Chained Machine Translation Using Morphemes as Pivot Language

Wen Li
Institute of Intelligent Machines, Chinese Academy of Sciences, University of Science and Technology of China
xtliwen@mail.ustc.edu.cn

Lei Chen
Institute of Intelligent Machines, Chinese Academy of Sciences
alan.cl@163.com

Wudabala
Institute of Intelligent Machines, Chinese Academy of Sciences
hwdbl@126.com

Miao Li
Institute of Intelligent Machines, Chinese Academy of Sciences
mli@iim.ac.cn

Abstract

As the smallest meaning-bearing elements of the languages which have rich morphology information, morphemes are often integrated into state-of-the-art statistical machine translation to improve translation quality. The paper proposes an approach which novelly uses morphemes as pivot language in a chained machine translation system. A machine translation based method is used therein to find the mapping relations between morphemes and words. Experiments show the effectiveness of our approach, achieving 18.6 percent increase in BLEU score over the baseline phrase-based machine translation system.

1 Introduction

Recently, most evaluations of machine translation systems (Callison-Burch et al., 2009) indicate that the performance of corpus-based statistical machine translation (SMT) has come up to the traditional rule-based method. In the corpus-based SMT, it is difficult to exactly select the correct inflections (word-endings) if the target language is highly inflected. This problem will be more severe if the source language is an isolated language with non-morphology (eg. Chinese) and the target language is an agglutinative language with productive derivational and inflectional morphology (eg. Mongolian: a minority language of China). In addition, the lack of large-scale parallel corpus may cause the sparse data problem, which will be more severe if one of the source language and the target language is highly inflected. As the smallest meaning-bearing elements of the languages which have rich morphology information, morphemes are the compact representation of words. Using morphemes as the semantic units in the parallel corpus can not only help choose the correct inflections, but also alleviate the data sparseness problem partially.

Many strategies of integrating morphology information into state-of-the-art SMT systems in different stages have been proposed. (Ramanathan et al., 2009) proposed a preprocessing approach for incorporating syntactic and Morphological information within a phrase-based English-Hindi SMT system. (Watanabe et al., 2006) proposed a method which uses Porter stems and even 4-letter prefixes for word alignment. (Koehn et al., 2007) proposed the factored translation models which combine feature functions to handle syntactic, morphological, and other linguistic information in a log-linear model during training. (Minkov et al., 2007) made use of the information of morphological structure and source language in postprocessing to improve SMT quality. (de Gispert et al., 2009) adopted the Minimum Bayes Risk decoding strategy to combine output from identical SMT system, which is trained on alternative morphological decompositions of the source language.

Meanwhile, the SMT-based methods are widely used in the area of natural language processing. (Quirk et al., 2004) applied SMT to generate novel paraphrases. (Riezler et al., 2007) adopted an SMT-based

Proceedings of the 8th Workshop on Asian Language Resources, pages 169–177,
Beijing, China, 21-22 August 2010.

method to query expansion in answer retrieval. (Jiang and Zhou, 2008) used SMT to generate the second sentence of the Chinese couplets.

As opposed to the above strategies, the paper proposes an approach that uses morphemes as pivot language in a chained SMT system, for translating Chinese into Mongolian, which consists of two SMT systems. First, Chinese sentences are translated into Mongolian morphemes instead of Mongolian words in the Chinese-Morphemes SMT (SMT_1). Then Mongolian words are generated from morphemes in the Morphemes-Mongolian SMT (SMT_2). The essential part of the chained SMT system is how to find the mapping relations between the morphemes and words, which is considered as a procedure of machine translation in our approach. More concretely, the first challenge of this approach is to investigate some effective strategies to segment the Mongolian corpus in the Chinese-Mongolian parallel corpus. And the second challenge is how to efficiently generate Mongolian words from morphemes. Additionally, on the one hand Mongolian words may have multiple kinds of morphological segmentations. On the other hand there is also the ambiguity of word boundaries in the processing of generating Mongolian words from morphemes. In order to solve these ambiguities, a SMT-based method is applied in that word context and morphemes context can be taken into account in this method.

The remainder of the paper is organized as follows. Section 2 introduces two methods of morphological segmentation. Section 3 presents the details of chained SMT system. Section 4 describes the experiment results and evaluation. Section 5 gives concluding remarks.

2 Morphological segmentation

Mongolian is a highly agglutinative language with a rich set of affixes. Mongolian contains about 30,000 stems, 297 distinct affixes. A big growth in the number of possible word forms may occur due to the inflectional and derivational productions. An inflectional suffix is a terminal affix that does not change the parts of speech of the root during concatenation, which is added to maintain the syntactic environment of the root. For instance, the Mongolian word "YABVGSAN" (walking) in the present continuous tense syntactic environment consists of the root "YABV" (walk) and the suffix "GSAN" (ing). Whereas, when a verb root "UILED" (do) concatenates a noun derivational suffix "BURI", it changes to a noun "UILEDBURI" (factory). According to that whether linguistic lemmatization (the reduction to base form) is considered or not, the paper proposes two methods of morphological segmentation. The two methods are tested on the same training databases.

The root lemmatization is concerned in the first method, which is called the SMT-based morphological segmentation (SMT-MS) in this paper. Given the Mongolian-morphemes parallel corpus, this method trains a Mongolian-morphemes SMT to segment Mongolian words. The root lemmatization is considered in the original morphological pre-segmented training corpus. So the SMT-based method can also deal with root lemmatization when it segments a Mongolian word. For instance, the Mongolian word "BAYIG_A" exhibits the change of spelling during the concatenation of the morphemes "BAI" and "G_A". We also investigate whether it is effective if those roots are identical to the original word forms. In other words, the root lemmatization is ignored in the second method, which takes the gold standard morphological segmentation corpus as a trained model of Morfessor (Creutz and Lagus, 2007) and uses the Viterbi decoding algorithm to segment new words. Therefore, this method is called the Morfessor-based morphological segmentation (Mor-MS). For instance, the word "BAYIG_A" will be segmented to "BAYI" and "G_A" instead of "BAI" and "G_A".

The mathematical description of SMT-MS is the same as the traditional machine translation system. In the Mor-MS method, the morphological segmentation of a word can be regarded as a flat tree (morphological segmentation tree), where the root node corresponds to the whole word and the leaves correspond to morphemes of this word. Figure 1 gives an ex-

ample. First, the joint probabilistic distribution (Creutz and Lagus, 2007) of all morphemes in the morphological segmentation tree are calculated. And then by using the Viterbi decoding algorithm, the maximum probability segmentation combination is selected.

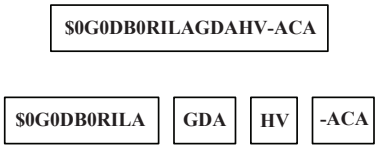

Figure 1: Morphological segmentation tree

3 Chained SMT system

3.1 Overview

In order to improve the performance of Chinese-Mongolian machine translation, the paper proposes an approach which incorporates the morphology information within a chained SMT system. More concretely, this system first translates Chinese into Mongolian morphemes instead of Mongolian words by the Chinese-Morphemes SMT. And then it uses the Morphemes-Mongolian SMT to translate Mongolian morphemes into Mongolian words. Namely, morphemes are regarded as pivot language in this system.

The chained SMT system consists of a morphological segmentation system and two phrase-base machine translation systems, which are given as follows:

- Morphological segmentation: segmenting Mongolian words (from the Chinese-Mongolian parallel corpus) into Mongolian morphemes and obtaining two parallel corpus: Chinese-Morphemes parallel corpus and Morphemes-Mongolian parallel corpus.
- SMT_1: training the Chinese-Morphemes SMT on the Chinese-Morphemes parallel corpus.
- SMT_2: training the Morphemes-Mongolian SMT on the Morphemes-Mongolian parallel corpus.

Figure 2 illustrates the overview of chained SMT system.

3.2 Phrase-based SMT

The authors assume the reader to be familiar with current approaches to machine translation, so that we briefly introduce the phrase-based statistical machine translation model (Koehn et al., 2003) here, which is the foundation of chained SMT system.

In statistical machine translation, given a source language f, the aim is to seek a target language e, such that $P(e|f)$ is maximized. The phrase-based translation model can be expressed by the following formula:

$$e^* = \arg\max_e P(e|f) = \arg\max_e \{P(f|e)P(e)\}$$

where e^* indicates the best result, $P(e)$ is the language model and $P(f|e)$ is the translation model. According to the standard log-linear model proposed by (Och and Ney, 2002), the best result e^* that maximizes $P(e|f)$ can be expressed as follows:

$$e^* = \arg\max_e \{\sum_{m=1}^{M} \lambda_m h_m(e, f)\}$$

where M is the number of feature functions, λ_m is the corresponding feature weight, each $h_m(e, f)$ is a feature function.

In our chained SMT system, SMT_1, SMT_2 and the SMT for morphological segmentation (namely SMT-MS in Section 2) are all phrase-based SMTs.

3.3 Features of Chained SMT system

As shown in Figure 2, Chinese is translated into Mongolian morphemes in SMT_1, which is the core part of the chained SMT system. Here morphemes are regarded as words. Therefore, morphemes can play important roles in SMT_1 as follows: the roots present the meaning of the word and the suffixes help select the correct grammatical environment. The word alignments between Chinese words and Mongolian morphemes are learned automatically by GIZA++. Figure 3 gives an instance of word alignment in SMT_1.

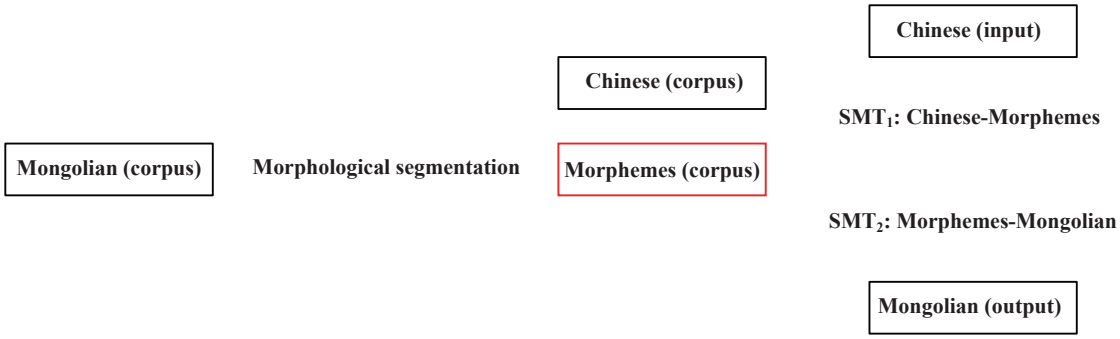

Figure 2: Morphemes as pivot language in Chained SMT system

We can see that the morphemes "BI","TAN" etc. are all regarded as words.

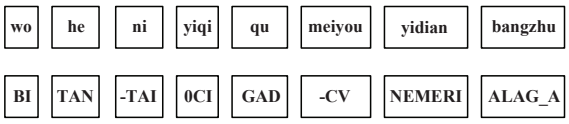

Figure 3: Word alignments between Chinese words and Mongolian morphemes in SMT_1

All the most commonly used features of standard phrase-based SMT, including phrase translation model, language model, distortion model and word penalty, are selected in SMT_1. These commonly used features determine the quality of translation together. The phrases of f and e are ensured to be good translations of each other in the phrase translation model $P(f|e)$. The fluent output is guaranteed in the language model $LM(e)$. The reordering of the input sentence is allowed in the distortion model $D(e, f)$. The translation is however more expensive with the more reordering. The translation results are guaranteed neither too long nor too short in the word penalty $W(e)$.

In SMT-MS and SMT_2, the task is to find the mapping relations between Mongolian morphemes and Mongolian words, which is considered as the word-for-word translation. Therefore, only phrase translation model and language model are considered. All the features weights are uniform distribution by default. Mongolian words may have multiple kinds of morphological segmentations. And there is the ambiguity of word boundaries in the processing of generating Mongolian words from morphemes. These ambiguities can be solved in SMT-MS and SMT_2 respectively, since the SMT-based method can endure mapping errors and solve mapping ambiguities by the multiple features which can consider the context of Mongolian words.

4 Experiments

4.1 Experimental setup

In the experiments, first we preprocess the corpus, such as converting Mongolian into Latin Mongolian and filtering the apparent noisy segmentation of the gold standard morphological segmentation corpus. And then we evaluate the effectiveness of the SMTs which find the mapping relations between the morphemes and their corresponding word forms. Namely, SMT-MS and SMT_2. As mentioned above, SMT_1 is the core part of the chained SMT system, which decides the final quality of translation results. So the evaluation of SMT_1 can be reflected by the evaluation of translation results of whole chained SMT system. Finally, we evaluate and analyze the performance of the chained SMT system by using the automatic evaluation tools.

The translation model consists of a standard phrase-table with lexicalized reordering. Bidirectional word alignments obtained with GIZA++ are intersected using the grow-diag-final heuristic (Koehn et al., 2003). Translations of phrases of up to 7 words long are collected and scored with translation probabilities and lexical weighting. The language model of morphemes is a 5-gram model with Kneser-Ney

smoothing. The language model of Mongolian word is 3-gram model with Kneser-Ney smoothing too. All the language models are built with the SRI language modeling toolkit (Stolcke, 2002). The log-linear model feature weights are learned by using minimum error rate training (MERT) (Och, 2003) with BLEU score (Papineni et al., 2002) as the objective function.

4.2 Corpus preprocessing

The Chinese-Mongolian parallel corpus and monolingual sentences are obtain from the 5th China Workshop on Machine Translation. In the experiments, we first convert Mongolian corpus into Latin Mongolian. In morphological segmentation, the gold standard morphological segmentation corpus contains 38000 Mongolian sentences, which are produced semi-automatically by using the morphological analyzer Darhan (Nashunwukoutu, 1997) of Inner Mongolia University. Moreover, in order to obtain the higher quality corpus, most of the wrong segmentation in the results of morphological analyzer are modified manually by the linguistic experts. However, there are still some wrong segmentation in the gold standard corpus. Therefore, we adopt a strategy to filter the apparent noisy segmentation. In this strategy, the sum of the lengths of all the morphemes is required to be equivalent to the length of the original word. After filtering, there are still 37967 sentences remained. In addition, the word alignment is vulnerable to punctuation in SMT-MS. So all punctuation of the gold standard morphological segmentation corpus are removed to eliminate some mistakes of the word alignment.

Meanwhile, since the Chinese language does not have explicit word boundaries, we also need to do the segmentation of Chinese words. The word segmentation tool ICTCLAS (Zhang, 2008) is used in the experiments.

4.3 Evaluation of SMT-MS and SMT_2

The tasks of SMT-MS and SMT_2 are to find the mapping relations between the morphemes and their corresponding word forms. Morphological segmentation is done by SMT-MS. Contrarily, SMT_2 is used to generate the words from morphemes. To evaluate the effectiveness of SMT-MS and SMT_2, we divide the filtered gold standard corpus into two sets for training (90%) and testing (10%) respectively. The correct morpheme boundaries are counted for SMT-MS evaluation, while the correct word boundaries are counted for SMT_2 evaluation. We use the two measures *precision* and *recall* on discovered word boundaries to evaluate the effectiveness of SMT-MS and SMT_2, where *precision* is the proportion of correctly discovered boundaries among all discovered boundaries by the algorithm, and *recall* is the proportion of correctly discovered boundaries among all correct boundaries. A high *precision* indicates that a morpheme boundary is probably correct when it is suggested. However the proportion of missed boundaries can not be obtained from it. A high *recall* indicates that most of the desired boundaries were indeed discovered. However it can not point out how many incorrect boundaries were suggested either. In order to get a comprehensive idea, we also make use of the evaluation method: *F-measure* as a compromise.

$$\textit{F-measure} = \frac{1}{\frac{1}{2}(\frac{1}{\textit{precision}} + \frac{1}{\textit{recall}})}$$

These measures assume values between zero and 100%, where high values reflect good performance. Therefore, we evaluate the SMT-based methods by incrementally evaluating the features used in our phrase-based SMT model.

Table 1 gives the evaluation results, where PTM denotes Phrase Translation Model, LW denotes Lexical Weight, LM denotes Language Model, IPTM denotes Inverted PTM, ILW denotes Inverted LW. Table 1(a) and Table 1(b) are corresponding to the evaluations of SMT-MS and SMT_2 respectively, where P, R and F denote the three measures, namely *precision*, *recall* and *F-measure*.

The results show that when we add more features incrementally, the *precision*, *recall* and *F-measure* are improved consistently. These indicate that the features are helpful for finding the mapping relations between morphemes and Mongolian words.

Table 1: Evaluation of SMT-MS and SMT_2

(a) Evaluation of SMT-MS

Feature	$P(\%)$	$R(\%)$	$F(\%)$
(1): PTM+LW	73.35	72.45	72.90
(2): (1)+LM	94.91	94.91	94.91
(3): (2)+IPTM+ILW	94.95	94.95	94.95

(b) Evaluation of SMT_2

Feature	$P(\%)$	$R(\%)$	$F(\%)$
(1): PTM+LW	75.86	60.04	67.03
(2): (1)+LM	95.13	89.92	92.45
(3): (2)+IPTM+ILW	95.13	90.02	92.51

Table 2: Evaluation of systems

(a) without MERT

	NIST	BLEU (%)
Baseline	5.3586	20.71
$Chain_1$	5.6471	23.91
$Chain_2$	5.6565	24.57

(b) with MERT

	NIST	BLEU (%)
Baseline	5.6911	24.13
$Chain_1$	5.7439	24.70
$Chain_2$	5.8401	25.80

4.4 Evaluation of chained SMT system

We use NIST score (Doddington, 2002) and BLEU score (Papineni et al., 2002) to evaluate chained SMT system. The training set contains 67288 Chinese-Mongolian parallel sentences. The test set contains 400 sentences, where each sentence has four reference sentences which are translated by native experts.

In the training phase, we convert Mongolian into Latin Mongolian. And while in the test phase, we convert the Latin Mongolian back into the traditional Mongolian words. We compare the chained SMT system with the standard phrase-based SMT. Table 2 gives the evaluation of experiment result of each system, where Baseline is the standard phrase-based SMT, $Chain_1$ is a chained SMT consisting of SMT-MS, SMT_1 and SMT_2, $Chain_2$ is also a chained SMT consisting of Mor-MS, SMT_1 and SMT_2. In Table 2(b), we use MERT to train the feature weights of the baseline system and the feature weights of SMT_1 in $Chain_1$ and $Chain_2$.

The experiment results show that both $Chain_1$ and $Chain_2$ are much better than the baseline system. The BLEU score is improved by 18.6 percent, from 20.71 (Baseline) to 24.57 ($Chain_2$). In addition, $Chain_2$ is better than $Chain_1$. We believe that it is essentially related to the different morphemes corpus of $Chain_1$ and $Chain_2$. The morphemes corpus of $Chain_1$ takes lemmatization into account, while the morphemes corpus of $Chain_2$ changes all morphemes to inflected forms which are identical to the original word forms. As the example in Section 2, the word "BAYIG_A" is segmented into "BAI+G_A" in $Chain_1$ and "BAYI+G_A" in $Chain_2$. Meanwhile, "BAI" is an independent Mongolian word in the corpus. So $Chain_1$ can not discriminate the word "BAI" from the morpheme "BAI".

As well known, the translation quality of SMT relies on the performance of morphological segmentation. We give the following example to intuitively show the quality of translation of the chained SMT system.

Example 1 *Table 3 gives four examples of translating Chinese into Mongolian. In each example, four reference sentences translated by native experts are also given. These examples indicate that the chained SMT system can help choose the correct inflections, and partly alleviate the data sparseness problem.*

In Table 3(a), the Mongolian word "HAGAS" (corresponding to the Chinese word "yiban") has multiple inflectional forms as follows:

Mongolian	*Chinese*
HAGAS-VN	*yi bande*
HAGAS-IYAR	*yiban de*
HAGAS-TV	*zaiban*
HAGAS-I	*ba ban*

From the above example, we can see that the baseline system translates the Chinese word "ban" to the incorrect inflection "HAGAS-TV, while $Chain_2$ translates it to the correct inflection "HAGAS" which is the morpheme of all the other inflections.

Table 3: Examples of translating Chinese into Mongolian

(a) Lexicalization of morphemes

Chinese	xianzai shi jiu dian ban .
Baseline	0D0 B0L YISUN CAG HAGAS-TV.
$Chain_1$	0D0 B0L YISUN CAG HAGAS-TV.
$Chain_2$	0D0 B0L YISUN CAG HAGAS B0LJV BAYIN_A.
References	0D0 YISUN CAG HAGAS B0LJV BAYIN_A. 0D0 YISUN CAG HAGAS. 0D0 YISUN CAG HAGAS B0LBA. 0D0 YISUN CAG HAGAS B0LJV BAYIN_A.

(b) Tense

Chinese	qunian zheshihou ni zai ganshenme ?
Baseline	NIDVNVN ENE HIRI CI YAGV HIJU BAYIHV BVI?
$Chain_1$	NIDVNVN ENE UYES CI YAGV HIJU BAYIHV BVI?
$Chain_2$	NIDVNVN ENE UY_E-DU CI YAGV HIJU BAYIG_A BVI?
References	NIDVNVN-V ENE UYE-DU TA YAGV HIJU BAYIG_A BVI? NIDVNVN ENE UY_E-DU TA YAGV HIJU BAYIBA? NIDVNVN JIL-VN ENE UYES TA YAGV HIJU BAYIBA? 0D0 NIDVNVN-V ENE UYE-DU TA YAGV HIJU BAYIG_A BVI?

(c) Syntax

Chinese	wo xiwang jinnian dong tian hui xiaxue .
Baseline	BI ENE JIL EBUL-UN EDUR-UN CASV 0R0JV B0L0N_A.
$Chain_1$	BI ENE EBUL-UN EDUR CASV 0R0HV-YI HUSEJU BAYIN_A.
$Chain_2$	BI ENE EBUL-UN EDUR CASV 0R0HV-YI HUSEJU BAYIN_A.
References	BI ENE EBUL CAS 0R0HV-YI HUSEJU BAYIN_A. ENE EBUL CASV 0R0HV-YI HUSEJU BAYIN_A. BI ENE EBUL CASV 0R0HV-YI HUSEN_E. BI ENE EBUL CAS 0R0HV-YI HUSEJU BAYIN_A.

(d) Out-Of-Vocabulary words

Chinese	wo guoqu chang yidazao chuqu sanbu .
Baseline	... URGULJI yidazao GADAN_A GARCV SELEGUCEN ALHVLABA.
$Chain_1$	... URGULJI BODORIHU-BER GADAGVR ALHVLAN_A.
$Chain_2$	... URGULJI ORLOGE ERTE GARCV ALHVLAN_A.
References	... URGULJI OROLGE ERTE GARCV AVHVDAG. ... URGULJI ORLOGE ERTE GADAGVR ALHVLADAG. ... YERU NI ORLOGE ERTE B0S0GAD GADAGVR ALHVLADAG. ... URGULJI OROLGE ERTE GARCV AVHVDAG.

In Table 3(b), the word "BAYIN" in the result of the baseline system indicates the past tense environment. However, the correct environment is the past continuous tense which is indicated by the word "BAYIN_A" appearing in the results of $chain_1$ and $chain_2$.

In Table 3(c), the baseline system translates "dongtian" into "EDUR-UN" as an attribute, while the correct translation should be "EDUR" as the subject of the object clause.

The statistical data-sets from word alignment corpus show that the vocabularies of the baseline system includes 376203 Chinese-Mongolian word pairs, while $Chain_1$ and $Chain_2$ contain 326847 and 291957 Chinese-Morphemes pairs respectively. This indicates that the chained SMT system can partly alleviates the data sparseness problem. As shown in Table 3(d), the baseline system can not translate the word "yidazao", while $Chain_1$ and $Chain_2$ can.

5 Concluding remarks

The paper proposes the chained SMT system using morphemes as pivot language for translating an isolated language with non-morphology into an agglutinative language with productive derivational and inflectional morphology. The experiment results show that the performance of the chained SMT system is encouraging. And the SMT-based method is quite effective for finding mapping relations between morphemes and words. When adding more features, the precision, recall and F-measure are all improved more obviously.

In the future, we will consider the confusion network or lattice of N-best translation results instead of one best translation result in the chained SMT system. Meanwhile, the distortion of morpheme order in Mongolian is still obscure and needs more investigation. And comparing our work with other language pairs, such as English-to-French translation, English-to-Spanish translation, and so on, is also concerned.

Acknowledgments

The authors would like to thank the anonymous reviewers for their helpful reviews. The work is supported by the National Key Technology R&D Program of China under No. 2009BAH41B06 and the Dean Foundation (2009) of Hefei Institutes of Physical Science, Chinese Academy of Sciences.

References

[Callison-Burch et al.2009] Callison-Burch, Chris, Philipp Koehn, Christof Monz, and Josh Schroeder. 2009. Findings of the 2009 workshop on statistical machine translation. In *StatMT*, pages 1–28.

[Creutz and Lagus2007] Creutz, Mathias and Krista Lagus. 2007. Unsupervised models for morpheme segmentation and morphology learning. *TSLP*, 4(1):1–34.

[de Gispert et al.2009] de Gispert, Adrià, Sami Virpioja, Mikko Kurimo, and William Byrne. 2009. Minimum bayes risk combination of translation hypotheses from alternative morphological decompositions. In *HLT*, pages 73–76.

[Doddington2002] Doddington, George. 2002. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *HLT*, pages 128–132.

[Jiang and Zhou2008] Jiang, Long and Ming Zhou. 2008. Monolingual machine translation for paraphrase generation. In *COLING*, pages 377–384.

[Koehn et al.2003] Koehn, Philipp, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *HLT-NAACL*, pages 48–54.

[Koehn et al.2007] Koehn, Philipp, Hieu Hoang, Alexandra Birch, Chris 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*, pages 177–180.

[Minkov et al.2007] Minkov, Einat, Kristina Toutanova, and Hisami Suzuki. 2007. Generating complex morphology for machine translation. In *ACL*, pages 128–135.

[Nashunwukoutu1997] Nashunwukoutu. 1997. An automatic segmentation system for the root, stem, suffix of the mongolian. *Journal of Inner Mongolia University*, 29(2):53–57.

[Och and Ney2002] Och, Franz Josef and Hermann Ney. 2002. Discriminative training and maximum entropy models for statistical machine translation. In *ACL*, pages 295–302.

[Och2003] Och, Franz Josef. 2003. Minimum error rate training in statistical machine translation. In *ACL*, pages 160–167.

[Papineni et al.2002] Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *ACL*, pages 311–318.

[Quirk et al.2004] Quirk, Chris, Chris Brockett, and William B. Dolan. 2004. Generating chinese couplets using a statistical MT approach. In *EMNLP*, pages 142–149.

[Ramanathan et al.2009] Ramanathan, Ananthakrishnan, Hansraj Choudhary, Avishek Ghosh, and Pushpak Bhattacharyya. 2009. Case markers and morphology: addressing the crux of the fluency problem in English-Hindi SMT. In *ACL-IJCNLP*, pages 800–808.

[Riezler et al.2007] Riezler, Stefan, Alexander Vasserman, Ioannis Tsochantaridis, Vibhu O. Mittal, and Yi Liu. 2007. Statistical machine translation for query expansion in answer retrieval. In *ACL*, pages 464–471.

[Stolcke2002] Stolcke, Andreas. 2002. SRILM - an extensible language modeling toolkit. In *Proc. Intl. Conf. on Spoken Language Processing*, pages 901–904.

[Watanabe et al.2006] Watanabe, Taro, Hajime Tsukada, and Hideki Isozaki. 2006. Ntt system description for the wmt2006 shared task. In *WMT*, pages 122–125.

[Zhang2008] Zhang, Huaping. 2008. ICTCLAS. http://ictclas.org/.