics.

is therefore not forced from the initial position. As a consequence, the rule states that it should only apply, if 'however' is not included in the left context of the reordering.

Rule 4 handles the placement of adverbials in a subordinate clause. Since the right context is subordinate and a verb phrase, the current sequences must also be subordinate. In contrast, the fifth rule deals with adverbials in a main clause, since the left context noun phrase is in a main clause.

A problem with the hand-aligned data used for rule-induction is that it is out of domain compared to the Europarl data used to train the SMT system. The hand-aligned data is news paper texts, and Europarl is transcribed spoken language from the European Parliament. Due to its spoken nature, Europarl contains frequent sentence-initial forms of address. That is, left adjacent elements that are not integrated parts of the sentence as in example (4).

This is not straightforward, because on the surface these look a lot like topicalized constructions, as in example (5). In topicalized constructions, it is an integrated part of the sentence that is moved to the front in order to affect the flow of discourse information. This difference is crucial for the reordering rules, since 'i' and 'have' should reorder in (5), but not in (4), in order to get Danish word order.

(4) mr president , i have three points .

(5) as president , i have three points .

When translating the development set, it became clear that many constructions like (4) were reordered. Since these constructions were not present in the hand-aligned data, the learning algorithm did not have the data to learn this difference.

We therefore included a manual, lexicalized rule stating that if the left context contained one of a set of titles (mr, mrs, ms, madam, gentlemen), the re-

ordering should not take place. Since the learning includes word form information, this is a rule that the learning algorithm is able to learn. To a great extent, the rule eliminates the problem.

The above examples also illustrate that local reordering (in this case as local as two neighboring words) can be a problem for PSMT, since even though the reordering is local, the information about whether to reorder or not is not necessarily local.

6.3 Reordering analysis

In this section, we will show and discuss a few examples of the reorderings made by the SPTO approach. Table 8 contain two translations taken from the test set.

In translation 1), the subject (bold) is correctly moved to the right of the finite verb (italics), which the baseline system fails to do. Moving the finite verb away from the infinite verb 'feature', however, leads to incorrect agreement between these. While the baseline correctly retains the infinite form ('stå'), the language model forces another finite form (the past tense 'stod') in the SPTO reordering approach.

Translation 2) illustrates the handling of adverbials. The first reordering is in a main clause, therefore, the adverbial is moved to the right of the finite verb. The second reordering occurs in a subordinate clause, and the adverbial is moved to the left of the finite verb. Neither of these are handled successfully by the baseline system.

In this case, the reordering leads to better word selection. The English 'aims to' corresponds to the Danish 'sigter mod', which the SPTO approach gets correct. However, the baseline system translates 'to' to its much more common translation 'at', because 'to' is separated from 'aims' by the adverbial 'principally'.

7 Conclusion and Future Plans

We have described a novel approach to word reordering in SMT, which successfully integrates syntactically motivated reordering in phrase-based SMT. This is achieved by reordering the input string, but scoring on the output string. As opposed to previous approaches, this neither biases against phrase internal nor external reorderings. We achieve an absolute improvement in translation quality of 1.1 % BLEU. A result that is supported by manual evaluation, which shows that the SPTO

1	S	based on this viewpoint , every small port and every ferry port which handles a great deal of tourist traffic <i>should</i> feature on the european list .
	B	baseret på dette synspunkt , ethvert lille havn og alle færgehavnen som håndterer en stor turist trafik <i>skal</i> stå på den europæiske liste .
	P	baseret på dette synspunkt , <i>skal alle de små havne , og alle færgehavnen som behandler mange af turister trafik</i> stod på den europæiske liste .
2	S	the rapporteur generally welcomes the proposals in the commission white paper on this subject but is apprehensive of the possible implications of the reform , which aims principally to decentralise the implementation of competition rules .
	B	ordføreren generelt bifalder forslagene i kommissionens hvidbog om dette emne , men er bekymret for de mulige konsekvenser af den reform , som <i>sigter hovedsagelig</i> at decentralisere gennemførelsen af konkurrencereglerne .
	P	ordføreren <i>bifalder generelt</i> forslagene i kommissionens hvidbog om dette emne , men er bekymret for de mulige konsekvenser af den reform , som især sigter mod at decentralisere gennemførelsen af konkurrencereglerne .

Table 8: Examples of reorderings. S is source, B is baseline, and P is the SPTO approach. The elements that have been reordered in the P sentence are marked alike in all sentences. The text in bold has changed place with the text in italics.

approach is significantly superior to previous approaches.

In the future, we plan to apply this approach to English-Arabic translation. We expect greater gains, due to the higher need for reordering between these less-related languages. We also want to examine the relation between word alignment method and the extracted rules and the relationship between reordering and word selection. Finally, a limitation of the current experiments is that they only allow rule-based external reorderings. Since the SPTO scoring is not tied to a source reordering approach, we want to examine the effect of simply adding it as an additional parameter to the baseline PSMT system. This way, all external reorderings are made possible, but only the rule-supported ones get promoted.

References

- Al-Onaizan, Y. and K. Papineni. 2006. Distortion models for statistical machine translation. In *Proceedings of 44th ACL*.
- Buch-Kromann, M., J. Wedekind, and J. Elming. 2007. The Copenhagen Danish-English Dependency Treebank v. 2.0. <http://www.isv.cbs.dk/~mbk/cdt2.0>.
- Callison-Burch, C., C. Fordyce, P. Koehn, C. Monz, and J. Schroeder. 2007. (Meta-) evaluation of machine translation. In *Proceedings of ACL-2007 Workshop on Statistical Machine Translation*.
- Charniak, E. 2000. A maximum-entropy-inspired parser. In *Proceedings of the 1st NAACL*.
- Cohen, W. 1996. Learning trees and rules with set-valued features. In *Proceedings of the 14th AAAI*.
- Collins, M., P. Koehn, and I. Kucerova. 2005. Clause restructuring for statistical machine translation. In *Proceedings of the 43rd ACL*.
- Crego, J. M. and J. B. Mariño. 2007. Syntax-enhanced n-gram-based smt. In *Proceedings of the 11th MT Summit*.
- Habash, N. 2007. Syntactic preprocessing for statistical machine translation. In *Proceedings of the 11th MT Summit*.
- Koehn, P. and C. Monz. 2006. Manual and automatic evaluation of machine translation between european languages. In *Proceedings on the WSMT*.
- Koehn, P., F. J. Och, and D. Marcu. 2003. Statistical phrase-based translation. In *Proceedings of NAACL*.
- Koehn, P. 2004. Pharaoh: a beam search decoder for phrase-based statistical machine translation models. In *Proceedings of AMTA*.
- Koehn, P. 2005. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of MT Summit*.
- Li, C., M. Li, D. Zhang, M. Li, M. Zhou, and Y. Guan. 2007. A probabilistic approach to syntax-based reordering for statistical machine translation. In *Proceedings of the 45th ACL*.
- Stolcke, A. 2002. Srilm – an extensible language modeling toolkit. In *Proceedings of the International Conference on Spoken Language Processing*.
- Wang, C., M. Collins, and P. Koehn. 2007. Chinese syntactic reordering for statistical machine translation. In *Proceedings of EMNLP-CoNLL*.
- Xia, F. and M. McCord. 2004. Improving a statistical mt system with automatically learned rewrite patterns. In *Proceedings of Coling*.
- Zens, R., F. J. Och, and H. Ney. 2002. Phrase-based statistical machine translation. In Jarke, M., J. Koehler, and G. Lake-meyer, editors, *KI - 2002: Advances in Artificial Intelligence. 25. Annual German Conference on AI*. Springer Verlag.
- Zhang, Y., R. Zens, and H. Ney. 2007. Improved chunk-level reordering for statistical machine translation. In *Proceedings of the IWSLT*.