The NiuTrans Machine Translation System for CWMT2011

Tong Xiao, Hao Zhang, Qiang Li, Qi Lu,

Jingbo Zhu, Feiliang Ren and Huizhen Wang

132#, Natural Language Processing Lab

Northeastern University, Shenyang, China 110819

E-mail: {xiaotong, zhujingbo, renfeiliang, wanghuizhen}@mail.neu.edu.cn

{zhanghao, liqiang, luqi}@ics.neu.edu.cn

Abstract: *This paper describes the NiuTrans system developed by the Natural Language Processing Lab at Northeastern University for the Patent Machine Translation Task at NTCIR-9. We present our submissions to the nine tracks of CWMT2011, and show several improvements to our core phrase-based and syntax-based engines, including: an approach to improving searching and modeling for tree-to-tree translation, better language modeling, and more features for phrase-based/syntax-based MT.*

Keywords: *natural language processing, statistical machine translation, phrase-based machine translation, syntax-based machine translation*

1 Introduction

We describe the *NiuTrans* system submitted to the nine tracks of the 7th China Workshop on Machine Translation by the Natural Language Processing Lab at Northeastern University. Our submissions were generated using two core translation systems implemented under the NiuTrans project[1]: 1) a phrase-based MT system (NiuTrans.Phrase) which is built based on our prior work in CWMT2009 (Xiao et al, 2009); 2) and a syntax-based MT system (NiuTrans.Syntax) which employs tree-to-tree approaches for training and decoding. Both the two systems are improved in several ways as will be described in this paper.

Some of our innovations focus on tree-to-tree translation. Since straightforward implementation of tree-to-tree system suffers from too few derivations explored in search space (Chiang, 2010), we focused on developing approaches to enhancing the searching ability for tree-to-tree MT. The basic idea is to use a coarse-grained model (or grammar) to search for more candidates, while use a fine-grained model (or grammar) to ensure accurate scoring. As both the two models can be obtained with simple grammar induction methods (GHKM-like extraction + Maximum Likelihood Estimation), this approach does not burden the rule extraction for our system. More interestingly, with careful thought to implementation, the improved system is able to run as fast as our tree-to-tree baseline. In addition to the improvement on the search problem, we also developed an interpolated translation model which jointly uses the features defined in either the coarse-grained model or the fine-grained model. Together, all these approaches provided a great improvement over the baseline, even outperformed the state-of-the-art phrase-based system on CWMT2011 Chinese-English News evaluation data, which suggests interesting directions for future work.

On language modeling, our improvements are two-fold. First, we develop a simple and effective approach to language model interpolation. By simply using multiple *n*-gram language models with different orders, our system yielded stable BLEU improvements on several tasks. On the development data of English-Chinese News task, it even achieved an absolute improvement of 0.8 BLEU points, and lead to our best ranking in this track. Moreover, we also developed a syntactic language model to improve grammaticality in MT output. This was inspired by our prior work on syntactic language modeling using Tree Substitution Grammars (TSGs) for string-to-tree systems (Xiao et al, 2011). Here we extended our previous work and made it work in the tree-to-tree MT environment.

---

[1] http://www.nlplab.com/NiuPlan/NiuTrans.html

All our improvements resulted in about twenty features for the submitted systems. While some of these features are not new at all, we describe them in detail and hope that the feature engineering work could ease the reproducing of our results. Also, we present the various settings of our systems, list the data sets used, and present automatic results with a summary of our CWMT2011 work.

2 Baseline Systems and Data Sets

2.1 The Phrase-based System (NiuTrans.Phrase)

Our phrase-based system follows the standard framework of phrase-based MT (Koehn et al., 2003) which models the translation process on non-syntactic word sequences instead of unigrams. Particularly, we focused on developing the system based on the Bracketing Transduction Grammar (BTG) (Wu, 1996). Under the BTG scheme, all possible reorderings are compactly represented with binary bracketing constraints, and the decoder can be easily implemented using the CKY parsing algorithm. Due to these advantages, we chose BTG for implementing our phrase-based system (NiuTrans.Phrase). In NiuTrans.Phrase, all phrases are compatible with word alignments and learned using the popular method described in (Koehn et al., 2003). For reordering, two models are involved, including: the maximum entropy-based lexicalized reordering model proposed in (Xiong et al., 2006) and the MSD[2] model proposed in (Koehn et al., 2007; Galley and Manning, 2008). We used a CKY-style decoder with cube pruning and beam search to decode new sentences under the BTG constraint. By default, the beam width was set to 30[3], and the distortion limit was set to 10. In addition to the reordering features, the NiuTrans.Phrase system adopts all standard features used in state-of-the-art SMT systems such as Moses (Koehn et al., 2007). In Section 3, we will give a more detailed description of the features used in our system.

2.2 The Syntax-based System (NiuTrans.Syntax)

The second of our baseline systems is a tree-to-tree system which can be regarded as an instance of the general framework of Synchronous Tree Substitution Grammars (STSGs) (Chiang and Knight, 2006). This type of systems has an assumption that any pair of source-language parse tree and target-language parse tree can be synchronously generated using a derivation of STSG rules (or tree-to-tree translation rules). The translation pipeline is in general divided into two phases – 1) obtaining translation rules and associated probabilities (*grammar induction*); 2) and searching for the "best" translation using the obtained rules (*decoding*). For grammar induction, the NiuTrans.Syntax system learns translation rules from the word-aligned bilingual text whose source and target-sides have been parsed using syntactic parsers. We used a bilingual version of the GHKM (Galley et al., 2004) and SPMT (Marcu et al., 2006) methods which have been successfully applied to tree-to-string/string-to-tree MT. Like in NiuTrans.Phrase, the rule probability is estimated using Maximum Likelihood Estimation. Since too large rules generally generalize poorly on the unseen data, we restricted ourselves to the rules of at most depth 3, and having at most 5 terminals and 4 non-terminals. Aside from this, we discarded all rules that appear less than 2 times to obtain a manageable rule set. For integrating *n*-gram language model into decoding efficiently, rules containing more than two variables or source word sequences were binarized using the synchronous binarization method (Zhang et al., 2006). The decoding was then carried out with the standard CKY parsing algorithm.

2.3 Data Preparation

The recourses we used were constrained to those provided for CWMT2011. All bilingual and monolingual data was pre-processed in a similar scenario.

---

[2] The term MSD comes from the three operations defined in the model: Monotone (M), Swap (S), and Discontinuous (D).

[3] In our implementation, beam width refers to the number edges (derivations) that are retained in each cell. This definition is different from that used in other systems such as the Chart-decoder in Moses, where beam width is defined to be the number of all edges that are accessed for a given chart cell.

- Chinese and Japanese sentences were segmented using the NEUNLPLab Chinese segmentation system[4] and the MeCab system[5], respectively. For processing English sentences, a rule-based English tokenizer was employed, and the case information was removed. As CWMT2011 provided the segmentation/tokenization result for other languages, no additional processing was performed on them.
- For Chinese-English/English-Chinese News tracks, all number/date/time entities were generalized to be unique symbols in order to alleviate data sparseness. These entities were then translated using an additional rule-based translation engine when we decoded test sentences. To achieve state-of-the-art performance, we further advanced the translation for person names (PER), organization names (ORG) and location names (LOC). We used three specialized translation modules to translate PER, ORG and LOC respectively, and then inserted their translations into the MT systems.
- To filter out noisy data, all sentence-pairs were cleaned up in terms of target-length/source-length ratio (ranges between 0.2 and 5.0).
- Bi-directional word alignments were performed on the bilingual sentences with GIZA++[6]. We also refined the alignments with the "grow-diag-final-and" method for obtaining symmetric word alignment.
- Parse trees on both the Chinese and English-sides were generated using the Berkeley parser (Petrov et al., 2006). To improve the phrasal coverage of transformation rules, all parse trees were binarized in a head-out manner.
- The *n*-gram language model was trained using our self-developed LM trainer which had been integrated in the open-source version of the NiuTrans system. By default, we used 5-gram language models for all the tracks.
- To recover the case information, we used the recaser in Moses SMT toolkit[7] which is based on heuristic rules and HMM models.

Table 1 shows the statistics of the training data used in different tracks. For weight tuning, we run a standard MERT program (Och, 2003) on the development data provided for each track. We also used 863-2005-writ and SSMT2007 evaluation data to perform held-out test for Chinese-English/English-Chinese News translation.

3 Improvements

3.1 Improving Searching Ability for Tree-to-tree MT

String-to-tree/tree-to-string models have been extensively investigated in Statistical Machine Translation (SMT) over the last few years. Tree-to-tree translation is no-doubt a promising next-step due to its use of the fine-grained syntactic information on both language sides. However, performance of tree-to-tree systems is still unsatisfactory on large-scale, well-established translation tasks. To our knowledge, it is rare to see that "tree-to-tree" participants could outperform their phrase-based or string-to-tree/tree-to-string counterparts in the CWMT-series competitions. In this section we present such a system.

We focus on improving the searching ability for tree-to-tree systems. In general, tree-to-tree systems suffer from their small search space (or derivation space) due to the syntactic constraints on both language sides and the structure divergence between languages. This has been recognized as one of the major problems with the tree-to-tree models. Some research group has explored the fuzzy method to alleviate this problem (Chiang, 2010). In our work we instead develop a simpler but more effective solution.

---

4 http://www.nlplab.com/
5 http://mecab.sourceforge.net/
6 http://code.google.com/p/giza-pp/
7 http://www.statmt.org/moses/

Table 1. Data sets used

Data Set	# sentences	# words (src / tgt)	*zh-en news*	*en-zh news*	*en-zh scie*	*jp-zh news*	*mn-zh dail*	*ti-zh gove*	*uy-zh news*	*ka-zh news*	*ki-zh news*
CLDC-LAC-2003-004	221K	2.6M/2.9M	√	√							
PKU C-E	195K	2.5M/2.9M	√	√							
XMU E-C Movie	169K	1.0M/1.1M	√	√							
HIT-IR E-C	99K	1.1M/1.1M	√	√							
HIT-MT E-C	49K	0.6M/0.5M	√	√							
Datum E-C	700K	19M/17M	√	√							
ICT Web C-E	2530K	47M/51M	√	√							
NEU C-E	1000K	12M/14M	√	√							
ISTIC E-C (2011)	818K	19M/18M			√						
NJU J-C	151K	2.7M/2.0M				√					
DUT J-C	111K	1.3M/1.0M				√					
PKU C-J	20K	3.6M/2.5M				√					
IMU C-MN	67K	0.8M/0.8M					√				
QHNU TI-C	102K	1.3M/1.0M						√			
XJU UY-C	50K	1.1M/1.1M							√		
XJU KA-C	50K	1.0M/1.0M								√	
XJU KI-C	50K	1.2M/1.0M									√
Reuters corpus (E)	10.7M	232M	√								
SogouCA (C)	19.4M	560M		√	√	√	√	√	√	√	√

For practical syntax-based MT systems, the translation problem can be stated as:

$$t^* = \arg\max_{t \in L(G)} \Pr_G(t \mid s)$$

where s is the (input) source string, t is the (output) target string, G is the grammar (or model) induced from the bilingual data, $L(G)$ is the target projection of the language for G. In the above formulation, output string t is scored using model $\Pr_G(t \mid s)$, but constrained in the string space of $L(G)$. Obviously this approach can be improved in two ways: 1) developing a better model ($\Pr_G(t \mid s)$) such as introducing more syntax/semantics into MT; and 2) enlarging the search space ($L(G)$). Here we focus on the latter one. The basic idea is to use a coarse-grained grammar *G-coarse* to do searching, while use the fine-grained grammar *G-fine* for modeling. Hence we reach a new formulation which is slightly different from the above one.

$$t^* = \arg\max_{t \in L(G\text{-}coarse)} \Pr_{G\text{-}fine}(t \mid s)$$

Obtaining the coarse-grained grammar is very trivial in our case: we remove the syntactic label and syntactic structure in tree-to-tree rules. For example, the following is a tree-to-tree transfer rule,

VP(VV(来自) $\mathbf{NP}_0$) → VP(VBG(coming) IN(from) $\mathbf{NP}_0$)

If the syntactic label and structure is removed from it, we have a new rule:

X(来自 $\mathbf{X}_0$) → X(coming from $\mathbf{X}_0$)

Obviously, it is a standard hierarchical phrase-based (Hiero) rule. Similarly, we can induce string-to-tree/tree-to-string rules from a given tree-to-tree grammar. Beyond this, the coarse-grained grammar can also be extracted from the bilingual data. Suppose that Hiero is chosen to be the coarse-grained model. We can straightforwardly obtain a default Hiero rule set using the Hiero rule extraction algorithm (Chiang, 2007). Of course, this newly-obtained rule set can enhance the coarse-grained grammar that is generated from the tree-to-tree grammar.

The implementation of decoder is straightforward within our approach: derivations are first generated using the coarse-grained model, and then scored using the fine-grained (tree-to-tree) model[8]. For unseen rules in fine-grained scoring, we assign them associated probabilities using a deleted interpolation from the coarse-grained model.

In addition to enlarging search space, the coarse-grained model also provides us more reliable estimates to the low-frequency events (rules). It is natural to make use of both the coarse-grained model and the fine-grained model to further advance our system. The solution is very simple: we jointly use them in decoding. In other words, we put together all the features of the two models, and combine them in a log-linear fashion (Och and Ney, 2002). The optimization can be trivially done by re-using the MERT component of our system.

Table 2 shows a comparison of our approach with other baseline systems on the development and held-out test sets of the Chinese-English news translation task. For comparison, we also report the BLEU score of an in-house re-implementation of the fuzzy tree-to-tree system (Chiang, 2010). To show a fair comparison, both the fuzzy tree-to-tree system and our improved system choose the string-to-tree grammar as the coarse-grained grammar. It is observed that the improved system stately outperforms the syntax-based baselines (i.e. naïve tree-to-tree, string-to-tree and fuzzy tree-to-tree) on the two test sets. It even achieves the best BLEU score on the test set of 2007, which outperforms the phrase-based baseline over 0.2 BLEU points.

Table 2. BLEU scores of various systems on C-E news track

System	**BLEU-SBP4** (Dev)	**BLEU-SBP4** (Test 2005)	**BLEU-SBP4** (Test 2007)
Phrase-based	**0.2889**	**0.2524**	0.2806
Naïve t2t	0.2717	0.2378	0.2655
Naïve s2t	0.2866	0.2470	0.2818
Fuzzy t2t	0.2841	0.2471	0.2780
Ours (t2t)	0.2857	0.2483	**0.2828**

3.2 Better Language Modeling

Our first improvement in language modeling is to interpolate multiple *n*-gram languages for MT decoding. This method can be regarded as an interpolation of the language models with different orders. Though very simple, it is effective in improving BLEU scores for some tasks. Table 3 shows the results on English-Chinese News translation task. The interpolated *n*-gram language model (ranging from bi-gram to 5-gram) outperforms the baseline about 0.2 BLEU points on all three of the data sets. On the development set, the improvement is even over 0.8 BLEU points.

Table 3. BLEU score of interpolated *n*-gram LM on E-C news track

System	**BLEU-SBP5** (Dev)	**BLEU-SBP5** (Test 2005)	**BLEU-SBP5** (Test 2007)
Phrase-based Baseline	0.3293	0.3826	0.3553
+ bi~4-gram LMs	**0.3382**	**0.3844**	**0.3572**

Moreover, we also develop a syntactic language model for our tree-to-tree system. This is inspired by our prior work (Xiao et al., 2011) which designed a TSG language model to evaluate the well-formedness of tree-output for string-to-tree systems. Here we extend our previous work in the tree-to-tree system. In the case of tree-to-tree MT, we first obtain a (coarse-grained) string-to-tree grammar using the method described in Section 3.1. We then train a TSG language model using the target-side of the string-to-tree grammar, as what is done in (Xiao et al., 2011). The resulting TSG-based LM is obviously applicable to the tree-to-tree system. We evaluate our syntactic language model on the Chinese-English News translation task (Table 4). The BLEU scores show that the TSG-based LM is helpful in improving our tree-to-tree system.

[8] There are several ways to do this. E.g. we can first generate *k*-best/forest translation candidates using the coarse-grained model, and then re-score and re-rank them using the fine-grained model. Also, we can intersect the coarse-grained searching with the fine-grained modeling in decoding. In our implementation we choose the second way for the efficiency consideration, though it requires more code changes.

Table 4. BLEU score of syntactic LM on C-E news track

System	BLEU-SBP4 (Dev)	BLEU-SBP4 (Test 2005)	BLEU-SBP4 (Test 2007)
Improved t2t Baseline	0.2857	**0.2483**	0.2828
+ TSG-based LM	**0.2875**	0.2482	**0.2842**

3.3 More Features

About twenty features are selected for generating our final submissions. We list them in Table 5. Although some of them are not new at all, we hope our description is helpful to anyone who'd like to replicate our results.

Table 5. Features Used in NiuTrans.Phrase and NiuTrans.Syntax for CWMT2011

ID	Phrase-based		Syntax-based (tree-to-tree)	
	Feature	Description	Feature	Description
1	$\Pr(t \mid s)$	Phrase trans-probability	$\Pr(t \mid s)$	Phrase trans-probability
2	$\Pr_{lex}(t \mid s)$	Lexical weight	$\Pr_{lex}(t \mid s)$	Lexical weight
3	$\Pr(s \mid t)$	Inverted $\Pr(t \mid s)$	$\Pr(s \mid t)$	Inverted $\Pr(t \mid s)$
4	$\Pr_{lex}(s \mid t)$	Inverted $\Pr_{lex}(t \mid s)$	$\Pr_{lex}(s \mid t)$	Inverted $\Pr_{lex}(t \mid s)$
5	$\Pr_{\mathrm{LM2}}(t)$	Bi-gram language model	$\Pr_{\mathrm{LM5}}(t)$	5-gram language model
6	$\Pr_{\mathrm{LM3}}(t)$	Tri-gram language model	Length(t)	# of target words
7	$\Pr_{\mathrm{LM4}}(t)$	4-gram language model	Count(R)	# of rules
8	$\Pr_{\mathrm{LM5}}(t)$	5-gram language model	WD	# of word deletions
9	Length(t)	# of target words	Bi-Lex	# of bi-lex links
10	Count(Phr)	# of phrases	$\Pr_{root}(r)$	root normalized probability
11	WD	# of word deletions	$\Pr_{tsg\text{-}lm}(r)$	TSG-based LM
12	Bi-Lex	# of bi-lex links	Count(Phr-R)	# of phrasal rules
13	$f_{BTG\text{-}ME}$	ME-based reordering feature	Count(Glue-R)	# of glue rules
14	$f_{M\text{-}previous}$	M orientation (previous)	Count(Lex-R)	# of lexical rules
15	$f_{S\text{-}previous}$	S orientation (previous)	Count(Low-Freq)	# of low-frequency rules
16	$f_{D\text{-}previous}$	D orientation (previous)	Count(Comp-R)	# of composed rules
17	$f_{M\text{-}following}$	M orientation (following)	Coarse $\Pr(t \mid s)$	Coarse-grained model
18	$f_{S\text{-}following}$	S orientation (following)	Coarse $\Pr_{lex}(t \mid s)$	Coarse-grained model
19	$f_{D\text{-}following}$	D orientation (following)	Coarse $\Pr(s \mid t)$	Coarse-grained model
20	N/A		Coarse $\Pr_{lex}(s \mid t)$	Coarse-grained model
21	N/A		Coarse $\Pr_{root}(r)$	Coarse-grained model

4 Results

After system update, we also tried various settings on our systems, such as enlarging beam width, smoothing translation tables and filtering reordering samples with source-side syntactic structure. To try these, we ran MERT over 50 times for the tasks we participated in. Then we selected the best performing systems on the development sets and submitted them for the final evaluation in CWMT2011. For all nine of the CWMT2011 tracks, we chose NiuTrans.Phrase as the primary system. We also submitted NiuTrans.Syntax to the Chinese-English and English-Chinese news tracks for comparison. For the Chinese-English news track, we chose string-to-tree model as the coarse-grained model for our improved tree-to-tree system; For the English-Chinese news track, we chose tree-to-string model as the coarse-grained model.

The evaluation results of our submissions are listed in Table 6. We see, first of all, that our tree-to-tree system achieved very promising result on the Chinese-English News translation task, even outperformed its phrase-based counterpart in such a large-scaled and noisy training environment. Although our tree-to-tree system still has a lower BLEU score on English-Chinese News translation task than NiuTrans.Phrase, it suggests interesting future directions due to the so many ways to improve upon the current approach. Also, our phrase-based system shows its robustness in most cases. In seven of the nine tracks, it achieved a BLEU score of over 0.30. It is a little bit amazing that the SMT systems could achieve good BLEU results on Minority language-Chinese translation tasks where only about 100K sentences were provided for training. This might be due to the strong correlation and consistency between the training data and the test data in these tasks.

We also compared our systems with others submitted systems to the same tracks[9]. We found that the Chinese-English/English-Chinese tasks were quite competitive and several systems were within a BLEU score of the top. This is a good situation for Chinese-English/English-Chinese translation and encourages more improvements on these tasks. Among all the participated systems, our (primary) system was ranked at 4, 2, 1 on the three evaluation sets[10], respectively. For other tasks, there seems a gap between our submitted systems and the top ones. This we attribute to the lack of character-based *n*-gram LM in our systems[11]. Since all these tasks share a property that the training data is relatively small (less than one million sentence pairs), it is expected that our system can be further improved with the use of large-scale character-based *n*-gram LM.

Table 6. Results of NiuTrans.Phrase and NiuTrans.Syntax in CWMT2011

Track	Entry	BLEU-SBP (4/5)	BLEU (4/5)	NIST (5/6)
zh-en news (progress)	Primary (*NiuTrans.Phrase*)	0.2245	0.2407	7.6715
	Contrast (*NiuTrans.Syntax*)	0.2276	0.2465	7.8043
en-zh news (current)	Primary (*NiuTrans.Phrase*)	0.3514	0.3701	10.090
	Contrast (*NiuTrans.Syntax*)	0.3366	0.3491	9.4846
en-zh news (progress)	Primary (*NiuTrans.Phrase*)	0.3357	0.3490	9.6791
	Contrast (*NiuTrans.Syntax*)	0.3239	0.3329	9.3083
en-zh scie	Primary (*NiuTrans.Phrase*)	0.3778	0.3934	10.374
jp-zh news	Primary (*NiuTrans.Phrase*)	0.4305	0.4438	10.661
	Contrast 1(*NiuTrans.Phrase*)	0.4279	0.4415	10.618
	Contrast 2(*NiuTrans.Phrase*)	0.4295	0.4432	10.660
mn-zh dail	Primary (*NiuTrans.Phrase*)	0.1725	0.1923	5.3735
	Contrast (*NiuTrans.Phrase*)	0.1962	0.2211	5.8040
ti-zh gove	Primary (*NiuTrans.Phrase*)	0.5174	0.5446	9.9268
uy-zh news	Primary (*NiuTrans.Phrase*)	0.4233	0.4572	10.059
ka-zh news	Primary (*NiuTrans.Phrase*)	0.3651	0.3919	9.1054
ki-zh news	Primary (*NiuTrans.Phrase*)	0.4130	0.4407	9.6019

4 Conclusion

We participated in all the nine tracks of CWMT 2011 Chinese-Foreign and Foreign-Chinese translation. Nonetheless, we show improvements on several tracks, including a win by BLEU-SBP on the English-Chinese News track. Most of them attribute to our recent research work on MT, including: an improved tree-to-tree MT approach, better language modeling, and feature engineering for phrase-based and syntax-based MT. Moreover, the results on small-scale tasks suggest a potential improvement with the use of character-based LM. Our phrase-based system (*NiuTrans.Phrase*) has been released under the open-source project *NiuTrans*, and our syntax-based system (*NiuTrans.Syntax*) will be open-source available in a few months. We hope that our code will be helpful when these results are replicated and brought to bear on related problems.

5 Acknowledgements

This work was supported in part by the National Science Foundation of China (60873091;

9 Thanks to CWMT2011 organizers for their early-released official result.

10 The three evaluation sets are: *zh-en news progress*, *en-zh news progress* and *en-zh new current*.

11 When target language is Chinese, word segmentation also affects language modeling. Character-based *n*-gram LM is an extreme case for segmentation. It is in general adopted to alleviate data sparseness, and maybe more helpful when the evaluation metric is also character-based.

61073140), Specialized Research Fund for the Doctoral Program of Higher Education (20100042110031), the Fundamental Research Funds for the Central Universities and Natural Science Foundation of Liaoning Province of China.

References

David Chiang and Kevin Knight. 2006. An introduction to synchronous grammars. In tutorial of ACL 2006.

David Chiang. 2007. Hierarchical Phrase-Based Translation. Computational Linguistics, 33(2):201-228.

David Chiang. 2010. Learning to translate with source and target syntax. In *Proc. of* ACL 2010, Uppsala, Sweden, pages 1443-1452.

Michel Galley and Christopher D. Manning. 2008. A Simple and Effective Hierarchical Phrase Reordering Model. In *Proc. of* EMNLP2008, Hawaii, pages: 848-856.

Philipp Koehn, Franz Josef Och and Daniel Marcu. 2003. Statistical Phrase-Based Translation. In *Proc. of* HLT/NAACL 2003, USA, pages 127-133.

Philipp Koehn, Hieu Hoang, Alexandra Birch Mayne, Christopher Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In ACL 2007 demonstration session.

Franz Och and Hermann Ney. 2002. Discriminative Training and Maximum Entropy Models for Statistical Machine Translation. In *Proc. of* ACL2002, Philadelphia, pages: 295-302.

Franz Och. 2003. Minimum Error Rate Training in Statistical Machine Translation. In *Proc. of* ACL2003, Japan, pages: 160-167.

Dekai Wu. 1996. A polynomial-time algorithm for statistical machine translation. In *Proc. of* ACL1996, California, USA, pages 152-158.

Tong Xiao, Rushan Chen, Tianning Li, Muhua Zhu, Jingbo Zhu, Huizhen Wang and Feiliang Ren. 2009. NEUTrans: a Phrase-Based SMT System for CWMT2009. In *Proc. of* the 5th China workshop on Machine Translation (CWMT), 2009, China, pages 40-46.

Tong Xiao, Jingbo Zhu and Muhua Zhu. 2011. Language Modeling for Syntax-based Machine Translation Using Tree Substitution Grammars: A Case Study on Chinese-English Translation. To appear in ACM Transactions on Asian Language Information Processing.

Deyi Xiong, Qun Liu and Shouxun Lin. 2006. Maximum Entropy Based Phrase Reordering Model for Statistical Machine Translation. In *Proc. of* ACL 2006, Sydney, pages 521-528.